



ANÁLISIS DESCRIPTIVO DE LA ACCESIBILIDAD A LOS CONTENIDOS AUDIOVISUALES DE WEBS DEL PARLAMENTO EUROPEO

Iris Serrat Roozen

TESIS DOCTORAL

DIRECTORAS DE TESIS
Rosa María Agost Canós
María Calzada Pérez

Mayo de 2019





Programa de Doctorado en Lenguas Aplicadas, Literatura y Traducción
Escuela de Doctorado de la Universitat Jaume I

ANÁLISIS DESCRIPTIVO DE LA ACCESIBILIDAD A LOS CONTENIDOS AUDIOVISUALES DE WEBS DEL PARLAMENTO EUROPEO

Memoria presentada por
Iris Serrat Roozen
para optar al grado de
doctora por la Universitat Jaume I

Iris Serrat Roozen

Rosa María Agost Canós

María Calzada Pérez

Castelló de la Plana, mayo de 2019

A mi madre y a mi hermana,
a mis sobrinos,
a mis hijos.

AGRADECIMIENTOS

En primer lugar quiero expresar mi más sincero y profundo agradecimiento a la directora de esta tesis doctoral, la Dra. María Calzada Pérez, por ser la persona que ha hecho posible que esta tesis salga adelante y con quien siempre estaré en deuda. Gracias por la confianza depositada en mí y por el derroche de paciencia que ha supuesto la redacción de esta tesis.

Asimismo, agradezco el apoyo, la comprensión y la disposición absoluta de mi también directora la Dra. Rosa María Agost Canós. Cuando la accesibilidad aún no estaba de moda apoyaste mis ideas, gracias. Lo que empezamos con aquel TFG sigue con esta tesis y el tiempo dirá dónde nos lleva.

A mi compañero de investigación desde los inicios, José Manuel Martínez Martínez, por su inestimable colaboración en la compilación del corpus EMPAC, pero sobre todo por su amistad y cariño.

Al Dr. Pedro Canales Ronda, por enseñarme a mirar los datos con otros ojos, por hacer de la estadística algo atractivo y sencillo y por compartir conmigo los primeros años de mi carrera docente.

A mis compañeros de departamento, con quienes comparto la pasión por la educación, gracias por todo vuestro apoyo, ayuda y comprensión, en especial, a Lola, Susan y Paul.

En el ámbito más personal quiero dedicar estas líneas a mis amigos de siempre, por vuestro apoyo, por las risas compartidas, por las confidencias y por hacerme sentir en casa cuando estoy con vosotros.

Gracias a mi familia, a mi madre y a mi hermana, sin vuestro apoyo y amor este largo camino hubiera sido insoportable. Gracias en especial a mi madre, por estar siempre que te he necesitado y por luchar para que tus hijas tuvieran el mejor de los futuros posibles.

Y por último, a las sonrisas más bonitas del mundo, las de Olivia y Leo.

RESUMEN

Observar las normas que subyacen a la subtitulación *online* generada por profesionales de la traducción en el ámbito institucional y determinar si están alineadas con el compromiso de acceso a la información por igual para todos los ciudadanos europeos es el reto que asume esta tesis.

Nuestro objetivo principal ha sido el de describir, en un entorno multilingüe como es el de la Unión Europea, las normas que se manifiestan en torno a varios aspectos esenciales de la subtitulación (número de caracteres, tiempo de permanencia en pantalla, segmentación de los subtítulos, etc.). De hecho, nuestro estudio conlleva el análisis empírico de un corpus representativo desde un enfoque descriptivo en el que damos prioridad al texto meta. El concepto de norma, es decir, las regularidades de comportamiento traductor que apreciamos de forma sistemática en los textos, nos permite describir cómo se subtitula en el seno de la Unión Europea, concretamente en el Parlamento Europeo, y también nos permite analizar la evolución histórica de las normas (técnicas) de subtitulación en el canal de televisión *online* EuroparlTV.

Para llevar a cabo nuestro objetivo y poder realizar el estudio sistemático y exhaustivo de las variables seleccionadas, hemos creado y analizado el Corpus EMPAC (EuroparlTV Multimedia Parallel Corpus) que reúne los subtítulos en inglés y en español de los vídeos emitidos en EuroparlTV entre 2009 y 2017.

ABSTRACT

The aim of this thesis is to observe the rules underlying online subtitling carried out by translation professionals in the institutional sphere and to determine whether or not these rules are aligned with the commitment to grant equal access to information for all European citizens.

Our main objective is to describe the observable rules relating to several essential aspects of subtitling (number of characters, time spent on screen, segmentation of subtitles, etc.) in a multilingual environment such as the European Union. Our study involves using a descriptive approach to conduct an empirical analysis of a representative corpus, wherein we prioritise the target text. The concept of norm - i.e., the regularities in translator behaviour to be found systematically in the texts - allows us to describe how subtitling is conducted within the European Union, specifically in the European Parliament, and also enables us to analyse the historical evolution of the norms (techniques) of subtitling in the online television channel EuroParlTV.

In order to achieve our objective and be able to carry out the systematic and exhaustive study of the selected variables, we created and analysed the EMPAC Corpus (EuroParlTV Multimedia Parallel Corpus) that gathers the English and Spanish subtitles of the videos broadcast on EuroParlTV between 2009 and 2017.

TABLA DE CONTENIDO

ÍNDICE DE TABLAS	15
ÍNDICE DE FIGURAS	17
INTRODUCCIÓN	19
CAPÍTULO 1	27
LOS ESTUDIOS DE TRADUCCIÓN Y LOS ESTUDIOS DESCRIPTIVOS DE TRADUCCIÓN	29
1.1. Los estudios de traducción	29
1.2. Escuela de la manipulación y el polisistema.....	33
1.3. La escuela de la manipulación y los EDT.....	35
1.4. El concepto de norma	36
1.5. Las normas en traducción audiovisual	41
1.5.1. Estudio de las normas en traducción audiovisual de Pedersen.	42
CAPÍTULO 2	45
EL ÁMBITO DE LA TRADUCCIÓN AUDIOVISUAL	47
2.1. La traducción audiovisual y la accesibilidad.....	47
2.2. Traducción audiovisual y estudios de traducción.....	49
2.3. La traducción audiovisual: un debate terminológico.....	53
2.4. La traducción audiovisual: definición y modalidades.....	54
2.5. Clasificaciones de modalidades de TAV.....	56
2.6. Modalidades de traducción audiovisual en Internet.....	59
2.6.1. La audiodescripción para personas ciegas.....	59
2.6.2. La subtitulación y la subtitulación para personas sordas y con discapacidad auditiva (SPS).....	62

CAPÍTULO 3	67
LA ACCESIBILIDAD EN LA ERA DE LA COMUNICACIÓN DIGITAL	69
3.1. La sociedad de la información y la accesibilidad	69
3.2. La accesibilidad en la sociedad de la información	73
3.2.1. Evolución del concepto de accesibilidad.....	73
3.2.2. El paradigma del Diseño para Todos	77
3.2.3. Las barreras de la comunicación y la información	80
3.2.4. Realidad de la discapacidad en nuestros días	82
3.3. Las reglas del juego: marco legal y normativo	85
3.3.1. Marco legal	86
3.3.2. Marco normativo en España y en Europa.....	88
CAPÍTULO 4	95
LA WEB Y LAS DIRECTRICES DE ACCESIBILIDAD	97
4.1. Las tres uves dobles: el W3C, la WAI y las WCAG	97
4.1.1. Pautas de accesibilidad al contenido web (WCAG): versión 1.0 y 2.0	99
4.1.2. Pautas de accesibilidad al contenido audiovisual: WCAG 1.0 y WCAG 2.0.....	101
CAPÍTULO 5	111
EVALUACIÓN DE LA ACCESIBILIDAD AUDIOVISUAL EN LA WEB	113
5.1. Evaluación de la accesibilidad del contenido audiovisual en la web	113
5.1.1. <i>How to meet WCAG 2.0</i> : definición y comprobación de los criterios de éxito	116
5.1.2. Modelo de evaluación adecuado a nuestro ámbito de estudio	119
CAPÍTULO 6	127
LA CALIDAD DE LA ACCESIBILIDAD AUDIOVISUAL EN LA WEB: GUÍAS Y RECOMENDACIONES	129
6.1. Calidad en la accesibilidad web al contenido audiovisual	129
6.2. Guías y recomendaciones en el ámbito de la subtitulación	130

6.3. Subtitulación: normas, guías y recomendaciones para la subtitulación online	133
CAPÍTULO 7	137
PARÁMETROS DE ESTUDIO	139
7.1. Los parámetros de estudio en subtitulación	139
7.2. Definición y análisis de los parámetros de estudio	140
7.2.1. Parámetro velocidad de lectura	141
7.2.2. Parámetro formato y posicionamiento	146
CAPÍTULO 8	159
LINGÜÍSTICA DE CORPUS COMO METODOLOGÍA CUANTITATIVA	161
8.1. La Lingüística de Corpus.....	161
8.2. La Lingüística de Corpus y el corpus lingüístico.....	162
8.3. Clasificación y tipos de corpus lingüísticos	164
CAPÍTULO 9	167
METODOLOGÍA DE COMPILACIÓN Y PREPARACIÓN DEL CORPUS DE ESTUDIO	169
9.1. Contextualización del corpus de estudio	169
9.2. Descripción del corpus EMPAC	174
9.3. Compilación y preparación del corpus de estudio.....	181
9.4. Obtención de los datos y metadatos	184
9.5. Procesamiento de los textos: transformación de los archivos en xml	187
9.6. Procesamiento lingüístico de los textos.....	192
9.6.1. Etiquetado morfosintáctico con TreeTagger.....	192
9.6.2. División del texto en frases.....	195
9.6.3. Indexación y alineado del corpus.....	196
9.6.4. Subida del corpus al servidor de CQPWeb	198
CAPÍTULO 10	199
METODOLOGÍA DE ANÁLISIS	201
10.1. Objetivos e hipótesis de trabajo.....	201

10.2. Estudio inicial del corpus EMPAC:	
accesibilidad al contenido audiovisual	204
10.2.1. Metodología de análisis para detectar si existen subtítulos	205
10.2.2. Metodología de análisis para detectar si son subtítulos convencionales o subtítulos para personas sordas	207
10.2.3. Metodología de análisis para determinar si existen audiodescripciones	208
10.3. Estudio central del corpus EMPAC	208
10.3.1. Metodología de análisis de los parámetros de estudio	212
10.3.2. Metodología de análisis de las causas de segmentación inadecuada	216
CAPÍTULO 11	225
ANÁLISIS DEL CORPUS	227
11.1. Análisis inicial del corpus EMPAC: accesibilidad al contenido audiovisual	227
11.1.1. Fase 1: Análisis para detectar si los vídeos incorporan subtítulos	228
11.1.2. Fase 2: Análisis para detectar si son subtítulos convencionales o subtítulos para personas sordas	235
11.1.3. Fase 3: Análisis para determinar si existen audiodescripciones ...	242
11.2. Estudio central del corpus EMPAC	243
11.2.1. Análisis del parámetro velocidad de lectura	244
11.2.2. Análisis del parámetro formato y posicionamiento	257
CAPÍTULO 12	407
CONCLUSIONES	409
BIBLIOGRAFÍA	431
ANEXOS	449

ÍNDICE DE TABLAS

Tabla 1. Ficha de evaluación contenido con subtítulo.	124
Tabla 2. Ficha de evaluación contenido con subtítulos convencionales o para personas sordas.	124
Tabla 3. Ficha de evaluación contenidos con audiodescripción.	124
Tabla 4. Guías para la subtitulación.	134
Tabla 5. Parámetro velocidad de lectura.	146
Tabla 6. Recomendaciones de segmentación de los subtítulos para la SPS online.	152
Tabla 7. Propuesta de análisis de causas de segmentación inadecuada.	155
Tabla 8. Parámetro formato y posicionamiento.	156
Tabla 9. Datos de EMPAC_EN.	174
Tabla 10. Datos EMPAC_ES.	174
Tabla 11. Distribución de los datos por año en EMPAC_EN.	175
Tabla 12. Distribución de los datos por año en EMPAC_ES.	175
Tabla 13. Distribución de los datos por tipo de texto en EMPAC_EN.	176
Tabla 14. Distribución de los datos por tipo de texto en EMPAC_ES.	176
Tabla 15. Distribución de los datos por tipo de texto y año en EMPAC_EN.	176
Tabla 16. Distribución de los datos por tipo de texto y año en EMPAC_ES.	177
Tabla 17. Distribución de los datos por categoría en EMPAC_EN.	179
Tabla 18. Distribución de los datos por categorías en EMPAC_ES.	179
Tabla 19. Distribución de los datos por categoría y año en EMPAC_EN.	179
Tabla 20. Distribución de los datos por categoría y año en EMPAC_ES.	180
Tabla 21. Archivos de subtítulo perdidos.	186
Tabla 22. Ficha para determinar si los vídeos incorporan subtítulos convencionales o para personas sordas.	207
Tabla 23. Ficha evaluación para determinar si los vídeos incorporan audiodescripción. ...	208
Tabla 24. Metadatos a nivel de texto.	209
Tabla 25. Metadatos a nivel de subtítulo.	210

Tabla 26. Metadatos a nivel de línea.....	210
Tabla 27. Ejemplo de análisis bivariable cps y años.....	214
Tabla 28. Búsquedas de segmentación inadecuada intersubtítulo para inglés.	219
Tabla 29. Búsquedas de segmentación inadecuada intersubtítulo para español.....	220
Tabla 30. Búsquedas de segmentación inadecuada intrasubtítulo para inglés	221
Tabla 31. Búsquedas de segmentación inadecuada intrasubtítulo para español.....	222
Tabla 32. Ejemplo visualización de resultados de segmentación intersubtítulo.....	224
Tabla 33. Datos vídeos sin subtítulo en Corpus EMPAC.	228
Tabla 34. Datos vídeos sin subtítulo por año en EMPAC_EN.....	229
Tabla 35. Datos de vídeos sin subtítulo por año en EMPAC_ES.....	229
Tabla 36. Datos de vídeos sin subtítulos por tipo de texto en EMPAC_EN y EMPAC_ES....	230
Tabla 37. Datos de vídeos sin subtítulos por tipo de texto en EMPAC_EN.	231
Tabla 38. Datos de vídeos sin subtítulos por tipo de texto en EMPAC_ES.....	233
Tabla 39. Datos de EMPAC_EN y EMPAC_ES con número de textos, subtítulos, palabras y duración.	243
Tabla 40. Resultados cps por tipo de texto en EMPAC_EN y EMPAC_ES.....	248
Tabla 41. Resultados pausa entre subtítulos por años en EMPAC_EN y EMPAC_ES.	251
Tabla 42. Resultados pausa entre subtítulos por tipo de texto en EMPAC_EN y EMPAC_ES.	252
Tabla 43. Resultados duración de los subtítulos en EMPAC_EN y EMPAC_ES.	252
Tabla 44. Resultados variable duración por años en EMPAC_EN y EMPAC_ES.	253
Tabla 45. Resultados duración por grupos y años en EMPAC_EN y EMPAC_ES.	256
Tabla 46. Resultados variable número de líneas por tipo de texto en EMPAC_EN y EMPAC_ES.	260
Tabla 47. Resultados análisis cpl por tipo de texto en EMPAC_EN y EMPAC_ES.	263
Tabla 48. Resultado de análisis de la segmentación intrasubtítulo en inglés.....	268
Tabla 49. Resultados de segmentación intersubtítulo en EMPAC_EN.....	268
Tabla 50. Resultado del análisis de segmentación intersubtítulo en inglés.	319
Tabla 51. Resultados análisis segmentación intrasubtítulo en español.....	335
Tabla 52. Resultados análisis segmentación intersubtítulo en español.....	335
Tabla 53. Resultados de segmentación intrasubtítulo inadecuada por categoría en EMPAC_ES.	375
Tabla 54. Resultados segmentación intersubtítulo inadecuada en EMPAC_ES.	387
Tabla 55. Resultados de análisis de causas de segmentación inadecuada por años en EMPAC_EN.	402
Tabla 56. Resultados de análisis de causas de segmetnación inadecuada por años en EMPAC_ES.	404

ÍNDICE DE FIGURAS

ÍNDICE DE FIGURAS

Figura 1. Estudios de Traducción: Homes (1972).....	30
Figura 2. The potency of translation norms.	39
Figura 3. Clasificación de modalidades de traducción audiovisual.....	57
Figura 4. Propuestas de taxonomía de modalidades y submodalidades de TAV.....	58
Figura 5. Sectores de comunicación e información. Acceplan, 2003.....	81
Figura 6. Alternativas equivalentes para el contenido sonoro y visual. WCAG 1.0, 1999.	102
Figura 7. Correspondencia entre WCAG y normas UNE. Olga Carreras.	106
Figura 8. Agentes que participan en la cadena de producción web.	109
Figura 9. Tabla de responsabilidades del Content Strategy Stakeholder.	109
Figura 10. Interfaz de How to meet WCAG 2.0.....	116
Figura 11. Selección de filtros criterio de éxitos 1.2.2.....	117
Figura 12. Cómo incluir subtítulos WCAG 2.0.	117
Figura 13. Ejemplo de causas para no alcanzar criterio de éxito.....	118
Figura 14. Ficha de evaluación subtítulos.	121
Figura 15. Ficha evaluación al contenido de Voces Merayo.....	123
Figura 16. Macrovariables de estudio propuestas por Tamayo.	140
Figura 17. Parámetro velocidad de lectura propuesto por Tamayo.....	141
Figura 18. Recopilación de distintos autores de velocidad de lectura máxima de subtítulos. Tamayo.....	143
Figura 19. Elemento web que define la hoja de estilo.....	148
Figura 20. Propuesta de segmentación por nodos.....	151
Figura 21. Ficha de cómo evaluar subtítulos de Voces Merayo.....	205
Figura 22. Estructura Excel All_Videos.....	207
Figura 23. Excel de EMPAC_EN con la información recuperada a nivel de subtítulo.	211
Figura 24. Excel con información recuperada a nivel de línea.....	212
Figura 25. Búsqueda con CQPweb de preposición al final de línea.....	218

Figura 26. Subtítulo en contexto en CQPweb.	237
Figura 27. Metadatos a nivel de texto en CQPweb.	239
Figura 28. Subtítulos alineados en CQPweb (ejemplo siglas EEUU).	241
Figura 29. Vídeo de EuroparlTV con información textual en pantalla.	241
Figura 30. Vídeo de EuroparlTV con información textual en pantalla subtitulada.	242
Figura 31. Resultado de análisis cps en EMPAC_EN Y EMPAC_ES.	245
Figura 32. Resultado análisis cps por años en EMPAC_EN.	246
Figura 33. Resultado análisis cps por años en EMPAC_ES.	247
Figura 34. Resultados análisis cps en 2017 en EMPAC_EN y EMPAC_ES.	247
Figura 35. Resultado pausa entre subtítulos en EMPAC_EN.	249
Figura 36. Resultado pausa entre subtítulos en EMPAC_ES.	249
Figura 37. Evolución de la pausa entre subtítulos entre 2009 y 2017 en EMPAC_EN.	250
Figura 38. Evolución de la pausa entre subtítulos entre 2009 y 2017 en EMPAC_ES.	250
Figura 39. Resultados de variable duración por tipo de texto en EMPAC_EN.	254
Figura 40. Resultados de variable duración por tipo de texto en EMPAC_ES.	254
Figura 41. Resultados duración por grupos en EMPAC_EN.	255
Figura 42. Resultados duración por grupos en EMPAC_ES.	255
Figura 43. Resultados de duración por grupos y cps en EMPAC_EN.	256
Figura 44. Resultados de duración por grupos y cps en EMPAC_ES.	257
Figura 45. Resultados de la variable número de líneas en EMPAC_EN.	258
Figura 46. Resultados de la variable número de líneas en EMPAC_ES.	259
Figura 47. Evolución de la variable número de líneas en EMPAC_EN.	260
Figura 48. Evolución de la variable número de líneas en EMPAC_ES.	260
Figura 49. Resultado de análisis de cpl en EMPAC_EN.	261
Figura 50. Resultado de análisis de cpl en EMPAC_ES.	261
Figura 51. Evolución de cpl en EMPAC_EN.	262
Figura 52. Evolución de cpl en EMPAC_ES.	262
Figura 53. Interfaz de CQPweb y el menú de búsquedas.	265
Figura 54. Visualización de los resultados de la búsqueda con CQPweb.	266
Figura 55. Visualización de resultados al filtrar con el <i>frequency breakdown</i>	266
Figura 56. Búsqueda restringida por categoría, tipo de texto, año en CQPweb.	267

INTRODUCCIÓN

La Declaración Universal de los Derechos Humanos, en el Artículo 27.1, afirma con claridad meridiana que toda persona, pertenezca a grupos mayoritarios o minoritarios de la sociedad, tiene derecho a «tomar parte libremente en la vida cultural de la comunidad, a gozar de las artes y a participar en el progreso científico y en los beneficios que de él resulten».

Esta plena integración en el mundo que habitamos está, irremediabilmente, vinculada a la accesibilidad, al desarrollo tecnológico y, especialmente, desde hace ya más de dos décadas, a Internet. De hecho, parece indiscutible que este medio ha supuesto un cambio vertiginoso en la organización social actual y ha generado profundas transformaciones en casi todas las esferas sociales hasta el punto que podemos hablar de una revolución comparable a la que posibilitó Gutemberg en el siglo XV gracias a la invención de la imprenta. A día de hoy, Internet es una herramienta indispensable para, entre otras cosas, trabajar, estudiar, relacionarse con la Administración y divertirse, y, sin duda, es también un instrumento que puede fomentar una convivencia más inclusiva, heterogénea y justa. En consecuencia, no ha de sorprender que realizar tareas cotidianas, como navegar por Internet y consultar los contenidos textuales y audiovisuales que se muestran en este medio, sea un derecho reconocido por ley. Y, sin embargo, es este un derecho cuya plena ejecución, por el momento, y a tenor de los datos públicos, todavía es una asignatura pendiente, tanto en España como en otros países europeos. Entendemos que para alcanzarlo es necesario, por un parte, eliminar las barreras erigidas en este canal de comunicación y, por otra parte, enseñar y fomentar la noción e importancia del denominado *Diseño para Todos*. Así pues, vencer la falta de sensibilización de la opinión pública y superar las lagunas en cuanto a la difusión de las medidas necesarias para el desarrollo de una sociedad de la información para todos, se revelan como escollos aún demasiado presentes que hay que salvar lo más pronto posible.

Esta tesis pretende sumar esfuerzos en este sentido, y para ello queremos poner de relieve la importancia de la accesibilidad y su vínculo indeleble con los Estudios de Traducción,

en general, y la traducción audiovisual, en concreto. Hemos pretendido evidenciar, de un lado, que el concepto de accesibilidad se ha hecho un hueco en nuestra disciplina y ha venido para quedarse y para ocupar, a nuestro entender, un lugar central. De hecho, tanto es así que podemos hablar de accesibilidad audiovisual, especialmente en Internet, puesto que crear contenidos Diseñados para Todos pasa por evitar las barreras de la comunicación (de especial relevancia en la sociedad actual) e incluir modalidades de traducción audiovisual tradicionales (subtitulación, doblaje, etc.) y otras modalidades emergentes (subtitulación para personas sordas, audiodescripción, rehablado, etc.) en los vídeos transmitidos en el entorno web. Por otra parte, pretendemos mostrar que los avances alcanzados en materia de accesibilidad a los contenidos audiovisuales se deben fundamentalmente al desarrollo de normas y leyes que amparan los derechos de los ciudadanos. En consecuencia, nos parece imprescindible colaborar en la consolidación del marco normativo y legal que impulse los avances en esta materia.

El estudio de la accesibilidad audiovisual en la web no ha suscitado, por el momento, el interés de muchos investigadores de nuestro ámbito. Probablemente esto es debido a que durante años los esfuerzos de asociaciones y grupos de interés se focalizaron en conseguir que los medios de comunicación, en particular, las televisiones, alcanzaran los porcentajes de emisiones con subtítulos y audiodescripciones previstos por la normativa vigente. Una vez alcanzados, la calidad de los subtítulos y las audiodescripciones adquirió el protagonismo y se generaron numerosos estudios y publicaciones que perseguían ser de utilidad para las distintas audiencias. Sin embargo, si bien en los últimos años han proliferado los trabajos sobre la accesibilidad audiovisual en otros medios (Orero, 2005a; Díaz Cintas 2007, entre otros), el estudio de contenidos audiovisuales en Internet es, a día de hoy, un terreno prácticamente ignoto.

Esta tesis nace con el propósito de contribuir a paliar este vacío. Sin embargo, no persigue objetivos de exhaustividad, sino que se presenta como una primera investigación con datos empíricos representativos sobre la accesibilidad a los contenidos audiovisuales a partir de los planteamientos de los Estudios Descriptivos de la Traducción (EDT) y con la ayuda de la metodología de corpus. Para este acercamiento incipiente a nuestro objeto de estudio nos valemos de un gran corpus *ad hoc* (*European Multimedia Parallel Corpus*, EMPAC, en sus versiones en inglés y español, de 5 millones de palabras) que hemos compilado de forma semiautomatizada desde cero a partir de los textos audiovisuales alojados en el canal de televisión *online* del Parlamento Europeo (EuroparlTV)¹ entre 2009 y 2017. Para la exploración del corpus y el análisis de la segmentación en particular hemos indexado

¹ Multimedia Center desde enero de 2018.

EMPAC con el *IMS Open Corpus Workbench*² (Evert and Hardie, 2011) lo que permite utilizar el procesador de consultas *Corpus Query Processor* (Christ, 1994) a través de la interfaz gráfica CQPWeb (Hardie, 2012). Todo ello posibilita la realización de búsquedas complejas aprovechando la información contenida tanto en nuestras etiquetas estructurales (metadatos, subtítulos, líneas y oraciones) como en los atributos posicionales de cada palabra (*token*, lema y categoría morfosintáctica).

Así, nuestra hipótesis de partida es que el contenido audiovisual (en inglés y español) alojado en el canal de televisión *online* del Parlamento Europeo, EuroParlTV, procura la accesibilidad a través del cumplimiento de los requisitos de la WAI (Iniciativa de Accesibilidad Web, sus siglas en inglés) y mediante la traducción multilingüe del contenido audiovisual; más en concreto, mediante el uso respetuoso de las normas o sugerencias emitidas por organismos o estudiosos en cuanto a los parámetros técnicos de velocidad de lectura, formato y posicionamiento. De hecho, en la Política de accesibilidad³ de la Unión Europea y en concreto de la web EUROPA, se afirma que todas las páginas web deben seguir las Pautas de Accesibilidad al Contenido (WCAG) según la versión 2.0 y procurar sin discriminación el acceso al contenido.

A fin de corroborar o refutar esta hipótesis, nos planteamos el siguiente objetivo general.

1. Determinar si los vídeos alojados en la web del Parlamento Europeo son accesibles y valorar si los subtítulos que acompañan a este material audiovisual respetan las normas (técnicas) de esta modalidad de traducción audiovisual.

Para alcanzar el objetivo general de esta tesis proponemos los siguientes objetivos específicos, primero en el plano teórico y seguidamente en el plano descriptivo:

Objetivos específicos en el plano teórico:

1. Presentar el estado de la cuestión sobre la accesibilidad; en concreto sobre la accesibilidad audiovisual en la web y definir el vínculo que la une a los estudios de traducción.
2. Revisar el marco normativo que regula la subtitulación y la subtítulos para personas sordas (SPS) y mostrar si las recomendaciones y las guías de buenas prácticas que proponen los organismos de estandarización y la academia se ven

² Diseñado por el *Institut für Maschinelle Sprachverarbeitung* de Stuttgart.

³ https://europa.eu/european-union/abouteuropa/accessibility_es

atendidas en este medio (Internet) y con este tipo de producto audiovisual (vídeos institucionales/traducción institucional).

3. Describir los parámetros técnicos de la subtitulación (velocidad de lectura y formato y posicionamiento, incluida la segmentación) y los valores asignados a cada parámetro y plasmar en qué medida estos valores están presentes en nuestro corpus EMPAC.

Objetivos específicos en el plano descriptivo:

1. Determinar si el contenido audiovisual alojado en EuroparlTV; concretamente, si el que hemos seleccionado para conformar el corpus EMPAC cumple con los requisitos de accesibilidad establecidos por la WAI a partir de las WCAG.
2. Plasmar la evolución de los parámetros y los valores asignados a cada parámetro entre 2009 y 2017, en consonancia con la naturaleza diacrónica de nuestro análisis en EMPAC_EN y EMPAC_ES.
3. Describir las coincidencias y divergencias respecto a los parámetros técnicos de la presentación de los subtítulos parámetros técnicos en EMPAC_EN y EMPAC_ES y su evolución desde el comienzo de las emisiones hasta 2017 mediante un análisis diacrónico.

Estructura

La consecución de los objetivos general y específicos de la presente tesis da lugar a una estructura de 12 capítulos: ocho de carácter teórico (capítulos 1-8), los dos siguientes de tipo metodológico (capítulos 9 y 10); un amplio capítulo analítico (capítulo 11); y un capítulo de conclusiones (capítulo 12).

Comenzamos nuestro trabajo enmarcando nuestra investigación en los Estudios de Traducción, (Capítulo 1), concretamente en los Estudios Descriptivos de Traducción, cuyo eje es el concepto de norma. Se trata de un concepto clave en nuestro trabajo puesto que nos proponemos evaluar el estado de las normas descriptivas y prescriptivas de la accesibilidad audiovisual. De un lado, la traducción audiovisual, que hoy ya ha superado los límites del pasado y puede ya caracterizarse como accesibilidad audiovisual de pleno derecho, se rige por normas (descriptivas o prescriptivas, emitidas desde los ámbitos académicos y formativos). De otro, el espíritu y los planteamientos descriptivos de los EDT son la vía más rápida para cumplir

los objetivos de la tesis por lo que quedan expuestos en este capítulo.

Con la norma y la descripción en mente, la tesis se adentra en el ámbito de la traducción audiovisual y en el capítulo 2 presentamos la Traducción Audiovisual (TAV) y su evolución histórica. Introducimos, por una parte, las modalidades tradicionales (doblaje, subtulado, etc.) y, por otra parte, las modalidades emergentes (Subtitulación para Sordos, SPS, Audiodescripción, AD, etc.) siendo estas últimas las que han acercado la TAV a la accesibilidad y han cambiado incluso la forma de entender la traducción al tiempo que han expandido sus fronteras.

Por tanto, a día de hoy la TAV trasciende su espacio primigenio y ya ha pasado a ser *accesibilidad audiovisual* de pleno derecho y por ello necesitamos acercarnos a este ámbito y a los conceptos que lo definen. Así, en el capítulo 3, presentamos los conceptos claves vinculados a la accesibilidad (*Diseño para Todos, barreras, brecha digital, etc.*); ponemos un énfasis especial en los aspectos relacionados con la accesibilidad al contenido audiovisual *online* (barreras de la comunicación) y presentamos el marco legal y normativo que posibilita el acceso a la información audiovisual a todas las personas. En este sentido, adquiere una relevancia especial la aprobación de la primera norma europea de accesibilidad para productos y servicios de tecnologías de la información y la comunicación que establece, entre otros, los requisitos de accesibilidad para el contenido web según las Pautas de Accesibilidad al Contenido (WCAG).

Además, en el caso de esta tesis, la accesibilidad se restringe al medio concreto de Internet y por eso en el capítulo 4 se centra en las cuestiones relacionadas con la accesibilidad en la web. Para ello, primero definimos qué es la red de redes y cuál es el espíritu que la caracteriza; seguidamente, mostramos cuál es su estructura y los organismos que la gestionan. Abordaremos en qué consisten las pautas de accesibilidad al contenido (WCAG) de forma que podamos vincular la accesibilidad web y la traducción audiovisual. Por último, en este capítulo hacemos especial hincapié en la importancia de la cadena de agentes que conforman el equipo de profesionales que se ocupan de la accesibilidad web y de la necesidad de incluir a profesionales de la traducción en la cadena.

En el capítulo 5 nos adentramos en la forma en la que se determina si los sitios web cumplen con los criterios de éxito en materia de accesibilidad audiovisual según las WCAG 2.0 y en él plasmamos qué indicaciones se dan en cuanto a la aplicación de la subtitulación para personas sordas o con discapacidad auditiva y la audiodescripción para personas ciegas o con discapacidad visual desde la WAI. Repasamos, de igual manera, propuestas de evaluación de estos contenidos y finalmente ofrecemos nuestra propuesta, una aportación personal a la

investigación en TAV que nos permitirá describir si los contenidos son accesibles según las WCAG.

Hemos constatado que las Pautas de Accesibilidad al Contenido Web (WCAG) se quedan en la superficie a la hora de ofrecer información para los desarrolladores de contenido web sobre cómo incorporar la SPS y la AD; por este motivo presentamos en el capítulo 6 las guías y recomendaciones sobre cómo hacer el material audiovisual *online* accesible, propuestas que nacen de los organismos competentes en esta materia, así como de empresas de reconocido prestigio. Nos fijamos, en concreto, en los estándares para la subtítulos *online* que velan por la calidad de esta práctica y que nos servirán como referencia para analizar los parámetros objeto de estudio en nuestra investigación.

Una vez mostradas las guías de referencia en nuestro estudio, en el capítulo 7 presentamos los parámetros y variables de estudio junto con las aportaciones de académicos de reconocido prestigio de nuestro ámbito, mientras que en el capítulo 8 definimos los aspectos fundamentales de la Lingüística de Corpus por ser la metodología de análisis de nuestro estudio puesto que entendemos que es especialmente adecuada para el estudio de las normas en el marco de los EDT.

Como toda investigación basada en corpus, la metodología comprende dos fases: la de compilación y la de análisis. Por ello, los capítulos 9 y 10 se han dedicado a este objetivo. Así, el capítulo 9 es fundamental en nuestra tesis doctoral puesto que describimos y contextualizamos nuestro corpus de estudio (*Corpus EMPAC*), y exponemos de forma detallada la metodología de compilación que está inspirada en la metodología del grupo de investigación de la Universitat Jaume I *European Comparable and Parallel Corpus of Parliamentary Speeches* (ECPC)⁴. EMPAC es un corpus robusto y muy valioso, pensamos, por su amplio tamaño, versatilidad genérica y formato informativo (con un XML múltiple, que abre innumerables vías de investigación). La transparencia descriptiva con la que presentamos el proceso compilador permite su replicabilidad y, por tanto, contribuye a reforzar prácticas investigadoras tan saludables como necesarias. En este sentido, destacamos además que EMPAC será de acceso libre a través de la web <https://empac-corpus.com/>

El capítulo 10 describe la metodología de análisis de nuestro estudio (estudio inicial y estudio central) y detallamos los objetivos e hipótesis de este estudio que ya han sido presentados en la Introducción. En el estudio central determinamos en qué medida los vídeos

⁴ El grupo de investigación ECPC es responsable de los siguientes proyectos de investigación con financiación ministerial: HUM2005-03756filo, FFI2008-01610, FFI2012-39389. También es responsable de los siguientes proyectos de financiación por parte de la Generalitat Valenciana o de la Universitat Jaume I: AICO17-082, UJI201-64, UJI2017-25).

ofrecidos en EuroparlTV⁵ son accesibles en consonancia con las WCAG 2.0. Por su parte, en el estudio central, analizamos los parámetros (técnicos) de nuestro estudio y de forma cuantitativa y cualitativa describimos cuál es la realidad de la accesibilidad en los vídeos de nuestro estudio.

Dedicamos un extenso capítulo 11 al análisis del corpus, en primer lugar, mostramos los resultados del análisis inicial que nos permitirá determinar si los vídeos alojados en EuroparlTV son accesibles. En segundo lugar, exploramos el Corpus EMPAC y mostramos los resultados obtenidos en EMPAC_EN y EMPAC_ES de forma que esto nos permite describir las regularidades en torno a los parámetros técnicos de nuestro corpus y contrastarlos con los valores definidos por la normativa vigente.

En la parte final de la investigación presentamos las conclusiones (capítulo 12) con el objetivo de resumir los resultados del análisis y proponer algunas líneas de investigación futuras.

Por último, en los Anexos pueden consultarse una serie de enlaces con toda la información a la que hemos ido haciendo referencia a lo largo de toda la investigación y que complementan los datos ofrecidos en la redacción de la misma.

⁵ En agosto de 2018 pasó a denominarse Multimedia Center.

CAPÍTULO 1

LOS ESTUDIOS DE TRADUCCIÓN Y LOS EDT

1.1. Los Estudios de Traducción

1.2. La Escuela de la Manipulación y el Polisistema

1.3. La Escuela de la Manipulación y los EDT

1.4. El concepto de norma

1.5. Las normas en traducción audiovisual

1.5.1. Estudio de las normas en traducción audiovisual de Pedersen

CAPÍTULO 1. LOS ESTUDIOS DE TRADUCCIÓN Y LOS ESTUDIOS DESCRIPTIVOS DE TRADUCCIÓN

Actualmente, la accesibilidad es un concepto transversal que se utiliza en diversos ámbitos, a pesar de que, en sus orígenes, se diera a conocer de forma muy destacada por la irrupción en el mundo de la arquitectura y el urbanismo. De hecho, también los Estudios Descriptivos de Traducción (EDT) lo han incorporado desde hace poco más de una década. La accesibilidad es, en realidad inherente a la propia existencia de la traducción, y se manifiesta de forma especialmente poderosa en el caso de la traducción audiovisual; tanto es así que han emergido nuevas modalidades de traducción audiovisual (TAV) para que, sobre todo personas con discapacidad sensorial tengan acceso a los contenidos audiovisuales. Dedicaremos este capítulo a enmarcar nuestro estudio, la accesibilidad de los contenidos audiovisuales, en los EDT y a desarrollar un concepto fundamental en nuestro trabajo, el concepto de norma, entendido como una regularidad de la práctica traductora.

1.1. Los estudios de traducción

En 1972, Holmes propuso denominar Estudios de Traducción⁶ o *Translation Studies* a la disciplina encargada de explicar, describir y predecir los fenómenos relacionados con el campo de la traducción. El autor presentó su propuesta en la ponencia seminal «The Name and Nature of Translation Studies» y con ella perseguía dotar de un rigor científico a la disciplina, evitar las fragmentaciones y recoger todas las vías de investigación posible. Esta nueva taxonomía se diferenciaba de las propuestas hasta entonces por su carácter comprensivo e innovador; por primera vez el foco cambió de ángulo y en lugar de posicionar la traducción como parte dependiente de otra rama del saber, emergió con entidad propia y sin dependencias directas. La propuesta organiza la traducción a partir de dos ejes centrales; uno enfocado a los estudios de traducción puros en el que se incluyen una rama teórica y una descriptiva; y un segundo eje central que recoge los estudios de traducción aplicados. El gráfico que mostramos a continuación representa la amplitud de su propuesta:

⁶ Holmes propone el nombre de *Translation Studies* en la sección dedicada a la traducción del Tercer Congreso Internacional de Lingüística Aplicada celebrado en Dinamarca en 1972.

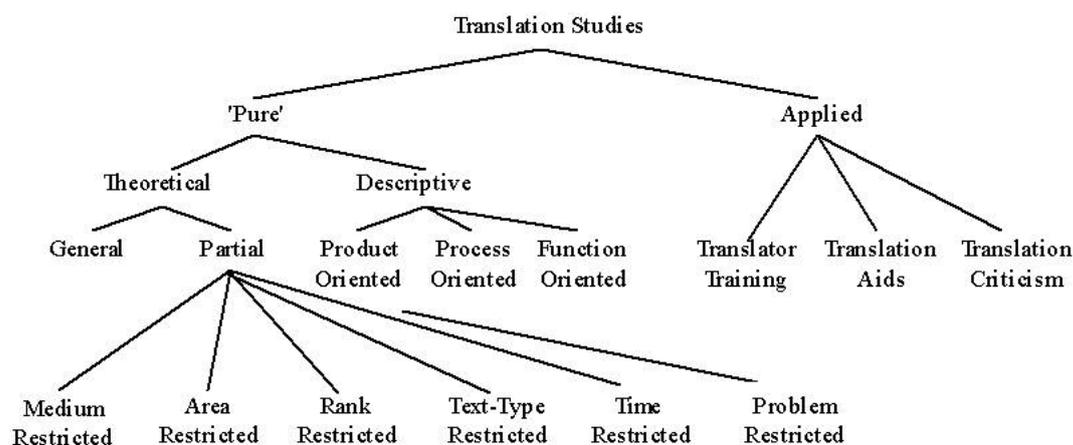


Figura 1. Estudios de Traducción: Holmes (1972).

Holmes definió la rama pura como una disciplina empírica y estableció los siguientes objetivos básicos (Holmes, 1988:71):

(1) to describe the phenomena of translating and translation(s) as they manifest themselves in the world of our experience, and (2) to establish general principles by means of which these phenomena can be explained and predicted.

La división de la rama pura en estudios descriptivos y estudios teóricos tiene su razón de ser en la consecución de los dos objetivos mencionados en la cita anterior respectivamente: describir el fenómeno de la traducción y establecer los principios generales de la disciplina. Estos objetivos básicos servirán a partir de entonces de guía a un gran número de investigadores en el ámbito de la traducción.

Por su parte, los estudios teóricos se dividen a su vez en estudios teóricos generales y parciales; los generales, dedicados a la búsqueda de una teoría general de la traducción que aglutine y recoja cualquier fenómeno relacionado con la traducción y el hecho traductor; y los parciales, que subdividió en: 1) teorías restringidas al medio; 2) teorías restringidas al área; 3) teorías restringidas al nivel; 4) teorías restringidas al tipo de texto; 5) teorías restringidas a la época; y por último 6) teorías restringidas a problemas específicos. El lazo que une las teorías parciales y las generales es bien visible, pues son muchos los estudios parciales necesarios para poder abordar una teoría general de la traducción siendo esta una muestra de la imbricación entre las ramas; rasgo inherente a la propuesta de Holmes.

Por otra parte, no hay que olvidar que el objetivo principal de la rama teórica es el de utilizar los resultados obtenidos en investigaciones de corte descriptivo y generar, a partir de estos, teorías basadas en datos empíricos que las hagan creíbles y objetivas. Y nuevamente esto refuerza la idea que comentábamos en el párrafo anterior, que no es otra que la estrecha

relación entre teorías parciales y generales y que pone de manifiesto el carácter integrador de la propuesta de Holmes. Como vemos, unas ramas beben de las otras, pero no hemos de olvidar que, como apunta el autor (2001:176):

It is perhaps appropriate to give first consideration to descriptive translation studies, as the branch of the discipline which constantly maintains the closest contact with the empirical phenomena under study.

A partir de los datos recogidos en estas investigaciones, la rama teórica podrá explicar y predecir el hecho traductor y la traducción.

Por su parte, los estudios descriptivos que nacen también de la rama pura, denominados por Holmes *Descriptive Translation Studies* (Estudios Descriptivos de Traducción, EDT), permiten estudiar la disciplina a partir de tres aproximaciones que, si bien están conectadas entre sí, ponen el foco en distintos factores propios de la traducción: 1) los estudios orientados hacia el producto, es decir, aquellos que describen y comparan los textos traducidos; 2) los estudios orientados hacia la función de la traducción en la cultura meta y 3) los estudios orientados al proceso, o lo que es lo mismo, aquellos relacionados con lo que pasa en la mente del traductor o en la verdadera *caja negra* (Király, 1995:42) durante el proceso de traducir.

Los estudios orientados al producto tienen su punto de partida en la descripción de traducciones individuales para posteriormente dar paso a una segunda etapa en la que se realiza un análisis comparativo de varias traducciones de un mismo texto. Estos materiales son la base para realizar investigaciones posteriores con corpus de traducciones más amplios que puedan aportar datos desde una perspectiva diacrónica o sincrónica. Por su parte, los estudios orientados hacia la función describen el papel del texto traducido en la situación sociocultural meta; son estudios centrados en el contexto, dejando el análisis del texto para otros enfoques. Estos estudios revelan, por ejemplo, qué tipo de textos se traducían en determinadas épocas y qué impacto producían en la cultura receptora. Por último, los estudios orientados hacia el proceso pretenden describir el complejo proceso mental que ocupa la mente del traductor. El impulso de este enfoque necesita del desarrollo de herramientas y métodos sofisticados que puede significar, en el futuro, el advenimiento de una nueva área de estudio denominada *Translation Psychology*, como bien señala Holmes (1988:72).

Además de la rama pura, con las subáreas teórica y descriptiva, Holmes postula también la existencia de una rama aplicada, *Applied Translation Studies*, que, por su parte, da cabida a trabajos de tipo principalmente prescriptivo cuyo objetivo es indicar cómo se ha de traducir, además de establecer recomendaciones que ayuden a mejorar la práctica de los

traductores profesionales y de los futuros traductores en formación. La primera aplicación de los Estudios de Traducción que señala Holmes es, como indica el autor, la misma que plantean generalmente el resto de disciplinas y que no es otra que la de la didáctica, pero matiza la diferencia existente entre los estudios aplicados a la traducción aprovechada como herramienta para la enseñanza de lenguas y la enseñanza de la traducción en sí, dado que, indiscutiblemente, estas aplicaciones tienen necesidades diferentes. Holmes presenta también estudios de tipo aplicado relacionados con las herramientas de ayuda tanto para la práctica profesional de los traductores como para estudiantes de traducción y aunque las demandas y necesidades son dispares, propone dos grandes grupos: herramientas de ayuda lexicográficas y terminológicas por un lado y gramáticas por otro. Por último, plantea la crítica de traducciones con el objetivo de reducir el sello intuitivo que la caracteriza y acercarla al análisis más objetivo. En resumen y como señala Hurtado (1996:152), la rama de los estudios aplicados «engloba las investigaciones en didáctica de la traducción, la traducción en didáctica de lenguas, la crítica y evaluación de traducciones y también la enseñanza de lenguas para traductores».

En los párrafos anteriores hemos pretendido revelar la fotografía del mapa de Holmes y mostrar que esta nueva configuración de los Estudios de Traducción permite explorar la traducción desde las diferentes ramas del saber científico. Por una parte, desde la rama teórica podemos explicar y predecir los fenómenos de la traducción, mientras que la rama aplicada es la plataforma indicada para la elaboración de materiales y herramientas de utilidad en el ejercicio de la traducción, y por último, la rama descriptiva recoge y regula los fenómenos empíricos estudiados. Pero sin duda, este recorrido nos permite plasmar el papel central de los Estudios Descriptivos de Traducción (EDT) como apuntan Rabadán y Fernández (2004:17):

«...el de la rama descriptiva es vital, ya que sobre ella descansan tanto la posibilidad de formular una teoría originada en la práctica de la traducción como las aplicaciones que de ella se pudieran derivar.»

Asimismo, queremos recoger aquí también la reflexión de Martí Ferriol (2006:47) quien considera que lo verdaderamente trascendente es que esta corriente, la de los EDT, marca un antes y un después en los Estudios de Traducción, pensamiento que compartimos plenamente. Y por otra parte, parece oportuno poner de relieve la imbricación de las tres ramas que recogemos en párrafos anteriores y que como apunta Hurtado (1999) es lógica y necesaria pues son los datos empíricos los que posibilitan la realización de estudios tanto aplicados como teóricos.

El trabajo que aquí presentamos se enmarca en la rama descriptiva, según Hurtado (1996) la rama que se ocupa de la investigación básica, en concreto en los estudios orientados al producto desde una perspectiva diacrónica, puesto que analizaremos un corpus de subtitulaciones publicadas entre 2009 y 2017 en la web del Parlamento Europeo. Seguiremos la metodología de los estudios descriptivos de traducción por lo que estableceremos los pasos necesarios para obtener, seleccionar y analizar los datos que pretendemos observar en un corpus de textos audiovisuales que describiremos en el capítulo dedicado a la metodología de compilación y preparación del corpus de estudio en el capítulo 9.

Por su relevancia y por ser el marco teórico en el que desarrollamos nuestro trabajo, en el siguiente apartado presentamos los EDT, su evolución y las contribuciones que distintos autores y traductólogos han hecho desde que comenzaran su andadura.

1.2. Escuela de la manipulación y el polisistema

A partir de la nueva denominación propuesta por Holmes para nuestra disciplina, Estudios de Traducción, y de la nueva estructura de esta, tuvo lugar una serie de conferencias en Lovaina (1976), Tel Aviv (1978) y Amberes (1980), que dieron lugar en 1985 y de la mano de Hermans a la publicación de *The Manipulation of Literature* en la que vieron la luz los trabajos realizados por un grupo heterogéneo de investigadores entre los que se encontraban: Bassnett, Van den Broeck, Hermans, Lambert, Lefevere y Toury, entre otros y que sirvió para dar nombre a un elenco de autores e investigadores que, desde la diferencia, compartían aspectos esenciales a la hora de abordar la investigación. La importancia de las aportaciones de unos y otros generó una diversidad terminológica, que en nuestra opinión, puede confundir y entorpecer la labor investigadora, para definir un mismo paradigma. El resultado de esta diversidad es que mientras unos hacen referencia a la escuela de la manipulación, otros hacen uso del término teoría del polisistema, la escuela de los Países Bajos, el eje israelo-neerlandés, etc. (Calzada Pérez, 2007:88). En cualquier caso, nosotros adoptaremos aquí la denominación de Escuela de la manipulación, no solo en recuerdo de un volumen que tuvo una influencia decisiva en el ámbito traductológico sino porque, además, dicha denominación apunta, ya desde 1985, al potencial explicativo de una metodología descriptiva y postula un enfoque más comprensivo que ensancha, en nuestra opinión, tanto el objeto de estudio como los fines que se persiguen en nuestro campo.

Por otra parte, es necesario destacar la figura de Itamar Even Zohar en alusión a la complejidad sistémica de la producción literaria y al decisivo estudio que este investigador

llevó a cabo por aquel entonces, a partir de la década de los setenta, y que dio en llamar la teoría del polisistema.

Esta aproximación tiene su origen en el formalismo ruso y el estructuralismo de Praga, cuya principal aportación fue la noción de *sistema* entendido como una estructura de varios niveles que aglutina elementos que se relacionan entre sí y que además se aplica tanto a obras y a géneros como a tradiciones literarias completas. De la pluma de Even-Zohar se aplicó esta idea al estudio de la literatura comparada, con el fin de explicar el dinamismo y heterogeneidad de un sistema. La contribución del autor (Even-Zohar, 1978, 1979, 1990, 1997, 2005) despertó el interés de una serie de teóricos de la traducción, entre los que destacó Gideon Toury, que se decidieron por describir los fenómenos que ocurrían en el seno de los polisistemas literarios.

Even-Zohar (1990:11-12) definió la noción de polisistema de la siguiente forma:

It is, therefore, very rarely a uni-system but is, necessarily, a polysystem--a multiple system, a system of various systems which intersect with each other and partly overlap, using concurrently different options, yet functioning as one structured whole, whose members are interdependent.

El autor concibe la literatura como un conjunto de sistemas en el que se establecen relaciones entre los elementos que la componen, y entre estos y el resto de sistemas que se encuentran en la sociedad. El sistema literario está a su vez constituido por otros subsistemas; un ejemplo de subsistema es la traducción, que además tiene entidad propia como apunta Even-Zohar (1990:46):

In other words, I conceive of translated literature not only as an integral system within any literary polysystem, but as a most active system within.

Asimismo, como mencionamos anteriormente, este sistema literario establece relaciones con otros sistemas que forman parte del polisistema sociocultural superior al que pertenece como pueden ser el sistema artístico o el sistema político.

Por su parte, Hermans (1985:11) define el polisistema como «a differentiated and dynamic conglomerate of systems characterized by internal oppositions and continual shifts» y son las relaciones de tensión que se generan dentro del polisistema las que lo hacen evolucionar de forma continuada.

En el artículo «The Position of Translated Literature Within the Literary Polisystem» que redactó Even-Zohar en 1978⁷ desarrolla su concepción de la traducción literaria como un

⁷ En Holmes, Lambert y Van den Broeck (1978:117-127)

subsistema del sistema literario de llegada en el que bien puede acoplarse o bien introducir elementos innovadores en el sistema. Si la literatura traducida ocupa una posición periférica, se construirá atendiendo a las normas de la literatura receptora, por el contrario, si ocupa una posición central contribuirá directamente en la creación innovadora de esa literatura e incluso se confundirá con esta; es en casos como este cuando la creación original y la creación traducida se abrazan hasta convertirse en una. En definitiva, y en relación a nuestro ámbito de estudio, la forma de traducción adoptada en una cultura dependerá de la posición que la literatura traducida ocupe en el polisistema. De la misma manera, como apunta Díaz Cintas (2001), al igual que hablamos de polisistema literario podemos hablar de un polisistema fílmico constituido por múltiples subsistemas. El autor propone cambiar el término *literatura* por el de *corpus fílmico* y aplica la teoría sistémica al ámbito audiovisual. De este modo, Díaz Cintas (2001:94-95) manifiesta que:

«...dentro del polisistema fílmico, la traducción de películas es simplemente un componente más que está en lucha constante con otros estratos o co-sistemas (obras en castellano) por conseguir la supremacía dentro del polisistema».

Even-Zohar observa el funcionamiento de la literatura traducida en distintas sociedades para posteriormente describir el sistema cultural jerárquico como un todo. Señala que la traducción será aceptable según las normas vigentes en el sistema cultural meta y entiende las normas como la regularidad en determinados comportamientos. Es Even-Zohar quien introduce el concepto de *norma* a comienzo de los setenta, concepto que desarrollará posteriormente Gideon Toury (1980, 1991) y que es la piedra angular de los EDT.

1.3. La escuela de la manipulación y los EDT

Hermans (1985:10-11) sintetiza el origen, finalidad y los objetivos principales de la escuela de la manipulación de la siguiente forma:

The group is not a school, but a geographically scattered collection of individuals with widely varying interests, who are, however, broadly in agreement on some basic assumptions – even if that agreement, too, is no more than relative, a common ground for discussion rather than a matter of doctrine. What they have in common is, briefly, a view of literature as a complex and dynamic system; a conviction that there should be a continual interplay between theoretical models and practical case studies; an approach to literary translation which is descriptive, target-oriented, functional and systemic; and an interest in the norms and constraints that govern the production and reception of translations, in the relation between translation and other types of text processing, and in the place and role of translations both within a given literature and in the interaction between literatures.

La descripción de Hermans, de carácter integrador, se apoya, como se muestra en la cita anterior, sobre una serie de premisas compartidas que el autor encabeza con una alusión a la complejidad sistémica que, por aquel entonces, ya había planteado con cierto detalle Itamar Even-Zohar con la teoría del polisistema que hemos introducido en el apartado anterior.

El carácter funcionalista es otra de las premisas apuntadas por Hermans dado que este grupo de investigadores centra su análisis en el estudio de las relaciones intertextuales y sobre todo intersistémicas (más que interlingüísticas) que se activan según el propósito para el que se ha concebido cada traducción.

Por lo demás, los integrantes de esta escuela inscriben su trabajo en la rama descriptiva, «an approach to literary translation which is descriptive» y, por tanto, describen la realidad sin pretender influir en ella. La finalidad, el objetivo del traductólogo, es describir y explicar los hechos a partir de la observación empírica. En apartados posteriores profundizaremos en este aspecto que consideramos de especial trascendencia por coincidir con el carácter de nuestro trabajo, que no es otro que el de servir como espejo de la práctica traductora, en concreto la práctica subtítuladora, en un contexto determinado como es el de las nuevas formas de comunicación en esta nueva sociedad europea. Los integrantes de esta escuela comparten un cuarto aspecto nuclear en su metodología investigadora: la orientación hacia la cultura meta. La posición predominante que, hasta el momento, ocupaba el texto origen en las investigaciones realizadas pasa ahora a ocupar un segundo plano y el texto traducido, y sobre todo el contexto en el que se le recibe, ocupan el papel protagonista.

Por último, Hermans apunta dos aspectos de los que participan los autores que forman parte de este mosaico académico que son: la naturaleza histórica, puesto que la función de una traducción en la cultura meta puede variar a lo largo del tiempo, y el carácter central del concepto de norma, aspecto que trataremos, como apuntábamos anteriormente, en el siguiente apartado puesto que el concepto de norma es, precisamente, una de las grandes aportaciones de la teoría del polisistema a los estudios descriptivos de traducción y a la escuela de la manipulación y piedra angular de nuestro análisis.

1.4. El concepto de norma

Son muchos los autores que han tratado el concepto de *norma*: Chesterman (1993, 1997), Nord (1991, 1998), Hermans (1991, 1999) y Rabadán (1991, 2002). Nosotros partiremos de la clasificación de las normas de Toury por ser la fuente de la que beben el resto de clasificaciones, además de por ser la más extendida. No obstante, aunque de forma

sucinta, repasaremos las propuestas de normas formuladas por otros autores en el ámbito de la traducción en general en un primer apartado y de la traducción audiovisual, en particular, en un segundo apartado. El repaso de estas propuestas nos permitirá establecer y explicar la metodología de estudio empleada en el análisis de nuestro corpus.

El concepto de norma se presentó en nuestro campo de estudio a partir de las publicaciones de Toury al final de los años setenta. El autor buscó definir en qué consistía la traducción a partir de generalidades que pudieran aplicarse en cualquier hecho traductor y lo concretó así (Toury, 1985:20):

...for the purpose of a descriptive study, a translation will be taken to be any target language utterance which is presented or regarded as a translation as such within the target culture, on whatever grounds...

La definición de Toury modificó el sentido del término «equivalencia» en el ámbito de la traducción y en cierta manera revolucionó los estudios de traducción. La equivalencia comenzó a entenderse como la relación existente entre dos textos en la medida en que se puede constatar que uno de ellos funciona como traducción del otro, en un sistema sociocultural determinado. Las investigaciones a partir de entonces se centrarían en describir qué tipo de relaciones se establecían entre un texto original y su traducción, y para determinar esas relaciones Toury incorporó la noción de norma, que adapta del campo de la sociología y psicología. Esta noción le permitió explicar las regularidades observadas en el comportamiento traductor.

Pero quizás antes de continuar ahondando en el concepto y naturaleza de las normas en traducción es necesario matizar qué es para Toury (1995:54) la traducción:

In its sociocultural dimension, translation can be described as subject to constraints of several types and varying degree...

Las constricciones o limitaciones socioculturales que apunta el autor forman un escala con dos extremos: en un polo se ubican las reglas generales y en el opuesto los rasgos idiosincrásicos. En el trayecto entre un polo y otro se ubican las normas aunque Toury introduce el siguiente matiz:

The norms themselves form a graded continuum along the scale: some are stronger, and hence more rule-like, others are weaker, and hence almost idiosyncratic. (ibid.)

Las normas ocupan, por tanto, una posición intermedia, pero no por eso una posición estática dado que están relacionadas con los cambios de estatus en la sociedad; aquello que nace como un rasgo idiosincrásico puede convertirse en un tipo de actuación cada vez más utilizada y acercarse así a la posición de regla general y al revés. En la medida en que una

norma exista como tal podremos detectar regularidades de comportamiento en situaciones repetidas del mismo tipo. A nuestro entender, los polos (idiosincrasia y reglas generales) son fáciles de detectar; podríamos hablar del polo blanco y del polo negro. Sin embargo, la gama de grises que une a estos extremos, y en la que se ubica la norma, es demasiado amplia y difusa y resultaría de gran utilidad establecer cuándo una regularidad puede ser entendida como norma.

En este mismo sentido, Martínez Sierra (2011:151) plantea dos cuestiones preliminares en torno al concepto de norma. Estas cuestiones nos resultan de gran utilidad a la hora de trazar nuestro camino hacia el concepto de norma y su delimitación en nuestro corpus. Por una parte, alude a la importancia de concretar el número de casos a partir del cual se puede empezar a hablar de que una regularidad sea considerada norma y, por otra parte, destaca la importancia de determinar y señalar cuál sería el estadio previo a la norma; es decir, identifica una tonalidad gris de entre esa gama de grises. Esa señal nos procura un alto en el camino, un punto de referencia y a su vez un nuevo punto de partida, pues puede establecer el lugar del que parten nuevas investigaciones. El autor, entiende que la norma nace de la decisión individual de un traductor de utilizar una estrategia de traducción (esto es la idiosincrasia de Toury) y pasa a ser entendida como norma cuando se aprecia que un traductor utiliza esa misma estrategia para resolver escollos similares regularmente, en este punto propone hablar de tendencia traductora, *regularities* en terminología de Pedersen (2011), y entiende que podremos hablar de norma cuando se constate que una determinada tendencia traductora se aprecia de forma recurrente.

La inclusión del concepto de tendencia traductora reduce la distancia entre la idiosincrasia y la norma, asimismo da sentido a aquellos estudios descriptivos que analizan dos o tres películas que, sin poder hablar de normas, sí arrojan información de gran utilidad para posibles estudios descriptivos a mayor escala que revelen las normas de traducción vigentes. El análisis descriptivo de obras audiovisuales utilizando la metodología del corpus puede bien ser un detector de normas de traducción de gran precisión. Del mismo modo, creemos que sería de gran interés establecer un punto intermedio entre la norma y el universal, como propone Pedersen (2011:30), quien en este sentido hace una propuesta muy interesante basada en las aportaciones de Toury (1995), Hermans (1999) y Chesterman (1997).

Pedersen (2011) reconoce que describir y descubrir normas es el objetivo fundamental de cualquier trabajo empírico desarrollado en el marco de los EDT y cita a Hermans (1999:79) para puntualizar que «in the descriptive paradigm norms provide the first level of abstraction and the first step towards explanation of the choices and decisions which translators make». La metodología de estudio que nos permite estudiar las normas está orientada al producto,

es decir, escrutamos el texto traducido para desvelar las normas que pudieran albergar. La metodología consiste, según Pedersen (2011:28), en seguir las fases siguientes:

... you take a text, look for regularities and patterns, try to generalize from your result, add more texts, verify your generalizations and if these are repeatable, you may have uncovered a norm.

Pero, ¿qué es una norma para Pedersen? Según el autor, las normas nacen de las decisiones de los traductores, sin embargo él establece una diferencia entre las normas prescriptivas y las descriptivas que entienden son:

Descriptive norms are the result of regularities of translational behaviour that have become so widespread and accepted that they have found an intersubjective existence. Prescriptive norms, on the other hand, are given to us by some authority. (2011:29)

En definitiva, las normas nos dicen qué hacen los traductores, cómo resuelven un problema determinado, por lo que las normas, en palabras del autor, representan la experiencia colectiva de los traductores, además de una valiosa herramienta para el aprendizaje de la traducción, así como una ayuda extraordinaria en el ejercicio profesional de la actividad.

Pedersen resume los términos que giran en torno al concepto de norma en el comportamiento traductor a partir de la figura que mostramos a continuación (2011:30).

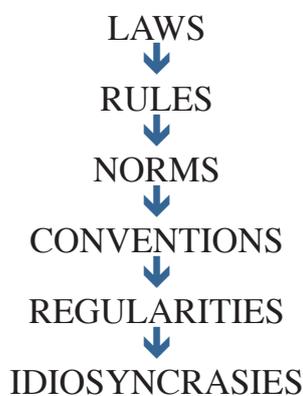


Figura 2. The potency of translation norms.

Según el autor, estos son los términos comúnmente más aceptados, aunque la diversidad terminológica genera que nos encontremos ante algún grado de solapamiento entre las categorías propuestas. Por esta razón, en ocasiones es complicado determinar en qué punto se pasa de una categoría a otra. Pedersen establece la diferencia entre idiosincrasia y regularidad de la siguiente forma (2011:31):

At the bottom of the cline, there are 'idiosyncrasies', which are fairly random, and are best used to describe the preferences of a single translator. The next level is called 'regularities' and these are recurrent patterns, which may or may not have been noticed

and shared by others. For the researcher, 'regularities' in the material are the first sign that there might be a norm at work.

Asimismo, el autor (*ibid.*) establece la diferencia entre convenciones y normas, y subraya que:

The main difference between norms and conventions are that while conventions tell you what people tend to do, norms tell you what you should do, because that is the way that people like you to do things.

Efectivamente, las normas ilustran lo que se espera que el traductor haga, pero como hemos comentado de forma reiterada, estas son también fenómenos dinámicos y en constante evolución, por lo que la historia de las normas es una historia de conflictos y luchas continuas en la que distintos comportamientos traductores se posicionan y recolocan una y otra vez. Hoy en día, los cambios constantes que acusa la profesión del traductor están relacionados con el desarrollo de las nuevas tecnologías y, como apunta Pedersen (2011:32): «When technology and societal values change, translation norms need to change with them.» Todo ello indica que, probablemente, estemos ante un momento de pugna entre normas.

La siguiente categoría que propone Pedersen es la regla y la identifica como «codified and authoritarian». En esta categoría se enmarcan las reglas gramaticales que son la regla de traducción más obvia e identificable. En el caso de la subtitulación pone el ejemplo del espacio que se deja entre el guion de intervención de un personaje y la palabra que aparece a continuación. Es una regla que a nadie se le ocurriría saltarse establecida por una entidad superior y que el traductor no puede cuestionar.

El último escalafón, o el primero, según se lea la figura 2 anterior sobre la potencia o fuerza de las normas de traducción, lo ocupan las leyes o universales de traducción, en función de la terminología del autor. En opinión de Pedersen, las leyes son de carácter universal, no son impuestas y sólo se revelan a través de la observación de los traductólogos. Las define como predictivas y probabilísticas y tienen la siguiente estructura: «if X, then the greater/the lesser the likelihood that Y» (Toury, 1995:265). A su vez, Pedersen nos recuerda que fue precisamente Toury quien formuló por primera vez este concepto y lo sitúa como el fin último de los EDT.

El análisis basado en corpus es de gran utilidad para describir las características de la lengua, así como para acometer un análisis descriptivo de las normas que gobiernan la traducción y como apunta Díaz Cintas (2005:14):

En lugar de lanzar ideas abstractas sin datos empíricos que las fundamenten, o de inventar ejemplos ad hoc para ilustrar un determinado punto que conviene al investigador de

turno, lo que ahora se sugiere es llevar a cabo un mapping de lo que verdaderamente ocurre cuando se traduce, para evitar caer en la teorización más absoluta. Sólo a partir de ejemplos reales, que existen y se han visto integrados en la sociedad de acogida, podemos extraer conclusiones que nos hagan avanzar en el conocimiento de la traducción. Y las normas son, precisamente, los útiles que nos ayudan en esta tarea.

Nuestro trabajo, recoge la reflexión del autor y pretende describir las normas en el ámbito de la TAV. Nos basamos en la metodología de la lingüística de corpus (que abordaremos en el capítulo 8) con el objetivo de mapear lo que sucede en la práctica traductora audiovisual en el entorno web. Por tanto, en el siguiente apartado presentaremos de forma sucinta el concepto de norma en TAV.

1.5. Las normas en traducción audiovisual

El ámbito de la traducción audiovisual, en particular, tiende a estar dominado por normas (Karamitroglou, 2000:14) y coincidimos con Martínez Sierra (2011:164) cuando afirma que existe la posibilidad de encontrar normas en cualquier modalidad de traducción audiovisual. En este sentido Delabastita (1989:205) apunta lo siguiente:

The translation of film, then, constitutes a typical situation where one can expect norms to guide the selection of actual behaviour in each specific historical set of circumstances. As a result of such norms the different possible modes of translation will not all be 'in free variation', as some of them will in the given conditions implicitly or explicitly be judged as 'better' solutions than the others.

De hecho, son ya muchos los autores que se han ocupado del estudio de las normas en traducción audiovisual, y con el fin de establecer qué normas son susceptibles de aparecer en esta especialidad de traducción, a continuación, nos referiremos a alguno de sus trabajos.

La investigación descriptiva en traducción audiovisual ha generado distintas propuestas de normas que se aprecian en las modalidades propias de este ámbito. Destacan las aportaciones de Goris (1993) por su transcendencia y por ser un pionero en su género; la propuesta de Ballester (2001) por ser de los primeros análisis descriptivos de normas en traducción audiovisual (doblaje y subtítulo) realizados en España y al que siguió por ejemplo la tesis de Martínez Sierra (2004) y Martí Ferriol (2006); la aportación de Karamitroglou (2000), para nosotros muy significativa pues su modelo de investigación de normas en traducción audiovisual incluye aspectos que hasta entonces no se habían incluido en el estudio de las normas en TAV.

Con todo y con eso, es Chaume (2003, 2004) quien realiza una revisión exhaustiva de las aportaciones traductológicas más importantes en el ámbito de la traducción audiovisual.

El grupo de estudios descriptivos con base polisistémica (es decir, aquellos que centran su atención en el impacto cultural que para la cultura meta tiene que los espectadores consuman nuevos formatos) lo divide en siete subgrupos. El que nos interesa a nosotros y en el que se enmarca nuestro trabajo es el que el autor denomina «estudios plenamente descriptivistas sobre las normas de traducción». Según Chaume, los autores que han centrado su labor en este ámbito son Goris (1993), Sokoli (2000), Karamitroglou (1998, 2000), Martí Ferriol (2003) y Roussou (2003). Posteriormente, la tesis de Martínez Sierra (2004) también tendría cabida en este grupo, como el mismo autor indica en su trabajo, si bien advierte que tanto en su investigación como en la de otros autores se pueden entrecruzar dos o más líneas de las planteadas por Chaume. Lo mismo ocurre con la tesis presentada por Martí Ferriol (2006), así como el estudio de las normas en televisión de Pedersen (2011). Este último trabajo es para nosotros de vital importancia por su innovadora propuesta del estudio de las normas técnicas en TAV y dedicamos el próximo apartado a explicar su aportación.

1.5.1. Estudio de las normas en traducción audiovisual de Pedersen.

La aportación de Pedersen (2011) se presenta como la más novedosa y enriquecedora para nuestro estudio dado que el autor propone un análisis de las normas en subtitulación a partir de una gran cantidad de textos mediante la combinación de estudios descriptivos y estudios basados en corpus. El objetivo del estudio es poner al descubierto las normas de traducción en subtitulación en la televisión en Dinamarca y extrapolar los resultados para compararlos con el panorama internacional. Pedersen investiga las normas técnicas, las estrategias de traducción utilizadas en la traducción de los referentes culturales y por último formula una serie de normas a partir de los resultados obtenidos.

El estudio que el autor realiza de las normas técnicas de subtitulación, nos sirve de inspiración en nuestro estudio, puesto que como señala el autor (2011:130):

The technical aspects are thus a facet of subtitled translations that cannot be disregarded, and before we move on to a more linguistic investigation into the norms of subtitling, we will thus have to make an investigation into the technical norms of subtitling.

En su análisis incluye tres parámetros: velocidad de lectura, densidad de los subtítulos y condensación. Nuestro estudio, sin embargo, analiza la velocidad de lectura y otros parámetros vinculados a la velocidad de lectura, así como otras cuestiones técnicas que están relacionadas con los aspectos especiales y temporales que definen la práctica de la subtitulación. Incluimos, además, el estudio de la segmentación de los subtítulos, que como apuntan diversos estudios, está relacionada con la velocidad a la que leemos y asimilamos

el texto escrito en el subtítulo. En el capítulo 7 definimos qué parámetros analizamos en esta tesis y ponemos de relieve la importancia de estos aspectos técnicos en la subtitulación.

Es importante recordar que no han sido pocos los esfuerzos realizados para intentar unificar los criterios que dominan esta práctica, pero la realidad se impone y estamos lejos de estandarizar esta modalidad de traducción audiovisual que se consume de forma global. Por ejemplo, La ESIST⁸ propone un código de buenas prácticas⁹ en subtitulación, redactado por Jan Ivarsson y Mary Carroll (1998) con la intención de que sea de aplicación en el mayor número de países posible; sin embargo, el código es ambiguo en ciertos aspectos y eso genera discordancias relevantes. Asimismo, Karamitrouglou (1998) en su artículo «A proposed set of subtitling standards in Europe» plantea una serie de pautas que permitan unificar criterios en los países de la Unión Europea para, de esta forma, cohesionar la práctica subtituladora (en las cuestiones conocidas como técnicas) en Europa:

Its aim is to provide a unifying formula based on thorough scientific research that could bridge the different subtitling conventions currently operating within the various European countries, to cater for the needs of the individual European viewer and address the European audiovisual audience market as a whole.

Estas propuestas junto con otras guías, recomendaciones y manuales de subtitulación que veremos en el capítulo 6 son de gran importancia en nuestro estudio pues nos permitirán señalar los puntos que analizaremos y describir en qué medida los subtítulos que componen nuestro Archivo cumplen con las indicaciones, pautas o recomendaciones señaladas en estas guías y que son referentes en la práctica subtituladora en Europa.

Los estudios descriptivos de traducción audiovisual, como señalamos en párrafos anteriores, nos permiten identificar normas de diversos tipos, pero también nos sirven para definir nuevas prácticas en el ámbito de la traducción audiovisual. Estas nuevas prácticas o modalidades que han surgido en el ámbito de la traducción audiovisual (subtitulación para personas sordas, audiodescripción para personas ciegas, etc.) han dado lugar a un nuevo concepto, el de accesibilidad audiovisual. Dicho concepto puede recoger a nuestro entender, todas las modalidades de traducción audiovisual puesto que todas ellas facilitan el acceso a la información del contenido audiovisual, con independencia de que la barrera sea lingüística o relativa a otros aspectos vinculados con la comunicación.

Por su parte, el desarrollo de las nuevas tecnologías ha redefinido la sociedad actual y entre otros cambios sustanciales encontramos que han surgido nuevas formas de acceder a los

⁸ ESIST, European Association for Studies in Screen Translation (www.esist.org)

⁹ The Code of Good Subtitling Practice se publicó originalmente en Ivarsson, Jan & Carroll, Mary. 1998. *Subtitling*. Simrishamn: TransEdit. Puede consultarse y descargarse en línea en (<http://www.esist.org/code-of-good-subtitling-practice/>)

contenidos. De entre ellos, destaca Internet, la forma más utilizada a día de hoy. Las mejoras tecnológicas también han procurado grandes avances en la metodología de lingüística de corpus (LC), puesto que las herramientas de análisis y compilación de corpus nos permiten estudiar una gran cantidad de textos de forma automatizada. Estos dos factores, Internet y los avances en las herramientas utilizadas en LC, nos han permitido estudiar las normas de traducción que se dan en Internet en el ámbito institucional de forma muy exhaustiva.

Este capítulo versa sobre los EDT y su concepto estrella, la norma. La norma nos ayuda a describir el ámbito de la traducción audiovisual y además nos permite descubrir nuevas prácticas en su seno, por ejemplo, surtitulado, subtitulación bilingüe, subtitulación para personas sordas, audiodescripción. Todo ello puede abarcarse bajo el paraguas de la accesibilidad audiovisual. Preferimos este concepto porque el concepto de accesibilidad audiovisual permite abarcar más espacio que la denominación de traducción audiovisual. La accesibilidad audiovisual nos permite poner el foco en la esencia de la traducción, es decir, en trasladar un mensaje/texto a una audiencia meta que por distintas razones (idioma, discapacidad sensorial, imposibilidad ambiental, etc.) no puede acceder a ella. Y desde este punto de partida, se ponen en práctica distintas técnicas (modalidades de traducción audiovisual) que consiguen que el producto audiovisual sea accesible para el mayor número de personas posibles. Por todo ello preferimos este concepto y este es el motivo por el cual entendemos que la presente tesis doctoral se inserta en este ámbito más amplio e integrador.

Hemos visto cómo el concepto de norma ocupa también una posición central en la accesibilidad audiovisual. De hecho, contamos con normas prescriptivas (guías y recomendaciones) desde distintos ámbitos y geografías que nos ayudarán a mapear la práctica subtituladora en distintos canales (TV, internet, etc.) e incluso nos permitirá identificar cambios que se han producido, con el paso del tiempo, en las normas predominantes. Por otra parte, las normas también nos ayudan a identificar nuevas prácticas (que forman ya parte de lo que se conoce como accesibilidad).

El apartado que sigue es precisamente sobre estas nuevas prácticas que se pueden considerar dentro del ámbito de la accesibilidad de modo que expondremos los aspectos más relevantes en torno al ámbito de la traducción audiovisual, con un énfasis especial en aquellos puntos relacionados con la accesibilidad.

CAPÍTULO 2

EL ÁMBITO DE LA TRADUCCIÓN AUDIOVISUAL

- 2.1. La traducción audiovisual y la accesibilidad
- 2.2. Traducción audiovisual y estudios de traducción
- 2.3. La traducción audiovisual: un debate terminológico
- 2.4. La traducción audiovisual: definición y modalidades
- 2.5. Clasificaciones de modalidades de traducción audiovisual
- 2.6. Modalidades de traducción audiovisual en internet
 - 2.6.1. La audiodescripción para personas ciegas
 - 2.6.1.1. Tipos de audiodescripción y campos de aplicación
 - 2.6.2. La subtitulación y la subtitulación para personas sordas y con discapacidad auditiva (SPS)
 - 2.6.2.1. Tipos de subtítulos y campos de aplicación

CAPÍTULO 2. EL ÁMBITO DE LA TRADUCCIÓN AUDIOVISUAL

A lo largo del presente capítulo definiremos la traducción audiovisual y la evolución que ha experimentado este campo en el marco de los Estudios de Traducción. Uno de nuestros objetivos es explicar por qué entendemos que la TAV es una forma de accesibilidad, concepto que ocupa una posición central en nuestra investigación y desde el que partimos para plasmar nuestra forma de entender la traducción en general y la traducción audiovisual en particular.

Del mismo modo, expondremos las distintas modalidades que conforman la TAV, desde las modalidades más tradicionales hasta las emergentes. Por ser el objeto central de nuestro estudio, analizaremos en mayor profundidad la subtítulos convencional y la subtítulos para sordos (SPS) y señalaremos las diferencias entre ambas.

2.1. La traducción audiovisual y la accesibilidad

Desde una aproximación dinámica y en consonancia con el objetivo de Delabastita (1989:214) de superar el enfrentamiento terminológico entre traducción y adaptación, el cual abordaremos en un apartado más adelante, Mayoral (2001c:46) propone que:

La definición del objeto de estudio de la traducción no es la definición de un proceso natural al que se le presupone inalterabilidad sino la definición de un proceso tecnológico que se encuentra en continua evolución y cambio. Nuestra función no es cerrar el paso a las nuevas realidades sino favorecerlas e impulsarlas. Necesitamos por tanto definiciones abiertas, que sea posible modificar tanto para contener las nuevas realidades (interpretación de lengua de signos, multimedia, producción de textos) como para desechar las que han dejado de mostrarse útiles y necesarias.

Este enfoque, al cual hemos apelado anteriormente, da cabida a un nuevo concepto de gran relevancia en traducción audiovisual que no es otro que el de la *accesibilidad*. A nuestro entender, la traducción audiovisual se asienta en el binomio traducción y accesibilidad, siendo este segundo término intrínseco a la actividad que nos ocupa. Tanto es así que, como afirma Díaz Cintas (2009:5), «Accessibility is a new key concept; an umbrella term that encompasses all associated new modes of translation.». No es baladí, por tanto, subrayar

que, en consonancia con la afirmación del autor, el término *accesibilidad* engloba y entrelaza varias modalidades de traducción. En nuestro estudio restringimos el término *accesibilidad* a la traducción audiovisual por lo que en el presente trabajo haremos referencia a la accesibilidad a los medios audiovisuales. En este sentido Chica Núñez y Rodríguez (2014:40) exponen lo siguiente:

Cuando la traducción tiene como objetivo la transmisión del conocimiento y, por lo tanto, la comunicación, a personas con diversidad funcional sensorial (visual o auditiva) se denomina traducción accesible. Dentro de ella se incluyen diferentes modalidades: la Interpretación en Lengua de Signos Española (ILSE), la Subtitulación para personas Sordas (SPS) y la Audiodescripción (AD).

En este mismo sentido, como afirma Díaz Cintas (2009), el concepto de accesibilidad ha entrado a formar parte de la TAV y como apunta Orero (2005:183) citando a Neves (2005):

AVT will inevitably need to follow the general trend in the audiovisual market and, rather than aim to cater for a general audience, audiovisual translation now finds itself focusing on the needs of smaller distinct audiences in order to respond to them in a more adequate manner.

Como vemos, la audiencia y las necesidades de grupos minoritarios han de tenerse en cuenta y además se convierten en la puerta de entrada de nuevas modalidades de traducción. El mercado audiovisual y los nuevos canales de comunicación están regulados para evitar la posible exclusión de grupos minoritarios por lo que la accesibilidad en TAV habita en el presente, pero cierto es que se trata de un presente en constante cambio a causa de la rapidez de los avances en el sector tecnológico; avances que, lógicamente, tienen impacto en el desarrollo y en la evolución de la disciplina. En este sentido, Remael y Neves (2007:4) señalan:

In the 21st century, AVT has grown exponentially and diversified to such an extent that it has exploded what were once considered to be its constraining technical features, its media-specific characteristics, its Euro-centric borders and its classic, interlingual translational and textual features. Indeed, AVT is mingling with multimedia translation and localisation, and increasingly determining our access to information and entertainment.

Las autoras indican que uno de los cambios más importantes se ha producido en el concepto mismo de la TAV. La transformación viene determinada por el cambio de rumbo al entender la TAV como una forma de accesibilidad: «the fairly recent trend of defining audiovisual translation as a form of media accessibility» (*ibid.*), siendo esta la misma línea que recorren diversos autores (Díaz Cintas, Orero, Remael, 2007:13) en traducción audiovisual.

Accessibility is a form of translation and translation is a form of accessibility, uniting all population groups and ensuring that all cultural events, in the broadest sense of the word, can be enjoyed by all.

Entendemos que la esencia de la traducción es trasladar una pieza de información a las personas que quieren acceder a ella. En este sentido, Díaz Cintas (2007:20) apunta:

Independientemente de que el obstáculo comunicativo sea de carácter lingüístico o sensorial, el objetivo del proceso traductor es exactamente el mismo: facilitar el acceso a una fuente de información y entretenimiento que de otra forma sería relativamente hermética. Desde esta perspectiva, la accesibilidad se convierte en un denominador común que subyace a todas las prácticas aquí mencionadas.

Entendido de esta manera, el autor propone ampliar el ámbito de acción del concepto de accesibilidad en traducción y, por tanto, en traducción audiovisual y argumenta que, en esencia, el doblaje o el subtítulo de un contenido audiovisual tienen tanto de accesibilidad como el subtítulo para personas sordas (SPS), la audiodescripción (AD) o la interpretación en lengua de signos (ILS). Díaz Cintas (*ibid.*) subraya que lo único que difiere son los perfiles de las audiencias. De hecho, si tenemos en cuenta que la mayor parte del contenido en internet está en inglés, la barrera más importante sería el idioma. El enfoque que propone sobre accesibilidad, con el que coincidimos plenamente, es quizás el más integrador y el que entendemos más adecuado para emprender el camino hacia el Diseño para Todos en el ámbito de la TAV.

Pero, ¿dónde ubicamos esta nueva forma de entender la traducción audiovisual en los Estudios de Traducción? El siguiente apartado lo dedicaremos a enmarcar de forma teórica esta nueva forma de entender la traducción audiovisual o la accesibilidad audiovisual.

2.2. Traducción audiovisual y estudios de traducción

A nadie se le escapa que los avances tecnológicos han reforzado el vínculo inherente entre traducción audiovisual y tecnología, y esto ha provocado, a su vez, que el concepto de traducción se haya de flexibilizar para dar respuesta a las nuevas necesidades. A este tenor, las nuevas formas y usos de la información y las necesidades que de ello se derivan han generado nuevas prácticas profesionales que entendemos se encuadran en el ámbito de estudio de la traducción audiovisual (TAV) y que pasamos a enmarcar desde una perspectiva teórica en los Estudios de Traducción.

Tomaremos prestado, para ubicar la TAV en los Estudios de Traducción, el enfoque de Díaz Cintas (2007:15) quien, por medio de la concepción que aportó Jakobson de traducción

(noción que, a su vez, este amplió a todo tipo de comunicación), posiciona perfectamente la traducción audiovisual en toda la amplitud que hoy requiere este tipo de traducción. Fue en 1959 cuando Jakobson propuso su taxonomía traductora en la que encontramos tres tipos de traducción: interlingüística, intralingüística e intersemiótica. Estos tres tipos de traducción fueron definidos por este autor (1959 [2000]:114) como mostramos a continuación.

- Intralingual translation or *rewording* is an interpretation of verbal signs by means of other signs of the same language.
- Interlingual translation or translation proper is an interpretation of verbal signs by means of some other language.
- Intersemiotic translation or *transmutation* is an interpretation of verbal signs by means of nonverbal sign systems.

En un principio, la interlingüística era el único tipo de traducción que se consideraba traducción en sí, pero como hemos anunciado anteriormente, los avances tecnológicos y la incursión del concepto de accesibilidad en la definición de traducción han forzado que, tanto teóricos como estudiosos de la traducción ensanchen el concepto. Sin embargo, no hay que olvidar que todavía hoy encontramos voces en contra de considerar algunas nuevas modalidades como parte de la traducción (Martínez Sierra, 2004:16):

Según concebimos el asunto, quisiéramos dejar patente nuestra reticencia a considerar el tipo intralingüístico como actividad traductora, puesto que entendemos que ésta ha de implicar, como mínimo, dos lenguas y dos sistemas culturales distintos en mayor o menor grado.

A nuestro entender, la taxonomía de Jakobson sí da cabida a todas las modalidades de traducción que conocemos hasta la fecha, incluidas las de TAV junto con las modalidades emergentes: la subtitulación para personas sordas (SPS), la audiodescripción para personas ciegas (AD), etc. Como se puede intuir, la SPS forma parte, sin lugar a dudas, del primer tipo (intralingüística) propuesto por Jakobson; mientras que la AD encaja, a nuestro parecer, en la tercera clasificación al ser un tipo de traducción intersemiótica, pero invirtiendo el orden: es una interpretación de los signos no verbales por medio de signos verbales que tiene sus orígenes en la *hipotiposis*, una figura retórica ya descrita en los tratados griegos. También presente en la tradición clásica encontramos el término *écfrasis* (Eco, 2003:110) para referirse a «cuando un texto verbal describe una obra de arte visual».

Por su parte, Steiner (1975/1980:65), al igual que Jakobson, amplía los límites de la traducción afirmando que «cualquier modelo de comunicación es al mismo tiempo un modelo de traslado, de transferencia vertical u horizontal de significado».

Esta forma menos convencional de entender la traducción, y que torna su significado hacia el concepto más moderno de traducción que encontramos actualmente, también se refleja en la afirmación de Nida (1959:429) cuando señala que hay que lograr la misma respuesta en los receptores meta del texto traducido que la que se produce en los receptores del texto original y esto es, como afirma Gala Rodríguez (2007:93), lo que procura, por ejemplo, la audiodescripción:

En la audiodescripción el objetivo es similar: apoyar la trama de la película (TO) mediante descripciones (TM) para conseguir el mismo efecto en la persona ciega (receptor del TM) que el que suscitó en la que vio la película (receptores TO).

Por otra parte, con el objetivo de hacerse eco de la realidad del momento, la propuesta tipológico-textual de Reiss (1977 y 1981) que tenía además como finalidad dar cabida a las actividades traductorales marginales en aquella época, resultó ser uno de los primeros pasos para dotar a la traducción audiovisual de entidad teórica, aunque la autora no menciona la traducción audiovisual explícitamente en su propuesta. Reiss distingue tres tipos de textos (informativos, expresivos y operativos) y un tipo adicional que entiende como una estructura superior que denomina *audio-medial text type*, entre los que apunta se encuentran «songs, comic strips, advertisements, medieval morality ballads, etc.» (1977:11). Es evidente, que la traducción audiovisual tiene cabida en este marco teórico, como señala Díaz Cintas (2003:34).

Las distintas aproximaciones que hemos descrito evidencian que la traducción audiovisual, tal y como la entendemos en el presente trabajo, no solo tiene cabida en los Estudios de Traducción, sino que ocupa, a nuestro entender, una posición cada vez más central y menos periférica. A nadie se le escapa que, como plantea Díaz Cintas (2009:1), la traducción audiovisual es una necesidad de la sociedad actual. Los medios de comunicación hacen uso de su potencial para distintos fines «to inform, arguably sometimes to misinform, to entertain, to sell and to educate», pero sobre todo para entretener, pero no solo los medios de comunicación, sino también y casi nos atreveríamos a decir, sobre todo, internet.

Por otra parte, por todos es sabido que los Estudios de Traducción han incorporado, dentro de la modalidad de traducción audiovisual, la accesibilidad y la formación de alumnos en SPS y AD. En este sentido Lorenzo y Pereira (2012:114) apuntan que:

... en la universidad, los estudiosos que ejercían su labor docente e investigadora en traducción audiovisual supieron ver la necesidad de formar traductores audiovisuales que incorporasen a sus conocimientos saber hacer accesibles a los receptores sordos los materiales con los que trabajaban.

El interés de las universidades españolas al igual que el de universidades europeas por incluir la accesibilidad audiovisual en los Grados y Postgrados en Traducción es evidente. Proyectos de investigación europeos financiados por la Unión Europea, como por ejemplo el proyecto de investigación liderado por Pilar Orero —*Understanding Quality in Media Accessibility*¹⁰— o el proyecto europeo MOOCAP —*A Pan-European Project on providing education on accesible design*¹¹— son un claro ejemplo de ello.

En este mismo sentido hay universidades que ofrecen programas de expertos en accesibilidad web en los que se incluye formación en subtitulación y audiodescripción como es el caso de la Universidad Carlos III de Madrid¹². Sería quizás recomendable que los cursos relacionados con la accesibilidad audiovisual que nacen desde los Estudios en Traducción incluyeran módulos de accesibilidad web para tener una formación más completa del ámbito de trabajo de los futuros profesionales o bien que expertos en traducción audiovisual y accesibilidad formaran parte del claustro que imparte estos títulos. Es decir, entendemos que un trabajo conjunto sería muy enriquecedor y satisfactorio para todas las partes.

Como hemos podido comprobar, la TAV y la accesibilidad están inexorablemente vinculadas, si bien es cierto que el camino recorrido es ya muy largo aún quedan cuestiones por resolver para que la accesibilidad audiovisual ocupe el lugar que le corresponde en los Estudios de Traducción. Hace poco más de una década Díaz Cintas (2007:20) se refería a la accesibilidad en TAV en estos términos:

[...] lo que nos resta por hacer es consolidar su integración en los departamentos universitarios de traducción e interpretación y promover y difundir las virtudes de la accesibilidad más allá de las fronteras de nuestra disciplina. Y es ahí precisamente donde reside uno de nuestros mayores desafíos. Por el momento se trata de un área con mínima proyección social, de la que muchos espectadores ni siquiera han oído hablar, nos guste o no.

Entendemos que estamos a menos de un paso de consolidar estas nuevas modalidades de traducción y quizás lo que resta por hacer es trasladar a la ciudadanía la existencia de las barreras de la comunicación. Visibilizar las barreras y las técnicas de traducción audiovisual

¹⁰ Proyecto *Understanding Quality in Media Accessibility* <http://pagines.uab.cat/umaq/>

¹¹ Proyecto MOOCAP *A Pan-European Project on providing education on accesible design* <http://gpii.eu/moocap/>

¹² http://portal.uc3m.es/portal/page/portal/postgrado_mast_doct/cursos_especializacion/Experto_Tecnologias_%20Accesibilidad_Web/Programa

(SPS y AD) que facilitan el acceso a los contenidos audiovisuales a las personas con discapacidad sensorial ha de ser objetivo prioritario.

Las causas más difíciles son aquellas que no se visibilizan; por eso quizás el gran reto de la accesibilidad audiovisual sea trabajar en la sensibilización social. A cualquiera le indigna que una persona en silla de ruedas no pueda acceder a un espacio físico; sin embargo, la reacción no es la misma cuando una persona con discapacidad auditiva quiere acceder (por derecho) a un contenido audiovisual y no puede. Introducir el concepto de accesibilidad audiovisual no hace más que reavivar el debate terminológico histórico en el que navega esta disciplina y consideramos importante ofrecer, en el siguiente apartado, un repaso sucinto a esta controversia.

2.3. La traducción audiovisual: un debate terminológico

Es por todos conocida la diversidad terminológica que acompaña a esta modalidad de traducción desde sus comienzos, pero no debemos olvidar que, como apunta Mayoral (2001:20):

Los estudios de traducción audiovisual se originaron en el estudio de la traducción cinematográfica. La incorporación de la traducción para televisión y vídeo llevó a la introducción de la denominación traducción audiovisual.

Sin embargo, antes de llegar a esta denominación más extendida, son muchos los términos utilizados para referirse a la traducción audiovisual. Como apuntan Díaz Cintas y Remael (2007:11): «The terminology used is very hesitant and can be confusing», pero según los autores, a partir de los años 90, en círculos académicos parece imponerse el término *audiovisual translation* (AVT) (*op. cit.*:12) «this coinage has the advantage of including the semiotic dimension and has taken root to such an extent that today its use is quite frequent». Ya en Agost (1996, 1999, entre otros) queda constancia del uso habitual del término *traducción audiovisual* (TAV).

Por su parte, Chaume (2004:30) recopila los diferentes términos que se utilizan para referirse a esta actividad: *film dubbing* (Fodor, 1976), *constrained translation* (Titford, 1982), *film translation* (Snell-Hornby, 1988), *film and tv translation* (Delabastita, 1989), *screen translation* (Mason, 1989), *media translation* (Eguíluz *et al.*, 1994), comunicación cinematográfica (Lecuona, 1994), traducción cinematográfica (Hurtado, 1994), traducción fílmica (Díaz Cintas, 1997), *multimedia translation* (Mateo, 1997). Los términos *constrained*

translation y *multimedia translation* son quizás los que más debate han suscitado, pero en este trabajo no entraremos en ese debate.

Sin embargo, la justificación del término traducción audiovisual con la que coincidimos plenamente es la que ofrece Orero (2004) en la presentación del libro *Topics in Audiovisual Translation*. En primer lugar, la autora destaca la importancia de consensuar un término para situar esta modalidad de traducción a la misma altura que el resto de modalidades y dotarla así de entidad propia para más tarde justificar su elección de la siguiente forma (2004:vii):

The title of the book has clearly opted for Audiovisual Translation since Screen Translation would leave out for example translations made for theatre or radio, and the term multimedia is widely perceived as related to the field of IT. Audiovisual Translation will encompass all translations –or multisemiotic transfer– for production or postproduction in any media or format, and also the new areas of media accessibility: subtitling for the deaf and the hard or hearing and audiodescription for the blind and the visually impaired.

Por su parte, como afirma Díaz Cintas (2007:13) en consonancia con la propuesta de Orero (*ibid.*) la acuñación de *traducción audiovisual* es mucho más común en español, convirtiéndose en el referente más usado, además de tener la ventaja de incluir la dimensión semiótica en su conceptualización y de establecer una relación directa con el concepto de accesibilidad con el que comenzábamos este apartado.

A modo de conclusión, resaltamos que, por todo lo expuesto anteriormente, en este trabajo utilizaremos el término *traducción audiovisual* como un término que da cabida a las diferentes prácticas traductoras que se implementan en los contenidos audiovisuales y que se convierten en vía de accesibilidad diversas de dichos contenidos. A continuación, pasamos a definir la traducción audiovisual y a introducir las clasificaciones de TAV más relevantes.

2.4. La traducción audiovisual: definición y modalidades

Las modalidades que configuran la TAV se caracterizan por tener que lidiar con las limitaciones espaciales y temporales del medio. Dichas limitaciones tienen un impacto en el trasvase lingüístico por lo que algunos estudiosos como Vinay y Darbelnet (1963) y García Yebra (1982) prefieren hablar de adaptación en lugar de traducción. Por otra parte, autores como Delabastita (1989, 1990), Lambert (1990) y Gambier (1996, 1998, 2003) han dedicado muchos esfuerzos a resaltar la importancia de la traducción en los medios de comunicación para reforzar su lugar en nuestra disciplina y dejar de lado el debate de si la traducción audiovisual es un caso de traducción o de adaptación, en el que nosotros no entraremos en este trabajo.

Pero, antes de adentrarnos en las clasificaciones de TAV propuestas hasta el momento y en las posibles modalidades que la conforman, es preciso indicar los rasgos definitorios de este campo. Mayoral apunta cuatro rasgos definitorios de la TAV (2001) para señalar qué es específico de la misma, aclarando previamente que al referirse a TAV (Mayoral, 1997) incluye no sólo «los productos transmitidos a través de cine, vídeo, DVD y televisión sino también los productos multimedia que reciben difusión a través de los ordenadores o, desde hace poco tiempo, de las consolas de videojuegos». No perdemos de vista que la flexibilidad necesaria para describir la TAV es tan importante como las características que identifican este ámbito y que según el autor (2001b:34-36) son cuatro:

La comunicación en el caso de la traducción audiovisual se realiza mediante múltiples canales y a través de diferentes tipos de señales. Esta convivencia da lugar, habitualmente, a la necesidad de una sincronización o ajuste.

En todo tipo de traducción audiovisual la traducción no la lleva a cabo tan sólo el traductor sino también toda una serie de protagonistas, como son los actores, director de doblaje, director de subtítulo, ajustadores, etc.

En los casos de subtítulo, *voice-over* (voces superpuestas), *half-dubbing* e interpretación o traducción simultánea el espectador percibe el producto audiovisual en al menos dos lenguas diferentes de forma simultánea (es un producto bilingüe), por los mismos canales (*voice-over*, *half-dubbing*, traducción simultánea) o por canales diferentes (subtítulo).

La traducción audiovisual tiene su propio repertorio de convenciones (normas) entre el producto traducido y el espectador.

Como ya hemos defendido en apartados anteriores, a nuestro entender, sería conveniente (para nosotros incluso imprescindible) incluir el concepto de accesibilidad en los rasgos definitorios de nuestro campo de estudio, pues en esencia, la traducción de cualquier contenido audiovisual lo que procura es precisamente eso, el acceso de las personas al mensaje original, es decir, lo convierte en un producto accesible.

Consideramos oportuno, por tanto, sumar la accesibilidad a las características compartidas por las modalidades de traducción audiovisual y ahora sí estamos en disposición de procurar definiciones acordes al objeto de estudio.

A partir de estas características encontramos definiciones que dan cuenta de lo que es la traducción audiovisual como en la definición propuesta por Chaume (2004:31):

La definición de traducción audiovisual incluye, a mi entender, las transferencias de textos verbo-icónicos de cualquier tipo transmitidos a través de los canales acústico y visual en cualquiera de los medios físicos o soportes existentes en la actualidad (pantalla de cine, televisor, ordenador, etc.).

Esta definición se ajusta, pensamos, a las exigencias propias del campo de la traducción audiovisual que, por su naturaleza cambiante, nos obliga a estar constantemente redefiniendo las modalidades que la incluyen, si bien, definiciones de este calado nos permiten incluir las realidades actuales y futuras.

Años más tarde Chaume (2013:14) propone la siguiente definición de la TAV:

La traducción audiovisual (TAV) es la denominación con la que los círculos académicos se refieren a las transferencias semióticas, interlingüísticas e intralingüísticas entre textos audiovisuales, tanto a las más consolidadas, como el doblaje y la subtitulación, como a las más novedosas, como el rehablado o la audiosubtitulación, por ejemplo.

En otro orden de cosas, en el marco de los Estudios de Traducción, la TAV suele tratarse como un tipo más de traducción especializada equiparable a la traducción jurídica o científica, pero coincidimos con Chaume (1999:210) al entender que esta modalidad de traducción se caracteriza por su modo de discurso híbrido (visual y acústico) y no se distingue por el campo temático ni por el tenor. La traducción especializada es aquella en la que existe un lenguaje de especialidad, una terminología especializada, pero es evidente que no es el caso de la traducción audiovisual. Según Chaume (1999:210-211):

De manera que partimos de la consideración de la traducción audiovisual como una modalidad de traducción que constituye un ámbito de investigación concreto que debe sustentarse en aquello que comparte con otras modalidades de traducción y particularizarse en aquellos rasgos que la singularizan frente al resto de modalidades.

Los rasgos que particularizan la TAV generan distintas modalidades de TAV que pasamos a describir en el siguiente punto.

2.5. Clasificaciones de modalidades de TAV

Los autores Hernández-Bartolomé y Mendiluce-Cabrera ofrecen en su artículo *New trends in audiovisual translation: the latest challenging modes* una recopilación de las distintas clasificaciones en el campo de la TAV desde los años 90 hasta el año de publicación del artículo, 2005. Agrupan las clasificaciones en una tabla (2005:104) dividida en cuatro grupos según el número de modalidades contempladas.

Chaume 2004	Gambier 2004	Díaz Cintas 2001 Linde and Kay 1999 Gambier 1996	Chaves 2000 Agost 1999 Luyken 1991
Dubbing	Dubbing	Dubbing	Dubbing
Subtitling	Intralingual subtitling	Subtitling	Subtitling
	Interlingual subtitling		
	Live or real-time subtitling	Live subtitling	
	Surtitling	Surtitling	
Voice over	Voice over or half dubbing	Voice over	Voice over
Half dubbing			
Simultaneous interpreting	Interpreting	Consecutive interpreting Simultaneous interpreting	Simultaneous interpreting
Narration		Narration	Narration (not in Agost)
Free commentary	(Free) commentary	Commentary	Free commentary
Sight translation	Simultaneous or sight translation		
Animation			
Multimedia translation			Multimedia translation (only in Agost)
	Scenario or script translation		
	Audio description		
	Multilingual productions		
		Multilingual broadcasting	

Figura 3. Clasificación de modalidades de traducción audiovisual.

En un primer grupo se encuentran las clasificaciones que diferencian cinco modalidades (Chaves, 2000, Agost, 1999, Luyken *et al.*, 1991) otros, como Díaz Cintas (2001), De Linde and Kay (1999), Gambier (1996) proponen hasta diez modalidades, y el mayor número de modalidades, hasta trece, es el propuesto por Gambier (2004). Este autor, ya en 2004, incluía la subtitulación intralingüística entre las modalidades de traducción audiovisual.

Posteriormente (Gambier, 2009) estudia la evolución de las modalidades de TAV y genera una nueva clasificación en la que distingue dos grandes grupos de modalidades de TAV: las modalidades dominantes y las modalidades *challenging types*. En este segundo grupo incluye las modalidades de subtitulado intralingüístico para personas sordas, subtitulado en directo y la audiodescripción, entre otras, y pone de relevancia el concepto de accesibilidad en TAV (2009:18):

More recently, three types of AVT have given rise to quite a large number of studies, namely, intralingual subtitling for the deaf and hard of hearing, audiodescription for the blind and visually impaired, and live subtitling (sometimes also called respeaking). The development of these modes can be explained by the umbrella concept of accessibility, or how to allow access to media for all.

Cabe resaltar aquí que Gambier se suma a la aproximación a la TAV desde la accesibilidad al igual que Díaz Cintas que también, en 2007, amplía el número de modalidades en consonancia con la evolución de la disciplina y las nuevas realidades profesionales de la TAV.

Las modalidades señaladas por Díaz Cintas (2001) en su primera clasificación, son posteriormente actualizadas por el autor (2007:15), quien incluye otras prácticas, como la SPS y la AD, para lo que se sirve de la taxonomía de Jakobson (1959). También Cerezo (2012:77) recurre a Jakobson puesto que le permite clasificar las modalidades «según se produzca la traducción entre códigos lingüísticos o entre códigos lingüísticos y otros códigos de significación».

		Canal acústico	Canal visual
Entre códigos lingüísticos	Trasvase interlingüístico	<ul style="list-style-type: none"> • Doblaje • <i>Voice-over</i>⁶² • Narración • Doblaje parcial • Comentario libre • Interpretación simultánea • Traducción a la vista • Audiosubtitulación 	<ul style="list-style-type: none"> • Subtitulación⁶³ • Subtitulación para sordos interlingüística • Sobretitulación interlingüística • Rehablado • Interpretación en lengua de signos
	Trasvase intralingüístico	<ul style="list-style-type: none"> • Doblaje entre dialectos de una misma lengua 	<ul style="list-style-type: none"> • Subtitulación para sordos intralingüística • Rehablado • Sobretitulación intralingüística
Entre código lingüístico y otros códigos de significación	Trasvase intersemiótico	<ul style="list-style-type: none"> • Audiodescripción 	

Figura 4. Propuestas de taxonomía de modalidades y submodalidades de TAV. Cerezo (2012:78).

La propuesta de taxonomía de Cerezo excluye modalidades que la estudiosa reconoce que existen (*fansubs*, *fandubs*, comentario libre, el doblaje parcial, la interpretación simultánea en festivales de cine, la localización de *software* o la traducción de páginas web) pero que no incluye explícitamente en su clasificación bien por no ser actividades profesionales, por no existir asignaturas en estos ámbitos en la formación universitaria en España o por no ser prácticas profesionales existentes en nuestra geografía. Nos parece que la propuesta de Cerezo (2012) es la más realista y la que mejor encaja con nuestra visión de la TAV en un contexto académico y con las modalidades objeto de nuestro estudio.

Como explicaremos en detalle en el capítulo 10 sobre la metodología seguida en nuestra tesis doctoral, el presente trabajo analiza un corpus de materiales extraídos de internet, por lo

que profundizaremos en las modalidades de traducción audiovisual que este medio requiere, según las Pautas de Accesibilidad para el Contenido Web (WCAG) 2.0: subtitulación para personas sordas y audiodescripción para personas ciegas por una parte y, por otra parte, en la subtitulación interlingüística, que es la modalidad escogida en la UE para cumplir con sus políticas lingüísticas y de accesibilidad.

Teniendo en cuenta que analizaremos la accesibilidad de contenidos audiovisuales en internet de una institución europea, restringiremos nuestro ámbito de estudio a la subtitulación convencional y a las modalidades que deberían incorporarse (según la normativa vigente) en este canal de comunicación, cada vez más hegemónico: subtitulación para personas sordas (SPS) y audiodescripción (AD).

2.6. Modalidades de traducción audiovisual en Internet

En este apartado incluimos las modalidades de TAV que desde la WCAG 2.0 proponen como técnicas adecuadas para generar contenidos accesibles y que están recogidas en las normativas de obligado cumplimiento en los Estados miembros de la Unión Europea: la subtitulación y la audiodescripción. La lengua de signos, no es, a día de hoy, de inclusión obligada en los contenidos audiovisuales emitidos en internet.

En cuanto a la subtitulación queremos apuntar, como señala Díaz Cintas (2004:50) de forma premonitoria, que:

In a context where the exchanges are so great, and in a period in which all the experts are in agreement that subtitles are bound to play an increasingly more important and prominent role in our societies, the elaboration of detailed studies which analyse the history and socio-cultural implications of this activity seems imperative. The predicted dominance of subtitling seems justified by virtue of cost—more economical than dubbing—, and execution speed since dubbing requires the participation of more professionals.

Es pues evidente que las predicciones del autor se han cumplido sobradamente y no solo la subtitulación es predominante en internet sino también ya en otros medios. Pasamos a describir las modalidades que hemos apuntado anteriormente.

2.6.1. La audiodescripción para personas ciegas

La audiodescripción (AD) se encuentra subordinada a las restricciones propias de la traducción audiovisual, ya que está comprometida por las imágenes y los diálogos, así como a los efectos de sonido del texto audiovisual. Estas restricciones que acompañan la práctica de la

audiodescripción requieren de la puesta en práctica de muchos de los recursos y competencias de los traductores audiovisuales¹³. Esta modalidad de traducción es especialmente exigente y como apuntaremos en el apartado sobre la cadena de la accesibilidad (cf. capítulo 4), la creación de audiodescripciones requiere de profesionales formados para ese propósito.

Actualmente encontramos diferentes definiciones de audiodescripción siendo, a nuestro modo de ver, la primera formulada en España por Navarrete (1997:71), la más breve y quizás poética. Navarrete define esta técnica en una frase: «El arte de hablar en imágenes».

Pero si nos adentramos en definiciones más técnicas encontramos la propuesta en la norma UNE 153020 de 2005 publicada por AENOR, *Audiodescripción para personas con discapacidad visual*, que la define como:

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.

Esta definición, en nuestra opinión, carece de una explicación clara y sencilla de la técnica audiodescriptiva, que sí encontramos en la ofrecida por Silverwood (1992:10):

AD is a commentary on scenery, costume, facial expression, body language and action inserted during the silent intervals of plays, films and television programmes to give information to blind and visually impaired people.

Por otra parte, la definición ofrecida por Ana Ballester (2007:152), incluye el concepto de audiencia meta para la que está concebida y que dota a la definición de todos los elementos necesarios para su correcta interpretación:

La audiodescripción es un servicio especialmente destinado a las personas ciegas o visualmente discapacitadas. Consiste en un comentario condensado que se teje alrededor de la banda sonora de un producto audiovisual (programa de TV, película, obra de teatro, etc.), y que explota las pausas para explicar la acción que se desarrolla en la escena, describir lugares así como personajes, su vestuario, lenguaje corporal y expresiones faciales, con la finalidad de aumentar la comprensión del texto audiovisual por parte del discapacitado visual.

Entendemos, por tanto, que la audiodescripción es un tipo de traducción intersemiótica que procura la información contenida en los signos no verbales de un texto fílmico para facilitar la comprensión de la película al espectador meta y que se realiza a través del canal acústico.

¹³ Informe del Centro Español de Subtitulado y Audiodescripción (CESyA) redactado por J. Díaz Cintas: “Competencias profesionales del subtitulador y el audiodescriptor”, 2006.

2.6.1.1. Tipos de audiodescripción y campos de aplicación

La audiodescripción como servicio que permite la accesibilidad a distintos entornos, productos y servicios, se da en distintas formas. Díaz Cintas (2007:17) en su artículo *Traducción audiovisual y accesibilidad*, establece tres categorías:

- a) AD grabada para la pantalla: de programas audiovisuales con imágenes en movimiento, como películas, series de televisión, documentales, espectáculos, etc., independientemente del soporte en el que se distribuyen o comercializan (televisión, cine, DVD, internet).
- b) AD grabada para audioguías: de obras estáticas como monumentos, museos, galerías de arte, iglesias, palacios, exposiciones entornos naturales y espacios temáticos en las que (normalmente) no hay imágenes en movimiento y en las que la experiencia táctil, o nuevas tecnologías que simulen este tipo de experiencia, tiene una gran importancia.
- c) AD en directo o semi-directo: de obras teatrales, musicales, ballet, ópera, deportes y otros espectáculos similares. También entran dentro de esta categoría los congresos y cualquier manifestación pública como los actos políticos.

El primer tipo de audiodescripción, «AD para la pantalla», es en el que ubicaríamos la audiodescripción grabada que establece las WCAG 2.0 en la pauta 1.2.3, mientras que la «AD en directo» correspondería a la pauta 1.2.5 ambas de nivel de conformidad AA y, por lo tanto, de aplicación obligatoria según la legislación vigente. Una vez establecida esta categorización pasamos a detallar los distintos tipos de audiodescripción empleados en diferentes campos de aplicación.

La audiodescripción en páginas web viene descrita en las WCAG 2.0 y para el contenido grabado se señala que:

La intención de este Criterio de Conformidad es proporcionar a las personas ciegas o con dificultades visuales acceso a la información visual de una presentación multimedia sincronizada. La audiodescripción se agrega al audio de la presentación e incluye toda la información sobre el vídeo que no está disponible en forma auditiva. En las pausas del diálogo, la audiodescripción proporciona información acerca de todas las acciones, personajes, cambios de escena y textos que aparezcan en pantalla que sean relevantes y que no estén descriptos o hablados en la pista de audio principal.

En cuanto a la guía de estilo o los estándares que hay que seguir, estos no vienen indicados en las WCAG 2.0, pero sí una propuesta de guía de estilo, «Standard techniques in audio description» de Clark (2001), y los recursos puestos a disposición de los desarrolladores web y expertos en accesibilidad multimedia de The design Studio, así como la referencia a la guía de audiodescripción ofrecida por el Description and Captioned Media Program (DCMP).

Todos estos materiales sirven de guía para la creación de materiales audiodescritos, pero consideramos oportuno contemplar a su vez las aportaciones que desde la academia se han hecho para mejorar esta técnica. Recordemos que la AD es el resultado de la interpretación de unas imágenes por parte de un traductor que ha de lidiar con las restricciones espacio-temporales del medio, servirse de los huecos de silencio en los que se incluye la información verbal necesaria para poder seguir la trama argumental, todo ello sin olvidarse de las características y necesidades del receptor, es decir, del destinatario de la audiodescripción. Generar un texto escrito para ser leído que traduzca las imágenes no parece ser tarea baladí.

2.6.2. La subtitulación y la subtitulación para personas sordas y con discapacidad auditiva (SPS)

Si bien no pretendemos caracterizar ni actualizar taxonomías previas de tipos de subtítulos, creemos que es conveniente hacer un breve recorrido a través de las propuestas más significativas en este sentido. Los subtítulos, independientemente del tipo que sean, recurren a parámetros coincidentes en muchos casos y clasificar y definir las distintas tipologías debe ayudarnos a detectar tanto las coincidencias como las divergencias entre distintos tipos de subtítulos.

A continuación, reflejaremos, a partir de las definiciones de autores de especial relevancia en el ámbito de la traducción audiovisual, qué entendemos por subtitulación y subtitulación para personas sordas y con discapacidad auditiva.

Comenzamos este recorrido echando la vista atrás y acercándonos a los orígenes de esta modalidad que Ivarsson (2004) explica de la siguiente forma:

It was not long after the invention of film that efforts were first made to convey the dialogue of the actors to the audience. They started with what we now call intertitles: texts, drawn or printed on paper, filmed and placed between sequences of the film.

Continuamos con una de las definiciones más tempranas que encontramos en la bibliografía de nuestro país, la de Agost (1999:17):

Consiste en la incorporación de subtítulos escritos en la lengua de llegada en la pantalla donde se exhibe una película en versión original, de manera que dichos subtítulos coincidan aproximadamente con las intervenciones de los actores en pantalla.

En esta definición, una de las primeras que se dan en España a esta modalidad de traducción audiovisual, es evidente que la autora hace referencia a la subtitulación interlingüística y a un contenido audiovisual concreto (las películas), dejando de lado otros tipos de subtitulación que aparecerán años más tarde y contenidos que también demandarán

la aplicación de esta técnica. Por su parte, la definición que aporta Díaz Cintas (2003:32) es lo suficientemente general para poder describir la subtitulación tal y como la entendemos hoy, sin especificar si hablamos de interlingüística o intralingüística. A partir de ahora nos referiremos a subtitulación para hablar de interlingüística y de subtitulación para sordos (SPS) para la intralingüística.

La subtitulación se puede definir como una práctica lingüística que consiste en ofrecer, generalmente en la parte inferior de la pantalla, un texto escrito que pretende dar cuenta de los diálogos de los actores, así como de aquellos elementos discursivos que forman parte de la fotografía (cartas, pintadas, leyendas, pancartas, etc.) o de la pista sonora (canciones, voces en off, etc.).

Como vemos, la evolución del concepto es evidente, pero lo que no varía es la relación que todos los autores establecen entre los componentes que articulan una obra subtitulada: la palabra oral, la imagen y los subtítulos. Estos tres componentes determinan las características de los distintos tipos de subtitulación, aunque hay unas características propias e inalterables propias de toda subtitulación. Díaz Cintas (*ibid.*) apunta que:

Los subtítulos han de estar sincronizados con la imagen y los diálogos, deben ofrecer un recuento semántico adecuado de los mismos y permanecer en pantalla el tiempo suficiente para que los espectadores puedan leerlos.

Sin tener en cuenta estos tres componentes, la subtitulación (o cualquier submodalidad) no puede ofrecer la calidad necesaria para que sea aceptada por el público meta.

Por su parte, Díaz Cintas (2008) define la subtitulación para personas sordas desde sus dos vertientes: lingüística y semántica, y técnica. Este doble enfoque nos resulta especialmente interesante puesto que explicita la particularidad de esta práctica traductora sin dejar ningún cabo suelto. El autor define el subtítulo, desde el punto de vista lingüístico y semántico, como una práctica sociolingüística que va del modo oral al escrito y que pretende dar cuenta de los aspectos que aquí presentamos (2008:160):

- **Lo que se dice:** los diálogos de los actores o personas que hablan en el programa audiovisual así como las canciones.
- **Quién lo dice:** indicando la persona que habla cuando sea necesario.
- **Cómo se dice:** la información suprasegmental que acompaña la entrega de los diálogos o monólogos como el énfasis, la entonación, el tono de voz, los acentos, los idiomas extranjeros, etc.

- **Lo que se oye:** los efectos sonoros que se escuchan en la pista sonora, como los ruidos ambientales y la música instrumental.
- **Lo que se ve:** aquellos elementos discursivos que forman parte de la fotografía y están en otros idiomas, como cartas, pintadas, leyendas, pantallas de ordenador, pancartas, etc.

Asimismo, Díaz Cintas señala respecto del tiempo de permanencia (2008:161) lo siguiente:

La información subtitulada se ha de presentar en pantalla el tiempo suficiente para que los espectadores puedan leerla, con un ritmo que permita una lectura cómoda del texto así como el disfrute de las imágenes, y se atenga a una velocidad de lectura adecuada para el público al que va dirigida (niños, adultos).

Con estas especificaciones propuestas por Díaz Cintas (2008) tenemos resumidos los puntos cardinales de la subtitulación y de la SPS, así como también la diferencia que existe respecto a la subtitulación convencional.

2.6.2.1 Tipos de subtítulos y campos de aplicación

Los subtítulos se pueden clasificar atendiendo a distintos aspectos y son varios los autores que han lanzado propuestas en este sentido (Gottlieb, 1997; Ivarsson y Carroll, 1998; De Linde y Key, 1999; Díaz Cintas y Remael, 2003, 2007, entre otros). Por nuestra parte, consideremos que la más completa es la propuesta por Bartoll (2008) por ser, en nuestra opinión, una de las más exhaustivas y que pretende dar cuenta de todos los tipos de subtítulos existentes. Bartoll asienta su taxonomía sobre la de Gottlieb (1997:71-72) y añade a los aspectos técnicos y lingüísticos una tercera categoría, los parámetros pragmáticos. Llegados a este punto, queremos especificar, como señala Orero (2008) que, aunque los avances tecnológicos hacen que las taxonomías se queden desfasadas rápidamente, sí son lo suficientemente flexibles para ser actualizadas con rapidez. La taxonomía de Bartoll (2008:269) considera en total 15 subparámetros y cada subparámetro contiene un número de variables. Aquí los presentamos de forma sintética:

- Parámetros lingüísticos
- Lengua: subtítulos interlingüísticos o intralingüísticos
- Densidad: subtítulos íntegros o reducidos

- Parámetros pragmáticos
- Destinatario: subtítulos para oyentes o para sordos
- Intención: subtítulos instrumentales (didácticos, *karaoke*) o documentales
- Tiempo de elaboración: subtítulos anteriores o simultáneos
- Autoría: subtítulos humanos (profesionales i aficionados [*fansubs*]); o subtítulos mecánicos
- Parámetros técnicos
- Opcionalidad: subtítulos opcionales o no opcionales
- Difusión: subtítulos proyectados, emitidos, automáticos o manuales
- Color: subtítulos monocromos o policromos
- Incorporación: subtítulos dinámicos (palabra a palabra, línea a línea, letra a letra, de parte a parte); subtítulos estáticos
- Posicionamiento: subtítulos (izquierda, derecha, centrados); sobretítulos (izquierda, derecha, centrados); laterotítulos
- Localización: subtítulos internos o externos
- Archivo: subtítulos disociables o indisociables
- Tipografía: fuente, estilo, tamaño
- Formato: *.txt, *.ssa, *.sub, *.vsf., etc.

Con esta estructura, Bartoll (2008) defiende que la inclinación por una variable en un parámetro puede suponer la alteración de otra variable en otro parámetro, de tal forma que podría generarse una relación de interdependencia entre parámetros que da cabida a casi cualquier tipo de subtítulo.

La subtitulación para sordos puede ser intralingüística o interlingüística según si hay o no cambio de lengua respecto al producto audiovisual (Pereira, 2005) y mientras que en la intralingüística se produce solo un cambio de modo (oral a escrito) en la interlingüística se produce un cambio de lengua y modo. Sin embargo, durante unos años, como afirma Díaz

Cintas (2003:40-41), la división tradicional entre ambas modalidades ignoró la subtitulación interlingüística para personas sordas:

Se trata de una categoría híbrida que recurre a parámetros y convenciones que se usan en ambos tipos de subtitulación [...] A nivel profesional este olvido ha significado que los sordos tienen que servirse de los mismos subtítulos interlingüísticos que los oyentes normales cuando son a todas luces inapropiados para sus necesidades [...].

Esta circunstancia que hoy parece estar superada en la mayor parte de los canales de difusión, se ignora en gran medida en las subtítulos que encontramos en internet. No perdemos de vista que los medios de difusión tradicional de los subtítulos han sido la TV, DVD, cine, teatro, etc; sin embargo, en los últimos tiempos y más allá de la subtitulación *amateur*, internet se erige como el canal de mayor relevancia y crecimiento exponencial de futuro para esta modalidad.

Este aspecto es especialmente importante dado que el canal por el que se emite el subtítulo tiene asociados una serie de aspectos técnicos que pueden variar en función del canal y las normas que lo gobiernan. El caso de internet, como canal de difusión de los subtítulos, es un terreno poco explorado e intuimos, situado a ojos de muchos, en un lugar de menos rango que el resto de canales, de ahí el poco interés suscitado entre los miembros de la comunidad científica.

En este capítulo hemos, por una parte, constatado la inherencia de la accesibilidad al concepto de traducción y la importancia de este cambio de enfoque en los estudios de traducción y, por otra parte, hemos ubicado la TAV en los estudios de traducción y realizado un breve recorrido por las cuestiones más relevantes entorno a esta modalidad de traducción. Con todo, hemos pretendido introducir las nuevas realidades y cuestionarnos los retos que tenemos pendiente afrontar.

Nos preguntamos, a partir de lo expuesto en este capítulo, la importancia de la accesibilidad en la web, las características que la definen y las personas que se benefician de ella. Esto nos conduce al siguiente capítulo que pretende, en primer lugar, explicar los conceptos claves de la accesibilidad y, en segundo lugar, definir el marco legal y normativo que procura el acceso a la información audiovisual (entre otros en internet) de todos los ciudadanos.

CAPÍTULO 3

LA ACCESIBILIDAD EN LA ERA DE LA COMUNICACIÓN DIGITAL

3.1. La sociedad de la información y la accesibilidad

3.2. La accesibilidad en la sociedad de la información

3.2.1. Evolución del concepto de accesibilidad

3.2.2. El paradigma del Diseño para Todos

3.2.3. Las barreras de la comunicación y la información

3.2.4. Realidad de la discapacidad en nuestros días

3.3. Las reglas del juego: marco legal y normativo

3.3.1. Marco legal

3.3.2. Marco normativo en España y Europa

CAPÍTULO 3. LA ACCESIBILIDAD EN LA ERA DE LA COMUNICACIÓN DIGITAL

En el capítulo anterior introdujimos el concepto de accesibilidad en los estudios de traducción y en la traducción audiovisual, pero es imprescindible, a nuestro entender, retomarlo y profundizar en la evolución del concepto de accesibilidad ligado a los nuevos avances tecnológicos. Entender cómo hemos llegado al momento en el que nos encontramos, en el que la accesibilidad parece estar en boca de muchos, nos descubre las bases sólidas sobre las que se construye la expansión de la accesibilidad.

En este capítulo pretendemos, por tanto, hacer un recorrido a través de alguno de los aspectos y conceptos más relevantes vinculados a la accesibilidad, con especial énfasis en aquellos puntos que afectan directamente a la accesibilidad a los contenidos audiovisuales en la web.

Para emprender este estudio partiremos de la relevancia de internet en la Sociedad de la Información y ahondaremos en la importancia de la accesibilidad y el Diseño para Todos en el campo de la comunicación. Pasaremos a continuación a describir el pilar sobre el que se apoya la accesibilidad y su consecución, es decir, el marco legal y normativo en España y en Europa sin el cual los avances en materia de accesibilidad serían impensables.

3.1. La sociedad de la información y la accesibilidad

Es comúnmente aceptado que el elemento catalizador del proceso de creación en la sociedad de la información ha sido el abrumador desarrollo tecnológico sobrevenido en las últimas décadas, pero sin duda, la capacidad de innovar de la sociedad y la forma de interaccionar entre los diferentes agentes que comparten el conocimiento también lo es (universidades, instituciones, ciudadanos, empresas, etc.). Estos factores son el pilar fundamental de esta nueva forma de comunicarse, de generar información y conocimiento; en definitiva, de esta nueva forma de ser sociedad.

Las tecnologías de la información y la comunicación (TIC) son, por tanto, las que permiten que todos los agentes interactúen de forma integrada en el proceso de intercambio de información, de ahí su relevancia. Castells (2000) sostiene que, en realidad, internet no es una tecnología, sino una nueva forma de organización de la producción; pero no sólo eso, sino también una nueva forma de la organización de los servicios operativos en la sociedad actual y en la economía. Es Castells (1998) quien define la sociedad de la información de la siguiente manera:

Nuevo sistema tecnológico, económico y social. Una economía en la que el incremento de productividad no depende del incremento cuantitativo de los factores de producción (capital, trabajo, recursos naturales), sino de la aplicación de conocimientos e información a la gestión, producción y distribución, tanto en los procesos como en los productos.

Es decir, según el autor, internet representa una nueva forma de organización de la producción y, en este mismo sentido, Valentí López (2002) apunta:

Lo que era antes la fábrica hoy lo es internet, pero con una diferencia, la fábrica se concentraba, fundamentalmente, en las actividades productivas de bienes, hoy en día internet incide sí en la producción pero también en la organización de los servicios, de los gobiernos, de actividades sociales como puede ser la educación, la salud, etcétera.

Por ende, es indiscutible la relevancia de las tecnologías de la información y la comunicación en la transición de la sociedad industrial a la sociedad de la información, hecho que se manifiesta especialmente en cómo se ha transformado la forma de crear, tratar y distribuir la información y sobre todo en la velocidad que las tecnologías han imprimido al proceso en sí.

En 2003 se celebró la Cumbre Mundial sobre la Sociedad de la Información (CMSI) en la que se concretó una declaración de principios¹⁴ y un plan de acción consensuado que manifiestan la visión compartida de lo que es y debe ser la sociedad de la información:

[...] una Sociedad de la Información centrada en la persona, integradora y orientada al desarrollo, en la que todos puedan crear, consultar, utilizar y compartir la información y el conocimiento, para que las personas, las comunidades y los pueblos puedan emplear plenamente sus posibilidades en la promoción de su desarrollo sostenible y en la mejora de su calidad de vida.

El término *sociedad de la información* se refiere a una nueva forma de organización de la economía y de la sociedad en la que el acceso ilimitado a la información, por parte de casi cualquier ciudadano y de forma instantánea, se convierte en su característica principal. Asimismo, entre los principios fundamentales de la mencionada declaración de principios está «el garantizar que las oportunidades que ofrecen las TIC redunden en beneficio de todos».

¹⁴ Declaración de principios de CMSI, recuperada en https://www.itu.int/dms_pub/itu-s/md/03/wsis/doc/S03-WSIS-DOC-0004!!PDF-S.pdf

Lógicamente para alcanzar este objetivo es necesario asegurar el acceso a la información por parte de todos los ciudadanos y evitar así la brecha digital.

De la definición anterior y del principio fundamental apuntado se desprende el compromiso implícito de convertir la brecha digital en una oportunidad digital para todos y aprovechar el desarrollo de las TIC con el fin de favorecer la inclusión en la sociedad actual.

Pero, ¿qué entendemos por brecha digital? Como apunta Valentí López (2002) las TIC no son igualitarias y la brecha digital viene a expresar la exclusión de la sociedad o parte de la sociedad del acceso a la información. Es una nueva forma de marginación que el autor denomina *marginalización digital*. El concepto de brecha digital implica la ausencia de acceso a las posibilidades que nos brindan las TIC de informarnos, comunicarnos, crear o aprender. Es importante destacar que la brecha digital se refiere a varias dimensiones como apunta Gómez Hernández *et al.* (2008:31):

La brecha digital tiene dimensiones sociales, lingüísticas, educativas o culturales: Entre países o regiones con o sin acceso a la Red, entre jóvenes y mayores, entre medio rural y medio urbano, entre géneros, entre profesionales y clases trabajadoras, entre lenguas con presencia y contenidos en la Red o no, entre personas con discapacidades físicas, psíquicas o perceptivas que limitan el acceso y las que no las tienen, etcétera.

En sus inicios el concepto se centraba en la posibilidad o no de las personas de acceder a ordenadores conectados pero, como vemos en la cita anterior, el concepto se ha ensanchado hasta incluir cuestiones lingüísticas, de discapacidad y de alfabetización digital. Que las personas con discapacidad no puedan acceder a un contenido o que la lengua en la que está disponible no sea comprensible para determinados grupos son las grietas que ensanchan la brecha. Según Camacho (2005):

El concepto de brecha digital se ha modificado a través del tiempo. En un principio se refería básicamente a los problemas de conectividad. Posteriormente, se empieza a introducir la preocupación por el desarrollo de las capacidades y habilidades requeridas para utilizar las TIC (capacitación y educación) y últimamente también se hace referencia al uso de los recursos integrados en la tecnología.

Los recursos integrados en la tecnología se refieren a la posibilidad de acceso que tienen las personas para utilizar los recursos disponibles en la web. Es decir, allí donde se aloje un contenido audiovisual al que un usuario no pueda acceder, hay brecha digital. Nos encontramos pues ante un mapa web agrietado por doquier.

En este mismo orden de cosas, y como viene recogido en los principios fundamentales acordados en la Cumbre Mundial Sobre la Sociedad de la Información¹⁵ (CMSI), el reto que

¹⁵ <http://www.itu.int/net/wsis/docs/geneva/official/dop-es.html>

asume esta sociedad de la información es el de lograr la equidad y la inclusión de todos los ciudadanos, siendo estas las señas de identidad que quedan reflejadas en los planes, acciones y políticas europeas que se han materializado hasta el momento, así como las planteadas para el futuro. De especial importancia es la iniciativa eEurope¹⁶ «Una sociedad de la información para todos» y los distintos planes de acción que desde 1999 Europa ha puesto en marcha con el objetivo de impulsar la sociedad de la información. Esta iniciativa presentada por la Comisión Europea pone de relieve el cambio que apuntábamos anteriormente de la sociedad industrial a la sociedad de la información y pretende asegurar la participación de todos en esta nueva economía basada en el conocimiento. Los objetivos principales de la iniciativa son:

1. Conseguir que todos los europeos entren en la era digital y estén conectados a la red.
2. Crear en Europa una cultura y un espíritu empresarial abiertos a la cultura digital.
3. Garantizar que el proceso no se traduzca en exclusión social y se gane la confianza del consumidor.

El tercer objetivo se refiere a la brecha digital y a las posibles diferencias de acceso a las TIC para unos y otros. Esta es, en la actualidad, una de las preocupaciones más importantes de los gobiernos y máximos organismos puesto que, para alcanzar una sociedad de la información equitativa e inclusiva, es necesario abordar la brecha digital de forma holística. En los inicios de la iniciativa, se hizo mucho hincapié en la accesibilidad referida a la infraestructura tecnológica dejando de lado otras dimensiones de vital importancia. Crovi Druetta (2002:20) propone tres dimensiones que conforman la brecha digital y apuesta por un enfoque integrador:

1. La dimensión tecnológica, ligada a proveer la infraestructura necesaria para que los países y personas se incorporen al paradigma de la sociedad de la información y el conocimiento (SIC).
2. La dimensión del conocimiento, vinculada a las habilidades y saberes que deben poseer los individuos, para que sean capaces de apropiarse de las herramientas de la convergencia tecnológica y a partir de ello cambiar sus prácticas culturales cotidianas (trabajo, educación, relaciones personales y entretenimiento).

¹⁶ <http://eur-lex.europa.eu/legal-content/ES/TXT/?uri=URISERV:124221>

3. La dimensión de participación, que significa que los recursos aportados por la convergencia tecnológica puedan emplearse en un contexto democrático, con un marco social y legal adecuado para que los individuos y las naciones tengan igualdad de oportunidades para manifestarse, participar e intervenir en las decisiones globales de las SIC.

Según la autora, la lucha contra la brecha digital ha de contemplar la infraestructura tecnológica, la accesibilidad y la igualdad de oportunidades por igual, de lo contrario la sociedad de la información y la accesibilidad, es decir, una sociedad de la información para todos, será difícilmente alcanzable.

Coincidimos con este enfoque integrador que, al igual que en traducción y especialmente en la modalidad de traducción audiovisual, refleja la necesaria imbricación de todos los elementos que conforman la obra traducida para su correcto funcionamiento, es decir, si falla un agente de la cadena ésta se rompe y no alcanzamos el objetivo deseado. Y efectivamente, uno de estos elementos es la accesibilidad; en nuestro caso, la accesibilidad al contenido audiovisual. Sin embargo, para ubicar correctamente la Traducción Audiovisual (TAV) en el ámbito de la accesibilidad, en el próximo apartado definiremos el concepto de accesibilidad en sentido *bottom-up*.

3.2. La accesibilidad en la sociedad de la información

En el apartado anterior hemos destacado la importancia del concepto de accesibilidad en la sociedad de la información, sin embargo nos parece necesario responder a preguntas que nos permitan echar la vista atrás para entender de forma holística qué es la accesibilidad. Así pues nos disponemos a contestar a las siguientes preguntas ¿dónde nace este concepto?, ¿qué entendemos por accesibilidad? y ¿qué tiene que ver la accesibilidad con la traducción? Para ello partiremos de la evolución del concepto de accesibilidad y abriremos la puerta a conceptos claves como son el Diseño para Todos (DpT), las barreras audiovisuales y las personas con discapacidad.

3.2.1. Evolución del concepto de accesibilidad

En sus inicios el concepto de accesibilidad era entendido fundamentalmente como la posibilidad mayor o menor de acceder al entorno físico. De hecho, aún hoy, para muchos ciudadanos tiene este único sentido debido principalmente a la invisibilidad de la propia

barrera audiovisual. Sin embargo, como mostraremos a continuación, esta noción ha quedado desterrada en el marco legal y normativo para dar paso a otras más acordes a la situación y necesidades actuales; necesidades estas que surgen como resultado del paso de la sociedad industrial a la sociedad de la información y la reorganización social y estamental asociada a ese cambio que explicábamos anteriormente. Quizás el hecho que mejor visibiliza este cambio es el relego del formato en papel en las diferentes formas de comunicación actual por el formato audiovisual. Este cambio no sólo tiene implicaciones en las actividades de ocio, a las que generalmente se asocia la sociedad de la información, sino que afecta directamente a diversos productos y servicios a cuyo acceso todo ciudadano tiene derecho, entre otros a la educación.

Pero, retomando el concepto en sus orígenes, y para plasmar de forma concienzuda su evolución, partiremos de una de las primeras definiciones que aparece en los foros que debaten el estado de la cuestión y que ofrece el *Concepto Europeo de Accesibilidad* (CCPT, 1996) que la define en los siguientes términos:

La accesibilidad es una característica básica del entorno construido. Es la condición que posibilita el llegar, entrar, salir y utilizar las casas, las tiendas, los teatros, los parques y los lugares de trabajo. La accesibilidad permite a las personas participar en las actividades sociales y económicas para las que se ha concebido el entorno construido.

El documento, *Concepto Europeo de Accesibilidad*, llega en respuesta a la petición de la Comisión Europea, en 1987, de definir los principios de diseño universal aplicables a edificios, infraestructuras y productos de la construcción y del equipamiento. Este sirvió como trabajo de referencia para armonizar la idea de la accesibilidad en Europa y aportar una base segura para establecer una norma europea de accesibilidad.

Es cierto que la definición no menciona la accesibilidad a productos y servicios, dado que está realizada con una fuerte impronta arquitectónica puesto que tomó como base el *Manual Europeo para un Entorno Construido Accesible*; sin embargo, la declaración¹⁷ que recoge esta definición basa su enunciación en los principios de Diseño Universal o Diseño para Todos (DpT), que es el concepto esencial en la idea actual de la accesibilidad y de ahí escogerlo como el punto de partida de nuestra idea de accesibilidad. No sería justo obviar que el documento se redactó cuando la sociedad de la información, aún en ciernes, no podía predecir los retos a los que tendría que hacer frente. No olvidemos que no fue hasta principios de los 90 cuando surgió internet.

¹⁷ La declaración fue aprobada el 2 de marzo de 1996.

Como adelantamos en el párrafo anterior, la importancia de este documento radica, a nuestro entender, en el ángulo desde el que aborda el concepto de accesibilidad, puesto que parte del principio sobre el que se sostiene una sociedad igualitaria, el principio de que todas las personas son distintas. El documento establece así la base sobre la que se asienta el diseño para todos y expone que la accesibilidad se consigue «integrando las distintas necesidades de las personas en instalaciones que pueda utilizar todo el mundo».

En el apartado siguiente profundizaremos en el diseño para todos y veremos cómo este concepto, a día de hoy, sí incluye el desarrollo no sólo de entornos (arquitectura y urbanismo), sino también de productos y servicios, además de rechazar la división de la población en personas capacitadas y discapacitadas.

Por su parte, en España, es el Libro Blanco del Plan de Accesibilidad ACCEPLAN (2003:20) el que nos hace reflexionar sobre otro de los conceptos que forman parte de la definición de accesibilidad; nos referimos al concepto de barrera y los diferentes tipos que encontramos según el ámbito:

La palabra accesibilidad puede entenderse en relación con tres formas básicas de actividad humana: movilidad, comunicación y comprensión. Todos, según sean nuestras capacidades funcionales o mentales, tropezamos con barreras en nuestra capacidad de movimiento, en nuestras comunicaciones o fuentes de información, y en nuestro alcance de comprensión de mensajes, instrucciones, instrumentos o sistemas (...). Los efectos de dichas barreras pueden llegar incluso a la exclusión social, a la discapacitación, a la estigmatización y a agravios psicológicos para las personas afectadas. La incapacidad de la sociedad para eliminar las barreras de movilidad, de comunicación y de comprensión es sintomática de la atención desigual que merecen las personas con capacidades reducidas. A la inversa, cada barrera al acceso que cae nos acerca un poco más a la consecución de una sociedad justa.

Sin duda, la ausencia de barreras es, por tanto, una condición necesaria para garantizar la presencia y la participación de las personas en la sociedad. Nuestro estudio abordará las barreras vinculadas a la comunicación en el apartado 3.2.3 con la pretensión de destapar aquellas con las que algunas personas suelen encontrarse en el uso de las TIC. Resulta muy obvio que el hecho de procurar la accesibilidad a las comunicaciones permite y potencia la inclusión de todos los ciudadanos, así que antes de tratar las barreras y su categorización habremos de preguntarnos: ¿qué entendemos por inclusión?

La inclusión, basada en el concepto de una sociedad para todos (Mank, 2000), nace en el ámbito educativo para extenderse a otros ámbitos de la sociedad. Mank afirma que la inclusión es un proceso en el que distinguimos tres niveles: la *presencia*, es decir, estar en la sociedad; la *participación*, alcanzada si se dan las condiciones necesarias para que la persona pueda

participar; y por último, el *desarrollo de las potencialidades*, para lo que hay que identificar y superar las barreras que lo impiden. En definitiva, una sociedad inclusiva se cimienta en el acceso igualitario a los sitios/lugares (movimiento), a la información (comunicación) y a los mensajes (comprensión). En caso de no alcanzar este grado de accesibilidad triforme estaremos favoreciendo la exclusión y acercándonos a la *marginalización digital* apuntada por Crovi Druetta.

Volviendo al concepto de accesibilidad y a las distintas vertientes reflejadas en su definición, Hernández y Montes (2002:21) señalan que el concepto de accesibilidad, posiblemente debido al aumento de su presencia en la sociedad, parece no gozar de la cohesión necesaria:

Como sucede con otros conceptos, cuyo uso se ha generalizado al expandirse su significación original, la noción de accesibilidad, hoy ampliamente aceptada, se ha incorporado al lenguaje cotidiano para designar múltiples aspectos de la realidad, y al banalizarse, se ha hecho más ambigua. Si, además, se pretende aplicarla a los fenómenos culturales, el riesgo de incurrir en imprecisiones puede parecer mayor.

Estas imprecisiones se traducen, en el ámbito de la accesibilidad, en una constricción del concepto que limita las formas de accesibilidad e ignora en muchos casos la necesidad de, por ejemplo, hacer accesibles los contenidos audiovisuales para todos los colectivos. Recordemos que todos tenemos derecho a acceder a la vida social y profesional por igual y hoy en día gran parte de esa vida social y profesional está ligada, de un modo u otro, a contenidos digitales, siendo un porcentaje considerable de estos contenidos multimedia.

Retomando la definición del concepto de accesibilidad y su enfoque en distintas áreas vinculadas con la creación y difusión de contenidos audiovisuales, aportamos a continuación la idea de accesibilidad en la web, de especial interés para nosotros puesto que la web es el canal de difusión en el que se enmarca nuestro trabajo.

La WAI (Web Accessibility Initiative) grupo de trabajo permanente del W3C (World Wide Web Consortium) que tiene el objetivo de facilitar el acceso a la web de las personas con discapacidad, propone en su *Guía Breve de Accesibilidad Web*¹⁸ (W3C), una definición que expande la naturaleza del concepto y refleja la necesidad y los beneficios que puede reportar una web accesible.

Hablar de accesibilidad web es hablar de un acceso universal a la web, independientemente del tipo de hardware, software, infraestructura de red, idioma, cultura, localización geográfica y capacidades de los usuarios.

¹⁸ <http://www.w3c.es/Divulgacion/GuiasBreves/Accesibilidad>

La accesibilidad web, a tenor de la gran cantidad de contenidos audiovisuales que se alojan en la web, ha de asumir la existencia de las barreras audiovisuales y aplicar las técnicas y tecnologías existentes para que el contenido multimedia sea accesible. No hemos de olvidar que la web, textual en sus inicios, es cada vez más audiovisual y las previsiones de su constante auge apuntan a que el consumo de estos contenidos *online* sea cada vez mayor.

En este orden de cosas, Voces Merayo (2008:3) señala que, en internet, «la implementación de la accesibilidad en los contenidos audiovisuales es mucho más compleja que en otros contenidos», dado que los contenidos audiovisuales necesitan un entorno de reproducción propio y éste ha de ser accesible, y a su vez que los contenidos en sí sean perceptibles. No obstante, la complejidad puede sortearse; las técnicas que lo permiten, al igual que en los contenidos presentes en otros medios son la audiodescripción (AD), la subtitulación para personas sordas (SPS) y la interpretación en lenguaje de signos (ILS). Si bien estas son las técnicas de aplicación para sortear las barreras a los contenidos audiovisuales, este trabajo abordará las que quedan enmarcadas dentro de nuestro ámbito de estudio (la traducción audiovisual y la accesibilidad), que son la SPS y la AD.

Por último, y puesto que nuestro estudio se enmarca en el ámbito de la accesibilidad web, queremos finalizar este apartado con la definición de accesibilidad aportada por el creador de la World Wide Web, Tim Berners-Lee:

El arte de garantizar que, tan amplia y extensamente como sea posible, los medios (como por ejemplo el acceso a la web) estén disponibles para las personas, tengan o no deficiencias de un tipo u otro.

Por tanto, uno de los conceptos clave en accesibilidad es el de Diseño para Todos, que entiende al individuo como parte de una comunidad diversa y no única, y que construye teniendo en cuenta los diferentes usuarios que conforman la sociedad y no según el ciudadano medio. Por su papel central en la confección de productos accesibles, en el próximo apartado abordaremos el concepto de Diseño para Todos.

3.2.2. El paradigma del Diseño para Todos

«El buen diseño capacita, el mal diseño discapacita»

Declaración de Estocolmo del EIDD

El concepto de Diseño para Todos o Diseño Universal (en este trabajo utilizaremos el primer término) se caracteriza por entender que únicamente existe una población, compuesta por personas con habilidades y necesidades distintas. Así pues, cuando se diseñen

entornos, productos o servicios se hará teniendo en cuenta a todas las personas evitando la discriminación sustentada en la diferencia. En este sentido Mace *et al* (1990:1). apuntan desde el ámbito de la arquitectura que:

Universal design means simply designing all products, buildings and exterior spaces to be usable by all people to the greatest extent possible. It is advanced here as a sensible and economical way to reconcile the artistic integrity of a design with human needs in the environment.

Es decir, el Diseño para Todos es una nueva forma de crear y planificar en la que se tiene en cuenta la diversidad desde el origen. El concepto de accesibilidad ha evolucionado, madurado y mutado hasta acercarse al paradigma del Diseño para Todos.

En 2004, el European Institute for Design and Disability (EIDD)¹⁹ establece en la *Declaración de Estocolmo* que:

El Diseño para Todos tiene como objetivo hacer posible que todas las personas dispongan de igualdad de oportunidades y de participación en cada aspecto de la sociedad. Para conseguir esto, el entorno construido, los objetos cotidianos, los servicios, la cultura y la información, es decir, todo lo que está diseñado o hecho por personas para ser utilizado por personas, ha de ser accesible y útil para todos los miembros de la sociedad y consecuente con la continua evolución de la diversidad humana.

Es decir, el Diseño para Todos tiene en cuenta al conjunto de personas que conforman nuestras sociedades a la hora de desarrollar entornos, productos o servicios. Deja de lado el concepto de *Diseño democrático*, en el que las necesidades de la mayoría son las únicas tenidas en cuenta a la hora de generar productos para posteriormente, y sólo en algunos casos, buscar soluciones para que los colectivos minoritarios puedan tener acceso a ese producto, servicio o entorno. Lo que procura el Diseño para Todos es diseñar teniendo en cuenta a todos los grupos poblacionales; de esta manera, una vez lanzado el producto no hay que pensar en cómo adaptarlo para determinados colectivos, puesto que se tiene en cuenta a todos los perfiles desde el origen del mismo y permite el acceso a un mayor número de personas y por qué no, ahorra costes innecesarios.

El Diseño para Todos se basa en siete principios que estableció, desde el Centro para el Diseño Universal de la North Columbia State University de EE.UU., un grupo de expertos que entienden que «the design of products and environments to be usable by all people, to

¹⁹ Poco después de su constitución en 1993, el European Institute for Design and Disability (EIDD) definió su objetivo básico: “Mejorar la calidad de vida mediante el *Design for All*”. Conceptos similares se han desarrollado paralelamente en otras partes del mundo. Los estadounidenses con la Americans with Disability Act, han contribuido a la evolución del Diseño Universal (*Universal Design*), mientras que el Diseño Inclusivo (*Inclusive Design*) ha ganado terreno en el Reino Unido. La Declaración de Estocolmo donde se trata de difundir este concepto, se aprueba en la Junta del EIDD del 9 de mayo de 2004.

the greatest extent possible, without the need for adaptation or specialized design». A partir de esta premisa propusieron²⁰:

1. *Equitable Use: The design is useful and marketable to people with diverse abilities.*
2. *Flexibility in Use: The design accommodates a wide range of individual preferences and abilities.*
3. *Simple and Intuitive Use: Use of the design is easy to understand, regardless of the user's experience, knowledge, language skills, or current concentration level.*
4. *Perceptible Information: The design communicates necessary information effectively to the user, regardless of ambient conditions or the user's sensory abilities.*
5. *Tolerance for Error: The design minimizes hazards and the adverse consequences of accidental or unintended actions.*
6. *Low Physical Effort: The design can be used efficiently and comfortably and with a minimum of fatigue.*
7. *Size and Space for Approach and Use: Appropriate size and space is provided for approach, reach, manipulation, and use regardless of user's body size, posture, or mobility.*

Como podemos observar, para satisfacer el tercer y cuarto principio han de ponerse en práctica la traducción y varias de sus modalidades. Igualmente, si el proceso de creación respeta todos los principios definidos por este grupo de expertos, hacemos del Diseño para Todos una realidad y alcanzamos la accesibilidad. Por tanto, como señala Madariaga Ortuzar (2004:499):

El diseño para todos, en el sentido más amplio del término, consiste en no crear entornos, productos y servicios específicos para personas con discapacidad, sino que cuando se fabrique o se desarrolle cualquier producto o servicio, se tengan en cuenta las necesidades e intereses de todos los posibles usuarios.

Por último y en este mismo sentido, la Convención de Naciones Unidas sobre los Derechos de las Personas con Discapacidad²¹, aprobada en 2006 y ratificada por España en 2007 entiende el diseño para todos como:

²⁰ Version 2.0-4/1/97 Copyright 1997, NC State University, The Center for Universal Design. Recuperada de http://www.ncsu.edu/ncsu/design/cud/about_ud/udprinciplestext.htm

²¹ <http://www.un.org/spanish/disabilities/default.asp?navid=13&pid=497>

El diseño de productos, entornos, programas y servicios que puedan utilizar todas las personas, en la mayor medida posible, sin necesidad de adaptación ni diseño especializado.

Cuando no se diseña para todos sino para la mayoría, aparecen las barreras; barreras que son más o menos visibles y, consecuentemente, a menor grado de visibilidad mayor grado de incidencia. Por desgracia, son muchos los tipos de barreras que existen, pero este trabajo se centrará en las que están relacionadas con la accesibilidad a la información, tanto las que se ven como las que no se ven.

3.2.3. Las barreras de la comunicación y la información

Con el fin de iniciar este apartado, tomaremos la definición de *barrera* propuesta por el Grupo de Accesibilidad Audiovisual del Comité Español de Representantes de Personas con Discapacidad (CERMI) en el documento «*Accesibilidad de la Televisión Digital para las Personas con Discapacidad*»:

Cualquier impedimento, traba u obstáculo que limita o impide el acceso, utilización, disfrute o interacción de manera digna, cómoda y segura con el entorno. (AENOR. Informe UNE 41500). En concreto se conoce como barreras de comunicación a aquellos impedimentos u obstáculos que limitan o impiden a las personas con discapacidad acceder a la información.

Identificar qué tipo de barreras existen nos ayuda a reflexionar acerca de por qué se producen y cómo sortearlas. El *Libro Blanco del Plan de Accesibilidad ACCEPLAN* (2003) define cuatro grandes grupos de barreras. Esta división coincide con los cuatro tipos esenciales de barreras que quedan plasmados en la normativa de accesibilidad y que vienen a ser las barreras arquitectónicas, las barreras urbanísticas, las barreras en el transporte y las barreras en las telecomunicaciones. Estas últimas son, lógicamente, las que afectan a nuestro ámbito de estudio y que se enmarcan en el sector de la comunicación y la información. Las barreras de la comunicación y la información se subdividen, según el *Libro Blanco del Plan de Accesibilidad* (2003:25), en los sectores reflejados en la siguiente figura.

Contenidos de la accesibilidad que se desarrollan en el sector de Comunicación e Información	
Foros tecnológicos	Muchos de los temas relacionados con la tecnología y, en algunas ocasiones, sus soluciones de accesibilidad, se tratan en foros tecnológicos en los que se deciden futuras acciones y modos de funcionamiento. Así, las definiciones de los sistemas GSM y UMTS, los protocolos TCP/IP y V.18, los sistemas de subtítulo, etc. son producto del trabajo realizado por grupos de expertos que proponen sus soluciones para el futuro

(Figura 5 continúa en la página siguiente)

(Figura 5 continúa de la página anterior)

Contenidos de la accesibilidad que se desarrollan en el sector de Comunicación e Información	
Informática	Comprende todos estos aparatos nuevos que se han introducido en nuestra sociedad en los últimos 20 años, dotándola de una capacidad sin precedentes en el ámbito del proceso de la información aplicada al trabajo o al de ocio: ordenadores personales, PDAs, agendas electrónicas, terminales informáticos, sistemas operativos (ej. Windows, Linux), programas (ej. Outlook, Word), herramientas de desarrollo (ej. Visual Basic), etc.
Internet	Apoyada en el mundo de la informática, de donde procede, y de las comunicaciones (especialmente la telefonía fija), internet se ha convertido en un servicio con entidad propia por el que se han instaurado nuevos paradigmas de comunicación e información como son el correo electrónico y la web. Se presta especial atención a las áreas en las que se incorporan nuevos servicios específicos que afectan a la ciudadanía y que son susceptibles de divergir tecnológicamente en el futuro como la banca, la sanidad, etc.
Lengua de Signos	La lengua de signos merece un subárea aparte, ya que es una de las más importantes herramientas de comunicación de las personas sordas y con discapacidad auditiva (de hecho es su lengua, y se intenta conseguir su aceptación como lengua oficial del Estado español). Sus peculiaridades y su marcado componente visual, hacen que los recursos que se pongan a su disposición en un futuro tengan características muy específicas que merecen ser estudiadas y agrupadas con el fin de conseguir un conjunto de acciones coherentes que ayuden a su implantación.
Señalización	La señalización es el sistema de información que sirve para orientarnos en nuestro tránsito por el entramado físico que compone nuestra sociedad. Nuestras carreteras, estaciones de metro, tren y autobús, los aeropuertos, las calles, o los grandes almacenes, por ejemplo, están llenos de señales que sirven para indicarnos dónde estamos, qué camino debemos tomar para ir dónde queremos, dónde está lo que necesitamos etc. Este subárea engloba los problemas que las señales representan para las personas con discapacidad y sus soluciones.
Telefonía fija	La telefonía fija nos lleva acompañando más 70 años y las últimas generaciones han disfrutado de sus beneficios sin reparar en ella. Estos teléfonos, los que tenemos conectados a un conector en las paredes de nuestros hogares han presentado tradicionalmente varias barreras para las personas con discapacidad y han permitido el desarrollo de nuevos servicios como el fax, los teléfonos de texto, servicios de teleasistencia y la conexión a internet. Esta subárea se ocupa de los problemas relacionados con la accesibilidad a los terminales y a los servicios relacionados, sin alcanzar a internet.
Telefonía móvil	La telefonía móvil ha significado la verdadera revolución de las comunicaciones de finales del siglo pasado. Supera el número de terminales de telefonía fija y facilita el uso de un gran número de nuevos servicios como son el acceso a internet, la transmisión de imágenes, los mensajes cortos, etc.
Contenidos	Denominamos contenido a toda la información depurada final, destinada a que un medio de comunicación la presente a un ciudadano (películas de cine, vídeos musicales, textos, fotos, etc.). La mayor parte de estos contenidos tienen ya un formato digital, lo que permite mucha flexibilidad a la hora de su presentación en formatos que resulten accesibles para personas con algún tipo de discapacidad. Este subárea abarca todos los problemas de accesibilidad de los contenidos, y sus soluciones como son el subtítulo, la audiodescripción, etc.
Dispositivos especiales: quioscos virtuales y máquinas expendedoras	Los quioscos virtuales y las máquinas expendedoras tienen su representación más cotidiana en los cajeros automáticos de los bancos, aunque hay muchos más que nos rodean, como son los quioscos de información virtual, las máquinas expendedoras de billetes, las máquinas de los parkings, etc. Estas máquinas son omnipresentes en nuestra sociedad y presentan problemas y soluciones especiales de accesibilidad que se tratan en esta subárea.

Figura 5. Sectores de comunicación e información. Acceplan, 2003.

En el subsector «Contenidos» se hace referencia a los contenidos digitales, en particular, vídeos y películas, y se presenta el subtítulo y la audiodescripción como solución a las barreras que se generan en torno a este tipo de contenido. Asimismo, internet presenta múltiples barreras que podrán sobrevenirse siguiendo las normas técnicas de referencia que procuran la accesibilidad web tanto en España como en Europa. Hablaremos de estas normas en detalle más adelante.

Las personas con algún tipo de discapacidad sensorial son las que generalmente han de sortear las barreras que acabamos de citar y es importante, según nuestro parecer, esbozar cuál es el contexto de estos grupos de población. Por tanto, en el siguiente apartado, arrojaremos datos básicos que nos sitúen frente a la realidad a la que se enfrenta un número nada desdeñable de personas en España y en el mundo.

3.2.4. Realidad de la discapacidad en nuestros días

Según recoge la Organización Mundial de la Salud (OMS) en su *Informe mundial sobre la discapacidad* (2011), aproximadamente el 15% de la población mundial vive con algún tipo de discapacidad. Este organismo define la discapacidad como parte de la condición humana y señala que «casi todas las personas sufrirán algún tipo de discapacidad transitoria o permanente en algún momento de su vida».

Por su parte, en España, la Encuesta de Discapacidad, Autonomía Personal y Situaciones de Dependencia (EDAD), publicada por el INE (2008)²² refleja que el número total de personas residentes en hogares españoles que declaran tener alguna discapacidad asciende a 3.847.900, lo que supone un 8,5% de la población española. En 2008, la edad media de las personas con discapacidad era de 64,3 años, mientras que en 1999 era de 63,3. En el grupo de edad de 65 y más años, la edad media se sitúa en 78,8 en 2008 frente a 77,1 en 1999. Según el *Libro Blanco del Envejecimiento Activo* hacia el 2050 las personas mayores supondrán el 32% de la población española, lo cual convertirá a España en una de las poblaciones más envejecidas del mundo. En este mismo orden de cosas, la Oficina Europea de Estadística, más conocida como Eurostat²³, apunta que la población europea habrá aumentado significativamente en 2050 y que la población estará más envejecida que en la actualidad.

²² El INE ha publicado tres encuestas sobre discapacidades. Los estudios se realizaron en 1986, 1999, y el más reciente en 2008.

²³ http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing#Past_and_future_population_ageing_trends_in_the_EU

La pregunta que sigue es ¿qué entendemos por discapacidad? Tomaremos la definición que la OMS propuso en 2001 para responder a la pregunta y establecer los límites que la enmarcan:

Es un término general que abarca las deficiencias, las limitaciones de la actividad y las restricciones de la participación. Las deficiencias son problemas que afectan a una estructura o función corporal; las limitaciones de la actividad son dificultades para ejecutar acciones o tareas, y las restricciones de la participación son problemas para participar en situaciones vitales.

Esta definición establece la discapacidad, además de como una condición de la salud de cada persona, como el resultado de la interacción entre las limitaciones de las personas y el medio en el que vivimos. Este nuevo enfoque reconoce el contexto social como un factor determinante en la discapacidad de las personas y es precisamente ese contexto social el que está en nuestras manos cambiar.

En este mismo sentido, la Convención Internacional sobre los Derechos de las Personas con Discapacidad (CDPD) se refiere a la discapacidad como:

Un concepto que evoluciona y que resulta de la interacción entre las personas con discapacidad y las barreras debidas a la actitud y al entorno que evitan su participación plena y efectiva en la sociedad, en igualdad de condiciones con las demás.

Es este un enfoque mucho más dinámico en el que la limitación en la realización de las actividades cotidianas o de la participación en la vida social determina también la condición de discapacidad. Asimismo, la CPCD se refiere a los distintos tipos de discapacidad en los siguientes términos:

Las personas con discapacidad incluyen a aquellas que tengan deficiencias físicas, mentales, intelectuales o sensoriales a largo plazo que, al interactuar con diversas barreras, puedan impedir su participación plena y efectiva en la sociedad, en igualdad de condiciones con las demás.

Las personas con deficiencias físicas, sensoriales y mentales se enfrentan, como ya hemos mencionado anteriormente, a diferentes tipos de barreras. Dado que nuestro trabajo se centrará en el acceso de las personas con discapacidad sensorial (visual y auditiva) a los contenidos audiovisuales, concretamente a los contenidos audiovisuales disponibles en la web, mencionaremos a continuación las características y datos de estos dos colectivos y las barreras de la comunicación a las que han de hacer frente.

Según las últimas estimaciones recogidas por la OMS cerca de 314 millones de personas en el mundo viven con discapacidad visual²⁴, 45 millones de las cuales son ciegas. La mayoría de las personas con discapacidad visual tienen edad avanzada y las principales causas de ceguera crónica son las cataratas, el glaucoma, la degeneración macular relacionada con la edad, la retinopatía diabética, el tracoma y las afecciones oculares infantiles, tales como las causadas por la carencia de vitamina A. Nos parece interesante resaltar que en los países desarrollados las principales causas de este tipo de discapacidad son: la degeneración macular relacionada con la edad y el envejecimiento de la población, así como la retinopatía diabética consecuencia de los malos y abusivos hábitos alimentarios, siendo esta una de las principales causas de la ceguera en el mundo desarrollado.

En España, según la última Encuesta de Discapacidad, Autonomía Personal y situaciones de Dependencia (EDAD) realizada por el INE, hay 979.200 personas afectadas por una discapacidad visual. Según esta misma encuesta, la edad y de nuevo la diabetes son consideradas los principales factores de riesgo.

Por su parte, los últimos estudios sobre la población europea muestran una clara tendencia al envejecimiento de la población, y esto se debe principalmente a dos factores: una mejor calidad de vida y al descenso de las tasas de natalidad. Las predicciones de la ONU (2001) sobre la edad de la población, muestran claramente esta tendencia y sitúan a España como el país con la población más envejecida del mundo en el año 2050. Si cruzamos estos datos con los del párrafo anterior, parece evidente que tanto la degeneración macular relacionada con la edad como la retinopatía diabética aumentarán significativamente el número de personas con discapacidad visual en España y en Europa.

Asimismo, la Organización Mundial de la Salud²⁵ señala que unos 360 millones de personas padecen pérdida de audición discapacitante, y una de cada tres personas mayores de 65 años sufre pérdida de audición (presbiacusia). En este orden de cosas, la OMS, en la campaña de sensibilización «Escuchar sin riesgos» (2015) apunta²⁶ que 1 100 millones de jóvenes de todo el mundo están en riesgo de sufrir pérdida de audición por realizar prácticas auditivas dañinas. Señala, a su vez, que más de 43 millones de personas de entre 12 y 35 años padecen una pérdida auditiva discapacitante debida a diferentes causas. En países de ingresos medianos y altos las causas serían:

²⁴ La ceguera se define como una agudeza visual de presentación inferior a 3/60, o una pérdida del campo visual a menos de 10°, en el mejor ojo, con la corrección disponible. Por discapacidad visual grave se entiende una agudeza visual de entre menos de 6/60 y 3/60, y por discapacidad visual moderada, una agudeza visual de menos de 6/18 a 6/60. Fuente: http://apps.who.int/gb/ebwha/pdf_files/EB124/B124_7-sp.pdf Fecha de acceso: febrero de 2018.

²⁵ <http://www.who.int/mediacentre/factsheets/fs300/es/>

²⁶ <http://www.who.int/topics/deafness/safe-listening/es/>

- Casi el 50% de la población está expuesta a niveles de ruido perjudiciales (dispositivos de audio personales como reproductores de MP3 y teléfonos inteligentes).
- Alrededor del 40% están expuestos a niveles de ruido potencialmente nocivos en clubes, discotecas y bares.

En el ámbito nacional, según la EDAD (2008) hay un total de 1 064 000 personas sordas y con algún tipo de discapacidad auditiva (es decir, un 2,3% de la población total).

Con los datos expuestos, parece razonable pensar que los medios de comunicación deberían concebirse y entenderse bajo el paradigma de la accesibilidad y el Diseño para Todos, pero la situación actual, aunque ha mejorado en los últimos años, muestra que este colectivo no está en igualdad de condiciones frente al resto de la sociedad. Las personas con discapacidad sensorial no tropiezan contra escaleras, ni aceras, sino contra el silencio y la invisibilidad. No *todos* tenemos el acceso garantizado a esta cultura visual en la que medios de comunicación, nuevas tecnologías y manifestaciones artísticas de diversa índole se mueven entre la imagen y la palabra. Garantizar el acceso a las nuevas tecnologías y medios de comunicación es una obligación adquirida por el Estado como quedó reflejado en la Constitución española (1978) y en los sucesivos desarrollos normativos que se han producido, pero aun hoy las personas con discapacidad continúan siendo muy vulnerables a la exclusión social.

3.3. Las reglas del juego: marco legal y normativo

La sociedad, cada vez más sensibilizada, demanda progresos en cuestiones de accesibilidad y asume la responsabilidad de exigirla a través de distintos mecanismos (Responsabilidad Social Corporativa, grupos de presión, asociaciones de partes implicadas). La información, reconocida como un valor indispensable en la actualidad, forma parte del conjunto de entornos, productos y servicios que han de ser diseñados para todos y así evitar la exclusión social. Internet es, actualmente, uno de los vehículos de información más utilizados por los ciudadanos; por tanto, parece lógico argüir que los poderes públicos tienen la obligación de procurar y garantizar los derechos de los ciudadanos en la sociedad de la información. Los medios de los que disponen las administraciones para minimizar las barreras en la web son varios: políticas de fomento, investigación, sensibilización, normalización y legislación. Sobre estas iniciativas normativas recae el peso de incluir a las

personas con discapacidad en las nuevas tecnologías y evitar así la *infoexclusión*²⁷. Veamos a continuación los instrumentos y mecanismos que se han puesto en marcha desde la llegada de la democracia y cómo nos sitúan en el mapa de la accesibilidad.

3.3.1. Marco legal

En España, la Constitución de 1978 recoge en diversos artículos (art. 20.1.d y art. 51.2) el derecho a la información de los ciudadanos, pero no es hasta el año 2002 cuando una ley, Ley de Servicios de la Sociedad de la Información y de Comercio Electrónico (LSSICE), hace referencia clara a la accesibilidad de las personas con discapacidad en la red. Por su parte, las Pautas de Accesibilidad al Contenido Web (WCAG) fueron aprobadas en 1994, pero hubieron de pasar diez años para que se legislara expresamente en esta materia y se comenzara a definir los niveles de accesibilidad y las fechas de cumplimiento. Por tanto, en España, es la Ley 34/2002, de 11 de julio, la arriba mencionada LSSICE, la que en su disposición adicional quinta expone que «los sitios web de las Administraciones Públicas adoptarán las medidas necesarias de acuerdo con los criterios de accesibilidad generalmente conocidos». Sin embargo, aunque la ley hace referencia expresa al derecho de las personas con discapacidad de acceder a internet, la redacción de esta disposición no define claramente los criterios de accesibilidad por lo que en la práctica su aplicación resultaba ambigua.

Posteriormente se aprueba la Ley 51/2003, de 2 de diciembre, de Igualdad de Oportunidades, No Discriminación y Accesibilidad Universal de las personas con discapacidad (LIONDAU) que pretende cubrir el vacío existente en España de un marco legal amplio y general. La disposición final séptima expone, respecto a los productos y servicios de la sociedad de la información, las «condiciones básicas de accesibilidad y no discriminación para el acceso y utilización de las tecnologías, productos y servicios relacionados con la sociedad de la información y medios de comunicación social». En ella se establece la obligación gradual y progresiva de que todos los entornos, productos y servicios sean abiertos, accesibles y practicables para todas las personas y dispone plazos y calendarios para la realización de las adaptaciones necesarias. Los instrumentos de planificación elaborados para la puesta en marcha de la LIONDAU fueron el «Plan Nacional de Accesibilidad 2004-2012» y el «II Plan de Acción para las personas con discapacidad 2003-2007».

²⁷ La *infoexclusión* remite al conjunto de procesos sociales que impiden, separan o dificultan el acceso de los individuos y grupos sociales a las tecnologías de la información y el conocimiento, excluyéndoles o situándoles en riesgo de exclusión, de las ventajas en cuanto a información, comunicación y conocimiento que poseen las TIC en la llamada Sociedad de la Información (Rodríguez Victoriano, 2003:12).

Cuatro años más tarde entra en vigor el Real Decreto 1494/2007, de 12 de noviembre, por el que se aprueba el «Reglamento sobre las condiciones básicas para el acceso de las personas con discapacidad a las tecnologías, productos y servicios relacionados con la sociedad de la información y medios de comunicación social». Este se inspira en los principios establecidos en la LIONDAU y define los criterios y condiciones básicas de accesibilidad y no discriminación en materia de sociedad de la información. En este sentido se especifica el grado de accesibilidad aplicable a las páginas de internet de las Administraciones Públicas o en aquellas con financiación pública y se establece como nivel mínimo obligatorio de cumplimiento las prioridades 1 y 2 de la Norma UNE 139803:2004²⁸. El reglamento dispone que a partir de 2009 las Administraciones Públicas y organismos financiados por ésta deberán cumplir con las prioridades arriba mencionadas.

La Ley 56/2007, de 28 de diciembre, de Medidas de Impulso de la Sociedad de la Información modifica la disposición adicional quinta de la Ley 34/2002, de 11 de julio, resuelve las ambigüedades de la LSSICE de 2002 y potencia los derechos de los ciudadanos en la sociedad de la información. El objetivo prioritario de esta nueva ley es conseguir que internet sea más accesible para las personas con discapacidad y de edad avanzada. Para la consecución de este fin se añadieron tres párrafos a la disposición final quinta de la LSSICE que se traducen en tres nuevas obligaciones. En el primer párrafo, y con el objetivo de que las páginas web incorporen de forma progresiva los criterios de accesibilidad, exhorta a las Administraciones Públicas a promover medidas de sensibilización, educación y formación. En segundo lugar, la Ley 56/2007 establece que los incumplimientos estarán sometidos al Régimen de infracciones y sanciones sobre las condiciones básicas para el acceso de las personas con discapacidad a las tecnologías, productos y servicios relacionados con la sociedad de la información y medios de comunicación social (Ley 49/2007), y por último, señala que las páginas de internet de las empresas que presten servicios al público en general de especial trascendencia económica, es decir, las que agrupen a más de cien trabajadores o su volumen anual de operaciones exceda los 6 millones de euros, deberán satisfacer a partir del 31 de diciembre de 2008, como mínimo, el nivel medio de los criterios de accesibilidad al contenido generalmente reconocidos: el nivel AA del W3C, que corresponde a las prioridades 1 y 2 de la Norma UNE 139803:2004.

²⁸ En España la legislación no permite que se haga referencia a documentos técnicos que no sean emitidos por organismos oficiales de normalización, como por ejemplo los redactados por AENOR (Asociación Española de Normalización y Certificación), encargada de generar las Normas UNE. En el caso de la accesibilidad web, los estándares, pautas y recomendaciones los dicta el *World Wide Web Consortium* (W3C), un organismo neutral y sin ánimo de lucro que procura el consenso entre los agentes que participan en la construcción de la Web. La adecuación de las pautas de este organismo se realizó a través de la Norma UNE 139803:2004: *Aplicaciones informáticas para personas con discapacidad. Requisitos de accesibilidad para contenidos en la Web*.

Días antes se aprobó la Ley 49/2007, de 26 de diciembre, por la que se establece el Régimen de infracciones y sanciones por el incumplimiento de la Ley 51/2003 LIONDAU mencionado en el párrafo anterior. Esta ley es imprescindible para garantizar el derecho a la igualdad de oportunidades de las personas con discapacidad; es por todos conocido que una ley sin sanción es entendida por la sociedad como de no aplicación. Las infracciones se clasifican en leves, graves y muy graves con multas de entre 301 euros y el millón de euros y el abono de las mismas no exime del cumplimiento de la Ley. Las muy graves podrán ir acompañadas de la supresión de ayudas y subvenciones.

Por último, el 3 de mayo de 2008 entró en vigor en España el Instrumento de ratificación de la Convención sobre los derechos de las personas con discapacidad, redactado en Nueva York el 13 de diciembre de 2006. El propósito de la Convención «es promover, proteger y asegurar el goce pleno y en condiciones de igualdad de todos los derechos humanos y libertades fundamentales por todas las personas con discapacidad, y promover el respeto de su dignidad inherente». El artículo 9, titulado Accesibilidad, declara:

A fin de que las personas con discapacidad puedan vivir en forma independiente y participar plenamente en todos los aspectos de la vida, los Estados Partes adoptarán medidas pertinentes para asegurar el acceso de las personas con discapacidad, en igualdad de condiciones con las demás, al entorno físico, el transporte, la información y las comunicaciones, incluidos los sistemas y las tecnologías de la información y las comunicaciones, y a otros servicios [...].

Por consiguiente, como hemos reflejado en este apartado, en la actualidad, España dispone de las medidas y mecanismos necesarios que procuran la accesibilidad, pero estos avances, en especial las leyes aprobadas en los últimos años, no hubieran sido posibles sin el impulso administrado por Europa. Veamos en el siguiente apartado el marco normativo nacional y europeo que nos ha situado en las actuales circunstancias.

3.3.2. Marco normativo en España y en Europa

A mediados de los años 80 se puso en marcha una serie de medidas políticas para el desarrollo de la sociedad de la información relacionadas con las TIC en los ámbitos de la educación y la innovación. Tras unos proyectos experimentales, el Ministerio de Educación y Cultura presentó en 1985 el Programa de Nuevas Tecnologías de la Información y la Comunicación (PNTIC). Este fue el primero de los programas y planes que se han ejecutado desde entonces hasta nuestros días. En 1988, se inició el primer Plan Nacional de Investigación y Desarrollo, en el que las TIC se incluían como un área de trabajo independiente. A partir de este primer plan, los planes nacionales de I+D han sido y son el eje principal para el desarrollo

de la sociedad de la información. En 1994, los retos impuestos por Bruselas pusieron en marcha una serie de acciones aisladas para la implantación de la sociedad de la información en España. Durante los siguientes años, las iniciativas y actuaciones de carácter innovador, y la creación de distintos foros, organismos y comisiones de trabajo especializadas, serían la semilla del plan de acción «INFO XXI: La sociedad de información para todos», presentado en enero de 2001.

Por su parte, Europa definió en el año 2000, en el Consejo de Lisboa, la estrategia de crecimiento para la próxima década con el objetivo de «convertirse en la economía basada en el conocimiento más competitiva y dinámica del mundo»²⁹. En junio de 2000, se aprobó el plan de acción «eEuropa2002: Una sociedad de la información para todos»³⁰, siendo éste el instrumento creado para la construcción de la sociedad de la información europea. Entre sus objetivos específicos destaca el de garantizar que las personas con discapacidad y de la tercera edad puedan beneficiarse plenamente de las nuevas tecnologías y de internet. El plan establece que:

[...] las páginas web del sector público y su contenido, en los Estados miembros y las instituciones europeas,³¹ deben diseñarse de manera que sean accesibles, a fin de que los ciudadanos con discapacidades puedan acceder a la información y aprovechar plenamente las posibilidades de la administración electrónica.³²

El plan de acción estableció la hoja de ruta de la UE en materia de accesibilidad a la sociedad de la información en Europa, a partir de ahí había que poner en marcha los mecanismos necesarios para que realmente se materializaran los objetivos. Con el fin de que las páginas web fueran, efectivamente, accesibles para la tercera edad y para las personas con discapacidad, el 25 de septiembre de 2001, la Comisión aprobó la comunicación «eEuropa 2002: Accesibilidad de los sitios web públicos y de su contenido»³³. La comunicación anima a las instituciones europeas y a los Estados miembros a impulsar que los sitios web sean accesibles y recomienda la adopción de las Pautas de Accesibilidad al Contenido (WCAG 1.0) redactadas por el WAI. Posteriormente, la resolución del Parlamento sobre esta Comunicación (2002) que apuesta por una internet accesible para todos, manifiesta en el punto 31 que «para que los sitios web sean accesibles es fundamental que satisfagan el nivel doble A y que se aplique en su totalidad la prioridad 2 de las Pautas WAI».³⁴ Con esto

²⁹ http://www.europarl.europa.eu/summits/lis1_es.htm

³⁰ http://www.europarl.europa.eu/summits/lis1_es.htm

³¹ El parlamento europeo es una institución europea por lo que el contenido mostrado en su página web ha de ser accesible.

³² http://ec.europa.eu/information_society/europe/2002/action_plan/pdf/actionplan_es.pdf

³³ <http://wma.comb.es/Upload/Documents/eEurope2002.pdf>

³⁴ http://www.europarl.europa.eu/pv2/pv2?PRG=DOCPV&APP=PV2&SDOCTA=18&TXTLST=1&TPV=DE-F&POS=1&Type_Doc=RESOL&DATE=130602&DATEF=020613&TYPEF=TITRE&PrgPrev=PRG@TITRE|APP@PV2|TYPEF@TITRE|YEAR@02|Find@web%20|FILE@BIBLIO02|PLAGE@1&LANGUE=ES

vemos el recorrido aproximado de uno de los objetivos marcados en el plan de acción, en este caso el de una internet para todos, que va desde la declaración de intenciones inicial a la Comunicación emitida por la Comisión en la que detalla cómo hacer que los sitios web sean accesibles.

Tras la estrategia de Lisboa, los ministros europeos se reúnen en Riga donde reconocen la existencia de una brecha digital que entienden prioritario estrechar, no sólo para ser una economía más competitiva, sino también para luchar contra la discriminación y el acceso desigual a las TIC. Por todo esto, la Declaración Ministerial de Riga (2006) sobre una sociedad de la información inclusiva fijó una serie de objetivos que debían alcanzarse antes de 2010, y de nuevo incorpora el compromiso de que todos los sitios web públicos sean accesibles. Para garantizar la accesibilidad se estableció que estos cumplieran con las Pautas de Accesibilidad al Contenido (WCAG), en aquel entonces la versión 1.0.

Europa continúa su andadura en pro de la accesibilidad con la «Comunicación i2010: Una sociedad de la información europea para el crecimiento y el empleo», en la que se determinan las políticas generales de la sociedad de la información y los medios de comunicación de forma integrada. Posteriormente, los estados miembros transponen las exigencias de manera desigual por lo que la Unión Europea se ve obligada a intervenir dada la fragmentación de esfuerzos y la falta de colaboración que presumiblemente harían que no se alcanzaran los objetivos de Riga. Por ello, y dada la importancia de las TIC, la «Iniciativa Europea para la inclusión digital» (2007) propone medidas para alcanzar los objetivos de la Comunicación i2010 y de Riga en materia de inclusión digital.

Los resultados de este llamamiento han fomentado el progreso en materia de accesibilidad, pero como quedó reflejado en el *Informe sobre la competitividad digital de Europa*:

[...] la Unión Europea aún presenta un retraso notable en el ámbito de la investigación y el desarrollo de las TIC con respecto a Estados Unidos, Japón o Corea del Sur. Por lo tanto, para mantener su competitividad, es importante que Europa cuente con una nueva agenda digital.³⁵

En diciembre de 2010, se aprobó la Comunicación «Estrategia Europea sobre Discapacidad 2010-2020: un compromiso renovado para una Europa sin barreras»³⁶ que tiene por objetivo que «cualquier ciudadano con discapacidad de la UE pueda utilizar un autobús sin problemas, conectarse a internet o manejar un DVD sin ayuda de otras personas, y por

³⁵ Informe sobre la competitividad digital de Europa: principales logros de la estrategia i2010 entre 2005 y 2009 [COM(2009) 390 final; no publicada en el Diario Oficial]

³⁶ <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0636:FIN:ES:PDF>

eso estudiará la posibilidad de proponer de aquí a 2012 una ley europea de accesibilidad». Es por tanto objetivo central la supresión de barreras que garanticen el derecho a las personas con discapacidad de integrarse y disfrutar plenamente de los servicios y recursos disponibles en y para la sociedad actual. En especial, la estrategia apunta a la necesidad de «fomentar la alfabetización, la capacitación y la inclusión digitales». Los servicios digitales incluyentes a los que se refiere describen una serie de acciones para garantizar que los contenidos electrónicos nuevos estén plenamente disponibles para las personas con discapacidad y que deben ajustarse, ahora sí, a las WCAG 2.0.

El 23 de diciembre de 2010 se dio un paso trascendental en materia de accesibilidad cuando la Unión Europea ratificó la Convención de la ONU sobre los derechos de las personas con discapacidad (2006).³⁷ El propósito de la Convención «es promover, proteger y asegurar el goce pleno y en condiciones de igualdad de todos los derechos humanos y libertades fundamentales por todas las personas con discapacidad, y promover el respeto de su dignidad inherente». Esta ratificación da un nuevo impulso a la accesibilidad en Europa y afianza su compromiso de asegurar el acceso de las personas con discapacidad a la información y comunicaciones.

Por último, en 2014 se produjo uno de los últimos avances de mayor impacto: la aprobación de la primera norma europea de accesibilidad para productos y servicios de tecnologías de la información y la comunicación. Esta norma, la EN 301.549 «Requisitos de accesibilidad adecuados para la contratación pública de productos y servicios TIC en Europa», establece los requisitos que garantizan la accesibilidad a la información desde diferentes dispositivos y se basa en las WCAG 2.0, además de establecer una metodología de evaluación que permite comprobar si se cumplen los requisitos establecidos por la misma. En España, la transposición de esta norma es la norma UNE 139803 «Requisitos de accesibilidad para contenidos en la Web».

Por su parte, en España, y volviendo al punto del camino donde nos habíamos quedado en cuanto a normativa en materia de accesibilidad, la escasez de resultados obtenidos en el marco del primer Plan de Acción INFOXXI y el retraso del país frente a Europa dieron pie a una nueva propuesta, el programa «España.es»³⁸ aprobado en julio de 2003, con una vigencia de dos años. Mientras, en Europa, se ponía en marcha el plan de acción «eAccessibility», cuyo objetivo era conseguir que todos los ciudadanos pudieran disfrutar de las ventajas que ofrece esta nueva Sociedad de la Información. Este plan fue renovado por

³⁷ <https://www.un.org/esa/socdev/enable/documents/tccconvs.pdf>

³⁸ www.campus.usal.es/~derinfo/derinfo/Espana.es/espana_es.pdf

«eEurope2005», cuyo propósito era acercar a los ciudadanos de la Unión a la Sociedad de la Información y de este modo continuar con la línea de trabajo de «eAccessibility» a través de «eInclusion».

Con el fin de que España se acercase a los objetivos europeos y conseguir la convergencia plena, se pone en marcha el «Plan Avanza (2006-2010)»³⁹, incluido en el Programa Ingenio 2010, que tiene por objeto impulsar las actividades de innovación, desarrollo e investigación (I+D+i). Asimismo, para contribuir a la convergencia con Europa y promover el desarrollo regional de la Sociedad de la Información se crea el Instituto Nacional de Tecnologías de la Información (INTECO),⁴⁰ que cuenta con el Centro de Referencia en Accesibilidad y Estándares Web, que «estudia el estado actual de la accesibilidad y el uso de estándares web de la Administración General del Estado, promoviendo acciones destinadas a homogeneizar los portales web de la administración en tecnologías del W3C», además de ser un referente en materia de accesibilidad web.

La Estrategia Española sobre Discapacidad 2014-2020, aprobada por el Consejo de Ministros el 14 de octubre de 2011, es actualmente la referencia en materia de discapacidad en el ámbito nacional y los objetivos de la misma son coincidentes con la Estrategia Europea sobre Discapacidad 2010-2020 y alcanzar así los objetivos marcados en la Convención Sobre Derechos de las Personas con Discapacidad. Para alcanzar la consecución de los objetivos marcados en las mencionadas estrategias hay que esperar hasta el 2014, año en que se aprueba el plan de acción de la «Estrategia Española sobre Discapacidad 2014-2020». Los cinco ejes de actuación principales son: igualdad para todas las personas, empleo, educación, accesibilidad y dinamización de la economía. El plan ocupa dos fases, la primera se desarrollará entre 2014 y 2016 y en la que se desarrollaran actuaciones tales como la de «promover la aplicación y desarrollo de la Estrategia Integral Española de Cultura para Todos» y el fomento de la accesibilidad en las tecnologías de la información y la comunicación.

Por último, apuntamos un dato que entendemos pone de relieve la importancia de trabajos como el que ahora presentamos: la Estrategia Europea sobre Discapacidad 2010-2020, apunta que «de media, en EU-27 solo el 5% de los sitios web públicos se ajustan completamente a las normas de accesibilidad de internet». Es decir, que los compromisos adquiridos por Europa y los estados miembros, en nuestro caso España, y que han sido

³⁹ Su objetivo se formula en términos de «conseguir la adecuada utilización de las TIC para contribuir al éxito de un modelo de crecimiento económico basado en el incremento de la competitividad y la productividad, la promoción de la igualdad social y regional, la accesibilidad universal y la mejora del bienestar y la calidad de vida de los ciudadanos».

⁴⁰ INTECO es miembro oficial del W3C desde el año 2009.

descritos en este apartado son fundamentales para la consecución del objetivo cardinal; una sociedad europea del conocimiento accesible para todos los individuos que la componen.

Como hemos recogido en este apartado, las medidas y acciones impulsadas por la UE demuestran su evidente implicación y compromiso por alcanzar la plena inclusión de los ciudadanos europeos en la sociedad de la información, si bien es cierto que los resultados logrados hasta el momento son merecedores de mayores esfuerzos futuros. Hemos recogido también en este apartado las estrategias, planes y actuaciones puestos en marcha en el ámbito nacional que nacen fruto de recomendaciones europeas. Con todo esto, evidenciamos que disponemos de los instrumentos necesarios para generar espacios virtuales accesibles cuyos contenidos sean para todos.

Por tanto, en este capítulo hemos puesto de manifiesto el lugar absolutamente central que tiene internet en la sociedad actual por lo que en el siguiente capítulo nos disponemos a definir la accesibilidad web a partir de las directrices de accesibilidad al contenido, así como a poner en relieve la importancia de la cadena de agentes que forman parte del proceso de creación de sitios web y la relación de todos estos aspectos con la traducción.

CAPÍTULO 4

LA WEB Y LAS DIRECTRICES DE ACCESIBILIDAD

4.1. Las tres uves dobles: El W3C, la WAI y las WCAG

4.1.1. Pautas de Accesibilidad al Contenido Web (WCAG)

4.1.2. Pautas de accesibilidad al Contenido audiovisual WCAG 1.0 y WCAG 2.0

4.1.2.1. Las WCAG 1.0 y la accesibilidad al contenido audiovisual

4.1.2.2. Las WCAG 2.0 y la accesibilidad al contenido audiovisual

4.1.2.3. La cadena de accesibilidad web: perfiles profesionales

CAPÍTULO 4. LA WEB Y LAS DIRECTRICES DE ACCESIBILIDAD

Es indiscutible que internet en la sociedad de la información ocupa una posición central, por lo que nos parece necesario presentar la consecución de la accesibilidad en este apartado y ahondar en los criterios que la gobiernan.

Asimismo, cabe señalar que en este capítulo, y dado su papel determinante en la accesibilidad de los contenidos audiovisuales, abordaremos en qué consisten las Pautas de Accesibilidad al Contenido en la Web (WCAG)⁴¹ tanto la primera como la segunda versión⁴², así como la norma UNE 139803 (2004 y 2012) que viene a ser una transposición normativa de las WCAG. Situaremos el foco en las pautas que hacen referencia al contenido audiovisual. A partir de las WCAG vincularemos la accesibilidad web con la traducción audiovisual, con la pretensión de reflejar el estado actual en materia de accesibilidad web y la importancia y beneficios que se desprenden, en nuestra opinión, de la inclusión del profesional de la traducción audiovisual en el ámbito del diseño y desarrollo web. Entendemos que para la consecución de este objetivo será necesario, al margen de superar muchos otros escollos, ampliar la cadena de agentes que participan en ella y visibilizar la importancia del trabajo que desarrollan.

4.1. Las tres uves dobles: el W3C, la WAI y las WCAG

La World Wide Web tiene sus orígenes en el Centre Européen de Recherche Nucléaire (CERN) en Ginebra. En este centro de investigación, con el objetivo de crear un sistema de comunicación entre los científicos de física nuclear de todo el mundo, en 1991, Tim Berners-Lee combinó dos tecnologías existentes, el hipertexto y el protocolo de comunicaciones de internet que dio lugar a la web. En la actualidad, este científico, lidera el World Wide Web Consortium (W3C), organismo internacional que promueve desde su creación en 1994

⁴¹ En junio de 2018 se publicaron las WCAG 2.1, pero son normativa de transición a las WCAG 3.0 y simplemente se añaden unos criterios de éxito que no están relacionados con nuestro estudio.

⁴² El 5 de junio de 2018 se hizo pública la versión definitiva de las WCAG 2.1, que sirven de transición hacia las futuras WCAG 3.0, pero puesto que los criterios de conformidad no han variado respecto a las versión 2.0, no la desarrollaremos en nuestro trabajo.

una web única cuyos beneficios estén disponibles para todos. Es precisamente Berners-Lee quien declaró que «el poder de la web está en su universalidad. Un acceso para todo el mundo independientemente de su discapacidad es un aspecto esencial»⁴³ y es también quien, desde sus inicios, lidera los movimientos para mejorar la web y hacerla más democrática y accesible.

El trabajo del W3C se divide en cuatro pilares básicos: arquitectura, interacción, tecnología y sociedad, y accesibilidad web. Precisamente, este cuarto pilar propició la creación de la Web Accessibility Initiative (WAI), entidad que persigue la eliminación de barreras para las personas con discapacidad. Este grupo de trabajo, que se creó en 1999, en coordinación con otras organizaciones, promueve la accesibilidad en cinco áreas: tecnología, guías, herramientas, educación y difusión e I+D. Pero quizás la tarea más conocida de la WAI sea el desarrollo de pautas de accesibilidad web para los diferentes componentes que la integran: Pautas de Accesibilidad para Herramientas de Autor (ATAG), Pautas de Accesibilidad para Herramientas de Usuario (UAAG) y Pautas de Accesibilidad al Contenido en la Web (WCAG). Todas son consideradas normas oficiales en la Unión Europea y se citan en la mayoría de las legislaciones sobre TIC de todo el mundo, en especial las WCAG que, por su repercusión y acogida, se han convertido en referente internacional aceptado. Puesto que nuestro trabajo analiza la accesibilidad al contenido audiovisual en la Web, nos centraremos en los aspectos más relevantes de las WCAG vinculados a nuestro ámbito de estudio, la traducción audiovisual.

Antes de comenzar con nuestra exposición, consideramos oportuno subrayar que para lograr la accesibilidad al contenido en la Web son diversos los componentes que desempeñan un papel importante. Según Moreno *et al.* (2008) estos componentes quedan agrupados en dos espacios diferenciados: el contenido en sí (textos, imágenes o sonidos) y el etiquetado (en HTML o similares) de la web. La interrelación entre contenido y etiquetado crea sinergias vitales para la consecución de la **accesibilidad** por lo que es muy importante entender y transmitir que, si uno falla, los demás también lo harán. Sin embargo, nosotros en este trabajo nos centraremos en la forma de hacer accesible el contenido audiovisual, puesto que otro tipo de investigación se escapa de nuestro ámbito de estudio y de nuestro conocimiento. Por lo tanto, en el próximo apartado presentaremos las Pautas de Accesibilidad para el Contenido Web (WCAG) desde sus orígenes hasta la versión actual con el objetivo de adentrarnos en el concepto de accesibilidad en el canal de difusión hegemónico de nuestros días, internet.

⁴³ <http://ares.cnice.mec.es/informes/17/contenido/43.htm>

4.1.1. Pautas de accesibilidad al contenido web (WCAG): versión 1.0 y 2.0

Como señalábamos anteriormente, las WCAG son las directrices para el desarrollo de contenidos accesibles y describen cómo hacerlo procurando no mermar el diseño de las páginas web. Las primeras pautas se aprobaron en 1999, las WCAG 1.0 y en 2008 se aprobó la segunda versión, las WCAG 2.0, que son en la actualidad de aplicación en Europa y en España y referente en el resto del mundo. Puesto que algunas páginas web siguen aplicando la primera versión y son, en definitiva, el pilar sobre el que se sustenta la segunda versión, pasamos a describir ambas pautas y plasmar así su recorrido histórico.

La versión 1.0 consta de 14 pautas, que son los principios generales para el diseño accesible, y cada pauta contiene puntos de verificación que describen cómo aplicarla (65 en total). A su vez, cada punto de verificación tiene asignado un nivel de prioridad que, como describe el Instituto Nacional de Tecnologías de la Comunicación (INTECO)⁴⁴ (2009:9), corresponden a:

- Prioridad 1: Todos los puntos de verificación que el desarrollador **tiene** que satisfacer; si no, algunos grupos de personas serán incapaces de acceder a la información de un sitio;
- Prioridad 2: El desarrollador **debe** satisfacerla; sin ello alguien encontrará muchas dificultades para acceder a la información;
- Prioridad 3: El desarrollador **puede** satisfacerla; de lo contrario, algunas personas hallarán dificultades para acceder a la información.

La guía diferencia tres niveles de adecuación que se corresponden con los niveles de prioridad, es decir, el nivel de adecuación A incluye los puntos de verificación de prioridad 1; el nivel AA, los puntos de verificación de prioridad 1 y 2; y el nivel de adecuación AAA, incluye las tres prioridades.

Las WCAG 1.0⁴⁵ tenían ciertas limitaciones; por una parte, se basaban en el supuesto de que HTML era la única tecnología existente y no incluían posibles nuevos usos y sobre todo eran más interpretables que las actuales, es decir, diferentes personas podían entender que la aplicación de una pauta comportaba unas obligaciones u otras. Ante estas imprecisiones, la WAI desarrolló unas nuevas pautas, las WCAG 2.0⁴⁶, aprobadas en diciembre de 2008, que

⁴⁴ Guía de Recomendaciones de Accesibilidad y Calidad Web (2009).

⁴⁵ <https://www.w3.org/TR/WAI-WEBCONTENT/>

⁴⁶ <https://www.w3.org/TR/WCAG20/>

son más fáciles de entender y que permiten evaluar la accesibilidad de forma más precisa, sin interpretaciones.

Por su parte, las WCAG 2.0 se organizan en cuatro principios generales que establecen que el contenido ha de ser: Perceptible, Operable, Comprensible y Robusto. Cada uno de estos principios recoge un conjunto de pautas (un total de 12 pautas) y cada pauta contiene una serie de criterios de éxito (61 en total). Los criterios de éxito están clasificados por niveles de conformidad (A, AA, AAA). En este sentido, es importante apuntar que los criterios han sido redactados para verificarse sin ambigüedad, bien de forma automática con la ayuda de una herramienta o bien de forma manual, verificación que ha de llevar a cabo por una persona.

Los criterios de éxito son fundamentales para conseguir que una página web sea accesible, por lo que la WAI ha redactado varios documentos que facilitan la comprensión y aplicación de cada uno de ellos. Así pues, cada criterio de éxito está vinculado a dos documentos: *Understanding WCAG 2.0*⁴⁷ y *How to meet WCAG 2.0*⁴⁸.

Understanding WCAG 2.0 es un documento que indica qué técnicas hay que utilizar para cumplir con cada criterio, además de explicar el criterio en sí, los beneficios de alcanzarlo y una lista de ejemplos con los principales errores. En cuanto a las técnicas que aconseja utilizar, que de ser aplicadas garantizan que se cumpla el criterio de éxito correspondiente, las divide según sean Técnicas suficientes (*Sufficient Techniques*), Técnicas complementarias (*Advisory Techniques*) y Fallos (*Failures*). Toda esta información está contenida en el documento *How to meet WCAG 2.0*.

La importancia del documento *How to meet WCAG 2.0* reside en que alberga las técnicas que desde la WAI consideran adecuadas para cumplir con los criterios de éxito. Las técnicas suficientes nos aseguran que en caso de ser utilizadas el criterio se cumplirá, pero no cierran la puerta a que los desarrolladores pongan en práctica otras técnicas que también cumplan el objetivo. Por otra parte, las técnicas complementarias, mejoran la accesibilidad, si bien no aseguran la consecución de un criterio. Por último y de especial importancia, el documento señala los fallos que define como las barreras de la accesibilidad web y que a su vez son de especial trascendencia para los evaluadores web. Sí hay un fallo entonces no se cumple el criterio de conformidad por lo que con el fin de facilitar la tarea a los evaluadores web, cada técnica propone un test que ayuda a determinar si la técnica se ha implementado con éxito. Profundizaremos en estos aspectos en el apartado correspondiente a la evaluación de la accesibilidad (capítulo 5).

⁴⁷ <http://www.w3.org/TR/UNDERSTANDING-WCAG20/>

⁴⁸ <http://www.w3.org/WAI/WCAG20/quickref/>

Por otra parte, es también de especial importancia el concepto de nivel de conformidad, puesto que las normativas se basan en ellos para cerciorarse de que una página, vídeo en nuestro caso, sea accesible, y mientras algunos criterios exigen un nivel de conformidad A, otros están asociados a niveles de conformidad AA o AAA. A continuación detallamos en qué consisten estos niveles que vienen definidos en las WCAG 2.0⁴⁹:

- **Level A:** *For Level A conformance (the minimum level of conformance), the Web page satisfies all the Level A Success Criteria, or a conforming alternate version is provided.*
- **Level AA:** *For Level AA conformance, the Web page satisfies all the Level A and Level AA Success Criteria, or a Level AA conforming alternate version is provided.*
- **Level AAA:** *For Level AAA conformance, the Web page satisfies all the Level A, Level AA and Level AAA Success Criteria, or a Level AAA conforming alternate version is provided.*

Como podemos apreciar, los niveles de conformidad de las WCAG 1.0 y 2.0 son prácticamente iguales; lo que diferencia a ambas versiones es la exigencia de alcanzar niveles de conformidad AA en un mayor número de criterios de éxito en las WCAG 2.0. respecto a las WCAG 1.0.

4.1.2. Pautas de accesibilidad al contenido audiovisual: WCAG 1.0 y WCAG 2.0

En cuanto a las pautas relacionadas con la accesibilidad al contenido audiovisual, mostramos a continuación la diferencia entre las WCAG 1.0 y las 2.0 para así plasmar la evolución que han experimentado y el aumento de la exigencia en la accesibilidad al contenido desde sus inicios hasta la ahora.

4.1.2.1. Las WCAG 1.0 y la accesibilidad al contenido audiovisual

En el caso de las WCAG 1.0 es la pauta 1, en los puntos de verificación 1.3 y 1.4, donde se recoge cómo hacer accesible el contenido sonoro y visual que puede aparecer en una página web.

⁴⁹ <http://www.w3.org/TR/WCAG20/#conformance>

WCAG 1.0 (1999)
Pauta 1. Proporcione alternativas equivalentes para el contenido sonoro y visual
Punto de verificación 1.3 Hasta que las aplicaciones de usuario puedan leer en voz alta automáticamente el texto equivalente de la banda visual, proporcione una descripción auditiva de la información importante de la banda visual de una presentación multimedia.
Punto de verificación 1.4 Para toda presentación multimedia tempodependiente (por ejemplo, una película o animación) sincronice alternativas equivalentes (por ejemplo, subtítulos o descripciones de la banda visual) con la presentación.

Figura 6. Alternativas equivalentes para el contenido sonoro y visual. WCAG 1.0, 1999.

En los documentos de apoyo⁵⁰ que acompañan a las WCAG 1.0, encontramos que estos dos puntos de verificación de la pauta 1 van asociados a las denominadas técnicas fundamentales (el equivalente a las técnicas de suficiencia de las WCAG 2.0 antes mencionadas) que proporcionan ejemplos y estrategias concretas de aplicación. Algunas de las matizaciones que presentan son las siguientes:

Las descripciones sonoras de la banda visual proporcionan una narración de los elementos visuales claves sin interferir con el sonido o el diálogo de una película. Los elementos visuales clave incluyen acciones, escenarios, lenguaje corporal, gráficos y el texto mostrado. Las descripciones auditivas son utilizadas, primordialmente, por las personas ciegas para seguir la acción y la información no auditiva en el material visual.⁵¹

Como podemos comprobar, las pautas están definiendo la técnica de audiodescripción que en la siguiente versión sí viene referida como tal.

Por otra parte, en cuanto a la pauta 1.3 que define cómo convertir la información sonora en información accesible para un mayor número de personas expone que:

Las presentaciones sonoras deben ir acompañadas por transcripciones del texto, equivalentes textuales de los eventos sonoros. Cuando estas transcripciones se presentan de forma sincronizada con la presentación visual, se denominan subtítulos y son utilizados por las personas que no pueden escuchar la banda sonora del material visual.

Añaden por último que «los subtítulos deberían incluir los diálogos y otros sonidos ambientales que ayuden a los espectadores a entender lo que está ocurriendo». De nuevo, estas aclaraciones acerca de cómo cumplir con el nivel de accesibilidad requerido en cuanto a la información sonora, remiten a una modalidad de traducción audiovisual, la subtitulación para personas sordas (SPS). Es evidente que ya en la primera versión de las pautas la accesibilidad web y la traducción audiovisual se daban la mano, pero es en la segunda versión donde queda especificado al referirse directamente a las técnicas de subtitulación y audiodescripción.

⁵⁰ http://www.discapnet.es/web_accesible/wcag10/WAI-WEBCONTENT-19990505_es.html

⁵¹ http://www.discapnet.es/web_accesible/tecnicas/core/WCAG10-CORE-TECHS-20001106.html#audio-information

4.1.2.2 Las WCAG 2.0 y la accesibilidad al contenido audiovisual

Por su parte, las WCAG 2.0, como comentábamos anteriormente, están estructuradas a partir de cuatro principios básicos. El primer principio, el de Perceptible⁵² indica que «la información y los componentes de la interfaz de usuario deben ser presentados a los usuarios de modo que ellos puedan percibirlos» y es este principio el que define cómo hacer el contenido accesible a partir de cuatro pautas que las WCAG 2.0⁵³ presentan del siguiente modo:

Pauta 1.1. Alternativas textuales: Proporcionar alternativas textuales para todo contenido no textual de modo que se pueda convertir a otros formatos que las personas necesiten, tales como textos ampliados, braille, voz, símbolos o en un lenguaje más simple.

Pauta 1.2. Contenido multimedia dependiente del tiempo: Proporcionar alternativas sincronizadas para contenidos multimedia sincronizados dependientes del tiempo.

Pauta 1.3. Adaptable: Crear contenido que pueda presentarse de diferentes formas (por ejemplo, con una disposición más simple) sin perder información o estructura.

Pauta 1.4. Distinguible: Facilitar a los usuarios ver y oír el contenido, incluyendo la separación entre el primer plano y el fondo.

La pauta 1.2 es la que define cómo hacer accesible el contenido audiovisual y comprende nueve criterios de éxito establecidos según el tipo de contenido. La diferencia frente a las WCAG 1.0 la encontramos en que por una parte, esta segunda versión, como comentábamos en epígrafes anteriores, explicita la técnica que ha de utilizarse para cumplir con los criterios de éxito y, por otra parte, aumenta la exigencia de cumplimiento. En las WCAG 2.0 se distingue entre multimedia pregrabada y en directo, además de incluir la lengua de signos como contenido alternativo. Los cinco primeros criterios de éxito propuestos por la WAI corresponden a niveles de conformidad A y AA por lo que todas las páginas web con contenidos audiovisuales han de cumplirlos, mientras que los últimos cuatro no son de cumplimiento obligatorio. Veamos a continuación los criterios de éxitos de la pauta 1.2 (Medios tempodependientes: proporcionar alternativas para los medios tempodependientes) divididos según los niveles de conformidad⁵⁴. Los primeros tres criterios de conformidad son nivel de conformidad A:

1.2.1. Sólo audio y sólo vídeo (grabado): Para contenido sólo audio grabado y contenido sólo vídeo grabado, se cumple lo siguiente, excepto cuando el audio o el vídeo es un contenido multimedia alternativo al texto y está claramente identificado como tal:

Sólo audio grabado: Se proporciona una alternativa para los medios tempodependientes que presenta información equivalente para el contenido sólo audio grabado.

⁵² <http://www.codexemplar.org/traduccion/pautas-accesibilidad-contenido-web-2.0.htm#text-equiv>

⁵³ <http://www.codexemplar.org/traduccion/pautas-accesibilidad-contenido-web-2.0.htm#text-equiv>

⁵⁴ <http://www.sidar.org/traduccion/wcag20/es/comprender-wcag20/media-equiv-av-only-alt.html>

Sólo vídeo grabado: Se proporciona una alternativa para los medios tempodependientes o se proporciona una pista sonora que presenta información equivalente al contenido del medio de sólo vídeo grabado.

1.2.2. Subtítulos (grabados): Se proporcionan subtítulos para el contenido de audio grabado dentro de contenido multimedia sincronizado, excepto cuando la presentación es un contenido multimedia alternativo al texto y está claramente identificado como tal.

1.2.3. Audiodescripción o Medio Alternativo (grabado): Se proporciona una alternativa para los medios tempodependientes o una audiodescripción para el contenido de vídeo grabado en los multimedia sincronizados, excepto cuando ese contenido es un contenido multimedia alternativo al texto y está claramente identificado como tal.

Con nivel de conformidad AA encontramos:

1.2.4. Subtítulos (en directo): Se proporcionan subtítulos para todo el contenido de audio en directo de los multimedia sincronizados.

1.2.5. Audiodescripción (grabado): Se proporciona una audiodescripción para todo el contenido de vídeo grabado dentro de contenido multimedia sincronizado.

Y finalmente, presentamos los últimos cuatro criterios que completan la pauta 1.2 y cuyo nivel de conformidad es AAA. Recordemos que según la normativa vigente no es necesario que una página web satisfaga estos criterios de conformidad:

1.2.6. Lengua de señas (grabado): Se proporciona una interpretación en lengua de señas para todo el contenido de audio grabado dentro de contenido multimedia sincronizado.

1.2.7. Audiodescripción ampliada (grabada): Cuando las pausas en el audio de primer plano son insuficientes para permitir que la audiodescripción comunique el significado del vídeo, se proporciona una audiodescripción ampliada para todos los contenidos de vídeo grabado dentro de contenido multimedia sincronizado.

1.2.8. Medio alternativo (grabado): Se proporciona una alternativa para los medios tempodependientes, tanto para todos los contenidos multimedia sincronizados grabados como para todos los medios de sólo vídeo grabado.

1.2.9. Sólo audio (en directo): Se proporciona una alternativa para los medios tempodependientes que presenta información equivalente para el contenido de sólo audio en directo.

Los criterios de éxito cuyo nivel de conformidad es A o AA son de aplicación obligatoria según la normativa vigente, es decir, del 1.2.1 a. 1.2.5, de este modo, como se puede apreciar en esta versión, las posibles ambigüedades de las pautas anteriores se desdibujan y directamente se señala la subtitulación y la audiodescripción como alternativa para los vídeos.

En este punto es oportuno presentar la Norma UNE 139803. *Requisitos de accesibilidad para contenidos en la Web (2004 y 2012)*, pues es el instrumento que nos permite transponer la normativa de las WCAG para adecuar las pautas a la legislación española. Esta norma

equivale a las WCAG 2.0 y, en este sentido, en su blog sobre accesibilidad, Olga Carreras⁵⁵ ha generado un recurso de gran utilidad en el que podemos acceder a la correspondencia entre los requisitos de la Norma UNE 139803, los puntos de control de la WCAG 1.0 y los criterios de éxito de las WCAG 2.0⁵⁶. Como explica Carreras (2009)⁵⁷ es una hoja Excel con cuatro pestañas organizada de la siguiente manera:

- a) Pestaña 1: Correspondencia organizada por prioridades tomando como referencia la Norma UNE 139803
- b) Pestaña 2: Correspondencia organizada por prioridades tomando como referencia las WCAG 1.0
- c) Pestaña 3: Correspondencia organizada por prioridades tomando como referencia las WCAG 2.0
- d) Pestaña 4: Correspondencia organizada por orden numérico tomando como referencia las WCAG 1.0

Según Carreras se trata de un recurso muy útil para la revisión de la accesibilidad conforme a las WCAG 2.0. En el caso de los criterios de éxito vinculados a la accesibilidad al contenido audiovisual la correspondencia organizada por prioridades tomando como referencia las WCAG 2.0 sería la siguiente:

	WCAG 2.0	UNE 139803	WCAG 1.0		WCAG 2.0	UNE 139803	WCAG 1.0
P1	1.1.1/1.2.1 1.2.9 (AAA)	4.4.3	1.1	P 2	1.2.2(A)/ 1.2.3(A)/ 1.2.4/ 1.2.5/ 1.2.7(AAA)	4.6.4	1.4
	1.1.1/2.1.1 /2.4.4	4.5.2	9.1		1.4.3/ 1.4.6 (AAA)	4.2.7	2.2(17)
	1.1.1/2.1.1 /2.4.4	4.5.3	1.2		1.4.3/ 1.4.6 (AAA)	4.2.13 (AAA)	2.2 (AAA) (17)
	1.2.2/1.2.3/ 1.2.4 (AA)/ 1.2.5(AA)/ 1.2.7(AAA)	4.6.4	1.4		1.4.4/ 1.4.8(AAA)	4.2.4	3.4

(Figura 7 continúa en la página siguiente)

⁵⁵ <http://olgacarreras.blogspot.com>

⁵⁶ Este documento es de acceso libre y de descarga gratuita y lo podemos encontrar en https://www.usableyaccessible.com/recurso_descargas.php

⁵⁷ <https://olgacarreras.blogspot.com/2009/03/correspondencia-entre-los-requisitos-de.html>

(Figura 7 continúa de la página anterior)

	WCAG 2.0	UNE 139803	WCAG 1.0		WCAG 2.0	UNE 139803	WCAG 1.0
	1.2.3/ 1.2.5 (AA)/ 1.2.7 (AAA)	4.6.3	1.3		1.4.5/ 1.4.9 (AAA)	4.2.6	3.1
	1.3.1/ H51/ H53	4.3.2	5.1		2.4.5/G63	4.5.7	13.3
	1.3.1/H43	4.3.3	5.2		2.4.6	N.C.E	N.C.E
	1.3.1 /H48	4.3.7 (AA)	3.6 (AA)		2.4.7	N.C.E	N.C.E
	1.3.1/ F43	4.2.9 (AA)	5.4 (AA)		3.1.2	4.4.5 (A)	4.1 (A)
	1.3.1 / 4.1.2/ H44	4.4.7(AA)	12.4(AA)		3.2.3/3.2.4/ 2.4.10(AAA)	4.5.8	13.4
	1.3.1/ 2.4.1 (10)	4.5.13 (AAA)	13.6 (AAA)		3.2.3 / 3.2.4	4.2.12	14.3
	1.3.1 / 2.4.10 (AAA)	4.3.5(AA)	3.5(AA)		3.3.3	N.C.E	N.C.E
	1.3.1 /F43	4.3.8(AA)	3.7(AA)		3.3.4	N.C.E	N.C.E
	1.3.2	4.2.8(AA)	5.3(AA)		N.C.E	4.4.6	6.5
	1.3.3	N.C.E	N.C.E		N.C.E (2)	4.2.5	3.3
	1.4.1	4.2.2	2.1		D.S.R	4.3.4	12.2
	1.4.2	N.C.E	N.C.E		D.S.R	4.1.1	11.1

Figura 7. Correspondencia entre WCAG y normas UNE. Olga Carreras.

La autora resalta en la tabla en verde los cambios de prioridad de la Norma UNE respecto a las WCAG 1.0; en azul, nuevos criterios de las WCAG 2.0 y con la leyenda N.C.E indica que el criterio no tiene correspondencia exacta mientras que con D.S.R señala que deja de ser requisito.

Una vez constatadas las técnicas que se han de aplicar tanto por exigencia de las WCAG y la norma UNE lo que nos preguntamos ahora es en qué medida las WCAG 2.0 establecen cómo aplicar estas técnicas y cómo evaluarlas.

Y esta es la pregunta que nos lleva a las guías de buenas prácticas que proponen las WCAG 2.0 para elaborar los contenidos alternativos propuestos (subtitulado y audiodescripción), y que abordaremos un poco más adelante.

En este sentido, a diferencia de las pautas, no existe normativa específica sobre la inclusión de estos contenidos en Web (Moreno *et al.* 2008:44). Sin embargo, en el caso de España sí que disponemos de documentos que regulan cómo han de generarse estos contenidos como es el caso de la Norma UNE 153010:2012 *Subtitulado para personas sordas y personas con discapacidad auditiva* y la Norma UNE 153020 *Audiodescripción para personas con discapacidad visual*. En este sentido, Voces Merayo (2009:250) apunta que se cuenta con una experiencia dilatada en la aplicación de la SPS y la AD en otros ámbitos audiovisuales (DVD, TV, cine) y que las normas técnicas UNE desarrolladas en España para estos ámbitos son también guías de utilidad para desarrolladores web. Por otra parte, Voces Merayo (2009:251) señala que «sería interesante analizar más en profundidad estas normativas en el ámbito de la web». Y eso es precisamente lo que nosotros pretendemos hacer en este trabajo. Examinar, por una parte, las recomendaciones y guías propuestas por la WAI y compararlas con las Normas UNE y, por otra parte, analizar si se cumplen los criterios de conformidad correspondientes al contenido audiovisual incluyendo las recomendaciones y normativas. Así, en primer lugar realizaremos una evaluación inicial para determinar si el contenido audiovisual cumple con los criterios de éxito y, seguidamente, realizaremos un análisis en mayor profundidad para determinar en qué medida se tienen en cuenta las recomendaciones (guías de estilo sugeridas desde la WAI) y las normas técnicas.

Por otra parte, la accesibilidad es un proceso complejo en el que participan distintos agentes y es trascendental entender cómo funciona y de qué manera los responsables de los contenidos (audiovisuales o no) son parte del mismo. En el siguiente apartado presentamos la imbricación de las distintas partes en un todo llamado accesibilidad web.

4.1.2.3. La cadena de accesibilidad web: perfiles profesionales

Crear un sitio web accesible y diseñado para todos requiere, por una parte, de la implicación de diversos agentes en el proceso y, por otra parte, de la coordinación de todos estos profesionales que participan en la creación, desarrollo y actualización del sitio. Nos referimos a lo que desde el W3C han denominado cadena de accesibilidad web, que dada su importancia e impacto en la creación de sitios web accesibles ocupa un lugar preferente en las propuestas del Consorcio.

En este sentido, el Consorcio ha redactado el documento colaborativo *Accessibility Responsibility Breakdown*⁵⁸ en el que se detallan los profesionales implicados en el ciclo de

⁵⁸ http://www.w3.org/community/wai-engage/wiki/Accessibility_Responsibility_Breakdown

vida de un proyecto web y el criterio de conformidad del que es responsable cada profesional para su desarrollo y posterior evaluación. El documento define los diez roles profesionales que han de participar en el desarrollo de una web y distribuye los 61 criterios de conformidad entre los diferentes roles profesionales. Esta iniciativa surge por la necesidad de ubicar en el lugar adecuado la accesibilidad en el diseño y creación web. El documento apunta en este sentido que:

Accessibility is not simply an extra requirement that can be added to the web development lifecycle. In order to show significant conformance results, accessibility must be incorporated in each existing link of the web production chain. The only way to successfully accomplish accessibility is to be able to plan it from the very start, and that means assigning responsibility to the various stakeholders and share the tasks in order to produce accessible content.

A continuación, mostramos la figura que aparece en el documento mencionado con los roles y funciones asignados a cada uno de los expertos que según el Consorcio forman parte de la cadena de accesibilidad:

STAKEHOLDERS IN A TYPICAL WEB PRODUCTION CHAIN	
Project management	For the purposes of this project, the project management role includes the tasks normally associated with production planning and the various related monitoring activities. While the project manager (PM) does not really have any implication when it comes to implementing the different Success Criteria from WCG 2.0, he or she plays a vital role in making sure every stakeholder understands what their role is when it comes to web accessibility.
Analysis	The analysis function covers tasks and related quality control normally associated with analysis of the project's strategic orientations, analysis of the options for technology platforms, or functional analysis of Web interfaces.
Architecture	The architecture function covers tasks and related quality control normally associated with the architecture of the information (Web content) and the architecture of the data.
Interaction Design/ Usability	The interaction design / usability function covers tasks and related quality control normally associated with the planning of web interfaces, content changes, interactivity and other interface-related contents of the pages
Graphic Design	The graphics design function covers tasks and related quality control normally associated with the graphic design of interfaces, the related graphic declinations, the specific design of navigation elements, context changes and other general design of the main content of the pages.
Content Strategy	The content strategy function covers tasks and related quality control normally associated with producing the site's textual contents, equivalent alternative for non-text content and other general text elements presented in the pages.
Search Engine Optimization	The search engine optimization function covers tasks and related quality control normally associated with providing text equivalents for non-text contents and making contents on a web page more easily indexable by search engines.

(Figura 8 continúa en la página siguiente)

(Figura 8 continúa de la página anterior)

STAKEHOLDERS IN A TYPICAL WEB PRODUCTION CHAIN	
HTML/CSS Prototyping	The HTML/CSS prototyping function covers tasks and related quality control normally associated with the production of all web site master templates (HTML and CSS).
Font-End Development	The front-end development function covers tasks and related quality control normally associated with the development of contribution tools, HTML and CSS integration, and the programming of proposed scripts and applications on the Web site.
Back-End Development	The back-end development function covers tasks and related quality control normally associated with the development of server side programming and database management.

Figura 8. Agentes que participan en la cadena de producción web.

Entre los roles profesionales que vienen definidos en la tabla precedente encontramos, en sexto lugar, el del profesional encargado de la *Content Strategy*. Este agente tiene bajo su responsabilidad 21 criterios de conformidad y su función está relacionada con la creación del contenido textual, las alternativas textuales y cualquier otro elemento textual que pueda aparecer en la web.

Principles	Applicable Success Criteria		
	A	AA	AAA
Perceivable	1.1.1 ↗ , 1.2.1 ↗ , 1.2.2 ↗ , 1.2.3 ↗ , 1.3.1 ↗ , 1.3.3 ↗	1.2.5 ↗	1.2.7 ↗ , 1.2.8 ↗
Operable	2.1.1 ↗ , 2.1.2 ↗ , 2.4.2 ↗ , 2.4.4 ↗	2.4.6 ↗	2.4.9 ↗
Understandable	3.3.1 ↗	3.1.2 ↗	3.1.3 ↗ , 3.1.4 ↗ , 3.1.5 ↗ , 3.1.6 ↗
Robust	--	--	--
Total (21)	11	3	7

Figura 9. Tabla de responsabilidades del Content Strategy Stakeholder.

Como mostramos en la figura anterior, entre estos 21 criterios de conformidad nueve de ellos se encuentra en el principio Perceptible, como por ejemplo los criterios 1.2.2 y 1.2.3, por lo que las alternativas textuales en forma de subtítulos y audiodescripción y el posterior control de calidad son responsabilidad de este profesional.

En este apartado que ahora concluimos se ha puesto de relevancia, no sólo la importancia del trabajo del traductor en el proceso de creación, desarrollo y evaluación de contenidos audiovisuales en la web y de su recomendable inclusión en el mismo, sino también el hecho que desde el WAI propongan entre los actores que participan en el proceso de creación de un sitio web, un profesional responsable de estos contenidos y de cómo evaluarlos, el *Content Strategy Stakeholder*. En nuestra opinión, esta figura debería desplegarse en dos e incorporar un profesional adicional de la traducción especializado en accesibilidad audiovisual con los

conocimientos y técnicas necesarias para generar estas alternativas textuales que son propias del campo de la traducción.

Asimismo, sería muy recomendable, en virtud de la calidad de la SPS y de la AD, ampliar o profundizar en la evaluación de los contenidos generados como alternativa textual, responsabilidad que recaería asimismo sobre esta nueva figura que proponemos incorporar en la cadena de accesibilidad web (y que, repetimos sería un profesional de la traducción). Cómo evaluar la calidad de la SPS y la AD en la web es también uno de los retos que asume este trabajo; cuestión que abordaremos en el capítulo 5.

En suma, en este apartado hemos descrito cómo la web tiene suficientes mecanismos para convertirse en un espacio accesible para todos y de qué manera las técnicas propuestas por las WCAG 1.0, las WCAG 2.0 y la Norma UNE 13980 ilustran la obligación de proporcionar alternativas equivalentes en los elementos multimedia que aparecen en las páginas web. Es decir, los contenidos alternativos que se proporcionan tienen que estar sincronizados y las técnicas de subtítulo y audiodescripción son las que permiten alcanzar esta exigencia. El vínculo entre accesibilidad web y accesibilidad audiovisual queda en este capítulo claramente definido.

En el siguiente capítulo pondremos el foco en cómo podemos determinar (evaluar) la accesibilidad audiovisual en la web según los organismos competentes en esta materia, y qué mecanismos de control existen para determinar si los sitios web cumplen los criterios de éxito relacionados con los contenidos audiovisuales.

CAPÍTULO 5

EVALUACIÓN DE LA ACCESIBILIDAD AUDIOVISUAL EN LA WEB

5.1. Evaluación de la accesibilidad del contenido audiovisual en la web

5.1.1. *How to meet WCAG 2.0*: definición y evaluación de los criterios de éxito

5.1.2. Modelo de evaluación adecuado a nuestro ámbito de estudio

CAPÍTULO 5. EVALUACIÓN DE LA ACCESIBILIDAD AUDIOVISUAL EN LA WEB

En el capítulo que introducimos ahora abordaremos la evaluación de la accesibilidad web, en particular la evaluación de los criterios de éxito vinculados al contenido audiovisual que describimos en el capítulo anterior. Explicaremos de qué manera han de implantarse estos contenidos alternativos según las WCAG, es decir, qué normas gobiernan esta práctica en la subtitulación online y cómo analizar si se cumplen los criterios de éxito. Asimismo, abordaremos las propuestas de evaluación y análisis que plantean expertos y organismos competentes en esta materia, para a partir de estas, definir nuestra propuesta de análisis que nos permita describir si los contenidos son accesibles según los criterios de éxito pautados en las WCAG.

5.1. Evaluación de la accesibilidad del contenido audiovisual en la web

Existen diferentes herramientas de evaluación automática y semiautomática⁵⁹ que permiten comprobar el grado de accesibilidad de una página web y determinar el nivel de adecuación de forma rápida y eficaz. Serrano (2008:246) las define de manera genérica como «un *software* específicamente creado con el fin de evaluar de manera automática la calidad de las páginas web, en cuanto a la accesibilidad de sus contenidos». Sin embargo, estas herramientas, que sí son de gran utilidad para comprobar el nivel de accesibilidad de gran parte de los criterios de conformidad, no permiten evaluar de forma automática si los problemas de perceptibilidad asociados a un contenido audiovisual han sido afrontados siguiendo las recomendaciones de la WAI.

La evaluación de la accesibilidad de un sitio web es una tarea compleja y a la que la WAI, por su importancia, ha dedicado no pocos recursos y tiempo. En julio de 2014, el W3C/WAI publicó la Metodología de evaluación de conformidad con la accesibilidad en sitios web (WCAG-EM)⁶⁰ con la que pretende armonizar la evaluación de sitios web

⁵⁹ <http://www.w3.org/WAI/ER/tools/index.html>

⁶⁰ <https://www.w3.org/WAI/eval/conformance.html>

de acuerdo con las WCAG 2.0. Esta metodología permite evaluar de forma exhaustiva un sitio web por lo que combinan el uso de herramientas de evaluación automáticas y la evaluación manual. La metodología es amplia, concienzuda y de gran interés, en especial, para profesionales del ámbito web, pero sin duda demasiado ambiciosa para los objetivos de nuestro trabajo.

Sin embargo, la WCAG-EM incluye otra metodología menos compleja y exhaustiva que ofrece la opción de realizar una evaluación previa menos exigente que aporta una visión global de la accesibilidad del sitio y que se ajusta mejor a nuestras necesidades. Nos referimos al *Easy Checks- A First Review of Web Accessibility*⁶¹ en la que se señalan los fallos más elocuentes de un sitio web y que define su función de la siguiente manera:

This page helps you start to assess the accessibility of a web page. With these simple steps, you can get an idea whether or not accessibility is addressed in even the most basic way.

Teniendo en cuenta esta filosofía y el papel central del organismo que la propone junto con el objetivo que persigue, entendemos que su estructura encaja sobradamente para tomar como referencia este sistema de evaluación. Nuestro interés se centra en evaluar la perceptibilidad del contenido audiovisual por lo que profundizaremos en este apartado de la evaluación y lo vincularemos al sistema de evaluación de Voces Merayo (2008), que presentaremos un poco más adelante, para así comenzar a perfilar los primeros pasos del sistema de evaluación preliminar con el que determinar si los contenidos son accesibles o no.

A continuación, pasamos a explicar cómo, a través del enlace *Easy Checks- A First Review of Web Accessibility*, se puede efectuar una primera evaluación de los contenidos multimedia. En primer lugar, constatamos que la evaluación es manual lo que puede ser un escollo añadido a la evaluación de contenidos, puesto que hay que invertir una cantidad de tiempo considerable desalentando así la consecución de esta tarea. En segundo lugar y en referencia a la subtitulación, indica que el primer paso consiste en determinar si el vídeo dispone de subtítulos (advierten que los subtítulos generados de forma automática no son suficientes para que el vídeo sea considerado accesible) y, en caso afirmativo, entonces ha de examinarse de forma manual si:

- (1) The captions seem in sync with the spoken content
- (2) The people who are speaking are identified when they speak

⁶¹ <https://www.w3.org/WAI/test-evaluate/preliminary/>

- (3) Important sound other than dialogue –e.g., footsteps approaching, doors closing, glass breaking– is included.

Aunque pudiera parecer una forma simplista de abordar la subtitulación es relativamente exigente puesto que, recordemos, es según indica este sistema de evaluación un proceso de evaluación manual.

Por su parte, en el caso de la audiodescripción la metodología indica que ha de describirse la información visual importante y pone como ejemplo los títulos que aparecen escritos en la imagen, el nombre del orador/locutor y las ilustraciones. De igual forma, recomienda una serie de materiales y guías para generar audiodescripciones.

Podríamos resumir las acciones que permiten validar el criterio de éxito de la siguiente forma:

Subtitulación	Audiodescripción
<ol style="list-style-type: none"> 1. Comprobar que los subtítulos están sincronizados con la imagen. 2. Comprobar que los personajes están identificados. 3. Comprobar la inclusión de información sonora relevante (pasos, cristales rotos, etc.). 	<ol style="list-style-type: none"> 1. Comprobar que la pista de audio del vídeo incluye la descripción de la información visual importante.

En el caso de la subtitulación entendemos que evalúa algunos de los aspectos más definitorios de esta modalidad de traducción si bien deja de lado, en este primer análisis, la inclusión de la información suprasegmental, la segmentación de los subtítulos y los parámetros técnicos en particular. Sin embargo, como veremos en próximos apartados, en las guías de subtitulación recomendadas en las WCAG, sí vienen recogidos estos aspectos que nosotros también incluiremos en nuestra propuesta de análisis por ser, a nuestro entender, fundamentales.

En el caso de la audiodescripción nos parece un análisis absolutamente insuficiente y poco preciso, que no da cuenta de aspectos fundamentales: ¿cuál es la información visual importante? ¿Cuándo se ha de incluir esa información? y una larga batería de preguntas sin resolver que convierten la evaluación de la calidad de esta técnica en un gran interrogante.

5.1.1. *How to meet WCAG 2.0*: definición y comprobación de los criterios de éxito

Por otra parte y de gran importancia en nuestro trabajo, el *Easy Checks- A First Review of Web Accessibility* nos remite al documento *How to meet WCAG 2.0*⁶² que proporciona toda la información necesaria para entender el criterio en sí mismo. Este último documento es un guía personalizable que nos permite navegar de forma sencilla por la información relacionada con cada criterio y acceder a toda la información necesaria para entenderlo, implementarlo y evaluarlo.

En este sentido, recordemos que el documento *How to meet WCAG 2.0*, descrito en el capítulo anterior, propone un test para determinar si se ha alcanzado el criterio de éxito en aquellos casos que sólo puede ser analizado de forma manual.

Figura 10. Interfaz de *How to meet WCAG 2.0*.

En esta primera captura podemos apreciar de qué manera se puede personalizar la búsqueda y lo sencillo que resulta acceder a la información que necesitamos. Este documento hipertextual centraliza toda la información y aligera enormemente el proceso de búsqueda de los criterios de conformidad.

Una vez accedemos a los criterios observamos que cada uno ofrece información con relación a las técnicas suficientes, técnicas complementarias y los fallos con los que está vinculado el criterio. A continuación y a modo de ejemplo, mostramos una captura de pantalla de la información que se muestra en el criterio 1.2.2.

⁶² https://www.w3.org/WAI/WCAG21/quickref/?versions=2.0&showtechniques=312¤tsidebar=%23col_customi-ze#readable

Selected filters: all success criteria and all techniques. [Clear filters](#) [Expand all sections](#) [Share](#)

1.2.2 Captions (Prerecorded) — Level A

Captions are provided for all prerecorded audio content in synchronized media, except when the media is a media alternative for text and is clearly labeled as such. [Understanding 1.2.2](#)

Hide techniques and failures for 1.2.2 Sufficient Advisory Failures

Sufficient Techniques

Note: Other techniques may also be sufficient if they meet the success criterion. See [Understanding Techniques](#).

- G93: Providing open (always visible) captions
- G87: Providing closed captions
- G87: Providing closed captions *using one of the following techniques*:
 - SM11: Providing captions through synchronized text streams in SMIL 1.0
 - SM12: Providing captions through synchronized text streams in SMIL 2.0
 - H95: Using the track element to provide captions
 - FLASH9: Applying captions to prerecorded synchronized media
 - SL16: Providing Script-Embedded Text Captions for MediaElement Content
 - SL28: Using Separate Text-Format Text Captions for MediaElement Content

Advisory Techniques

- Providing a note saying "No sound is used in this clip" for video-only clips ([future link](#))
- Using SMIL 1.0 to provide captions for all languages for which there are audio tracks ([future link](#))
- Using SMIL 2.0 to provide captions for all languages for which there are audio tracks ([future link](#))

Failures

- F8: Failure of Success Criterion 1.2.2 due to captions omitting some dialogue or important sound effects
- F74: Failure of Success Criterion 1.2.2 and 1.2.8 due to not labeling a synchronized media alternative to text as an alternative
- F75: Failure of Success Criterion 1.2.2 by providing synchronized media without captions when the synchronized media presents more information than is presented on the page

[SHARE](#) | [BACK TO TOP](#)

Figura 11. Selección de filtros criterio de éxitos 1.2.2.

Por otra parte, advertimos que en la esquina superior derecha hay una pestaña «Understanding 1.2.2» que nos remite al documento *Understanding WCAG 2.0* en el que se explica en detalle por qué es importante aplicar este criterio y qué colectivos se benefician de su implementación. Encontraremos también ejemplos de cómo generar subtítulos y guías de referencia para, en este caso, la subtitulación.

Si queremos ampliar información sobre las técnicas suficientes vinculadas con un criterio de conformidad, al hacer clic sobre el vínculo accedemos a toda la información relacionada con esta técnica.



[Techniques for WCAG 2.0](#)

Contents
Intro
Previous: Technique G86
Next: Technique G88

G87: Providing closed captions

Important Information about Techniques

See [Understanding Techniques for WCAG Success Criteria](#) for important information about the usage of these informative techniques and how they relate to the normative WCAG 2.0 success criteria. The Applicability section explains the scope of the technique, and the presence of techniques for a specific technology does not imply that the technology can be used in all situations to create content that meets WCAG 2.0.

On this page:

- [Important Information about Techniques](#)
- [Applicability](#)
- [Description](#)
- [Examples](#)
- [Resources](#)
- [Related Techniques](#)
- [Tests](#)

Figura 12. Cómo incluir subtítulos WCAG 2.0.

Todas las técnicas suficientes ofrecen el mismo tipo de información estructurada en los siguientes apartados: *important information about the techniques, applicability, description, examples, resources, related techniques and tests*. Los apartados los podemos encontrar en el menú derecho, como podemos apreciar en la figura anterior. Para nuestro trabajo los apartados más importantes son la descripción de la técnica, los ejemplos, las guías de buenas prácticas recomendadas por la WCAG y el test para comprobar que la técnica ha sido implementada con éxito.

Sobre cómo evaluar si la página web satisface el criterio 1.2.2 *Subtítulos (grabados)* establecen un sencillo protocolo de tres pasos y en caso de no cumplir con el paso 2 o 3 no superaría el test de accesibilidad de este criterio:

1. Visualizar todo el material con los subtítulos activados
2. Comprobar que todo diálogo va acompañado de subtítulos.
3. Comprobar que toda la información sonora relevante está subtitulada.

Y en cuanto a los Fallos que hemos de evitar dado que son entendidos como barreras de la comunicación (puesto que dejaría de ser contenido perceptible), en la siguiente captura de pantalla resaltamos describe en qué consiste el fallo F8.⁶³

Description

This describes a failure condition for all techniques involving captions. If the "caption" does not include all of the dialogue (either verbatim or in essence) as well as all important sounds then the 'Captions' are not real captions.

NOTE: Captions sometimes simplify the spoken text both to make it easier to read and to avoid forcing the viewer to read at very high speed. This is standard procedure and does not invalidate a caption.

Examples

Failure Example 1

Examples of text streams that are not captions include:

- text that contains the dialogue (possibly simplified dialogue) but that does not describe important sounds
- text that omits dialogue during portions of the material

Figura 13. Ejemplo de causas para no alcanzar criterio de éxito.

⁶³ <https://www.w3.org/WAI/WCAG21/Techniques/failures/F8>

Nos parece muy interesante resaltar la información que viene en la nota de la descripción del fallo (F8). En ella se sobreentiende que los subtítulos no se pueden presentar a velocidades de lectura rápida porque impediría la lectura cómoda de los mismos. La información relativa a los parámetros técnicos de la subtitulación es absolutamente obviada en los criterios de éxito, sin embargo, encontramos información al respecto en las guías de referencia propuestas por la WCAG, pero como decíamos antes no así en las Pautas. Abordaremos en mayor profundidad las guías de buenas prácticas en relación a los parámetros técnicos en el capítulo 6.

En el caso de la audiodescripción se establece también un protocolo de tres pasos, y en caso de no cumplir con el primer o tercer paso, no superaría el test de accesibilidad de este criterio:

1. Comprobar que el vídeo tiene una pista de audio que incluye descripciones.
2. Escuchar la pista mientras se visualiza el vídeo.
3. Comprobar que los huecos de silencio contienen la descripción de la información visual importante.

Los documentos que nos permiten describir si los contenidos multimedia cumplen con las normas establecidas (criterios de éxito): *Easy Checks- A First Review of Web Accessibility* y *How to meet WCAG 2.0* se diferencian primordialmente en que el primero ofrece una serie de pasos que conforman una evaluación preliminar, mientras que el segundo procura información más detallada de cómo aplicar la técnica. Tomaremos ambos documentos de referencia para configurar una evaluación preliminar.

Como hemos visto el protocolo de evaluación manual propuesto en el *Easy Checks- A First Review of Web Accessibility* y *How to meet WCAG 2.0* permite determinar si la técnica indicada en el criterio se ha aplicado, es decir, si el contenido audiovisual viene acompañado de una subtitulación o audiodescripción, pero no nos permite detectar si la calidad del subtítulo o del texto audiodescrito es idóneo para alcanzar el objetivo de su implementación.

5.1.2. Modelo de evaluación adecuado a nuestro ámbito de estudio

Con el objetivo de crear un modelo de evaluación que le permitiera determinar el grado de cumplimiento en materia de accesibilidad de los contenidos audiovisuales en la web, Voces Merayo (2010:251) realiza una propuesta pionera en este ámbito que abordaremos

a continuación. Asimismo, el autor, señala en su estudio que las normas técnicas UNE en materia de accesibilidad audiovisual son instrumentos de gran ayuda para los desarrolladores y señala que sería interesante analizar en profundidad su grado de aplicación en la web.

Nuestra intención es atender a la propuesta de Voces Merayo (*ibid.*) y analizar en mayor profundidad algunos aspectos de la Norma UNE de SPS en la web (dejaremos el análisis de la AD para futuros estudios) y proponer un sistema de evaluación que nos permita describir las características de los subtítulos a partir de estas normas y otras guías recomendadas por el Consorcio.

El modelo de evaluación de la accesibilidad audiovisual de Voces Merayo (2010:256) está compuesto por tres elementos que representan distintos niveles de concreción: módulos, parámetros e indicadores. El sistema de evaluación está dividido en dos módulos: (1) accesibilidad lógica del reproductor y (2) accesibilidad del contenido audiovisual; y seis parámetros, que ocupan un segundo nivel de concreción y son los que definen qué se va a evaluar:

1. Para la accesibilidad lógica del reproductor:
 - (a) Accesibilidad técnica
 - (b) Operabilidad del reproductor

2. Para la accesibilidad del contenido audiovisual propone los cuatro principios que definen las WCAG 2.0
 - (a) Perceptibilidad
 - (b) Operabilidad
 - (c) Inteligibilidad
 - (d) Robustez

En el último nivel de concreción, los indicadores fijan cómo se evalúa cada parámetro de cada módulo. A modo de ejemplo, mostramos a continuación los campos de cada indicador en la metodología de evaluación de Voces Merayo (2010:257).

Definición	Una frase que defina sin ambigüedad.
Ejemplo	Ejemplo de un enlace de un contenido con subtitulación.
Examen	Existe subtitulación.
Procedimiento	Se reproduce el contenido y se activa la subtitulación en el caso de que exista.
Resultado	El significado de los valores es: PARA SORDOS BÁSICA NO

Seguiremos este sistema de evaluación y utilizaremos estas fichas en nuestro análisis preliminar puesto que aglutina las acciones de control planteadas en el *Easy Checks- A First Review of Web Accessibility* y *How to meet WCAG 2.0*, si bien es cierto que se hace necesario adaptarlas ligeramente por las razones que desvelamos a continuación.

En primer lugar, dado que nuestro estudio está centrado en las técnicas que procuran que el contenido audiovisual en sí sea accesible pasamos a detallar únicamente los parámetros e indicadores para la evaluación de la accesibilidad del contenido.

El autor considera que los parámetros de Perceptibilidad e Inteligibilidad contienen indicadores relacionados con estos contenidos. En relación al parámetro perceptibilidad propone tres indicadores: (1) subtitulación, (2) audiodescripción y (3) lengua de signos. La definición de los indicadores de este parámetro consiste en concretar si el contenido audiovisual incorpora subtitulación, audiodescripción o lengua de signos, y el procedimiento para verificarlo consiste en reproducir manualmente el contenido y constatar si se activa o no.

A continuación mostramos la ficha de cómo evaluar si el contenido incorpora subtítulos de Voces Merayo (2010:268).

Definición	Se evalúa si el contenido audiovisual incorpora, de algún modo, una subtitulación.
Ejemplos	El contenido que se presenta en el NCAM http://ncam.wgbh.org/webaccess/ccforflash/ccplayermain.html incluye una subtitulación.
Examen	¿Existe una subtitulación en el contenido audiovisual?
Procedimiento	Se reproduce el contenido y se activa la subtitulación en el caso de que exista.
Resultado	<p>PARA SORDOS BÁSICA NO</p> <p>El significado de los valores es:</p> <p>PARA SORDOS: existe una subtitulación en la que se reproducen todos los efectos sonoros y suprasegmentales que aparecen en el contenido.</p> <p>BÁSICA: existe una subtitulación en la que se transcriben los diálogos y narraciones que aparecen en el contenido.</p> <p>NO: carece de subtitulación</p>

Figura 14. Ficha de evaluación subtítulos.

Sin embargo, el resultado de la evaluación, recogida en el campo Resultado no es coherente con la pregunta que se ha de responder en el *Examen* ni con el *Procedimiento* en sí. La pregunta es cerrada: ¿existe subtitulación? Y el examen consiste en reproducir el vídeo y comprobar si se activan los subtítulos, pero los valores planteados en el resultado no se pueden comprobar con ese procedimiento.

Por otra parte, nos parece destacable que haya incluido estos valores puesto que si nos fijamos en el significado observamos que coinciden en gran parte con la propuesta de evaluación preliminar del documento *Easy Checks- A First Review of Web Accessibility* y *How to meet WCAG 2.0* y nos son de gran utilidad. Como podemos observar el valor subtitulación BÁSICA se refiere a una transcripción de diálogos y transcripciones; el valor subtitulación PARA SORDOS ha de contener información como «la identificación del parlante, los efectos de sonido y otros elementos suprasegmentales» (Voces Merayo, 2010:119).

Por su parte, el parámetro Inteligibilidad propuesto por Voces Merayo (2010) contiene a su vez tres indicadores: (1) sigue las recomendaciones de audiodescripción UNE 153010, (2) sigue las recomendaciones de subtitulación UNE 153020 y (3) sigue las recomendaciones de lengua de signos UNE 139804. Estos indicadores evalúan el grado de cumplimiento de estas normas y dependiendo de si el subtítulo cumple con todas, algunas o ninguna de las recomendaciones recogidas en la norma, considera que el contenido es accesible o parcialmente accesible o si por el contrario no es accesible.

Por una parte, consideramos que, por coherencia con las WCAG 2.0, los indicadores del parámetro *Inteligibilidad* se incluyen en el parámetro *Perceptibilidad* y suponen simplemente un nivel de concreción más dentro de ese indicador. Por otra parte, analizar si la subtitulación y audiodescripción siguen todas las recomendaciones de las normas UNE requiere, a nuestro entender, de un análisis muchísimo más extenso y necesariamente en combinación con las guías de buenas prácticas sugeridas desde la WAI.

Por último, Voces Merayo (2010:280) genera unas fichas de evaluación que le permiten recoger los resultados del análisis y determinar si el contenido es accesible o no.

Accesibilidad contenido audiovisual	
Perceptibilidad	
Subtitulación	PARA SORDOS BASICA NO
Audiodescripción	SI NO
Lenguaje de signos	SI NO
Operabilidad	
Funciones accesibles por teclado	SI NO PARCIALMENTE
Descripción de los elementos de interacción	SI NO
Activación/desactivación recursos accesibilidad	VISIBLE INVISIBLE NO
Inteligibilidad	
Cumple normativa 153010	SI NO PARCIALMENTE
Cumple normativa 153020	SI NO PARCIALMENTE
Cumple normativa 139804	SI NO PARCIALMENTE
Robustez	
Identificación técnica	[FORMATO] [CÓDEC AUDIO] [CÓDEC VIDEO]
Formato compatible	SI NO
Códecs compatibles	SI NO
Evaluación de la accesibilidad del contenido	
Valoración	ACCESIBLE NO ACCESIBLE PACIALMENTE ACCESIBLE

Figura 15. Ficha evaluación al contenido de Voces Merayo (2010:280).

La parte relativa a la accesibilidad del contenido audiovisual en sí se encuentra en el parámetro perceptibilidad e inteligibilidad; sin embargo, por las razones indicadas en el epígrafe anterior, nosotros evaluaremos con esta metodología la accesibilidad al contenido que exigen los criterios de éxito obligatorios para procurar la accesibilidad, es decir, la subtitulación y la audiodescripción.

Es cierto, sin embargo, que como apunta el autor (2010:281) su modelo de evaluación es:

(...) a diferencia de otros sistemas que se basan en una lista de preguntas, se ha definido de forma completa y sin ambigüedad cada uno de los indicadores, de manera que: (1) se trata de un análisis que se puede replicar por parte de otros investigadores: (2) se puede utilizar por parte de equipos de trabajo sin riesgo de confusión [...]

Nos parece un modelo de gran utilidad y lo replicaremos en el estudio previo para determinar si nuestro material dispone de subtitulación, SPS y AD. Por su parte, dado que las WCAG 2.0 no establecen que la LS sea de obligada aplicación, no la tendremos en cuenta en nuestro análisis.

Las fichas que generamos para evaluar la accesibilidad del contenido audiovisual son:

Ficha 1

Tabla 1. Ficha de evaluación contenido con subtítulo.

Definición	Se evalúa si el contenido incorpora subtítulos.
Examen	Se comprueba si existe subtitulación del contenido audiovisual.
Procedimiento	Se ejecuta un <i>script</i> que descarga todos los vídeos y registra cuáles disponen de un archivo SRT con la subtitulación.
Resultado	El significado de los valores es: SÍ NO

Ficha 2

Tabla 2. Ficha de evaluación contenido con subtítulos convencionales o para personas sordas.

Definición	Se evalúa si los subtítulos son convencionales o SPS.
Examen	Se comprueba si los subtítulos incorporan información sonora relevante .
Procedimiento	Analizar todos los archivos SRT y examinar si hay información sonora relevante entre paréntesis o corchetes con CQPweb. Examinar si hay información sonora relevante en mayúsculas con CQPweb.
Resultado	SUBTITULACIÓN CONVENCIONAL SPS El significado de los valores es: SUBTITULACIÓN CONVENCIONAL: si los subtítulos únicamente recogen la información textual. SPS: si los subtítulos incorporan (entre paréntesis o corchetes) información sonora relevante.

Ficha 3

Tabla 3. Ficha de evaluación contenidos con audiodescripción.

Definición	Se evalúa si el contenido incorpora audiodescripción.
Examen	Se comprueba si existe audiodescripción del contenido audiovisual.
Procedimiento	Examinar la web y el HTML en busca de enlaces o elementos de vídeo.
Resultado	El significado de los valores es: SÍ NO

En este apartado hemos efectuado un repaso de las propuestas de evaluación vertidas desde la WAI y las hemos alineado con la propuesta de evaluación de contenidos audiovisuales de Voces Merayo (2008). Esto nos ha permitido generar unas fichas de evaluación para determinar (según los criterios establecidos en las WCAG 2.0) en qué medida los vídeos del canal de televisión del Parlamento Europeo⁶⁴, a partir del cual hemos conformado nuestro

⁶⁴ <https://multimedia.europarl.europa.eu/en/home?referer=%22www.europarltv.europa.eu%22>

corpus, son accesibles. Por tanto, el estudio previo de nuestro análisis se basa en un sistema de evaluación anteriormente testado (en otro ámbito) y que hemos adecuado a nuestras necesidades.

Por otra parte, nuestro estudio central, tratará el análisis de dos parámetros (técnicos) que incluyen, en nuestro estudio, siete variables. Por tanto, en el siguiente capítulo, realizaremos un repaso teórico de cada uno de estos aspectos y expondremos las recomendaciones vertidas desde la academia y especialmente desde el sector profesional. Uno de los objetivos de este trabajo es describir en qué medida los parámetros que analizamos están alineados con las normativas y guías de buenas prácticas que detallamos en el siguiente capítulo. Por ende, la intención es averiguar el grado de adhesión a las normas prescriptivas que encontramos en las guías y recomendaciones para subtitular en un entorno como la web y si se ha producido alguna variación a lo largo del tiempo en las normas que subyacen a la subtitulación en la web.

CAPÍTULO 6

LA CALIDAD DE LA ACCESIBILIDAD AUDIOVISUAL EN LA WEB: GUÍAS Y RECOMENDACIONES

- 6.1. Calidad en la accesibilidad web al contenido audiovisual
- 6.2. Guías y recomendaciones en el ámbito de la subtitulación
- 6.3. Subtitulación: normas, guías y recomendaciones

CAPÍTULO 6. LA CALIDAD DE LA ACCESIBILIDAD AUDIOVISUAL EN LA WEB: GUÍAS Y RECOMENDACIONES

La ausencia de estándares básicos para el ejercicio de la subtitulación puede conducir a una mala praxis que se traduce en una pésima calidad, o lo que es lo mismo, en una falta de comprensión del contenido y, por lo tanto, en una barrera para la comunicación. Es cierto que una mayor estandarización y unificación de los criterios proporcionaría mayores índices de calidad y de accesibilidad, pero para materializar este objetivo quizá sería recomendable flexibilizar la definición de los parámetros para evitar así la disparidad de criterios que se aplican.

En este capítulo, de una parte, recogemos tanto las recomendaciones como las guías que proponen, desde distintos organismos, empresas y autores destacados de nuestro ámbito de estudio; de otra, identificamos los parámetros de estudio de nuestro trabajo. La accesibilidad, o es de calidad o no es accesible, especialmente en lo que nosotros consideramos parámetros definitorios de la subtitulación que son los que analizaremos en nuestro corpus y especificamos en este apartado.

6.1. Calidad en la accesibilidad web al contenido audiovisual

Un código de buenas prácticas de subtitulación y audiodescripción *online* sería muy beneficioso para todos los actores que participan en el proceso de desarrollo web, en especial para los receptores de la información. No obstante, la realidad es otra y a día de hoy no hay una guía que estandarice estas prácticas en el entorno web, pero sí es cierto que, gracias a la redacción de normativas y leyes para alcanzar la igualdad de oportunidades, como presentamos en capítulos anteriores, encontramos guías y normas que definen y prescriben cómo se han de aplicar las técnicas de subtitulación para personas sordas y las técnicas de audiodescripción para personas ciegas.

Por el contrario, en el caso de la subtitulación convencional no existen normativas de aplicación, por lo que las convenciones varían según el país, la cadena de televisión, la agencia, etc. En consecuencia, la práctica de esta técnica carece de uniformidad,

especialmente en aquello relacionado con las consideraciones espaciales y temporales. Sin embargo, consideramos importante señalar que tanto desde el ámbito profesional como desde el académico se ha dotado a esta práctica de multitud de guías de estilo internas, códigos de buenas prácticas o recomendaciones a partir de trabajos de investigación en este campo, que dibujan de alguna manera los parámetros que definen la práctica de la subtítulos convencional y han sido la fuente de la que han bebido las recomendaciones de la SPS. Por tanto, ambas modalidades comparten la mayor parte de los parámetros, si bien, los valores asignados para cada tipo de audiencia difieren.

6.2. Guías y recomendaciones en el ámbito de la subtítulos

De los códigos, guías o recomendaciones con más repercusión en el ámbito de la subtítulos, queremos mencionar en primer lugar, por la posición indiscutible que ocupa en la literatura de nuestro ámbito, el compendio de directrices que quedó recogido en el *Code of Good Subtitling Practice*⁶⁵ de Ivarsson and Carroll (1998:157-159). Este conjunto de normas es sin duda el que más referencias acumula, probablemente por su carácter generalista y porque los principios fundamentales son, todavía hoy, de aplicación. Casi de forma simultánea Karamitroglou (1998) con la pretensión de estandarizar la práctica de la subtítulos en Europa redactó *A proposed set of subtitling norms in Europe*⁶⁶, mientras que unos años más tarde, el proyecto DTV4ALL⁶⁷ financiado por la Unión Europea contaba entre sus objetivos con la redacción de guías que favorecieran la estandarización de los servicios que garantizan el acceso a la información mediante el asesoramiento «a las entidades participantes sobre los servicios de accesibilidad consolidados y los emergentes para que velen por la estandarización de estos servicios y su diálogo ante diversas entidades».

En el panorama internacional y nacional Díaz Cintas (2003) propuso unos parámetros para la práctica de la subtítulos con la intención de favorecer la calidad y homogeneidad. Posteriormente, en *Audiovisual Translation: Subtitling* (Díaz Cintas y Remael, 2007), obra de absoluta referencia en esta modalidad de traducción audiovisual, formulan una serie de recomendaciones con el objetivo de establecer unos estándares mínimos, y en la que queda reflejada una clara declaración de intenciones de no inmiscuirse en la tradición subtítuladora establecida en los distintos países (*ibid.*:80):

⁶⁵ Este código fue aprobado por la ESIST (European Association for Studies in Screen Translation).

⁶⁶ <https://translationjournal.net/journal/04stndrd.htm>

⁶⁷ <http://www.psp-dtv4all.org/>

In our opinion, these guidelines should not be understood as an invasion on any country's or company's subtitling tradition, but rather a declaration of good intentions aspiring to set some minimum standards in the profession and to safeguard subtitlers' rights.

Este planteamiento, no invasivo, hacia prácticas consolidadas en otros países nos parece muy acertado puesto que, como apuntábamos en el epígrafe que introduce este apartado, entendemos que favorecería el consenso y la estandarización de la subtitulación. En cuanto a la estructura de los estándares propuestos por Díaz Cintas y Remael (*ibid.*) están agrupados en tres categorías: (1) dimensión espacial, (2) dimensión temporal y (3) puntuación y otras convenciones. En definitiva, son una redistribución y ampliación de las recomendaciones planteadas por Ivarsson y Carroll (1998) junto con la actualización de ciertos parámetros que por el impacto del avance de las nuevas tecnologías y el correr de los años había provocado en la práctica traductora.

Por su parte, la SPS y la AD en España sí disponen de normativa específica. En el caso de la SPS encontramos la norma UNE 153010 publicada en 2003 y actualizada en 2012 «Subtitulado para personas sordas y personas con discapacidad auditiva». La actualización se realizó con el fin de adaptarse a las nuevas tecnologías y, a día de hoy, se considera norma general de subtitulado de aplicación tanto para DVD, televisión general o internet. La norma (2012:5) explicita que:

Se contemplan todos los escenarios de comunicación audiovisual que permiten la incorporación del subtitulado y se consideran tanto el subtitulado en directo como el grabado, indicándose los casos en los que los requisitos no sean de aplicación común.

Es decir, que esta norma es de aplicación en el entorno web y por tanto nos sirve de guía para definir los parámetros relativos a la práctica de la subtitulación. Igualmente, queremos resaltar el hecho de que en la redacción de la Norma colaboraron asociaciones de personas sordas, empresas de subtitulado y medios audiovisuales, universidades y organismos estatales para así garantizar la homogeneidad e intentar establecer unos requisitos mínimos de calidad en la subtitulación para personas sordas.

Por otra parte, como indicábamos en el apartado anterior, la web no dispone de una normativa específica para la inclusión de los contenidos alternativos, pero sí hace referencia directa a unas guías, además de las consideraciones que hay que tener en cuenta para superar el criterio de éxito. Los guías referidas en las WCAG 2.0 que velan por las buenas prácticas para el subtitulado y la audiodescripción son: la página web del experto en accesibilidad Clark⁶⁸ (2004) que recoge buenas prácticas para el subtitulado y la audiodescripción, y las

⁶⁸ <http://joelclark.org/access/captioning/bpoc/>
<http://joelclark.org/access/description/ad-principles.html>

recomendaciones de apoyo redactadas por el Described and Captioned Media Program⁶⁹ (DCMP).

En cuanto a la página web de Clark, en la que se ofrecen especificaciones y recomendaciones, entendemos que no sirven de referente en nuestro estudio puesto que la información relativa a la accesibilidad *online* que ofrece por una parte, no está actualizada y, por otra parte, profundiza sobre todo en cuestiones técnicas de desarrollo web que se escapan a nuestro ámbito de estudio. Nos referimos a las recomendaciones disponibles en el documento *Best practices in online captioning*⁷⁰ en las que revela que en 2003-2004 formó parte del proyecto TILE⁷¹ con el objetivo de generar una guía específica de subtitulación *online*, pero sin alcanzar los objetivos propuestos. Por tanto, las guías que tomaremos como referencia propuestas en las WCAG 2.0 serán las *Media Accessibility Guidelines* del DCMP. Este organismo está financiado por el Departamento de Educación de los EE.UU. y administrado por la Asociación Nacional de Sordos. Estas guías las utilizan tanto profesionales como *amateurs* de todo el mundo y se han redactado para asegurar el acceso igualitario a la información y a partir de la experiencia, investigación y opinión de los consumidores como apuntan es su web⁷².

The Keys to Access guidelines for captioning and description were developed by the Described and Captioned Media Program (DCMP) to ensure that the accessibility features of its media were of the high quality required to provide equal access. These guidelines are based on recorded research findings, the input and guidance of consumers and experts in the field of accessible media, and decades of experience by DCMP staff.

Por otra parte, por su larga experiencia y prestigio, nos servirán asimismo de referente las guías de estilo de la BBC (2009) que, además, están siempre actualizadas en su versión en línea⁷³. Recordemos que la BBC comenzó a subtítular a través del teletexto en 1979 y en 2008 subtítulaba ya el 100% de su programación. Asimismo, desde 2012 y a través de BBC iPlayer ofrece todos los vídeos a la carta y descargables con subtítulos. Pero en la BBC no solo se han preocupado por la cantidad de programas subtítulados, sino que la calidad siempre ha guiado su camino. En el documento *Understanding The Diverse Needs Of Subtitle Users In A Rapidly Evolving Media Landscape* (2015) encontramos la descripción del recorrido que la BBC ha hecho en materia de investigación y normativa interna para asegurar la calidad de

⁶⁹ <http://www.captioningkey.org/>

⁷⁰ <https://joelclark.org/access/captioning/bpoc/introduction.html>

⁷¹ <https://barrierfree.ca/tile/>

⁷² <http://keystoaccess.org/#1>

⁷³ http://www.bbc.co.uk/guidelines/futuremedia/accessibility/subtitling_guides/online_sub_editorial_guidelines_vs1_1.pdf

los subtítulos además de una descripción del trabajo que desempeña actualmente el equipo de investigación en BBC R&D y cómo estructuran su investigación (2015:3)

Our work breaks down into three main areas, developing affordable and sustainable ways of increasing the availability of subtitles across all platforms, carrying out user research to understand the quality of the current user experience of subtitles for the audience and research aimed at utilising the characteristics of new platforms to improve the experience.

Es decir, la investigación contempla la forma de incrementar la disponibilidad de subtítulos en todas las plataformas, la calidad de los subtítulos a partir de la experiencia de los espectadores y por último abrir nuevas líneas de investigación para descubrir de qué manera las nuevas tecnologías pueden mejorar la experiencia de visualizar un contenido subtítulo.

6.3. Subtitulación: normas, guías y recomendaciones para la subtitulación online

En el apartado anterior definíamos, entre otras, las guías de recomendaciones propuestas por la WCAG y otros organismos. En nuestro análisis, tomaremos como referencia tres documentos: la *DCMP Captioning Key*, la Norma UNE 153010 de subtitulación y las *BBC Subtitle Guidelines* sin olvidarnos, en la reflexión teórica que acompaña a cada parámetro, de otras propuestas de calado académico y profesional. Obviamente no podemos dejar de lado la fuente académica de la que beben la mayor parte de las recomendaciones en este campo, el *Code of Subtitling Practice* (1998) junto con las aportaciones de distintos autores que desde la academia han tratado de aportar calidad a la creación de contenidos audiovisuales accesibles (Díaz Cintas y Remael, 2007; Martí Ferriol, 2012; Swakowska, 2016a y 2016b; entre otros). Consideramos también oportuno, con el afán de acercarnos a la realidad, tener en cuenta las guías de subtitulación que Netflix ha creado para estandarizar la práctica de la subtitulación en todos sus contenidos y en los distintos idiomas. Netflix es una de las mayores plataformas de vídeo en *streaming* y un fenómeno mundial, por lo que merece ser considerada en un trabajo de estas características.

Describiremos, por una parte, la estructura de las guías citadas que nos permiten dibujar el panorama nacional y europeo en cuanto a recomendaciones para la subtitulación *online*: norma UNE, *BBC guidelines* y *DCMP Captioning Key* y comprobaremos los parámetros generales en los que se incide para posteriormente analizar en mayor profundidad los parámetros en los que centraremos nuestro análisis.

En la siguiente tabla recogemos las consideraciones previstas en las guías:

Tabla 4. Guías para la subtitulación.

NORMA UNE 153010:2012	DCMP GUIDELINES	BBC OFCOM
1. Presentación del subtítulo en pantalla: aspectos visuales <ol style="list-style-type: none"> Posición de los subtítulos Posición de los efectos sonoros Número de líneas de texto Líneas distintas por personajes Número de caracteres por línea 2. Presentación del subtítulo en pantalla: aspectos temporales <ol style="list-style-type: none"> Velocidad de exposición del texto del subtítulo Sincronismo 3. Identificación de los personajes 4. Efectos sonoros 5. Información textual y voz en <i>off</i> 6. Música y canciones 7. Criterios editoriales <ol style="list-style-type: none"> División del texto Criterios gramaticales y ortográficos Lenguas oficiales del estado Personajes con habla específica (idiolecto) Abreviaturas y símbolos Numeración Literalidad 	1. Text <ol style="list-style-type: none"> Case Font Line division Caption Duration Caption Placement 2. Language Mechanics <ol style="list-style-type: none"> Spelling and Capitalization Punctuation and Grammar 3. Presentation Rate 4. Sound Effects and Music 5. Speaker Identification 6. Special Considerations <ol style="list-style-type: none"> Intonation, Play on Words, No Audio Foreign Language, Dialect, Slang, and Phonetics 	1. Editing 2. Timing 3. Subtitle breaks 4. Keeping in sync 5. Line-breaks 6. Matching shots 7. Identifying speakers 8. Colours 9. Presentation 10. Intonation and emotion 11. Accents 12. Difficult speech 13. Inaudible speech and silence 14. Hesitation and interruption 15. Cumulatives 16. Humour 17. Children's subtitling 18. Music and songs 19. Sound-effect labels 20. Numbers

La tabla anterior evidencia, por una parte, que las tres guías son exhaustivas y coincidentes en la mayor parte de los parámetros y, por otra parte, pone de manifiesto que cada organismo ha decidido organizar la información en grupos o categorías diferentes según un nivel de concreción. De igual manera, nos brinda información suficiente para aseverar que la coincidencia en los parámetros que definen esta práctica traductora son el primer paso hacia la estandarización de la SPS *online* y hacia una norma preliminar en este campo. En este sentido, Martínez Sierra (2011:156) afirma que las normas son afirmaciones de la conducta traductora y añade:

Con el fin de que dichas afirmaciones se muestren de manera sistemática y de que no constituyan una selección aleatoria de observaciones, éstas deberán mostrarse en forma de generalizaciones que sean aplicables a una clase o subclase particular de fenómenos y que se puedan testar intersubjetivamente (es decir, objetivamente entre varios sujetos).

Queda claro pues que una de las normas que gobiernan la subtitulación *online* es la de tener en cuenta una serie de parámetros establecidos a partir de guías, normativas o recomendaciones y procurar de este modo la calidad de los subtítulos. El primer objetivo de toda subtitulación es que los subtítulos se puedan leer, es decir, que el espectador disponga de tiempo suficiente para procesar la información escrita y que a su vez disfrute de tiempo para ver el contenido visual del vídeo. Para que estos objetivos se cumplan, el pausado y la segmentación de los subtítulos tienen que ser adecuados, de ahí que nuestro trabajo se centre en el estudio de estos parámetros.

En este capítulo hemos presentado las guías que, de forma prescriptiva, establecen cómo se debe subtítular y que nos sirven de referente para el análisis de la subtitulación *online*, así como los intentos que desde la academia se han hecho en este sentido. En el siguiente capítulo, y a partir de estas normativas (incluiremos también las recomendaciones que se vierten desde la academia) presentaremos los parámetros objeto de estudio en esta modalidad de TAV y definiremos, esto es, enmarcaremos de forma teórica, los que nosotros vamos a investigar.

CAPÍTULO 7

PARÁMETROS DE ESTUDIO

7.1. Los parámetros de estudio en subtitulación

7.2. Definición y análisis de los parámetros de estudio

7.2.1. Parámetro velocidad de lectura

7.2.1.1. Variable velocidad de lectura en caracteres por segundo

7.2.1.2. Variable pausa entre subtítulos

7.2.1.3. Variable tiempo mínimo y tiempo máximo de permanencia en pantalla

7.2.2. Parámetro formato y posicionamiento

7.2.2.1. Variable número de líneas

7.2.2.2. Variable número de caracteres por línea

7.2.2.3. Variable segmentación en unidades de sentido

CAPÍTULO 7. PARÁMETROS DE ESTUDIO

En este apartado realizaremos un brevísimo apunte relacionado con las clasificaciones de los parámetros de estudio en subtitulación y nos centraremos en enmarcar teóricamente los parámetros que vamos a estudiar en este trabajo atendiendo a las propuestas de las guías y recomendaciones que, como expusimos en el capítulo anterior, hemos tomado como referencia en nuestro estudio.

7.1. Los parámetros de estudio en subtitulación

Los parámetros objeto de análisis en subtitulación y SPS han sido clasificados por distintos autores (Karamitrouglou, 1998; Díaz Cintas y Remael, 2007; Bartoll, 2008; Arnáiz, 2012; Tamayo, 2015a y 2015b, entre otros.) En el caso de la subtitulación, Bartoll (2008) propuso una concienzuda taxonomía en la que años más tarde se basó Arnáiz (2012) para crear una clasificación de la SPS. La propuesta de Arnáiz incluye una nueva categoría a la de Bartoll en la que aglutina los parámetros específicos de la SPS y que son imprescindibles para su análisis, puesto que como apunta Arnáiz (2012:107) el estudio de trabajos en el ámbito de la SPS y de las normativas existentes,

[...] desvelan la existencia de una serie de elementos específicos del SPS, condicionados fundamentalmente por la representación de la información sonora adicional, que no aparecen representados en el trabajo de Bartoll.

Pero no solo los elementos de la información sonora son específicos de la SPS, puesto que existen otros aspectos que difieren de la subtitulación convencional, como es el caso de la velocidad de lectura, y que la autora también incorpora a su taxonomía.

Tanto las taxonomías mencionadas como estudios y trabajos previos desvelan las prácticas habituales en el desarrollo de esta modalidad de traducción, y permiten al investigador delimitar y enmarcar su ámbito de estudio para generar a su vez propuestas de análisis de la subtitulación. En este sentido, en su tesis doctoral, Tamayo (2015a:76) presenta una serie de parámetros que reúnen las convenciones y características técnicas que han de tenerse en cuenta a la hora de crear e investigar en SPS y, como apunta la autora, la selección

de las variables está cimentada en publicaciones previas de subtítulos convencionales y SPS (Karamitroglou, 1998; Lorenzo y Pereira, 2001; Neves, 2005, 2009; Díaz-Cintas y Remael, 2007; Arnaiz, 2012, Pereira, 2010; Lorenzo, 2010; Cambra *et al.* 2009, 2013, entre otros); lo que de nuevo da cuenta de que, lógicamente, la bibliografía producida en SPS bebe de los trabajos y estudios producidos en subtítulos convencionales.

En el trabajo de tesis doctoral de Tamayo (2015a:6) encontramos, por una parte, un análisis cuantitativo de la cantidad de SPS ofrecida por la TDT española dirigida al público infantil y juvenil y, por otra parte, un estudio cualitativo basado en datos cuantitativos sobre los porcentajes observados sobre las prácticas habituales en los once parámetros objeto de su estudio que mostramos a continuación:

V1	Velocidad de lectura	V7	Identificación de los personajes
V2	Formato y posicionamiento del subtítulo	V8	Léxico
V3	Convenciones ortotipográficas	V9	Sintaxis
V4	Emociones de los personajes	V10	Música ambiental
V5	Sonidos emitidos por los personajes	V11	Música argumental
V6	Efectos sonoros		

Figura 16. Macrovariables de estudio propuestas por Tamayo (2015b:368).

Consideramos importante señalar que, aunque con una organización distinta, la Norma UNE, las guías de la BBC y de la DCMP, también incluyen estos parámetros.

Sin embargo, nuestro análisis no pretende abarcar los once parámetros o macrovariables, sino centrarse en los dos primeros parámetros, es decir, la velocidad de lectura y el formato y posicionamiento y sus correspondientes variables que expondremos en apartados posteriores.

7.2. Definición y análisis de los parámetros de estudio

Los parámetros que hemos resuelto estudiar son, a nuestro entender, los que forman parte de las dimensiones espacio-temporales propuesta por Díaz Cintas y Remael (2007) y los parámetros sobre los que se sustentan los principios básicos de la subtítulos, como mencionamos anteriormente. En este sentido, Díaz Cintas y Remael (2007:95) consideran que:

It is very frustrating and disconcerting to see how the subtitle disappears from the screen when we have not yet finished reading it, [...] the typical occasion in which we feel that we have 'read' rather than 'watched' the film.

Opinamos que, por muy elocuentes y ricas que sean las propuestas de traducción de los subtítulos, si no tenemos tiempo de leerlos, no podemos ni disfrutarlos ni tan siquiera, en ocasiones, entenderlos.

A continuación, definiremos, por una parte, los parámetros objeto de estudio en nuestro trabajo y las variables que incluyen y, por otra parte, realizaremos una revisión teórica de cada aspecto para así delimitar los valores que comprende cada variable.

7.2.1. Parámetro velocidad de lectura

Tamayo (2015a:86), como podemos observar en la figura inferior, define este parámetro a partir de dos variables, la velocidad del subtítulo y la pausa entre subtítulos y atribuye una serie de valores posibles indicando a su vez cuál es el que cumple la norma según la documentación que toma como referencia. A continuación, y a partir de nuestra revisión teórica indicaremos cuáles son los valores que serán referentes en nuestro estudio.

PARÁMETRO	VARIABLE	VALORES POSIBLES
Velocidad de lectura (V1)	Velocidad en caracteres por segundo (V1.1)	- ≤ 10
		- 11
		- 12
		- 13
		- 14
		- 15 (máximo recomendado por la UNE)
		- 16
		- 17
		- 18
		- 19
	- ≥ 20	
	Pausa entre subtítulos (V1.2)	- ≤ 4 cuadros (0,16 segundos)
		- > 4 cuadros (0,16 segundos)

Figura 17. Parámetro velocidad de lectura propuesto por Tamayo (2015a:86).

7.2.1.1. Variable velocidad en caracteres por segundo

En cuanto a la primera variable, son muchos los estudios que dan información significativa al respecto, en especial los últimos estudios de recepción. En este sentido, Romero-Fresco (2015:335) introduce el concepto *viewing speed*, velocidad de visionado que define como:

[...] viewing speed is regarded as the speed at which a given viewer watches a piece of audiovisual material, which in the case of subtitling includes accessing the subtitle, the accompanying images and the sound, if available.

En el estudio de recepción que el autor llevó a cabo en el marco del proyecto europeo DTV4ALL⁷⁴ desveló que con una velocidad de visionado de 120 wpm (*words per minute*) los participantes pasaban el 40% del tiempo leyendo y el 60% visualizando la imagen, si bien al aumentar la velocidad a 160 wpm la proporción de tiempo entre subtítulo e imagen era equilibrada. Sin embargo, al incrementar la velocidad a 200 wpm los participantes invertían el 80% del tiempo en leer el subtítulo y solo un 20% en visualizar la imagen. A partir de los nuevos datos que arrojó este estudio se considera, a día de hoy, que la velocidad de lectura óptima para un público adulto está más cerca de las 160 wpm que de las 120 wpm.

Recordemos, sin embargo, que la velocidad de lectura está basada en «la regla de los seis segundos» (d'Ydewale *et al.*, 1987) que, tras ser revisada en varias ocasiones, sigue siendo de aplicación. En este sentido Díaz Cintas y Remael (2007:96) explican:

According to this rule, an average reader can comfortably read in six seconds the text written on two full subtitle lines, when each line contains a maximum of 37 characters, i.e. a total of 74 characters.

Esto se traduce, según los autores, en 130 wpm⁷⁵ o 12 cps (caracteres por segundo). En este sentido, Martí Ferriol (2013) y Romero-Fresco (2009) apuntan que 144 wpm equivale a 12 cps, mientras que 160 wpm equivale a 15 cps y 180 wpm 18 cps (Martí Ferriol, 2012 y Díaz Cintas, 2008). Por su parte, Díaz Cintas (2012: 276) establece que 130 wpm equivale a 12 cps, 160 wps a 15 cps y 180 wpm a 17 cps. Añade además que 15 cps es la velocidad habitualmente utilizada en la industria televisiva mientras que 17 cps es la velocidad que se suele utilizar para DVD e internet, incluso a veces más rápido, pero recordemos que los últimos estudios de recepción indican que la velocidad de lectura óptima, en subtitulación convencional, es de 160 wpm o 15 cps.

En el presente trabajo mediremos la velocidad de lectura en cps puesto que medirla en palabras por minuto es menos preciso que hacerlo en caracteres por segundo; de igual manera, convertir los wpm en cps no resulta sencillo ni parece que se haya alcanzado consenso. Lo que sí es evidente es que los cps son la medida más utilizada para determinar la velocidad de lectura como revela Szarkowska (2016), quien gracias al proyecto SURE comprobó que la mayor parte de las compañías trabajan con cps y que la regla de los seis segundos se mantiene.

⁷⁴ <http://www.psp-dtv4all.org/>

⁷⁵ Este dato aparece también en la página 276 de Díaz Cintas, J. (2012). «Subtitling» en *The Routledge Handbook of Translation Studies*. Londres: Routledge. <https://doi.org/10.4324/9780203102893>

En cuanto a la famosa regla, la autora señala (2018:2): «While the origins of the rule are difficult to trace [...] it is the golden standard recommended in subtitling textbooks», y apunta que lo que sí ha variado es el número de caracteres por línea y de caracteres por segundo. Cada país aplica una velocidad de lectura diferente y añade que los países escandinavos establecen una velocidad de 12 cps, en Europa central entre 15-16 cps y los servicios globales de *streaming online* entre 17-20 cps, como es el caso de Netflix. Es curioso que en la guías de estilo para subtitulación de esta empresa, en el caso de la subtitulación al inglés, recomiendan que la velocidad de los subtítulos sea de 20 cps para adultos y 17 para niños, mientras que en las guías de español para adultos indican una velocidad de 17 cps y 13 cps para niños. Nos llama la atención puesto que la velocidad de lectura recomendada en España es de 15 cps para ambos públicos.

En el caso de la SPS, la regla de los seis segundos se mantiene en las normativas de casi todos los países, sin embargo, son muchos los estudios y autores (d'Ydewalle *et al.*, 1987; Linde and Kay, 1999; Neves, 2008; Lorenzo, 2010, entre otros), que defienden que este tiempo de permanencia en pantalla no es suficiente para las personas sordas. En su tesis doctoral, Tamayo (2015a:84) recoge en la siguiente tabla las velocidades de lectura más efectivas según algunos autores:

	ADULTOS NORMO-OYENTES	ADULTOS SORDOS	NIÑOS NORMO-OYENTES	NIÑOS SORDOS
REGLA	<ul style="list-style-type: none"> • Regla de los 6 segundos (d'Ydewalle <i>et al.</i>, 1987) • 66 % de la velocidad del discurso (de Linde y Kay, 1999) 	<ul style="list-style-type: none"> • Regla de los 9 segundos (Neves, 2009) 	<ul style="list-style-type: none"> • Regla de los 9 segundos 	<ul style="list-style-type: none"> • Regla de los 8 segundos (Lorenzo, 2010b)
WPM	<ul style="list-style-type: none"> • $\approx 150 - 180$ (Díaz Cintas y Remael, 2007). 	<ul style="list-style-type: none"> • $\approx 90 - 120$ (Neves, 2009) 	<ul style="list-style-type: none"> • $\approx 90 - 120$ (Karamitroglou, 1998) 	<ul style="list-style-type: none"> • ≈ 60 (Baker <i>et al.</i>, 1984; Baker, 1985; de Linde y Kay, 1999)
CPS	<ul style="list-style-type: none"> • ≈ 16 (Gottlieb, 2009) 	<ul style="list-style-type: none"> • ≈ 15 (Ofcom, 2006; AENOR, 2012) 	<ul style="list-style-type: none"> • ≈ 15 	<ul style="list-style-type: none"> • ≈ 12 (presente estudio)

Figura 18. Recopilación de distintos autores de velocidad de lectura máxima de subtítulos. Tamayo (2015a:84).

En su estudio, Tamayo establece que la velocidad de lectura de referencia en su análisis sea de 15 cps, valor que recogemos por ser el más habitual y por aparecer tanto en la Norma UNE como en la DCMP⁷⁶ que señalan lo siguiente:

1. *Lower- to middle-level educational videos should be captioned at a presentation rate range of 120–130 words per minute (wpm). Upper-level educational videos can be captioned slightly above the 120–130 range.*
2. *Adult special-interest videos require a presentation rate of 150–160 wpm. The presentation rate can be increased if heavy editing radically changes the original meaning, content, or language structure.*
3. *Children’s cultural movies should be captioned at a rate of 150 wpm.*
4. *Adult cultural movies should be captioned at a near verbatim rate, but no captioning should remain on screen less than 2 seconds or exceed 225 wpm.*

Asimismo, según Pedersen (2017: 223)⁷⁷:

The time spent reading subtitles increases with reading speeds, so already at 15 cps, viewers spend on average about two thirds of their time in the subtitle area and at 16.5 cps, they spend 80% of their time reading subtitles. That may not be so much of a problem for relatively static genres such as interviews or news, but for a feature film, that leaves very little attention for the on-screen action.

El autor plantea, en su propuesta de análisis de la calidad de subtítulos, penalizar velocidades por encima de los 15 cps, si bien es cierto que la tendencia apunta a generar subtítulos entre 16 y 20 cps. Según Pedersen (*ibid.*), 20 cps equivale a unas 240 wpm «which is a level where most people would probably do nothing else but reading subtitles (or stop using them)». Por lo que entiende que subtítulos con velocidades superiores a 20 cps han de ser considerados como error en la subtitulación y no como una penalización.

Tras esta revisión teórica, la variable velocidad de lectura en nuestro estudio considerará tres valores: 15 cps, puesto que es la velocidad de lectura recogida en la norma UNE y en las guías DCMP, así como por ser la velocidad reconocida por la academia como la más habitual; un segundo valor que comprenderá todos aquellos subtítulos con velocidades de lectura entre 16 y 20 cps, que aunque exceden de la velocidad de lectura óptima recogida en la normativa, como hemos visto, tanto la industria televisiva como las plataformas que ofrecen servicios *online* apuntan a valores que se mueven entre los 16 y los 20 cps; y por último, el grupo de los subtítulos imposibles, aquellos de 21 cps o más.

⁷⁶ <https://dcmp.org/learn/5-captioning-guidelines-for-the-dcmp>

⁷⁷ Pedersen, J. (2017). «The FAR model: assessing quality in interlingual subtitling» en *The Journal of Specialised Translation*, 28. Recuperado en https://www.jostrans.org/issue28/art_pedersen.php

7.2.1.2. *Variable pausa entre subtítulos*

En la descripción de esta segunda variable, Tamayo (2015a:280) recoge un dato pocas veces analizado pero de gran importancia a nuestro entender que no es otro que el de la pausa entre subtítulos. Nos recuerda que no solo es importante la velocidad de los subtítulos sino que también juegan un papel central las pausas entre subtítulos puesto que el ojo humano necesita tiempo para detectar la presencia de un nuevo subtítulo y cita a Sponholz (2003:24):

At least four frames should be inserted between two consecutive subtitles in order to avoid the effect of a subtitle overlay. This time break is necessary to signal to the brain the disappearance of one subtitle and the appearance of another.

Tamayo, tras revisar las sugerencias de los investigadores más prominentes en este campo que han investigado este parámetro (Ivarsson, 1992; Castro Roig, 2001 o Sponholz, 2003), apuesta por un valor medio y se inclina por establecer un valor de 4 fotogramas (0,16 segundos) de pausa entre subtítulos.

Nos resulta curioso, sin embargo, que ninguna de las guías que utilizamos como referencia mencionan este parámetro, aun así y teniendo en cuenta que son muchos los autores que lo contemplan, además de Tamayo (Ivarsson, 1992; Karamitroglou, 1998; Castro Roig, 2001; Mayoral, 2001; Díaz Cintas, 2001 y González Iglesias, 2012), incorporaremos el análisis de la pausa entre subtítulos en nuestro análisis.

Nosotros nos inclinamos, al igual que los autores que mencionamos en párrafo anterior por un valor medio de 0,16 segundos, pero añadiremos la pausa 0, dado que nos interesa saber si se generan subtítulos sin pausa alguna entre ellos, lo que aumenta la velocidad de lectura requerida para el espectador, y un tercer valor que recoja las pausas que superan los 0,16 segundos.

7.2.1.3. *Variable tiempo mínimo y tiempo máximo de permanencia en pantalla*

La dimensión temporal va acompañada de otra variable que, aunque Tamayo no ha incluido en este parámetro, consideramos de especial relevancia y proponemos estudiar. Nos referimos al tiempo mínimo y tiempo máximo que un subtítulo habría de permanecer en pantalla. La famosa regla de los seis segundos habla por sí misma de la duración máxima de un subtítulo de dos líneas y, por su parte, Díaz Cintas y Remael (2007:89) suscriben esta norma puesto que entienden que es la forma de evitar la relectura del subtítulo. Opinan de igual manera Koolstra *et al.* (2002) quienes pusieron a prueba la regla y concluyeron nuevamente que seis segundos es el tiempo adecuado para que no se produzca una relectura.

Por otra parte, también hay razones para establecer un tiempo mínimo de permanencia en pantalla, y como señalan Díaz Cintas y Remael (2007):

At the other end of the scale, to avoid flashing subtitles on screen and guarantee that viewers have enough time to read the content, the ideal minimum exposure time for a subtitle is commonly agreed at one second.

Ivarsson y Carrol (1998) también apuntan que el tiempo de permanencia mínimo es de un segundo y Karamitroglou (1998) lo establece en 1 y 1/4 de segundo coincidiendo con la recomendación de la guía del DCMP; sin embargo, la BBC propone unos 0,3 segundos por palabra, pero entendemos que la longitud de las palabras puede ser muy diferente por lo que podría producirse el temido efecto *flash*.

Por su parte, la norma UNE no define este aspecto, de modo que, teniendo en cuenta que la norma comúnmente aceptada es que el tiempo mínimo de permanencia sea de 1 segundo, tomaremos este valor para la variable tiempo mínimo de permanencia.

En la siguiente tabla resumimos las variables que analizaremos de este parámetro y los valores que pueden contener.

Tabla 5. Parámetro velocidad de lectura.

Parámetro	Variable	Valores posibles
Velocidad de lectura	Velocidad en caracteres por segundo	15- Norma UNE y DCMP 16-20 – Netflix ≥21
	Pausa	0 segundos ≤ 0,16 segundos > 0,16 segundos
	Duración	≤1 segundo 1 y 6 segundos > 6 segundos

7.2.2. Parámetro formato y posicionamiento

El parámetro que nos ocupa incluye siete variables; Tamayo considera que estas están relacionadas con los aspectos espaciales de la subtitulación y señala entre otras la segmentación por ser este un aspecto lingüístico directamente relacionado con las cuestiones espaciales. Este parámetro aglutina los siguientes aspectos: segmentación en unidades de sentido, sincronización, número de líneas, distribución y forma en pantalla, número de caracteres por línea, posición y, finalmente, color y fuente. Sin embargo, en el presente

estudio nosotros estudiaremos de forma exhaustiva y automatizada el comportamiento de tres de ellas (número de líneas, número de caracteres por línea y segmentación en unidades de sentido) que son las que integran, a nuestro entender, el proceso de traducción y por tanto responsabilidad directa del subtitulador. En cuanto al resto de variables (sincronización, distribución y forma en pantalla, posición así como color y fuente) explicamos a continuación de forma sucinta los aspectos fundamentales.

En cuanto a la sincronización, como apuntan las tres guías que hemos tomado como referencia así como académicos de reconocido prestigio en este ámbito, se debe procurar que el subtítulo coincida con la locución y con la información sonora, aunque sin olvidar como indican Ivarsson y Carroll (1998:76) que «readability should not be sacrificed for the sake of synchronisation». Para la distribución y forma en pantalla, se prefiere que estén abajo centrados, tanto en el caso de la BBC («BBC online subtitles should be centred below the display and above the controls of the Embedded Media Player») como en el de la norma UNE («Los subtítulos de todo aquello que no sean efectos sonoros deben aparecer centrados en la parte inferior de la pantalla excepto cuando oculten información relevante»). Sin embargo, el DCMP *Captioning Key* recomienda alinearlos a la izquierda. Por su parte, el proyecto DTV4ALL desveló que la mayor parte de países europeos también optan por los subtítulos abajo centrados. En cuanto a la posición de los efectos sonoros debe ubicarse, según la norma UNE, en la parte superior derecha, pero como indica Arnáiz (2015), en un estudio de recepción con *eye-tracking*, los subtítulos que aparecen en esta posición no se procesan correctamente y España es, además, el único país que los coloca alejados del subtítulo que da cuenta de la información lingüística.

Por último, para la variable color y fuente, todas las guías apuntan que lo más recomendable es distinguir a los personajes por colores y sobre la fuente recomiendan que sea de la familia *sans-serif* (Arial, Helvetica). Anteriormente apuntamos que esta investigación no se centrará en estos aspectos, pero sí aprovechamos este espacio para indicar que son las hojas de estilo que alojan el vídeo, las que determinan el tipo de letra y el tamaño. Nosotros, para llevar a cabo nuestro análisis, descargamos de internet los archivos con los subtítulos en formato SRT, es decir, texto plano y estos no contienen información alguna del formato. Es el reproductor el que nos permite escoger un formato determinado así como el tamaño de la letra. Por otra parte, cada vez viene siendo más habitual que el espectador personalice los subtítulos, de modo que lo que impera es que cada uno pueda escoger aquello que le resulte más fácil de leer.

En nuestro caso, pudimos inspeccionar el elemento desde la web al activar el vídeo y acceder así a la información que buscábamos en este sentido.

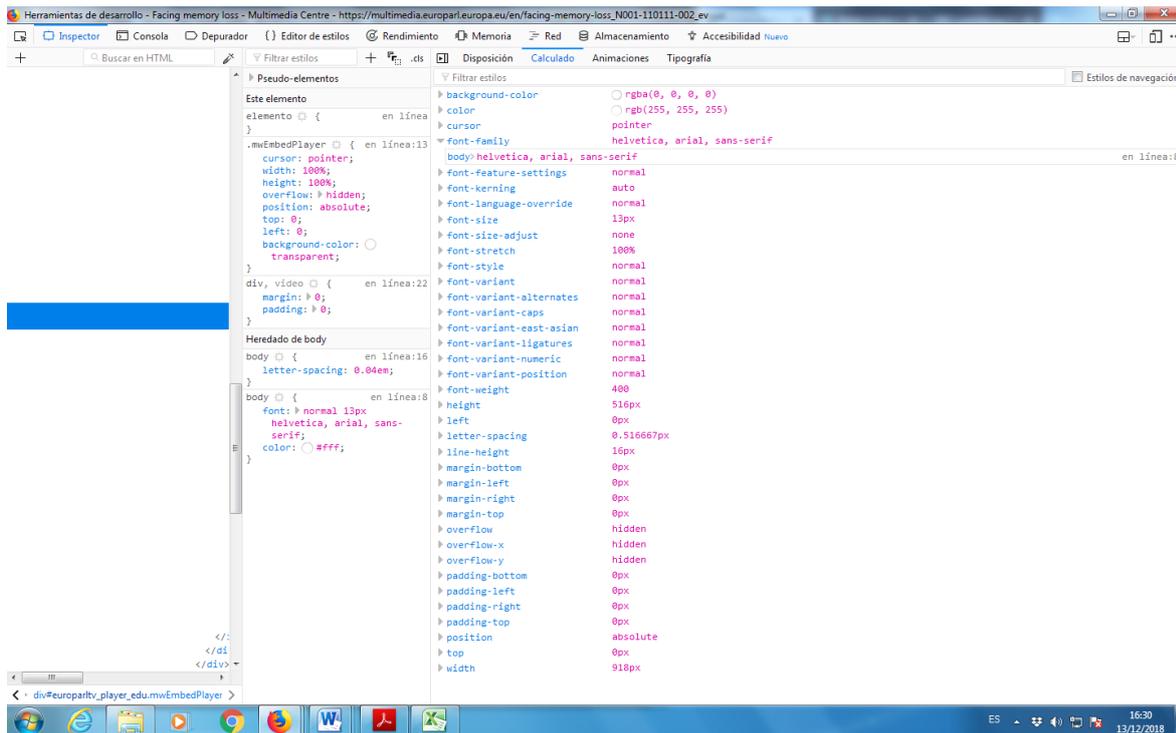


Figura 19. Elemento web que define la hoja de estilo.

7.2.2.1. Variable número de líneas

Por regla general, los subtítulos en España deben tener un máximo de dos líneas aunque la norma UNE 153010 establece que, en casos excepcionales, se pueden usar tres líneas. En la guía del DCOMP y de la BBC también encontramos información en este sentido en la que se indica que es conveniente no usar más de dos líneas. Encontramos la excepción a esta regla en países bilingües donde los subtítulos se muestran a la vez en las dos lenguas oficiales por lo que aparecen hasta 4 líneas de subtítulo (dos líneas por idioma). El debate que podemos encontrar en torno a esta variable se sitúa en la preferencia hacia los subtítulos de una o de dos líneas. La reflexión de Díaz Cintas y Remael (2007:85-86) a este respecto es que «the ideal number of lines and their positioning on screen is an issue a bit closer to the jurisdiction of the subtitlers»; pero por otra parte, apuntan al consenso en cuestiones básicas tales como que, si toda la información de un subtítulo cabe en una línea, conviene no utilizar dos.

Encontramos en la bibliografía autores que esgrimen los beneficios tanto de dar preferencia a subtítulos de una línea como a subtítulos de dos líneas (Brondeel, 1994; Koolstra *et al.*, 2002; Lomheim, 1992) pero quizás si la preferencia la situamos en facilitar la lectura del subtítulo habremos de recurrir a ambos tipos de subtítulos (Díaz Cintas y Remael,

2007:93) y como apunta Tamayo (2015a:91) «se debería alcanzar un equilibrio en favor del dinamismo y la comprensión del texto».

7.2.2.2. *Variable número de caracteres por línea*

En cuanto al número de caracteres por línea, en un estudio reciente conducido por Szarkowska (2016:7) se apunta que:

The most frequent range provided by the subtitlers was between 37 to 42 characters per line [...] The number of characters per line reported by the subtitlers in this study shows a slight increase compared to classic sources on subtitling, e.g. by Prof. Géry d'Ydewalle and his colleagues (see d'Ydewalle, Van Rensbergen and Pollet, 1987 or d'Ydewalle, Praet, Verfaillie, and Rensbergen, 1991) who used 32 characters as the maximum displayed for 6 seconds.

Es evidente que el incremento del número de caracteres por línea implica un aumento en la velocidad de lectura y como señala la autora (2016:21) el ejercicio de condensación típico de la subtitulación se reduce significativamente con el aumento del número de caracteres por línea (de 32 a 42) y de la velocidad de lectura (de 10-12 cps a 15-18 cps).

La norma UNE 153010, por su parte, establece el número máximo de caracteres por línea en 37, coincidiendo con la recomendación de la BBC; sin embargo, el documento del DCMP propone no superar los 32 caracteres por línea. En cuanto a Netflix eleva el número de caracteres por línea a 42 para todas las lenguas. Como podemos observar, estos datos son un reflejo de los recogidos por Szarkowska en su estudio (2016) y nos ayudan a definir los valores que asignaremos a esta variable de estudio. En nuestro trabajo consideramos cuatro valores: 32, 37, 42 y más de 42 caracteres por línea.

7.2.2.3. *Variable segmentación en unidades de sentido*

Empezamos, pues, definiendo el primer parámetro estudiado en profundidad en esta investigación: la segmentación en unidades de sentido que se preocupa de la segmentación del material lingüístico que aparece en los subtítulos de dos líneas y entre los subtítulos cuando se da el caso de que la oración se extiende entre subtítulos. Es importante destacar que la segmentación tiene un peso especialmente importante en nuestro análisis puesto que nos proponemos revelar, a partir de una propuesta de análisis de la segmentación, si los traductores siguen las normas dominantes en este aspecto o si, por el contrario, las opciones que toman difieren de lo establecido por la norma a la hora de segmentar. Nuestra propuesta permite comparar parcialmente las versiones de subtítulos en inglés y en español puesto

que incluye una serie de categorías compartidas en ambas lenguas. Pero, antes de continuar, definiremos qué es para nosotros la segmentación en subtitulación que no es otra cosa que, como apuntan Díaz Cintas y Remael (2007:172):

In subtitling, segmentation is the division of the ST dialogue, narration, etc. into sections or segments –subtitles– that the viewers can understand at a glance. Segmenting is done on two levels. A sentence may have to be distributed over the two available lines of the subtitle –line breaks– or it may run on, into two or more subtitles.

Los autores señalan que las pautas para la segmentación son las mismas entre líneas que entre subtítulos. Sin embargo, cuando la segmentación se produce entre subtítulos, se hace necesario recordar que las oraciones complejas son más difíciles de procesar y se recomienda partirlas en estructuras más simples. Nosotros analizaremos tanto la segmentación entre las líneas del subtítulo, que es donde se producen la mayor parte de las segmentaciones, como entre subtítulos.

La segmentación, en nuestra opinión, es un aspecto de gran interés en la subtitulación. Por una parte, está directamente relacionada con distintas variables técnicas (velocidad de lectura, número de caracteres por línea) y, por otra parte, es el punto en el que el traductor pone a prueba su destreza lingüística. Las constricciones propias de la técnica hacen que no sea tarea fácil conseguir que las piezas encajen y a su vez favorecer la lectura y comprensión del subtítulo.

Obvia decir que la segmentación es una pieza más del engranaje de la subtitulación y que esta debe funcionar para que el producto final sea de calidad. En este sentido, Perego (2008:211) realizó un estudio en el que señala que:

This study is a reflection on the possible effects of line-breaks on subtitle readability and usability. It offers evidence that subtitle processing might be enhanced by careful segmentation of lines, i.e. a segmentation that is made in accordance with the principal rules of syntax.

De acuerdo con esto, tanto desde el entorno académico como empresarial coinciden en apuntar que los subtítulos no solo han de ser fáciles de leer sino que, lógicamente, deben respetar las reglas sintácticas. Como apunta Esté (2011) en su tesis doctoral:

Readability issues include the technical aspects of subtitle presentation on a macro-level (e.g. font, size, colour, and area of display) as well as micro-level aspects such as sentence structure, word choice and semantically coherent line breaks.

Igualmente, la mayor parte de normativas o libros de referencia (BBC Guidelines, 2018; Díaz Cintas & Remael, 2007; Ivarsson & Carroll, 1998; Karamitroglou, 1998; Ofcom, 2017; Perego, 2008), albergan entre sus recomendaciones indicaciones de cómo ha de realizarse la

segmentación de los subtítulos. La premisa desde la que se parte es que no se deben separar unidades lingüísticas; las diferencias entre las recomendaciones las encontramos en el nivel de concreción de las mismas. Díaz Cintas y Remael, (2007:173) expresan:

When we segment a sentence, we force the brain to pause its linguistic processing for a while [...] we should try to force this pause on the brain at a point where the semantic load has already managed to convey a satisfactorily complete piece of information.

Para Karamitrouglou (1998:6) es preferible que la segmentación se produzca, a ser posible, en los nodos sintácticos más altos, pero las restricciones que acompañan a la subtitulación no hacen de esto una tarea fácil. Hay que procurar no dividir unidades de sentido completo, por lo que es recomendable que la división se produzca entre el nodo N1 y N4. El siguiente diagrama propuesto por Karamitrouglou (1998:6) permite ver con claridad su propuesta de segmentación de subtítulos por nodos:

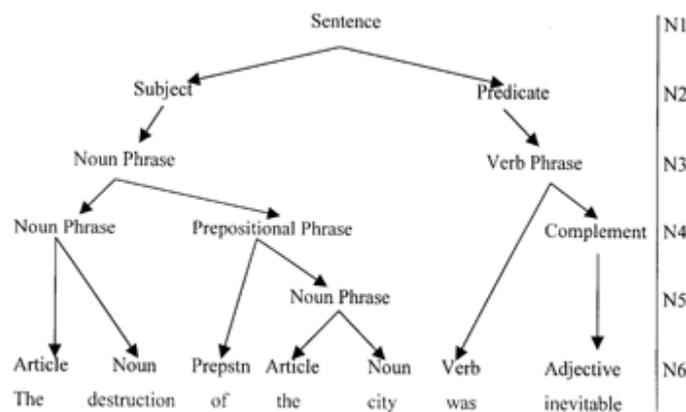


Figura 20. Propuesta de segmentación por nodos.

Asimismo, encontramos planteamientos similares de cómo segmentar por parte de empresas y organismos competentes en esta materia (BBC *Guidelines*, 2017; DCMP *Captioning Key*, 2017; Netflix *Time Text Style Guides*, 2018; norma UNE 153010, 2012) y todos parten del árbol de Karamitrouglou que ha sido y sigue siendo válido y tenido en cuenta por la mayoría. La diferencia, como apuntábamos en epígrafes anteriores, radica en la concreción de las unidades de sentido que no se deben romper.

Por otra parte, es cierto que no hay hasta la fecha abundantes estudios sobre la segmentación de los subtítulos desde los EDT que determinen si las normas que rigen este parámetro siguen existiendo o que permitan descubrir posibles cambios en la forma de segmentar subtítulos. Sin embargo, desde los estudios de recepción sí que encontramos trabajos recientes que ayudan a descubrir la percepción que los usuarios tienen de la

segmentación aportando información muy valiosa a este campo de estudio. En un trabajo interesantísimo de estas características, Gerber-Morón y Szarkowska (2018:2) señalan que «Previous research on subtitle segmentation, including studies with eye tracking, has been limited and inconclusive»; por este motivo, formula un análisis a partir de las diez unidades lingüísticas más comunes que vienen recogidas tanto en las guías que tomamos como referencia en nuestro estudio como en la mayor parte de guías y libros de referencia. El trabajo de las autoras comprueba si al romper estas unidades lingüísticas, que vendría a ser la ruptura de uno de los nodos sintácticos que no han de ser vulnerados según el árbol de nodos de Karamitroglou (1998), los receptores encuentran más dificultades para procesar la información. Por la solidez de su propuesta y el paralelismo existente con el trabajo de Tamayo, al definir la variable *segmentación* tendremos en cuenta estos análisis y las indicaciones que Díaz Cintas y Remael (2007) hacen en este sentido.

Antes de definir cómo vamos a estudiar la segmentación mostramos las recomendaciones que encontramos en las guías de referencia de nuestro trabajo:

Tabla 6. Recomendaciones de segmentación de los subtítulos para la SPS online.

Recomendaciones Segmentación de los subtítulos para SPS online		
Norma UNE	DCMP	BBC
1. Aprovechar pausas interpretativas y silencios.	1. Do not break a modifier (adjectives or adverbs) from the word it modifies.	1. Article and noun (e.g. the + table; a + book).
2. Aprovechar las pausas gramaticales o los signos de puntuación.	2. Do not break a prepositional phrase.	2. Preposition and following phrase (e.g. on + the table; in + a way; about + his life).
3. Escribir en la línea inferior las conjunciones y los nexos.	3. Do not break a person's name nor a title from the name with which it is associated.	3. Conjunction and following phrase/clause (e.g. and + those books; but + I went there).
4. No separa en líneas los sintagmas nominales, verbales y preposicionales.	4. Do not break a line after a conjunction.	4. Pronoun and verb (e.g. he + is; they + will come; it + comes).
	5. Do not break an auxiliary verb from the word it modifies.	5. Parts of a complex verb (e.g. have + eaten; will + have + been + doing).

Comprobamos que las recomendaciones son muy similares y sólo, de nuevo, varía el nivel de concreción y por tanto la exigencia, pues las ambigüedades en este sentido no favorecen a la segmentación en sí. Por su parte, Netflix, en su guía de estilo para la creación de subtítulos, describe nueve principios básicos que deben respetarse en la segmentación de los subtítulos. Los tres primeros indican que se ha de segmentar un subtítulo después de un

signo de puntuación y antes de conjunciones y preposiciones, mientras que para los siguientes seis señala que la segmentación del subtítulo no debe separar: un sustantivo de un artículo, un sustantivo de un adjetivo, un nombre propio del apellido, un verbo con preposición de la preposición y un verbo de un auxiliar, pronombre reflexivo o de la negación.

En nuestro estudio, para determinar las posibles causas de segmentación inadecuada, beberemos de distintas fuentes y propuestas de análisis para finalmente generar nuestra clasificación de causas de segmentación inadecuadas.

En cuanto a los criterios marcados desde la norma UNE, estos respetan la propuesta de Karamitroglou, pero añaden la partición del nodo 4 en el caso de sintagmas verbales, nominales y preposicionales, por lo que Tamayo (2015a:98) incluye esta consideración en la variable *segmentación* y define diez causas de segmentación inadecuada que le permiten definir si en el material que ella analiza la segmentación es adecuado o no. Al añadir, sin especificar, la partición de sintagmas verbales y nominales, se genera la ambigüedad que antes mencionábamos, por lo que en nuestra propuesta, seremos más concretos.

Por su parte, Gerber-Morón y Szarkowska (2018), en un estudio de recepción para determinar si la segmentación sintáctica facilita la lectura y comprensión de los subtítulos, proponen analizar diez unidades lingüísticas, las más habituales en lengua inglesa. Estas unidades coinciden con las unidades de sentido que no se deberían romper según las guías de referencia y las indicaciones de investigadores del ámbito como explicamos en epígrafe anterior. A partir de estas unidades llevan a cabo un estudio de recepción en el que se estudia la respuesta de los espectadores a distintas formas de segmentar el subtítulo (2018:5):

In one version, the two lines were segmented in accordance to subtitling standards, using syntactic rules to keep linguistic units on a single line (SS version). In the other version, syntactic rules were not followed and linguistic units were split between the first and the second line of the subtitle (NSS version).

Estas diez propuestas de segmentación no sintáctica se producen en los nodos 5 y 6 que Karamitroglou (1998) recomendaba evitar por lo que entendemos sería también una propuesta de diez causas de mala segmentación válidas para el análisis de subtítulos en inglés.

Tanto la propuesta de análisis de Tamayo (2015a) como la de Gerber-Morón y Szarkowska (2018) son exhaustivas y relativamente coincidentes de modo que, junto con las indicaciones recogidas en la obra de Díaz Cintas y Remael (2007) y la norma UNE, nos basaremos en ellas para generar nuestra propuesta de análisis de la segmentación.

Los valores que Tamayo (2015a:96) propone en su análisis son:

1. No poner al final de línea una preposición
2. Partición de locución preposicional
3. Pronombre al final
4. Partición de perífrasis verbal
5. Determinante al final
6. Adverbio al final
7. Partición de locución conjuntiva.
8. Conjunción al final
9. Partición del sintagma nominal
10. Partición del sintagma verbal

Las categorías que Gerber-Morón y Szarkowska (2018:5) proponen en su estudio son:

1. Indefinite article + noun
2. Definite article + noun
3. To + infinitive
4. Compound
5. Auxiliary + lexical verb
6. Sentence + sentence
7. Preposition
8. Possessive
9. Adjective + noun
10. Conjunction

Esta última clasificación es, en algunos puntos, más concreta que la anterior, pero a su vez, deja de lado algunas unidades que entendemos son indivisibles y bastante frecuentes en la lengua inglesa y que tendremos en cuenta en nuestro análisis.

Las consideraciones de Díaz Cintas y Remael (2007:175-178) nos servirán de pivote entre las clasificaciones anteriores para generar así una propuesta de análisis de la segmentación que nos permita comparar la versión española e inglesa del corpus. Con este objetivo en mente, hemos generado seis categorías con sus correspondientes subcategorías en ambas lenguas.

Tabla 7. Propuesta de análisis de causas de segmentación inadecuada.

CATEGORIAS GRAMATICALES	SEGMENTACIÓN ESPAÑOL	SEGMENTACIÓN INGLÉS
1. PREPOSICIÓN	1. <i>Prep</i> Preposición final de línea	1. <i>Prep</i> Preposición final de línea
2. DETERMINANTES	2. <i>Det_Indef</i> Artículo indefinido + SN 3. <i>Det_Def</i> Artículo definido + SN 4. <i>Det_Dem</i> Demostrativos + SN 5. <i>Det_Pos</i> Posesivos + SN	2. <i>Det_Indef</i> Artículo indefinido + NP 3. <i>Det_Def</i> Artículo definido + NP 4. <i>Det_Dem</i> Demostrativos + NP 5. <i>Det_Pos</i> Posesivos + NP
3. FORMAS VERBALES COMPUESTAS	6. <i>Verb_Comp</i> Partición verbos compuestos 7. <i>Verb_Perif</i> Partición de perífrasis verbal 8. <i>Verb_PredNom</i> Partición del predicado nominal del atributo	6. <i>Verb_Aux</i> Partición auxiliar + verbo 7. <i>Verb_Inf</i> Partición <i>To</i> + <i>infinitivo</i> 8. <i>Verbo frasal</i> Partición de un verbo frasal
4. CONJUNCIÓN	9. <i>Conj</i> Conjunción al final	9. <i>Conj</i> Conjunción al final
5. FRASES	10. <i>Frase</i> Frase completa + Frase completa	10. <i>Frase</i> Frase completa + Frase completa
6. SINTAGMA	11. <i>Sintg_Adj</i> Partición del sintagma nominal 12. <i>Sintg_Adv</i> Partición del sintagma adverbial	11. <i>Sintg_Adj</i> Partición del sintagma nominal 12. <i>Sintg_Adv</i> Partición del sintagma adverbial
7. PARTICIÓN NOMBRE COMPUESTO	13. <i>Nom_Comp</i> Partición nombre compuesto	13. <i>Nom_Comp</i> Partición nombre compuesto

Una vez establecida la definición y las variables de este parámetro, y siguiendo el modelo de representación de Tamayo (2015a), proponemos la siguiente tabla que recoge toda la información que analizaremos en este parámetro:

Tabla 8. Parámetro formato y posicionamiento.

PARÁMETRO	VARIABLE	VALORES POSIBLES
Parámetro formato y posicionamiento	Número de líneas	- 1 línea - 2 líneas - Más de dos líneas
	Número de caracteres por línea	- 32 - 37 Norma UNE - 42 Netflix ≥ 43
	Causas de segmentación inadecuada	- Preposición al final de línea - Determinante final de línea - Partición formas verbales compuestas - Conjunción al final de línea - Frase-Frase - Partición del sintagma - Partición nombre compuesto

En este capítulo hemos realizado la revisión teórica en torno a los parámetros objeto de estudio que nos permite realizar un análisis descriptivo de los mismos y reflejar la realidad de la práctica subtituladora *online* y descubrir si la SPS ofrecida respeta las normas prescritas y aceptadas en este ámbito.

En el siguiente capítulo, presentaremos la lingüística de corpus como metodología de estudio que, como apunta Baker (1993), está en sintonía con los estudios descriptivos y permite desvelar las características de la lengua traducida y las regularidades en la conducta traductora. En este mismo sentido, Corpas Pastor (2012:12) indica:

La metodología de corpus posibilitaba por primera vez el estudio detallado y sin complejos de la traducción como fenómeno y, en consonancia con ello, de las características intrínsecas de la lengua traducida.

A la vez que señala que (2012:14):

[...] el análisis de corpus también puede arrojar resultados muy interesantes en cuanto al estudio de las normas de traducción y la caracterización de diversos estilos traductores.

En especial, como apunta Baker (1992), la lingüística de corpus como metodología es especialmente adecuada para la investigación de las normas propuestas por Toury (en concreto las matriciales y textuales), pues que para detectar regularidades y normas es

necesario analizar un corpus de textos traducidos, pero ese corpus debe ser «a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria» (Bowker y Pearson, 2002:9).

CAPÍTULO 8

LINGÜÍSTICA DE CORPUS COMO METODOLOGÍA CUANTITATIVA

- 8.1. La lingüística de Corpus
- 8.2. La lingüística de Corpus y el corpus lingüístico
- 8.3. Clasificación y tipos de corpus lingüísticos

CAPÍTULO 8. LINGÜÍSTICA DE CORPUS COMO METODOLOGÍA CUANTITATIVA

La comunidad científica no acaba de ponerse de acuerdo en cuanto a la percepción de la Lingüística de Corpus. En este sentido, Taylor (2008:180) señala que mientras unos se refieren a ella como metodología, otros a disciplina, a paradigma o a herramienta. Y así son muchos los autores (McEnery y Wilson, 1996; Meyer, 2002; Bowker y Pearson, 2002) que la consideran una metodología de análisis; sin embargo, autores como Sinclair (1991) y Simpson y Swales (2001) la entienden como una técnica en la que el corpus es absolutamente central. Este apartado nos servirá a modo de presentación de la Lingüística de Corpus (LC) ya que en él realizaremos un breve recorrido por algunos de sus aspectos más relevantes, lo cual nos permitirá enmarcar nuestro trabajo en lo que, para nosotros, es una metodología de análisis que encaja a la perfección con los EDT y que nos permite detectar regularidades en el conjunto de los textos concretos que componen nuestro corpus.

8.1. La Lingüística de Corpus

Es ineludible el carácter central del corpus en la LC, razón por la que comenzaremos definiendo este concepto en el marco de esta metodología. Por todos es conocido que un corpus es una colección de textos compilados de una forma concreta que permite realizar estudios relacionados con la lengua. Una de las figuras indiscutibles en este ámbito, Sinclair, lo define primero (1991:171) como «a collection of naturally occurring language text, chosen to characterize a state or variety of a language», para años más tarde, con el impacto de los avances tecnológicos en la LC, el autor (2005) abordarlo del siguiente modo:

A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research.

Por su parte, Baker (1995:226) entiende el corpus en los siguientes términos:

I intend to use corpus to mean any collection of running texts (as opposed to examples/ sentences), held in electronic form and analyzable automatically or semi-automatically rather than manually).

Es decir, para la autora un corpus es un conjunto de textos orales o escritos que pueden ser procesados por ordenador y que tienen una extensión suficiente para ser representativos. Baker (1993), pionera de los Estudios de Traducción Basados en Corpus (ETBC) pretende describir la naturaleza de la lengua traducida analizando una gran cantidad de textos resultado del trasvase lingüístico para así poder establecer cuáles son las normas que definen la traducción. Realizar un análisis de esta envergadura requiere del uso de herramientas de gestión de corpus que se revelan como indiscutibles para este nuevo enfoque. Baker (1995) entiende que hay una serie de rasgos que comparten todas las traducciones, los *universales de traducción*, que están en un nivel superior al de la norma, pues estos rasgos son inherentes a la traducción. Somos conscientes del interés que tiene el examen de los universales, pero en esta ocasión no son el objeto de nuestro estudio, puesto que aquí nos centramos en la detección de regularidades y normas de traducción en el marco de los Estudios Descriptivos de la Traducción; sin embargo, la LC entendida como metodología se ajusta perfectamente a nuestras necesidades.

Por último, queremos dejar cuenta de la definición de corpus que goza, a nuestro entender, de mayor aceptación, que no es otra que la que se propone en el Expert Advisory Group on Language Engineering Standards (EAGLES) (1996): «A collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language».

La intención de este trabajo es analizar la subtitulación (al inglés y al español) de vídeos generados bajo el paraguas del Parlamento Europeo y emitidos por su canal de televisión, EuroparlTV (que recientemente ha cambiado de nombre y pasa a ser el Multimedia Center), como forma de representar la subtitulación *online* realizada por profesionales de la traducción en las instituciones y organizaciones europeas. Las razones de nuestra elección son la inquietud por descubrir la calidad de la subtitulación de los contenidos audiovisuales ofrecidos desde sitios web que tienen un compromiso adquirido con la accesibilidad y habrían de ser ejemplo de las políticas que aprueban y defienden.

8.2. La Lingüística de Corpus y el corpus lingüístico

La LC ha evolucionado al compás del desarrollo tecnológico y con cada avance las posibilidades de analizar la lengua han crecido en distintos sentidos. Resumimos a continuación los cambios sobrevenidos en la LC como resultado de los avances tecnológicos. Comenzamos el repaso con los primeros corpus digitalizados pero antes de empezar no debemos olvidar como apunta Muñoz Cabrera, (2017:18) quien afirma que:

Aunque la lingüística de corpus como tal es relativamente reciente (tanto si es considerada como metodología o como disciplina), su práctica es tan antigua como la investigación lingüística. La lingüística histórica siempre se ha fundamentado en el corpus.

Nuestro recorrido histórico comienza con la primera generación de corpus. En esta etapa se establecen los principios de la compilación de corpus electrónicos con la creación del *Corpus de Brown* (1964), un corpus de un millón de palabras extraídas de 500 textos publicados en EE.UU. Posteriormente, en los años 70 se creó el *LOB* (Lancaster-Oslo-Bergen) corpus, que utilizó una metodología similar a la del *Brown Corpus* pero en inglés británico. Sin embargo, el corpus más relevante de esta época, liderado por John Sinclair, fue el *English Lexical Studies*, corpus de un millón de palabras de textos orales y escritos, que posteriormente le permitió realizar un estudio que, por primera vez, demuestra estadísticamente las concurrencias. Esta primera generación de la LC se caracteriza por el tamaño de los corpus (hasta un millón de palabras) y por ser los primeros corpus procesados con ordenador, sin olvidar que existía una parte de procesamiento que era manual puesto que los avances tecnológicos no permitían otra cosa.

En la segunda generación de corpus lingüísticos, conocida como *megacorpus*, por su gran tamaño, destaca el corpus *Cobuild* dirigido por Sinclair (1987) que aunque en sus inicios no fuera un megacorpus (contaba con 7 millones de palabras) fue clave en el desarrollo de la LC porque se desarrollaron herramientas para el manejo del corpus y se utilizó con fines más allá de los académicos. El *Cobuild* es el actual *Bank of English* (BOE) con más de 200 millones de palabras. Esta generación se caracteriza por la incorporación de los ordenadores personales en la investigación y por la apertura a otros ámbitos de la compilación de estos corpus.

Por su parte, la tercera generación según Muñoz Cabrera (2017:18):

[...] se inicia en la década de los noventa y destaca por la amplia capacidad de almacenamiento, la recopilación, no solo a nivel escrito, sino a nivel oral y transcrito, y por su accesibilidad en Internet. La particularidad de esta fase reside en que ya es posible analizar otras lenguas (además de la inglesa), la aparición de herramientas que permiten realizar múltiples funciones, y el empleo de interfaces de fácil uso, incluso para los lingüistas con limitadas nociones de informática.

Actualmente, nos encontramos en la era de los *gigacorpus*, pues la unión del *Big Data* y de la LC permite generar corpus de billones de palabra que los convierte en corpus que, definitivamente, son representativos de una lengua. Sin embargo, analizar corpus de este tamaño requiere de herramientas de gestión de corpus diferentes a las existentes, por lo que el alojamiento de estos corpus pasa a estar en Internet y herramientas como CQPweb (Hardie, 2012) o SketchEngine (Kirgariff, 2013) son imprescindibles para trabajar con

corpus de estas dimensiones. Por otro lado, si bien la etapa en la que nos encontramos se caracteriza por la abrumadora dimensión de los corpus, no hemos de olvidar que también se crean corpus para estudiar patrones lingüísticos en contextos específicos más reducidos, de modo que distintos tipos de corpus cohabitan en el espacio investigador actual.

8.3. Clasificación y tipos de corpus lingüísticos

En lo que respecta a la clasificación de corpus existe un amplio repertorio de clasificaciones tanto desde la LC como desde los Estudios de Traducción basados en Corpus; sin embargo, este apartado no pretende mostrar las múltiples taxonomías que describen los tipos de corpus, sino más bien mencionar las más significativas que nos permitan clasificar nuestro corpus de estudio.

A nuestro juicio, las taxonomías de corpus más destacables son las de García Izquierdo (2015) y Faya (2005), si bien nos centraremos en la primera por lo que la trataremos con mayor detalle a continuación. No obstante, no queremos dejar de mencionar otras propuestas de gran relevancia que se han generado desde la LC, como por ejemplo las clasificaciones que ofrecen Sardinha (2000:340-341) y Procázková (2006:5) o, desde los ETBC, la clasificación de Bowker y Pearson (2002:45-46), que desde su sencillez nos permite ver una clasificación clara a partir de tres parámetros: número de idiomas (monolingüe, bilingüe y multilingüe), especialización de los textos (generales y especializados) y la relación entre ellos (paralelos o comparables). Importantísima es también la propuesta de Laviosa (2003:34-38), extensamente citada, que divide los corpus (principalmente traductores) a partir de seis parámetros que, de forma jerárquica, permiten una descripción muy detallada.

La tipología de corpus de García Izquierdo (2005:4) bebe de las propuestas de Corpas (2001), García Izquierdo y Monzó (2003) y Baker (1995), entre otros, y distingue los siguientes tipos:

- a) Según el porcentaje y la distribución de los diferentes tipos de texto: corpus grande o extenso, corpus equilibrado, corpus piramidal, corpus monitor, corpus paralelo y corpus comparable.
- b) Según la especificidad de los documentos contenidos en el corpus: corpus general, corpus especializado, corpus canónico, corpus periódico o cronológico y corpus diacrónico.

- c) Según la cantidad de texto que se reúne en cada uno de los documentos que componen el corpus: corpus textual, corpus de referencia y corpus léxico.
- d) Según la codificación y anotación: corpus no anotado, corpus anotado.
- e) Según la documentación que acompaña al corpus: corpus documentado y corpus no documentado.

Por otra parte, nos parece muy reveladora la propuesta de clasificación de Faya (2015:350-352), en la que plantea una clasificación en forma de árbol que da cabida tanto a la Lingüística Contrastiva como a los Estudios de Traducción. Su taxonomía se divide en dos partes: la primera establece los tipos de corpus según lo que la autora denomina aspectos formales (medio de difusión, grado de amplitud del análisis, momento temporal, posibilidad de adición, extensión de los textos, publicación y temática tratada). La segunda, centrada en aspectos lingüísticos, establece tipos de corpus según el número de lenguas, la composición de originales y traducciones y la relación que se establece entre ellos, la dirección de la traducción y, finalmente, la alineación⁷⁸.

En este capítulo hemos abordado brevemente la metodología de la lingüística de corpus en la que enmarcamos nuestro estudio. Llegados a este punto, el siguiente capítulo lo dedicamos a la metodología de compilación de nuestro corpus y a la contextualización de los textos objeto de estudio.

⁷⁸ Para ver una explicación extensa de esta clasificación, consúltese el artículo «Propuesta de clasificación corpus textuales» de Faya (2015). Puede consultarse en línea en <https://uvadoc.uva.es/bitstream/10324/16449/1/Propuesta-clasificaci%C3%B3n-corpus.pdf>

CAPÍTULO 9

METODOLOGÍA DE COMPILACIÓN Y PREPARACIÓN DEL CORPUS DE ESTUDIO

- 9.1. Contextualización del corpus de estudio
- 9.2. Descripción del corpus EMPAC
- 9.3. Compilación y preparación del corpus de estudio
- 9.4. Obtención de los datos y metadatos
- 9.5. Procesamiento de los textos: transformación de los archivos en .xml
- 9.6. Procesamiento lingüístico de los textos
 - 9.6.1. Etiquetado morfosintáctico con *TreeTagger*
 - 9.6.1.1. Procesamiento previo
 - 9.6.1.2. Etiquetado morfosintáctico
 - 9.6.1.3. Postproceso
 - 9.6.2. División del texto en frases
 - 9.6.3. Indexación y alineado del corpus
 - 9.6.4. Exportación al servidor de CQPWeb

CAPÍTULO 9. METODOLOGÍA DE COMPILACIÓN Y PREPARACIÓN DEL CORPUS DE ESTUDIO

Este trabajo consiste en el análisis de las normas técnicas de subtitulación *online* a través del estudio del corpus EMPAC, un conjunto de textos audiovisuales del portal del Parlamento Europeo. El primer paso de nuestro análisis consiste en compilar un corpus electrónico sobre la base de la metodología de la lingüística de corpus. Sirva este capítulo para explicar cómo hemos compilado el corpus y de qué información lo hemos dotado para poder realizar análisis desde distintas perspectivas.

9.1. Contextualización del corpus de estudio

Para comenzar, y atendiendo a las clasificaciones de corpus que describimos en el capítulo anterior, parece oportuno explicar que EMPAC es un corpus textual (puesto que no es representativo de una lengua), escrito, especializado (describe un sublenguaje), diacrónico y compuesto por textos completos que son del dominio público. Es a su vez un corpus multilingüe (contiene más de una lengua; en nuestro caso, inglés y español), paralelo y unidireccional (todas las traducciones son a partir de la misma lengua origen, el inglés) y alineado. Es este último sentido, Baker (2006:9) aclara que la alineación puede producirse en el nivel de la palabra o de la frase; sin embargo, las nuevas tecnologías permiten alinear en el nivel del subtítulo, como ocurre en este trabajo. Por último, resta añadir que es, según la clasificación de García Izquierdo (2005), un corpus anotado y documentado puesto que incluye metadatos y etiquetado morfosintáctico que brinda información relevante sobre el texto.

El diseño del corpus EMPAC se construye con la intención de analizar la subtitulación de vídeos emitidos en Internet, generados por profesionales y con el objetivo de estudiar su evolución a lo largo de los años. De hecho, uno de los propósitos de nuestro estudio nace de la propuesta que Díaz Cintas (2012:276) lanzó a propósito de la necesidad de examinar el aumento de la velocidad de lectura de los subtítulos en distintos ámbitos:

The current state of affairs is one in which subtitling reading speeds tend to be considerably higher for the cinema and the DVD than for the television. The same can be applied to professional online subtitling, although, as a relative new comer, little research has been conducted in this field to verify it.

Consideramos que no hay mejor espacio para encontrar material de las características descritas por el autor que las páginas web de las instituciones europeas que emiten vídeos con subtítulos en varias lenguas.

EUROPA es el portal oficial de la Unión Europea y a él se encuentran conectadas todas las páginas web de las instituciones y agencias europeas. La Unión Europea mantiene un compromiso con la accesibilidad, esta conexión implica que, *de facto*, las webs vinculadas a este portal se adhieren al compromiso de esta web con respecto a su política de accesibilidad, tal y como queda reflejado en la siguiente cita extraída de su web:

Todas las páginas oficiales de las instituciones de la UE deben seguir unas directrices internacionales de accesibilidad a los contenidos virtuales. Es decir, que el mayor número posible de personas, sin discriminación, debe poder consultar y entender los textos, imágenes, formas, sonidos, etc.⁷⁹

Como indicamos en el capítulo 4, la conformidad con las WCAG 2.0 supone que los vídeos dispongan de subtitulación y de audiodescripción, de modo que en las páginas de las instituciones y organismos europeos el material audiovisual ha de ser accesible para todas las personas. Por la relevancia que tienen en la vida de los ciudadanos europeos las instituciones de la UE⁸⁰, decidimos interesarnos por ellas en nuestro estudio. Y así, con el fin de detectar en cuáles los vídeos ocupaban un lugar central y valorar la accesibilidad de los mismos, consultamos todas las sedes web de las siguientes instituciones: el Parlamento Europeo, el Consejo Europeo, el Consejo, la Comisión Europea, el Tribunal de Justicia de la Unión Europea, el Banco Central Europeo y el Tribunal de Cuentas.

Finalmente decidimos analizar la web del Parlamento Europeo puesto que aquí se aloja el canal de televisión *online* EuroparlTV que, como declaró el 17 de septiembre de 2008 el entonces presidente del PE (Hans-Gert Pöttering) durante la presentación del canal, surgió en un momento en el que se hacía necesario acercar el Parlamento a los ciudadanos por medio de un formato moderno y creativo y justo antes de unas elecciones de gran importancia:

As we approach the European elections of June 2009, EuroparlTV should be an excellent Internet tool for citizens, especially young people, to keep themselves informed about the activities and decisions of the directly-elected European Parliament - decisions which have an impact on the everyday lives of almost 500 million European Union citizens.

⁷⁹ https://europa.eu/european-union/abouteuropa/accessibility_es

⁸⁰ https://europa.eu/european-union/about-eu/institutions-bodies_es

Por aquel entonces fue definido como un canal único en el mundo, en esencia por ofrecer los vídeos subtítulos a más de 20 lenguas, característica que, a día de hoy, sigue siendo definitoria de este canal aunque las lenguas oficiales de la actualidad son ya 24.

Dos años más tarde, en octubre de 2010, se presentó una versión mejorada de la web de EuroParlTV con el objetivo de mejorar los siguientes aspectos:

Para poder informar a los usuarios de las actividades que lleva a cabo el Parlamento Europeo, esta institución ha introducido una serie de mejoras en el sitio Web EuroParlTV que incluyen tanto una nueva estética de la página, como el acceso directo a programas, un buscador más veloz y la integración en las redes sociales.

Los cambios afectaron principalmente a la visualización y a la potencia del motor de búsqueda. Sin embargo, la estructura y las secciones de EuroParlTV cambiaron tan solo ligeramente para introducir, eso sí, contenidos que alimentaran la curiosidad de los más jóvenes. Se pretendía, igualmente, que la web pudiera usarse con fines educativos a fin de favorecer el acercamiento del PE al aula. De esta manera, a partir de entonces los contenidos de EuroParlTV se dividen en tres canales: Noticias, Parlamento Joven y Descubre el Parlamento.

Los tres canales cubren toda la información relacionada con el Parlamento Europeo, pero, sin duda, el de mayor relevancia es el canal Noticias, donde se abordan asuntos de interés general, actualidad e información de última hora a través de noticias, entrevistas o reportajes. Por su parte, Parlamento Joven, que se caracteriza por su mayor dinamismo y por pretender acercar el PE al público juvenil, encuentra en Blink⁸¹, un personaje virtual, la mejor forma de explicar a los niños y adolescentes la idea de Europa y de su cometido; sin embargo, este canal dejó de emitirse en 2014. Por último, en Descubre el Parlamento, encontramos información sobre la fundación de Europa, desde sus inicios, hasta la composición del Parlamento en la actualidad. Con el paso del tiempo, estos canales se fueron desdibujando para dar paso a un amplio repertorio de programas que quedaron repartidos en cinco secciones: Background, Discovery, History, Interviews y News. A modo de ejemplo, mencionaremos que Discovery contiene todos los contenidos que se crearon para Parlamento Joven junto con los perfiles de los eurodiputados, mientras que History y Background emite una oferta vinculada con la historia de la creación de la UE y de los problemas más acuciantes de este espacio geopolítico en la actualidad.

Sin embargo, esta estructura ha dado paso, recientemente, a otra completamente nueva. El 4 de enero de 2018 EuroParlTV pasó a ser el Centro Multimedia del Parlamento⁸² y desde

⁸¹ https://multimedia.europarl.europa.eu/en/az-b-for-blink-ii_J002-0101_ev (Ejemplo de *Blink*).

⁸² <https://multimedia.europarl.europa.eu/es/home>

ese día, la información a la que acceden los ciudadanos está ordenada de forma muy distinta. El comunicado de prensa de la presentación de este centro resume los cambios de la nueva plataforma de la siguiente manera:

La nueva plataforma cuenta con una interfaz más intuitiva y un motor de búsqueda avanzado para encontrar con facilidad todos los recursos audiovisuales del PE [...]. El Centro Multimedia permite ver, descargar y compartir recursos audiovisuales con un solo clic. Además, se podrá descargar cualquier material, sin necesidad de registrarse en el sitio.

Queda claro que, en esta ocasión, los cambios han sido significativos, puesto que ponen a disposición de los ciudadanos toda la actividad parlamentaria en 24 lenguas y durante 24 horas al día, e incluye tanto las sesiones parlamentarias como las reuniones de las comisiones. Huelga decir que las modificaciones han sido introducidas paulatinamente, puesto que, desde finales de 2016, concretamente en julio de 2016, fecha en la que se adjudicó el concurso internacional para desarrollar y prestar servicios a EuroparlTV, comenzaron a evidenciarse los primeros cambios. Un consorcio de empresas del Grupo Lavinia⁸³ ha sido la encargada de desarrollar esta nueva plataforma en la que no solo se facilita el acceso a todo el material sino que también se han cambiado las secciones y la estructura que tenía EuroparlTV desde 2008 hasta 2017.

El Parlamento Europeo, a partir de un procedimiento de licitación internacional, asignó la creación del renovado canal de televisión web a dos empresas; Twofour se hizo cargo de la plataforma, el alojamiento de datos y las emisiones, mientras que ICF Mostra del contenido. Esta última empresa, encargada, entre otras funciones, de generar los subtítulos de EuroparlTV creó, según nos indicó el responsable⁸⁴ del contenido en sus inicios, los primeros subtítulos incluso antes de que EuroparlTV comenzara su andadura: «They started in September 2008 when EPTV was launched– actually even before that, because a number of files are from the prototype before the actual launch of EPTV». Estos subtítulos no siempre se traducían al inglés, puesto que en esta fase de prueba todavía no se había implementado el sistema de traducción pivote (a partir del inglés) que se pondría en marcha cuando EuroparlTV comenzó oficialmente. Esto viene a explicar la diferencia en el número de textos (vídeos) de los que disponemos subtítulos en español y en inglés y que explicaremos con más detalle en el capítulo 11. ICF Mostra se encargó de generar los subtítulos de los vídeos de EuroparlTV desde los inicios hasta octubre de 2016 cuando CPSL, empresa que forma parte del grupo Lavinia, tomó las riendas de las traducciones y subtítulos de lo que ahora pasa a denominarse Multimedia Center. Resulta interesante

⁸³ <https://www.lavinia.tc/noticias/nace-el-multimedia-centre-del-parlamento-europeo-de-la-mano-de-lavinianext/>

⁸⁴ Andreas Dielt, Web Content Manager EuroparlTV hasta finales de 2011

constatar que ha habido un cambio significativo en la forma de hacer subtítulos desde que esta nueva empresa se responsabilizara de la subtitulación de los contenidos, pero pasaremos a explicar estas cuestiones en el apartado de análisis.

Por otra parte, la traducción audiovisual en el contexto institucional nos ha parecido un marco de indiscutible valor para analizar el cumplimiento de los requisitos de accesibilidad establecidos por las WCAG 2.0 y describir la subtitulación profesional en Internet, lo cual posibilita la comparación con la subtitulación en otros canales (TV, cine, DVD, etc.). Como apunta Fuentes-Luque (2015: 14), las instituciones u organizaciones públicas:

[...] have a duty to make all the information relating to their activities public and accessible in response to their citizens' right to know and understand what is happening within them.

El autor señala que estas instituciones han cambiado significativamente su forma de transmitir la información reduciendo drásticamente el material imprimible para pasar a convertir el contenido audiovisual en su forma de comunicación predominante. Las razones para que este cambio se produjera son que «audiovisual and multimedia platforms provide an attractive, immediate, user-friendly, readily updatable way of delivering information».

Fuentes-Luque (2015) realiza un estudio, el primero a su entender, de la traducción del material audiovisual en las páginas web de instituciones internacionales. El autor persigue el objetivo de describir, por una parte, hasta qué punto estas instituciones hacen uso de material audiovisual y, por otra parte, analizar aspectos relacionados con la traducción (2015:17).

The focus will be on the conventions and technical aspects followed and applied. In any case, as it has been noted above, this is a starting point for further and wider research on the topic.

Este primer estudio muestra los resultados del análisis de los siguientes sitios web: EuroparlTV, European Commission Webcast Portal, Europe by Satellite (EbS), United Nations Webcast. Entre las conclusiones de este estudio, cabe destacar el hecho de que, por una parte, parece indiscutible que las instituciones consideran primordial el uso del material audiovisual. Por otra, el autor las anima a trabajar de forma conjunta para establecer estándares comunes a la hora de implementar la TAV con el fin de asegurar la calidad y la accesibilidad del contenido.

En definitiva, de un lado, Díaz Cintas anima al estudio de los aspectos técnicos de la subtitulación en internet mientras que Fuentes-Luque invita a hacerlo en las webs de instituciones y organismos internacionales por el compromiso tácito que estas tienen con la accesibilidad y el derecho a la información de todas las personas.

Nuestro corpus, como revelamos al principio de este apartado, está formado por los subtítulos en inglés y en español de vídeos del Parlamento Europeo y el análisis profundiza en aspectos técnicos y lingüísticos de los subtítulos en este canal, el EuroparlTV.

9.2. Descripción del corpus EMPAC

El EuroparlTV Multimedia Parallel Corpus (EMPAC)⁸⁵ es, según la taxonomía que describimos anteriormente, un corpus bilingüe (en inglés y en español) de subtítulos de los vídeos distribuidos por el Multimedia Center, que anteriormente (hasta julio de 2018) se denominaba canal EuroparlTV. A partir de procedimientos inspirados en la metodología del grupo de investigación de la Universitat Jaume I *European Comparable and Parallel Corpus of Parliamentary Speeches* (ECPC)⁸⁶, para cada vídeo se ha recogido el archivo de los subtítulos, se han anotado los metadatos del texto (título, fecha de publicación, tipo de contenido, URL, etc.) y la estructura del documento (compuesto por subtítulos y líneas), se ha añadido información lingüística (mediante procesos de tokenización, lematización y análisis morfosintáctico) obtenida con *TreeTagger*⁸⁷ (Schmid, 1994, 1995), se ha segmentado el texto en oraciones con NLTK (Bird *et al.*, 2009), y, finalmente, se han alineado los subtítulos entre ambos idiomas (inglés y español). El resultado es un corpus compuesto aproximadamente de 4 000 textos, más de 200 000 subtítulos, 2,5 millones de palabras y una duración cercana a las 280 horas para cada lengua. La riqueza de los datos, metadatos y anotaciones con el que lo hemos dotado nos ha permitido analizar distintos parámetros inherentes a la subtitulación.

A continuación describiremos, en datos estadísticos, el corpus EMPAC, primero en cifras totales, para luego explicar la organización y distribución de los datos que contiene según las diferentes formas en las que podremos investigar con este material.

Tabla 9. Datos de EMPAC_EN.

Textos	Subtítulos	Tokens (Componentes léxicos)	Duración
3812	224.036	2.448.772	274 horas y 12 min

Tabla 10. Datos EMPAC_ES.

Textos	Subtítulos	Tokens (Componentes léxicos)	Duración
3922	243.660	2.463.944	283 horas y 2 min

⁸⁵ <https://cqpweb.empac-corpus.com/>

⁸⁶ El grupo de investigación ECPC es responsable de los siguientes proyectos de investigación con financiación ministerial: HUM2005-03756filo, FFI2008-01610, FFI2012-39389. También es responsable de los siguientes proyectos de financiación por parte de la Generalitat Valenciana o de la Universitat Jaume I: AICO17-082, UJI201-64, UJI2017-25).

⁸⁷ <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

El número total de subtítulos de nuestro corpus asciende a 467 694 que aglutinan entre ambos 4912716 palabras; es decir, casi cinco millones de palabras en medio millón de subtítulos.

La distribución de estos datos en años queda reflejada en la siguiente tabla para la versión en inglés, EMPAC_EN:

Tabla 11. Distribución de los datos por año en EMPAC_EN.

Year_En	texts	subtitles	tokens	duration
2009	355	23120	247647	40,14
2010	744	48533	510721	57,31
2011	563	34054	372787	39,9
2012	592	37392	417046	43,29
2013	484	34704	385908	38,97
2014	356	21532	244466	25,08
2015	361	13939	156020	16,57
2016	203	6996	77264	8,28
2017	154	3766	36913	4,18

La distribución de los datos en años queda reflejada en la siguiente tabla para la versión en español, EMPAC_ES:

Tabla 12. Distribución de los datos por año en EMPAC_ES.

Year_ES	texts	subtitles	tokens	duration
2009	464	41536	403839	50,92
2010	757	50286	493231	59,33
2011	566	34452	348906	39,99
2012	595	38420	390905	43,38
2013	483	35288	361153	38,93
2014	346	19031	200571	21,83
2015	358	13918	149752	16,49
2016	194	6710	73089	7,86
2017	159	4019	42498	4,46

Podemos observar que 2010 fue el año de mayor producción y que los últimos dos años, 2016 y 2017, el descenso es considerable. La razón del descenso en las emisiones podría estar relacionada con el cambio de empresa que gestiona los contenidos del ahora Multimedia Center, puesto que como comentábamos anteriormente fue en 2016 cuando se asigna al nuevo grupo empresarial.

Por su parte, los contenidos, como anunciábamos arriba, están etiquetados también por tipo de texto (Background, Discovery, History, Interviews y News), lo que nos permite generar subcorpus o análisis por tipo de texto. La distribución del total de datos por tipo de texto en inglés es la que sigue:

Tabla 13. Distribución de los datos por tipo de texto en EMPAC_EN.

Type_EN	texts	subtitles	tokens	duration
Background	675	99353	1091938	126,47
Discovery	386	15900	169765	18,05
History	111	4365	46781	5,64
Interview	405	30670	339448	36,39
News	2235	73748	800840	87,17

La distribución del total de datos por tipo de texto en español aparece debajo:

Tabla 14. Distribución de los datos por tipo de texto en EMPAC_ES.

Type_ES	texts	subtitles	tokens	duration
Background	683	109054	1089441	127,18
Discovery	393	16632	167214	18,42
History	113	4608	46634	5,79
Interview	427	34156	344235	40,32
News	2306	79210	816420	91,49

Creemos interesante, asimismo, exponer la distribución de los tipos de texto por años:

Tabla 15. Distribución de los datos por tipo de texto y año en EMPAC_EN.

Year_EN	type	texts	subtitles	tokens	duration
2009	Background	100	10688	113795	23,48
	Discovery	49	1665	17854	2,26
	History	12	349	3487	0,63
	Interview	39	5445	58665	6,84
	News	155	4973	53846	6,94
2010	Background	126	19573	209609	23,1
	Discovery	78	3130	33133	3,48
	History	16	804	8328	0,96
	Interview	56	4162	43626	4,93
	News	468	20864	216025	24,84
2011	Background	100	14748	161903	17,33
	Discovery	45	2376	24756	2,56
	History	11	650	6993	0,78
	Interview	43	2745	30049	3,23
	News	364	13535	149086	16

(Tabla 15 continúa en la página siguiente)

(Tabla 15 continúa de la página anterior)

Year_EN	type	texts	subtitles	tokens	duration
2012	Background	135	21880	243369	25,39
	Discovery	51	2182	23095	2,36
	History	11	637	6765	0,75
	Interview	54	3596	40090	4,1
	News	341	9097	103727	10,69
2013	Background	107	20576	228074	23,15
	Discovery	47	2181	23490	2,33
	History	8	518	5459	0,59
	Interview	43	2949	33184	3,4
	News	279	8480	95701	9,5
2014	Background	47	8671	99611	10,17
	Discovery	41	1575	17279	1,78
	History	49	1226	14037	1,73
	Interview	63	4650	53094	5,43
	News	156	5410	60445	5,98
2015	Background	26	1993	22409	2,36
	Discovery	48	1827	19653	2,2
	Interview	62	4698	54037	5,61
	News	225	5421	59921	6,4
2016	Background	16	687	7513	0,85
	Discovery	17	590	6607	0,67
	Interview	34	2057	23401	2,46
	News	136	3662	39743	4,29
2017	Background	18	537	5655	0,64
	Discovery	10	374	3898	0,42
	History	4	181	1712	0,21
	Interview	11	368	3302	0,38
	News	111	2306	22346	2,53

A continuación, sintetizamos la distribución de los datos por años en la versión en español:

Tabla 16. Distribución de los datos por tipo de texto y año en EMPAC_ES.

Year_ES	type	texts	subtitles	tokens	duration
2009	Background	113	21401	207851	26,35
	Discovery	49	1884	18418	2,26
	History	14	538	5128	0,78
	Interview	60	8870	85275	10,81
	News	228	8843	87167	10,73

(Tabla 16 continúa en la página siguiente)

(Tabla 16 continúa de la página anterior)

Year_ES	type	texts	subtitles	tokens	duration
2010	Background	131	20692	200775	24,35
	Discovery	80	3221	31569	3,59
	History	16	805	7784	0,96
	Interview	57	4245	41302	5,03
	News	473	21323	211801	25,41
2011	Background	100	14661	147795	17,14
	Discovery	46	2446	23706	2,62
	History	11	666	6849	0,78
	Interview	43	2749	27397	3,23
	News	366	13930	143159	16,22
2012	Background	135	22159	223976	25,39
	Discovery	51	2242	21948	2,36
	History	11	662	6630	0,75
	Interview	54	3607	36377	4,1
	News	344	9750	101974	10,78
2013	Background	107	20854	210326	23,15
	Discovery	47	2261	22656	2,33
	History	8	526	5302	0,59
	Interview	43	2973	30428	3,4
	News	278	8674	92441	9,47
2014	Background	37	6006	63067	6,92
	Discovery	41	1591	16730	1,78
	History	49	1230	13159	1,73
	Interview	63	4669	48815	5,43
	News	156	5535	58800	5,98
2015	Background	24	1955	21079	2,29
	Discovery	48	1844	19434	2,2
	Interview	62	4708	50528	5,61
	News	224	5411	58711	6,38
2016	Background	16	704	7646	0,85
	Discovery	17	590	6593	0,67
	Interview	32	1885	20262	2,25
	News	129	3531	38588	4,08
2017	Background	20	622	6926	0,74
	Discovery	14	553	6160	0,62
	History	4	181	1782	0,21
	Interview	13	450	3851	0,45
	News	108	2213	23779	2,44

Por último, dado que los datos y las etiquetas de nuestro corpus proceden del HTML de la web del Parlamento Europeo, comprobamos que la estructura original incluía también una

etiqueta denominada *category* que incorporamos a nuestro xml con la información de cada archivo SRT. Esta etiqueta apunta al área temática del vídeo (EU_affairs, Economy, Others, Security, Society, World) que, por cierto, no estaba en el formato original del EuroparlTV. Podemos ver la distribución de la información por categoría de la versión en inglés del corpus en la siguiente tabla:

Tabla 17. Distribución de los datos por categoría en EMPAC_EN.

Category_EN	texts	subtitles	tokens	duration
EU_affairs	147	4270	45033	4,84
Economy	44	1433	14701	1,6
Others	3455	213222	2334045	261,21
Security	43	1217	13318	1,47
Society	83	2649	28395	3,12
World	40	1245	13280	1,47

Asímismo, debajo se muestran las cifras equivalentes en la versión en español del corpus:

Tabla 18. Distribución de los datos por categorías en EMPAC_ES.

Category_ES	texts	subtitles	tokens	duration
EU_affairs	146	4249	46384	4,8
Economy	45	1551	16258	1,72
Others	3569	232878	2347850	270,83
Security	41	1210	13034	1,43
Society	81	2660	28445	3,1
World	40	1112	11973	1,31

A continuación, en la siguiente tabla (con la división de las categorías por años), vemos cómo, efectivamente, la etiqueta categoría se introdujo en 2016, que viene a coincidir con el cambio de grupo empresarial en la gestión de los contenidos del EuroparlTV. En la versión en inglés la distribución es la siguiente:

Tabla 19. Distribución de los datos por categoría y año en EMPAC_EN.

Year_EN	category	texts	subtitles	tokens	duration
2009	Others	355	23120	247647	40,14
2010	Others	744	48533	510721	57,31
2011	Others	563	34054	372787	39,9
2012	Others	592	37392	417046	43,29

(Tabla 19 continúa en la página siguiente)

(Tabla 19 continúa de la página anterior)

Year_EN	category	texts	subtitles	tokens	duration
2013	Others	484	34704	385908	38,97
2014	Economy	1	73	772	0,07
	Others	355	21459	243694	25,01
2015	Others	361	13939	156020	16,57
2016	EU_affairs	73	2738	30382	3,16
	Economy	21	678	7395	0,84
	Security	38	1096	12174	1,33
	Society	49	1680	18496	1,98
	World	22	804	8817	0,97
2017	EU_affairs	74	1532	14651	1,69
	Economy	22	682	6534	0,69
	Others	1	21	222	0,02
	Security	5	121	1144	0,14
	Society	34	969	9899	1,14
	World	18	441	4463	0,51

La división de las categorías por años en la versión en español puede consultarse debajo:

Tabla 20. Distribución de los datos por categoría y año en EMPAC_ES.

Year_ES	category	texts	subtitles	tokens	duration
2009	Others	464	41536	403839	50,92
2010	Others	757	50286	493231	59,33
2011	Others	566	34452	348906	39,99
2012	Others	595	38420	390905	43,38
2013	Others	483	35288	361153	38,93
2014	Economy	1	74	726	0,07
	Others	345	18957	199845	21,76
2015	Others	358	13918	149752	16,49
2016	EU_affairs	71	2696	29971	3,09
	Economy	21	678	7355	0,84
	Security	36	1089	11723	1,3
	Society	46	1616	17383	1,89
	World	20	631	6657	0,75
2017	EU_affairs	75	1553	16413	1,71
	Economy	23	799	8177	0,81
	Others	1	21	219	0,02
	Security	5	121	1311	0,14
	Society	35	1044	11062	1,22
	World	20	481	5316	0,56

En los estudios basados en corpus, resulta especialmente esclarecedor presentar por separado métodos de compilación y métodos de análisis. Es lo que hacemos en este trabajo. En el siguiente apartado describiremos los primeros y, para facilitar la lectura de la tesis, dejaremos la metodología del análisis para el siguiente capítulo.

9.3. Compilación y preparación del corpus de estudio

Los textos en formato digital que conforman nuestro corpus se han descargado de la web del Parlamento Europeo.⁸⁸ Si bien el formato digital ha comportado una serie de ventajas en cuanto a la recopilación del material, la complejidad derivaba de la gran cantidad de archivos que pretendíamos descargar, en concreto los subtítulos correspondientes a todas las secciones de EuroParlTV desde 2009 hasta 2017 y de la extracción de sus metadatos a partir de la información contenida en los HTLM.

Parece obvio que la principal ventaja de que los subtítulos estén en formato electrónico y que sean descargables es que evitamos tener que transcribirlos. En consecuencia, ahorramos tiempo y evitamos cometer errores en el proceso, además de la posibilidad de obtener muchos más datos que si la recopilación fuera un proceso manual. Sin embargo, el reto es conseguir descargar toda esa ingente cantidad de información de forma totalmente automática y eficiente. Un rastreador web o *webcrawler* (término que utilizaremos en nuestro estudio) estándar puede descargar mucho contenido de una web, pero va a descargar bastante material innecesario haciendo el proceso ineficiente en cuanto a tiempo (va a tardar más y va a haber que emplear más tiempo durante el proceso de descarga, y luego habrá que invertir tiempo en buscar y organizar la información descargada) y espacio (descarga una gran cantidad de datos para cada vídeo que son innecesarios para nuestro estudio). Con nuestro método, una versión refinada y adaptada a nuestras necesidades de la metodología del grupo de investigación European Comparable and Parallel Corpus of Parliamentary Speeches (ECPC), que explicaremos en más detalle en este apartado, conseguimos obtener los datos primarios de forma eficiente y organizada. Como el volumen de datos es tan grande, tanto la descarga, como el procesamiento de los textos deben hacerse de forma automática,

El primer paso que dimos fue analizar los HTLM en los que se encontraban alojados los vídeos y detectamos que, además de la URL con el archivo SRT, contienen información de especial relevancia para nuestro estudio por lo que, con la intención de dar respuesta a las preguntas planteadas en los objetivos de esta tesis, definimos qué información adicional

⁸⁸ <https://multimedia.europarl.europa.eu/es/home>

recuperar (idioma del subtítulo, título, descripción/resumen del contenido del vídeo, fecha de publicación, número de subtítulos, duración, etc.). Una vez ubicamos la información adicional, el reto es convertir el archivo original SRT en un XML con la información adicional de los HTLM que lo enriquecen y dotan de gran valor para poder llevar a cabo análisis más precisos y ambiciosos.

Así pues, para procesar el corpus EMPAC hemos preparado unas herramientas de compilación, (*EMPAC ToolKit*) que permiten descargar y procesar el corpus de forma automatizada. Para aglutinar estas herramientas de compilación hemos contado con la ayuda de un especialista en PLN, quien ha generado *scripts* para automatizar el proceso a partir de las instrucciones detalladas que le aportamos.

El *EMPAC ToolKit* está concebido para compilar y alinear las versiones del corpus en inglés y en español a nivel de texto y subtítulo. Puesto que parte de nuestro estudio se basa en un análisis cuantitativo de datos numéricos incorporamos también una serie de herramientas que transforman los XML en archivos de Excel, lo que nos va a permitir realizar una descripción estadística del corpus y llevar a cabo análisis de varianza y tablas cruzadas con el programa de análisis estadístico IBM SPSS. A continuación mostramos el contenido de *EMPAC ToolKit*:

```
|— README.md
|— add_token_info.py
|— add_token_info.sh
|— check_which_files_were_converted.py
|— check_which_videos_had_subtitles.py
|— check_which_videos_had_subtitles.sh
|— compare_files_in_directories.py
|— describe_corpus.py
|— describe_corpus.sh
|— download_srt.sh
|— europartv_redownloader.py
|— europartv_subtitle_downloader.py
|— from_srt_to_xml.py
|— from_xml_to_excel.py
|— from_xml_to_excel.sh
|— get_alignments.py
|— get_sentences.py
|— get_sents.sh
```

- |— get_wrong_segmentation.py
- |— get_wrong_segmentation.sh
- |— pipeline.sh
- |— pre_treetagger.sh
- |— queries_wrong_segmentation_inter_en.txt
- |— queries_wrong_segmentation_inter_es.txt
- |— queries_normalization_inter.txt
- |— queries_wrong_segmentation_intra_en.txt
- |— queries_wrong_segmentation_intra_es.txt
- |— queries_normalization_intra.txt
- |— redownload_srt.sh
- |— reencode_corpus.sh
- |— requirements.txt
- |— transform_srt_to_xml.sh
- |— treetagger.sh

Aparecen aquí, principalmente, programas escritos en Python y Bash (sin interfaz gráfica) que, al ejecutarlos desde la línea de comandos, nos permiten automatizar todas las fases de la compilación de EMPAC-EN y EMPAC-ES: descarga de los archivos y extracción de información metatextual de la web, etiquetado, indexación, alineado (a nivel texto y subtítulo) y, finalmente, extracción de información sobre las distribuciones de distintos fenómenos observados en el corpus. Como hemos indicado arriba, esta metodología está inspirada en el trabajo del equipo ECPC, pero orientado a nuestros propósitos concretos.

Hemos organizado las etapas que comporta el procesamiento del corpus EMPAC en cuatro apartados, que nos permiten describir y explicar las fases fundamentales de la compilación y preparación del corpus EMPAC con más agilidad.

1. Obtención de los datos y metadatos
2. Procesamiento de los textos (transformaciones de SRT a XML y VRT)
3. Procesamiento lingüístico de los textos (etiquetado morfosintáctico, anotación de frases/división del texto en frases, alineado)
4. Subida del corpus al servidor de CQPWeb

9.4. Obtención de los datos y metadatos

El primer paso consiste en analizar la estructura del documento fuente, el HTML, y ver en qué elemento se almacena la información que conforma nuestro corpus (archivos SRT) además de la información que lo enriquece (metadatos). En lugar de utilizar un *web crawler* estándar como HTTRACK⁸⁹, hemos generado un *script*, *europartv_subtitle_downloader.py*, más eficiente para los intereses de nuestra investigación. Esta herramienta por una parte descarga y transfiere los SRT y deja de lado la información que no nos interesa mientras que, por otra parte, nos permite extraer información metatextual del HTML para después estructurarla en formato Excel para su utilización posterior. Otra ventaja fundamental de nuestra herramienta es que registra si para el vídeo en cuestión se pueden recuperar los subtítulos en español e inglés (y podemos guardar un registro de los vídeos que tienen subtítulos o no) y descarga el material en carpetas organizadas que tienen sentido para nuestro estudio, cosa que no puede hacer un *web crawler* estándar, ya que estos descargan la información con la estructura de directorios web atendiendo a criterios de eficiencia para servir la información a los usuarios, pero no para realizar una investigación de estas características.

Antes de continuar con la explicación de esta fase de la compilación del corpus, consideramos necesario describir de nuevo, en este punto, qué es un archivo SRT puesto que es el archivo que contiene la información textual de nuestro corpus. Un SRT es un archivo en texto plano, que contiene los subtítulos y los tiempos de entrada y salida; es un estándar *de facto* en vídeos para la web para ofrecer subtítulos cerrados, es decir, que permite usar un mismo vídeo y simplemente superponer el texto del subtítulo sobre la imagen. De esta forma, en vez de tener muchos vídeos con los subtítulos incrustados tienes un solo vídeo con versiones de los subtítulos que son más ligeros que los vídeos.

```
1
00:00:02,480 --> 00:00:04,760
SEPA is the Single Euro Payments Area.
```

```
2
00:00:04,920 --> 00:00:08,560
If you find making payments across Europe
risky, expensive and slow.
```

⁸⁹ <https://www.httrack.com/>

La primera línea es un número que por convención indica la posición del subtítulo, en la siguiente línea vienen los códigos de tiempo de entrada y de salida (hora, minuto, segundo, *frame* o milisegundos de entrada), luego una o dos líneas de texto que quedan separados por una línea en blanco del siguiente subtítulo.

Una vez descrito el archivo SRT podemos continuar con la descripción del proceso de compilación. En esta primera fase se ejecuta el *script europarltv_subtitle_downloader.py* para descargar de forma automática todos los archivos SRT. Con esta herramienta se realiza una búsqueda en el repositorio de Multimedia Centre según la definición de los parámetros (fecha, idioma, url, categoría y tipo de programa, etc.) que consideramos necesarios para el análisis de nuestro corpus y que describimos en los siguientes apartados. Descargamos los subtítulos en formato SRT y extraemos la información metatextual en una hoja Excel.

Los metadatos obtenidos en esta etapa contienen la siguiente información organizada en las siguientes etiquetas:

- lang: idioma del subtítulo en código ISO de dos letras (ej. en para inglés);
- title: título del vídeo;
- video_url: URL del texto audiovisual;
- type: tipo de programa;
- category: categoría del texto audiovisual;
- date: fecha de publicación del texto audiovisual en la plataforma;
- description: breve descripción/resumen del contenido del texto;
- length: duración del texto audiovisual en formato hora HH:MM:SS;
- srt_url: URL del archivo SRT;
- id: ID único basado en el nombre del archivo SRT

En el proceso de descarga de los SRT conseguimos hacernos con 3 878 archivos en inglés de los que 3 812 correspondían a vídeos válidos; en español de los 3 974 documentos SRT descargados solo 3 922 eran subtítulos útiles para nuestro trabajo. Los SRT despreciados, que correspondían a archivos vacíos, no encontrados o repetidos, aparecen en la tabla que sigue.

Tabla 21. Archivos de subtítulo perdidos.

ES_LOST		EN_LOST	
1.	P001-90618-002-S-ES	1.	P005-00094-EN.srt
2.	NI01-PUB-171114-LUX-ES	2.	NI01-PUB-171114-LUX-EN.srt
3.	NI02-PUB-170216-VERH-ES	3.	NI02-PUB-170216-VERH-EN.srt
4.	NI01-PUB-170918-ARMS-ES	4.	NI01-PUB-170905-NKOR-EN.srt
5.	NI01-PUB-170503-PRAG-ES	5.	NI01-PUB-170918-ARMS-EN.srt
6.	NI01-PUB-170905-NKOR-ES	6.	NI01-PUB-170228-EMIS-EN.srt
7.	NI01-PUB-170228-EMIS-ES	7.	NI01-PUB-170503-PRAG-EN.srt
8.	NI01-PUB-170308-FEMM-ES	8.	NI01-PUB-170214-EMIS-EN.srt
9.	NI01-PUB-170220-BROK-ES	9.	NI01-PUB-170220-EURO-EN.srt
10.	NI01-PUB-170220-EURO-ES	10.	NB02-PUB-161213ORI-EN.srt
11.	NI01-PUB-170214-EMIS-ES	11.	NB01-PUB-161214INT-EN.srt
12.	NB01-PUB-170111INT-ES	12.	NB01-PUB-170111INT-EN.srt
13.	NB02-PUB-161213ORI-ES	13.	N06-ESN-170201INT-EN.srt
14.	NB01-PUB-161213INT-ES	14.	NB01-PUB-1170118INT-EN.srt
15.	NB01-PUB-161214INT-ES	15.	N04-ESN-170201INT-EN.srt
16.	N06-ESN-170201INT-ES	16.	N05-ESN-170201INT-EN.srt
17.	NB01-PUB-1170118INT-ES	17.	N03-ESN-170201INT-EN.srt
18.	N04-ESN-170201INT-ES	18.	N02-ESN-170201INT-EN.srt
19.	N05-ESN-170201INT-ES	19.	N01-PUB-171205-SAKH-EN.srt
20.	N03-ESN-170201INT-ES	20.	N01-PUB-171121-EASO-EN.srt
21.	N02-ESN-170201INT-ES	21.	N01-PUB-170927-ROMA-EN.srt
22.	N01-PUB-171205-SAKH-ES	22.	N01-PUB-170720-KEXT-EN.srt
23.	N01-PUB-171121-EASO-ES	23.	N01-PUB-170720-KEXT-EN.srt
24.	N01-PUB-170927-ROMA-ES	24.	N01-PUB-170707-CMUPNV-EN.srt
25.	N01-PUB-170704-CUBA-ES	25.	N01-PUB-170704-CUBA-EN.srt
26.	N01-PUB-170614-ROAM-ES	26.	N01-PUB-170614-ROAM-EN.srt
27.	N01-PUB-170608-OBSO-ES	27.	N01-PUB-170608-OBSO-EN.srt
28.	N01-PUB-170210INT-ES	28.	N01-PUB-170210INT-EN.srt
29.	N01-PUB-170206INT-ES	29.	N01-PUB-170206INT-EN.srt
30.	N01-PUB-170208INT-ES	30.	N01-PUB-170208INT-EN.srt
31.	N01-PUB-170130INT-ES	31.	N01-PUB-170130INT-EN.srt
32.	N01-PUB-170131INT-ES	32.	N01-PUB-170131INT-EN.srt
33.	N01-PUB-170124INT-ES	33.	N01-PUB-170124INT-EN.srt
34.	N01-PUB-170113INT-ES	34.	N01-PUB-170113INT-EN.srt
35.	N01-PUB-161214INT-ES	35.	N01-ESN-170201INT-EN.srt
36.	N01-ESN-170201INT-ES	36.	N01-ESN-170201INT-EN.srt
37.	N003-161206-001-ES	37.	N005-160914-001-EN.srt
38.	N001-161129-001-ES	38.	N003-161206-001-EN.srt
39.	N001-161130-001-ES	39.	N001-ESN-080617-EN.srt
40.	N001-161123-003-ES	40.	N001-ESN-080617-EN.srt

(Tabla 21 continúa en la página siguiente)

(Tabla 21 continúa de la página anterior)

ES_LOST		EN_LOST	
41.	N001-121029-001-ESX	41.	N001-91119-002-EN.srt
42.	N001-120621-001-ES_MF	42.	N001-91118-002-EN.srt
43.	N001-110317-002-ES	43.	N001-91118-EN.srt
44.	E001-ESN-170428INT-ES	44.	N001-91117-001-EN.srt
45.	E01-ESN-170626-ES	45.	N001-91105-010-EN.srt
46.	D001-ESN-000317INT-ES	46.	N001-161129-001-EN.srt
47.	B01-ESN-170207INT-ES	47.	N001-121029-001-ENX.srt
48.	B001-ESN-280217-ES	48.	N001-120621-001-EN_MF.srt
49.	B001-ESN-170123INT-ES	49.	E001-ESN-170428INT-EN.srt
50.	B001-ESN-170126INT-ES	50.	E01-ESN-170626-EN.srt
51.	B001-ESN-130317-ES	51.	D01-ESN-170706-EN.srt
		52.	D01-ESN-170707-EN.srt
		53.	B02-ESN-171208-EN.srt
		54.	B01-ESN-170207INT-EN.srt
		55.	B01-ESN-170901-EN.srt
		56.	B01-ESN-171006-EN.srt
		57.	B001-ESN-280217-EN.srt
		58.	B001-ESN-170123INT-EN.srt
		59.	B001-ESN-170126INT-EN.srt
		60.	B001-ESN-170126INT-EN.srt
		61.	B001-ESN-130317-EN.srt
		62.	B001-ESN-130317-EN.srt
		63.	B001-0091-EN.srt

9.5. Procesamiento de los textos: transformación de los archivos en xml

En esta fase el objetivo es transformar los archivos SRT en XLM a fin de estructurar los textos e incorporar los metadatos extraídos en la fase previa y añadir nuevos que enriquezcan sustancialmente el etiquetado. El *script* que nos permite realizar esta acción es *from_srt_to_xml.py*.

Transformar los archivos a XML permite, como apuntan Serrat y Martínez (2012:68):

[...] describir con precisión la estructura de los documentos, anotar información sobre los participantes en la situación comunicativa y facilitar el procesamiento del corpus en fases posteriores.

La estructura de nuestro etiquetado está basada en las características técnicas de los subtítulos que nos permitirán analizar de forma cuantitativa los parámetros objeto de estudio

en este trabajo. Con ese fin, se genera un XML al que se le añade la siguiente información (metadatos) a tres niveles: texto, subtítulo y línea. La estructura del documento contiene una serie de atributos organizados a partir de un elemento raíz, el texto, y la información a nivel subtítulo y a nivel línea. A continuación describimos la información de cada nivel:

a) Metadatos a nivel texto:

- id: ID único para el texto
- category: categoría del texto audiovisual (EU affairs, Economy, Security, Society, World, Others)
- type: tipo de programa (News, Interview, Background, Discovery, History, Others)
- record: la primera parte del ID que identifica el documento entre las lenguas
- version: la segunda parte del ID que identifica la versión del documento por lo general en referencia al idioma.
- date: fecha de la publicación
- day: día de la publicación
- month: mes de la publicación
- year: año de la publicación
- number of subtitles: número de subtítulos en el texto
- duration: duración del texto audiovisual en formato hora HH:MM: SS
- number of one-liners: número de subtítulos de una línea
- number of two-liners: número de subtítulos de dos líneas
- number of n-liners: número de subtítulos de más de dos líneas
- lang: idioma del subtítulo en código ISO de dos letras

b) Metadatos a nivel subtítulo:

- id: ID único para el subtítulo
- number/position: número del subtítulo en el texto, se empieza a contar desde 1
- begin: código de tiempo para la entrada del subtítulo
- end: código de tiempo para la salida del subtítulo
- duration: duración del subtítulo en pantalla
- number of lines: número de línea en el subtítulo

- number of characters: número de caracteres en el subtítulo
- characters per second: número de caracteres por segundo
- pause: tiempo transcurrido entre la salida de un subtítulo y la entrada del siguiente.

c) Metadatos a nivel línea:

- id: ID único para la línea
- line number/position: número de la línea en el subtítulo, se empieza a contar desde 1
- number of characters: número de caracteres por línea
- text: el texto que aparece en la línea

Queremos subrayar, en este punto, la importancia de la identificación de los textos, subtítulos y líneas con un ID. Todo ello nos permite identificar de forma unívoca en el corpus completo el elemento concreto al que hacemos referencia en todo momento. Otra ventaja consiste en que a la hora de hacer el alineado a nivel subtítulo podemos hacer referencia directa al elemento en cuestión que queremos alinear. Por otra parte, en CQPweb es obligatorio que cada texto tenga un ID único. Por último, tener ID en los tres niveles (texto, subtítulo, línea) facilita hacer agrupaciones de resultados (a la hora de describir) por subtítulos, por líneas y por textos.

A continuación y a modo de ejemplo, vamos a mostrar cómo quedaría transformado un SRT en XML. Sin embargo, por cuestiones de espacio mostraremos solo dos subtítulos de uno de nuestros textos. En la parte superior del cuadro que presentamos a continuación, tenemos dos subtítulos en SRT y, a continuación, presentamos la visualización del resultado al transformarlo en XML:

SUBTÍTULOS FORMATO SRT

1 00:00:02,480 --> 00:00:04,760 SEPA is the Single Euro Payments Area.
2 00:00:04,920 --> 00:00:08,560 If you find making payments across Europe risky, expensive and slow.

XML CON LOS METADATOS

```
<?xml version='1.0' encoding='utf-8'?>
<text category="Others" date="2011-12-21 14:59:00" day="21" description="Fraud and
hidden charges should be a thing of the past when making EU cross-border payments in a
new deal on the Single Euro Payment Area. And more of today's news." duration="00:02:55"
id="N001_111221_EN_MF" lang="en" month="12" n_liners="0" n_lines="73" one_
liners="7" record="N001-111221" srt_url="https://www.europarlv.europa.eu/programme-
files/2777/srt/n001-111221-en_mf.srt" subtitles="40" title="The News: Payments without
borders" two_liners="33" type="News" version="" video_url="https://www.europarlv.europa.
eu/programme/others/the-news-payments-without-borders" year="2011">
  <subtitle begin="00:00:02,480" chars="38" cps="18.97" duration="2.003"
end="00:00:04,760" id="7f53d31b32e80f9ace77d1bcdc2b36fa" n_lines="1" no="1"
pause="0">
    <line chars="38" id="fcbc0fb2f6eccdd9e07db2d01a177938" no="1">SEPA is the Single
Euro Payments Area.</line>
  </subtitle>
  <subtitle begin="00:00:04,920" chars="67" cps="16.77" duration="3.996"
end="00:00:08,560" id="4be0f79232c4ff77f258ec00b2b00651" n_lines="2" no="2"
pause="0.002">
    <line chars="41" id="4c0662a14ead2fec5459bb5ef6384044" no="1">If you find making
payments across Europe</line>
    <line chars="26" id="864339894e6cb9b2855d3408200c52ce" no="2">risky, expensive and
slow.</line>
  </subtitle>
  ...
</text>
```

Ahora, desglosamos toda la información de los subtítulos de nuestro ejemplo organizada según los tres niveles de concreción (texto, subtítulo, línea):

1) Metadatos a nivel texto

- id: N001_111221_EN_MF
- category: "Others"
- type: News"
- record: N001-111221
- versión: = video_url="https://www.europarlv.europa.eu//programme/others/the-news-payments-without-borders"
- date: 2011-12-21
- day: 21
- month: 12
- year: 2011

- number of subtitles: 40
- duration: 00:02:55
- number of one-liners: 7
- number of two-liners: 33
- number of n-liners: 70

Como se puede observar, y aún a riesgo de repetirnos (pero con el afán de apuntalar la transparencia del proceso), del archivo que hemos descargado tenemos la siguiente información: un nombre de archivo que lo identifica, el tipo de programa (en este caso News), la categoría, la url del vídeo, la fecha de publicación, el número de subtítulos de ese vídeo y su duración, además del número de subtítulos de una línea o de dos líneas. Ahora nuestro texto raíz, ese archivo SRT que contenía información limitada en texto plano, pasa a ser un XML con mucha más información de interés para un estudio de estas características.

De igual modo, también añadimos información a nivel de subtítulo con datos relacionados con el tiempo de entrada y salida del subtítulo, la duración del subtítulo, el número de caracteres, el número de caracteres por segundo y la pausa entre subtítulos. La información desglosada del subtítulo del ejemplo la visualizamos de la siguiente forma:

2) Metadatos a nivel subtítulo

1
00:00:02,480 --> 00:00:04,760
SEPA is the Single Euro Payments Area.

- id: 4be0f79232c4ff77f258ec00b2b00651
- number/position: 1
- begin: 00:00:02,480
- end: 00:00:04,760
- duration: 2.003
- number of lines: 1
- number of characters: 38
- characters per second: 18.97
- pause: 0

Por último, los que incorporamos relacionados con la línea son: la posición de la línea en el subtítulo, el número de caracteres por línea y el texto del subtítulo.

3) Metadatos a nivel línea

- id: `fcbc0fb2f6eccdd9e07db2d01a177938`
- line number/position: 1
- number of characters: 38
- text: SEPA is the Single Euro Payments Area.

Una vez hemos organizado y añadido al XML la información metatextual necesaria para el análisis del corpus, describimos a continuación el procesamiento lingüístico de los textos.

9.6. Procesamiento lingüístico de los textos

Este apartado contiene información sobre cómo se ha realizado el etiquetado morfosintáctico, la división del texto en frases y la indexación y alineado del corpus.

9.6.1. Etiquetado morfosintáctico con *TreeTagger*

Para el etiquetado morfosintáctico del corpus se ha utilizado *TreeTagger* (Schmid, 1994) y el conjunto de *scripts* escritos en Python que se ofrece con esta herramienta (*TreeTaggerWrapper*) para producir directamente los archivos en formato VRT a la vez que continúan siendo documentos XML bien formados. El formato VRT es el que se emplea para la indexación con *Corpus WorkBench* (CWB) para así realizar búsquedas avanzadas con el *Corpus Query Processor* (CQP) (Christ,1994).

El etiquetado lingüístico de los subtítulos se ha realizado en tres fases: procesamiento previo, etiquetado morfosintáctico, y postproceso.

9.6.1.1 *Procesamiento previo*

Antes de etiquetar la información lingüística es recomendable normalizar el texto que se va a procesar para obtener un tokenizado (la división del texto en palabras) más preciso

con *TreeTagger*. La normalización consiste básicamente en buscar y reemplazar caracteres unicode especiales por sus equivalentes ASCII/UTF-8 más habituales. Esta es una práctica habitual en la preparación del corpus para tareas de NLP como la Traducción Automática.

Con ese objetivo, se pasa un *script pre_treetagger.py* que forma parte de la normalización⁹⁰ de los caracteres en XML a nivel de texto.

9.6.1.2. Etiquetado morfosintáctico

En esta fase etiquetamos los textos con *TreeTagger* para lo que utilizamos el *script treetagger.py* del *Wrapper of the TreeTaggerWrapper* que permite tokenizar y etiquetar morfosintácticamente con *TreeTagger* textos en formato XML cargando una única vez el módulo de *TreeTagger*. Una vez finalizada esta fase, tenemos un archivo en VRT para poder trabajar con el *Open Corpus Workbench*. Este formato representa el texto con una palabra por línea, y cada línea está separada por tabuladores palabra, categoría gramatical y lema.

```
If      IN    if
you     PP    you
find    VVP   find
making  VVG    make
payments NNS    payment
across  IN    across
Europe  NP    Europe
```

9.6.1.3 Postproceso

En esta última fase del etiquetado morfosintáctico añadimos, con el *script add_token_info.py* diseñado para este fin, el número de tokens en los tres niveles de nuestro corpus: texto, subtítulo y línea. En archivo contiene ahora también información sobre el número de tokens además de la información morfosintáctica en un formato que nos permite operar con el *IMS Open Corpus Workbench*. En la siguiente página ilustramos el resultado de un archivo etiquetado en esta fase:

⁹⁰ Lista completa de búsquedas y reemplazos https://github.com/chozelinek/wottw/blob/master/pre_treetagger.py#L106

```
<?xml version='1.0' encoding='utf-8'?>
<text category="Others" date="2011-12-21 14:59:00" day="21" description="Fraud
and hidden charges should be a thing of the past when making EU cross-border
payments in a new deal on the Single Euro Payment Area. And more of today's
news." duration="00:02:55" id="N001_111221_EN_MF" lang="en" month="12"
n_liners="0" n_lines="73" one_liners="7" record="N001-111221" srt_url="https://
www.europarltv.europa.eu/programme-files/2777/srt/n001-111221-en_mf.srt"
subtitles="40" title="The News: Payments without borders" two_liners="33"
type="News" version="" video_url="https://www.europarltv.europa.eu//programme/
others/the-news-payments-without-borders" year="2011" tokens="471">
<subtitle begin="00:00:02,480" chars="38" cps="18.97" duration="2.003"
end="00:00:04,760" id="7f53d31b32e80f9ace77d1bcdc2b36fa" n_lines="1" no="1"
pause="0" tokens="8">
<line chars="38" id="fcbc0fb2f6eccdd9e07db2d01a177938" no="1" tokens="8">
SEPA NP SEPA
is VBZ be
the DT the
Single NP Single
Euro NP Euro
Payments NNS payment
Area NP Area
. SENT .
```

```
</line>
</subtitle>
<subtitle begin="00:00:04,920" chars="67" cps="16.77" duration="3.996"
end="00:00:08,560" id="4be0f79232c4ff77f258ec00b2b00651" n_lines="2" no="2"
pause="0.002" tokens="13">
<line chars="41" id="4c0662a14ead2fec5459bb5ef6384044" no="1" tokens="7">
```

```
If IN if
you PP you
find VVP find
making VVG make
payments NNS payment
```

```

across IN across
Europe NP Europe
</line>

```

```

<line chars="26" id="864339894e6cb9b2855d3408200c52ce" no="2" tokens="6">
risky JJ risky
, , ,
expensive JJ expensive
and CC and
slow JJ slow
. SENT .
</line>
</subtitle>
...
</tex>

```

9.6.2. División del texto en frases

Etiquetar principio y final de frase es necesario, además de porque suele ser la unidad mínima con sentido completo, también porque es indispensable para poder para analizar no solo la segmentación *intrasubtítulo* (entra las líneas de un mismo subtítulo) sino también la segmentación *intersubtítulo*, (entre un subtítulo y otro que continua con la misma oración).

El etiquetado a nivel de frase lo hacemos en un XLM diferente dado que en un XML válido y bien formado no podemos tener áreas marcadas por elementos que se solapan, así pues debemos generar un XML diferente para anotar a nivel de frase. Los elementos de un XML pueden estar unos dentro de otros, pero no entrecruzarse. En el caso de que nuestro texto estuviese compuesto de párrafos y frases podríamos tener la anotación de los dos elementos dentro del mismo documento porque las oraciones siempre serán parte del elemento párrafo. Y una frase no comienza en un párrafo y acaba en el siguiente. Es una forma de almacenar distintas capas de información de forma separada. Es más flexible, limpio y sostenible. Además, de esta forma se puede añadir tantas capas de información como se quiera. Se las puede modificar por separado, usarlas o no, sin tener que reconstruir o mantener una estructura de XML válida de gran complejidad. A continuación, mostramos la visualización del XML con la división en frases:

```
<?xml version='1.0' encoding='utf-8'?>
<text id="N001_111221_EN_MF" sentences="25">
<s id="d5d2007f70ea5399d07d0f80c8a2465e" no="1" tokens="8">
SEPA
is
the
Single
Euro
Payments
Area
.
</s>
<s id="1d384ff73b04c76c02c95c2cb78567a5" no="2" tokens="13">
If
you
find
making
payments
across
Europe
risky
,
expensive
and
slow
.
</s>
...
</text>
```

9.6.3. Indexación y alineado del corpus

Indexar un corpus nos permite analizar y estudiar fenómenos lingüísticos concretos en corpus de gran tamaño de forma rápida. Ello se debe a que una estructura de datos indexados almacena el registro de estos y así la información se ordena y centraliza en un índice. En definitiva y con este fin, a cada palabra se le asigna un identificador que permite recuperarla

rápidamente cuando interrogamos el corpus con búsquedas sobre aspectos concretos que queremos estudiar.

La indexación de nuestro corpus la hemos realizado con el *IMS Open Corpus Workbench*⁹¹(CWB), una colección de herramientas de código abierto creada para gestionar corpus de gran tamaño que dispone, entre su conjunto de herramientas, de un procesador de consultas muy potente, el *Corpus Query Processor* (CQP). Este procesador permite realizar búsquedas complejas y, como señala Sierra *et al.* (2017:58), «soporta expresiones regulares y es capaz de encontrar atributos de palabras (por ejemplo, etiquetas de parte de la oración) así como elementos entre etiquetas estructurales». A nuestro entender, este es el gestor de corpus más adecuado para nuestro estudio dado que es un software libre, abierto y flexible, frente a opciones de pago como (*Sketch engine*, *Wordsmith*) o más limitadas y lentas (*AntConc*). Además, es, hasta la fecha, lo más novedoso en gestión de corpus y cuenta, a su vez, con una comunidad relativamente sólida que brinda apoyo a través de una lista de correo y cuenta con documentación tanto para administradores como usuarios bastante completa. Y, por último, nuestro corpus cuenta con una gran cantidad de etiquetas estructurales, y el análisis de la segmentación de los subtítulos requiere de la combinación de elementos posicionales y estructurales en las búsquedas.

Por otra parte, CWB dispone de una interfaz gráfica (CQPWeb; Hardie, 2012) que permite utilizar el procesador CQP a través del navegador. Para explorar el corpus conviene conocer la sintaxis de las expresiones regulares de CQP⁹² para así poder realizar búsquedas de formas léxicas, lemas o construcciones. Para las búsquedas de construcciones complejas en las que aparezca información sobre la categoría e información morfológica es necesario familiarizarse con el etiquetado de cada lengua, pues las etiquetas dependen del *tagset*⁹³ de cada una.

Para indexar nuestro corpus con CWB utilizamos el *script reencode_corpus.sh* que tiene la función de indexar el corpus con el conjunto de herramientas de CWB. Finalmente, para alinear los archivos VRT, empleamos el *script get_alignments.py*. Este último *script* alinea el texto de cada par de lenguas, a partir del código de tiempo de los subtítulos (TCR) y genera un texto plano con la información del alineado.

⁹¹ Diseñado por el *Institut für Maschinelle Sprachverarbeitung* de Stuttgart.

⁹² http://cwb.sourceforge.net/files/CQP_Tutorial.pdf

⁹³ <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/english.par.gz>.
<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/spanish.par.gz>

Ejemplo de un archivo alineado:

N001_100420_010_EN:1 N001_100420_010_ES:1
N001_100420_010_EN:2 N001_100420_010_ES:2
N001_100420_010_EN:3 N001_100420_010_ES:3
N001_100420_010_EN:4 N001_100420_010_ES:4
N001_100420_010_EN:5 N001_100420_010_ES:5
N001_100420_010_EN:6 N001_100420_010_ES:6
N001_100420_010_EN:7 N001_100420_010_ES:7
N001_100420_010_EN:8 N001_100420_010_ES:8
N001_100420_010_EN:9 N001_100420_010_ES:9
N001_100420_010_EN:10 N001_100420_010_ES:10

La alineación se incorpora en ambas direcciones al corpus indexado con la herramienta *cwb-align-import*, lo cual permite la visualización del subtítulo y de sus equivalentes en otras lenguas además de búsquedas paralelas.

9.6.4. Subida del corpus al servidor de CQPWeb

Podemos encontrar la información en detalle de cómo realizar este paso en el manual del administrador de CQPweb.⁹⁴ Básicamente consiste en subir a la carpeta adecuada del servidor donde se aloja CQPweb los índices y el archivo de registro. Este último contiene la localización en el sistema de archivos de la carpeta con el índice y algunos metadatos que usa CQP como la declaración de atributos posicionales, estructurales y alineado, el nombre del corpus, etc. Una vez hemos subido la carpeta hay que realizar una serie de pasos de configuración con la interfaz gráfica: declarar los atributos, posicionales y estructurales, definir la visualización de los subtítulos, los contextos, la información lingüística, generar las listas de frecuencias de palabras, a nivel texto, subcorpus y corpus, que se usan después para hacer más rápidas la búsqueda de colocaciones, *keywords*, distribuciones, y por último definir el acceso al corpus por grupos de usuarios, etc.

Este apartado nos ha servido para contextualizar, describir y explicar la forma en la que hemos compilado el Corpus EMPAC. En el siguiente capítulo describimos la metodología de análisis de nuestro estudio, antesala del capítulo del análisis del corpus en el que presentaremos el resultado de nuestro estudio.

⁹⁴ <http://cwb.sourceforge.net/files/CQPwebAdminManual.pdf>

CAPÍTULO 10

METODOLOGÍA DE ANÁLISIS

10.1. Objetivos e hipótesis de trabajo

10.2. Estudio inicial del corpus EMPAC: accesibilidad al contenido audiovisual

10.2.1. Metodología de análisis para detectar si existen subtítulos

10.2.2. Metodología de análisis para detectar si son subtítulos convencionales o subtítulos para personas sordas

10.2.3. Metodología de análisis para determinar si existen audiodescripciones

10.3. Estudio central del corpus EMPAC

10.3.1. Metodología de análisis de los parámetros de estudio

10.3.1.1. Metodología de análisis del parámetro velocidad de lectura

10.3.1.2. Metodología de análisis del parámetro formato y posicionamiento

10.3.2. Metodología de análisis de las causas de segmentación inadecuada

10.3.2.1. Diseño de búsquedas para las causas de segmentación inadecuada con CQPweb

10.3.2.2. Extracción automatizada de resultados de las causas de segmentación inadecuada

CAPÍTULO 10. METODOLOGÍA DE ANÁLISIS

10.1. Objetivos e hipótesis de trabajo

Esta tesis gira en torno al concepto de accesibilidad y su estrecha relación con los Estudios de Traducción, en concreto con la modalidad de Traducción Audiovisual en el entorno web. Nuestro ámbito de estudio más concreto es la web de EuroparlTV (Multimedia Center desde agosto de 2018)⁹⁵, en la cual el Parlamento Europeo publica vídeos informativos sobre su realidad con el objetivo de acercarla a la ciudadanía europea. Por una parte, hemos escogido este entorno web por la posición absolutamente central que ha adquirido en los últimos años en la sociedad de la información y, por otra parte, por el impacto que Internet ha generado en el ámbito concreto de la traducción. En este orden de cosas, consideramos que el material que hemos decidido analizar (es decir, los vídeos generados desde el Parlamento Europeo) pueden responder perfectamente a la inquietud académica por describir la situación de la traducción audiovisual en el contexto institucional y por compararla con otros ámbitos a partir de las normativas y guías que particularizan a esta modalidad.

EuroparlTV, la televisión por Internet del Parlamento Europeo, nació con el espíritu de facilitar el acceso a la actividad de la Eurocámara y de acercar la realidad de la institución a través de cuatro canales disponibles en más de veinte lenguas; las 24 oficiales. Se creó, en palabras del gran impulsor de este proyecto⁹⁶, con el objetivo de “romper la brecha de comunicación entre las instituciones comunitarias y los ciudadanos europeos”. El concepto de brecha es clave en el contexto de la sociedad de la información en la que habitamos y fundamental en nuestro estudio, por lo que dedicaremos una parte importante de nuestro trabajo a profundizar en él y en la relación que guarda con la accesibilidad y la traducción. Nos preguntamos si la brecha digital se da especialmente en los contenidos audiovisuales. ¿Cómo se manifiesta? ¿Cómo evitarla?

⁹⁵ <https://multimedia.europarl.europa.eu/es/home?referer=%22www.europarltv.europa.eu%22>

⁹⁶ El eurodiputado español del Partido Popular Europeo Alejo Vidal Cuadras.

Partimos de la hipótesis de que, dada la legislación vigente, el contenido audiovisual alojado en la web del Parlamento Europeo (el foro legislativo y democrático más determinante de la UE) procura su accesibilidad de, al menos, dos formas: a través de del cumplimiento de los requisitos de la WAI y mediante la consolidación de las normas más extendidas en el ámbito de la traducción multilingüe del contenido audiovisual. En concreto, estas dos formas de accesibilidad se reflejarán en el uso de los parámetros técnicos (velocidad de lectura y formato y posicionamiento) de un corpus de EuroparlTV especialmente diseñado para este trabajo (el corpus EMPAC, en sus versiones en inglés y en español). La definición de estos parámetros, se construye a partir de las normas que se dan en TAV entendiendo que, independientemente del canal de difusión, se tienen en cuenta de igual manera.

Por tanto, y teniendo en cuenta la redacción de esta hipótesis, las primeras preguntas que parece necesario plantearse en un trabajo de estas características sería: ¿qué entendemos por norma?, ¿qué relación se establece entre la accesibilidad (entendida como nuevo paradigma en traducción) y las normas en TAV?, ¿las normas que gobiernan la práctica de la subtitulación en nuestro corpus son iguales a las que se asumen para la TV, DVD, cine, etc.? Y en cuanto a la segunda cuestión central de nuestro estudio nos preguntamos: ¿qué se entiende por accesibilidad y por accesibilidad audiovisual? ¿Cuál es la legislación vigente en los diversos estratos normativos (nacionales y europeos) ¿Sustenta efectivamente dicha legislación el planteamiento de nuestra hipótesis de partida en los términos esbozados arriba? ¿Podríamos decir que hay brecha digital en los contenidos audiovisuales?

De entre las dos formas de garantizar accesibilidad de la web del Parlamento Europeo que se destacan en nuestra hipótesis, los requisitos de la WAI quedan recogidos en las WCAG 2.0 y obligan a la adecuación de todo material para su recepción por parte de las personas con discapacidad sensorial e indican que los vídeos han de contener audiodescripciones para personas ciegas y subtitulación para personas sordas. Preguntas naturales en este sentido serían: ¿de qué manera y hasta qué punto se aplica la normativa WAI? ¿Mediante qué técnicas? ¿Son estas técnicas comparables a las que se utilizan en cine y TV? ¿Cómo se puede evaluar la accesibilidad al contenido audiovisual en la web?

Por lo que respecta a la accesibilidad facilitada por la traducción multilingüe de contenido audiovisual (la segunda de las formas de accesibilidad explicitadas en nuestra hipótesis), son numerosos los trabajos que se centran en la descripción de normas lingüísticas y son diversos los autores que han realizado propuestas y estudios respecto a los parámetros del subtulado (como se ha visto en el capítulo 7); entre ellos, Gottlieb (1997), Ivarsson y Carroll (1998), Karamitroglou (1998), Díaz Cintas (2007). No obstante, las normas relacionadas con los parámetros técnicos resultan determinantes en este tipo de

traducción, en especial en la modalidad de subtitulación, y su estudio brinda material de suficiente relevancia como para convertirlo en el epicentro de esta tesis. Nótese que, entre los aspectos técnicos, nosotros situamos aquellos lingüísticos estrechamente relacionados con los técnicos; es decir, la segmentación. Por ello, parece lógico proponer toda una serie de cuestiones vinculadas a las normas técnicas y restringir su examen a un corpus (EMPAC) especialmente diseñado para nuestros propósitos: ¿la velocidad de lectura es acorde a los estándares comúnmente aceptados? ¿La velocidad de lectura es más exigente en su evolución histórica? ¿Hay diferencias significativas entre EMPAC_EN (versión en inglés de nuestro corpus) y EMPAC_ES (versión en español de nuestro corpus)? ¿De las variables estudiadas (velocidad en caracteres por segundo, pausa y duración), cuál tiene más impacto en este parámetro? Por lo que respecta al parámetro formato y posicionamiento nuestras preguntas serían: ¿los subtítulos son de una o dos líneas, tal y como define la norma? ¿El número de caracteres por línea ha aumentado entre 2009 y 2017? ¿Hay subtítulos que superen el número máximo de caracteres permitidos? ¿Cuál es la tendencia? Y en torno al aspecto lingüístico de nuestro estudio nos preguntamos: ¿la segmentación en inglés afecta a la segmentación en español? ¿Afecta el uso de *Master Files* a la traducción en español? ¿Existe en el corpus analizado como parte de esta tesis un uso coincidente de las normas más extendidas en cuanto a la segmentación? ¿Hay diferencias significativas en la segmentación entre 2009 y 2017? ¿Qué unidades de sentido son las que más dificultad pueden presentar?

Todas estas preguntas investigadoras son la antesala de la determinación de los objetivos de este trabajo.

De nuestras primeras preguntas se desprende el siguiente objetivo general:

Determinar si los vídeos alojados en la web del Parlamento Europeo son accesibles y valorar si los subtítulos que acompañan a este material audiovisual respetan las normas (técnicas) de esta modalidad de traducción audiovisual.

Para alcanzar el objetivo general de esta tesis proponemos los siguientes objetivos específicos, primero en el plano teórico y seguidamente en el plano descriptivo:

Objetivos específicos en el plano teórico:

1. Presentar el estado de la cuestión sobre la accesibilidad; en concreto sobre la accesibilidad audiovisual en la web y definir el vínculo que la une a los estudios de traducción.

2. Revisar el marco normativo que regula la subtítulos y la subtítulos para personas sordas (SPS) y mostrar si las recomendaciones y las guías de buenas prácticas que proponen los organismos de estandarización y la academia se ven atendidas en este medio (Internet) y con este tipo de producto audiovisual (vídeos institucionales/traducción institucional).
3. Describir los parámetros técnicos de la subtítulos (velocidad de lectura y formato y posicionamiento, incluida la segmentación) y los valores asignados a cada parámetro y plasmar en qué medida estos valores están presentes en nuestro corpus EMPAC.

Objetivos específicos en el plano descriptivo:

1. Determinar si el contenido audiovisual alojado en EuroParlTV; concretamente, si el que hemos seleccionado para conformar el corpus EMPAC cumple con los requisitos de accesibilidad establecidos por la WAI a partir de las WCAG.
2. Plasmar la evolución de los parámetros y los valores asignados a cada parámetro entre 2009 y 2017, en consonancia con la naturaleza diacrónica de nuestro análisis en EMPAC_EN y EMPAC_ES.
3. Describir las coincidencias y divergencias respecto a los parámetros técnicos de la presentación de los subtítulos parámetros técnicos en EMPAC_EN y EMPAC_ES y su evolución desde el comienzo de las emisiones hasta 2017 mediante un análisis diacrónico.

10.2. Estudio inicial del corpus EMPAC: accesibilidad al contenido audiovisual

Las fases que comprenden la metodología de análisis de nuestro estudio son dos: un estudio inicial, en el que realizamos tres acciones que nos permitirán determinar si el contenido audiovisual cumple con los criterios de éxito de las WCAG 2.0, y un estudio central, en el que abordaremos el examen de los parámetros técnicos de la subtítulos, definidos en el capítulo 7. Sirva este apartado para plasmar la metodología de análisis del estudio inicial.

Uno de los objetivos fundamentales de esta tesis es determinar si el contenido audiovisual alojado en EuroParlTV cumple con los requisitos de accesibilidad establecidos por la WAI a

partir de las WCAG. Por tanto, en el primer estudio de esta tesis abordamos la evaluación de la accesibilidad de los contenidos audiovisuales que hacen referencia a la subtitulación y a la audiodescripción. En definitiva, nuestro estudio inicial consiste en tres fases en las que determinamos (1) si los vídeos incorporan subtítulos o no; (2) qué tipos de subtítulos son estos, cuando existen (convencionales o SPS); y (3) si los vídeos incorporan audiodescripciones.

Como puede observarse esta metodología se nutre de las investigaciones de Voces Merayo (2010:268), cuyo trabajo se reseñó en el capítulo 5 y que, a su vez, recoge propuestas de *Easy Checks—A First Review of Web Accessibility y How to meet WCAG 2.0*.

Describimos a continuación la metodología de análisis de estas tres fases.

10.2.1. Metodología de análisis para detectar si existen subtítulos

Como abordamos en la revisión teórica, el sistema de evaluación de Voces Merayo (2010) incluye unas fichas que recogen información sobre qué se quiere evaluar, cómo y cuál puede ser el resultado del examen al que se somete el contenido audiovisual. En siguiente figura se muestra la propuesta de Voces Merayo (2010:268):

Definición	Se evalúa si el contenido audiovisual incorpora, de algún modo, una subtitulación.
Ejemplos	El contenido que se presenta en el NCAM http://ncam.wgbh.org/webaccess/ccforflash/ccplayermain.html incluye una subtitulación.
Examen	¿Existe una subtitulación en el contenido audiovisual?
Procedimiento	Se reproduce el contenido y se activa la subtitulación en el caso de que exista.
Resultado	<p>PARA SORDOS BÁSICA NO</p> <p>El significado de los valores es:</p> <p>PARA SORDOS: existe una subtitulación en la que se reproducen todos los efectos sonoros y suprasegmentales que aparecen en el contenido.</p> <p>BÁSICA: existe una subtitulación en la que se transcriben los diálogos y narraciones que aparecen en el contenido.</p> <p>NO: carece de subtitulación</p>

Figura 21. Ficha de cómo evaluar subtítulos de Voces Merayo.

El procedimiento que propone el autor para examinar si el vídeo dispone de subtitulación o no es, en nuestra opinión, inadecuado puesto que la pregunta que se plante en el examen es cerrada y la respuesta solo podría ser sí o no; sin embargo, como podemos ver en la ficha el “resultado” incluye hasta tres posibilidades: para sordos, básica y no. A partir de su propuesta nosotros hemos configurado la siguiente ficha de evaluación que, además, se ajusta al estudio de un corpus de las dimensiones del que proponemos:

Definición	Se evalúa si el contenido incorpora subtítulos
Examen	Se comprueba si existe subtitulación del contenido audiovisual
Procedimiento	Se ejecuta un <i>script</i> que descarga todos los vídeos y registra cuáles disponen de un archivo SRT con la subtitulación
Resultado	El significado de los valores es: SÍ NO

En esta fase sólo hemos contestamos a la pregunta de si el vídeo dispone de subtitulación mientras que, en una segunda fase, el examen que abordaremos más adelante consistirá en analizar si la subtitulación es convencional o es SPS.

Para iniciar nuestro primer examen ejecutamos el script *check_which_videos_had_subtitles.py* mediante el que se comprueba si los vídeos del corpus EMPAC incorporan subtítulos en inglés y/o español o en ningún idioma. En concreto, obtenemos así el Excel *All-videos* con la totalidad de programas publicados en EuroparlTV entre 2009 y 2017 y sus SRT correspondientes (en inglés y español) así como la siguiente información asociada: título del vídeo, tipo de programa, fecha de publicación en la plataforma, ID con el nombre del archivo, idioma del subtítulo, categoría del texto, y el campo (*has_srt*) que indica si el vídeo contiene subtítulo (VERDADERO) o no (FALSO). Todos estos datos son la fuente de la que se nutre la primera fase del estudio inicial.

El documento Excel completo de *All-videos* puede encontrarse en los Anexos. A continuación, mostramos una captura de dicho documento:

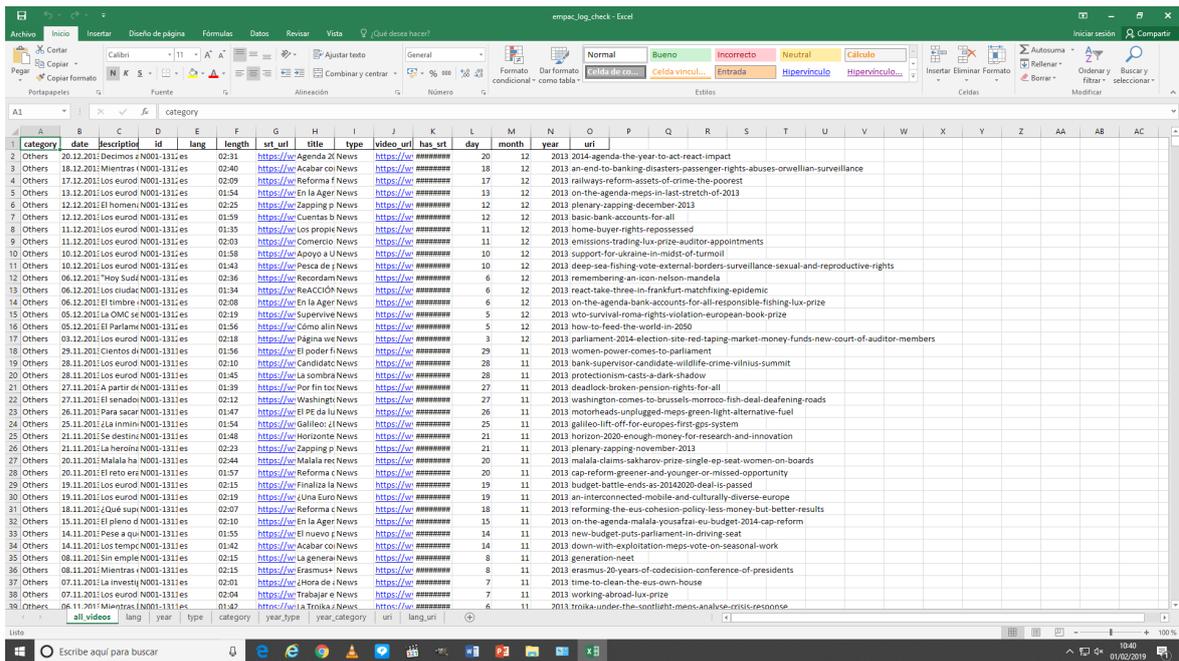


Figura 22. Estructura Excel All_Videos.

10.2.2. Metodología de análisis para detectar si son subtítulos convencionales o subtítulos para personas sordas

Una vez tenemos el número de vídeos que incorporan subtítulos, la siguiente fase del análisis consiste en analizar si estos últimos son convencionales o si se hallan específicamente diseñados para personas sordas. En la siguiente ficha recogemos la forma en la que llevaremos a cabo el análisis

Tabla 22. Ficha para determinar si los vídeos incorporan subtítulos convencionales o para personas sordas.

Definición	Se evalúa si los subtítulos son convencionales o SPS.
Examen	Se comprueba si los subtítulos incorporan información sonora relevante.
Procedimiento	Analizar todos los archivos SRT y examinar si hay información sonora relevante entre paréntesis o corchetes con CQPweb. Examinar si hay información sonora relevante en mayúsculas con CQPweb.
Resultado	SUBTITULACIÓN CONVENCIONAL SPS El significado de los valores es: SUBTITULACIÓN CONVENCIONAL: si los subtítulos únicamente recogen la información textual. SPS: si los subtítulos incorporan (entre paréntesis o corchetes) información sonora relevante.

Como indicamos en el capítulo de compilación, los archivos que contienen los subtítulos de nuestro corpus están en SRT porque es el formato escogido para su publicación en la web de EuroparlTV.

A partir de la estructura básica de los SRT podemos analizar con búsquedas avanzadas en CQPweb si los archivos contienen únicamente información textual (subtítulos convencionales) o si incorporan también información sonora relevante (SPS).

10.2.3. Metodología de análisis para determinar si existen audiodescripciones

La tercera fase de nuestro estudio inicial examina si los contenidos audiovisuales incorporan audiodescripción. Recordemos que el criterio de éxito 1.2.3 (los vídeos deben incorporar audiodescripción) es de obligado cumplimiento y un vídeo no se puede considerar accesible si no dispone de audiodescripción.

Tabla 23. Ficha evaluación para determinar si los vídeos incorporan audiodescripción.

Definición	Se evalúa si el contenido incorpora audiodescripción
Examen	Se comprueba si existe audiodescripción del contenido audiovisual
Procedimiento	Se examinar la web y el HTML en busca de enlaces o elementos de vídeo.
Resultado	El significado de los valores es: SÍ NO

De modo que, con el objetivo de determinar si los vídeos incorporan audiodescripción, examinaremos la web y su código fuente.

10.3. Estudio central del corpus EMPAC

Entre los objetivos de nuestro trabajo está el de presentar los valores recomendados para los parámetros técnicos de la subtitulación y describir en qué medida estos valores están presentes en nuestro corpus de estudio. Sirva este estudio central para alcanzar este objetivo.

El análisis central de EMPAC, basado en la propuesta de Tamayo (2015a) (véase capítulo 7) y se centra en estudiar dos de los once parámetros que propone la autora: el de velocidad de lectura (con sus variables velocidad en caracteres por segundo, pausa entre subtítulos y duración mínima y máxima de los subtítulos) y el de formato y posicionamiento (especialmente en cuanto al número de líneas, el número de caracteres por línea y las causas de segmentación inadecuada).

Para llevar a cabo el análisis de todas las variables, como explicamos en el capítulo sobre compilación, hemos extraído la información en Excel para posteriormente poder realizar análisis estadísticos con el programa SPSS. En cuanto a la variable concreta de “causas de segmentación inadecuada”, además de estos datos en bruto, ya presentamos en el capítulo 7 (a partir de bibliografía sobre todo procedente del ámbito de los estudios de recepción) nuestra propia clasificación de categoría y subcategorías sobre la que diseñaremos búsquedas avanzadas para extraer información con la herramienta CQPweb.

En cualquier caso, recordemos (véase capítulo 9) que la extracción de datos (por ejemplo, número de palabras por subtítulo, duración, velocidad en caracteres por segundo, etc.) conducentes al estudio de todas las variables se hizo posible gracias a la conversión de los archivos (en español e inglés) de SRT a XML con el *script from_srt_to_xml.py* (véase capítulo 9) Por su parte, los metadatos extraídos a nivel de texto, subtítulo y línea son los que siguen:

Tabla 24. Metadatos a nivel de texto.

Metadatos a nivel de texto:
· <i>id</i> : ID único para el texto
· <i>category</i> : categoría del texto audiovisual (EU affairs, Economy, Security, Society, World, Others)
· <i>type</i> : tipo de programa (News, Interview, Background, Discovery, History, Others)
· <i>record</i> : la primera parte del ID que identifica el documento entre las lenguas
· <i>version</i> : la segunda parte del ID que identifica la versión del documento por lo general en referencia al idioma.
· <i>date</i> : fecha de la publicación
· <i>day</i> : día de la publicación
· <i>month</i> : mes de la publicación
· <i>year</i> : año de la publicación
· <i>number of subtitles</i> : número de subtítulos en el texto
· <i>duration</i> : duración del texto audiovisual en formato hora HH:MM:SS
· <i>number of one-liners</i> : número de subtítulos de una línea
· <i>number of two-liners</i> : número de subtítulos de dos líneas
· <i>number of n-liners</i> : número de subtítulos de más de dos líneas
· <i>lang</i> : idioma del subtítulo en código ISO de dos letras

Tabla 25. Metadatos a nivel de subtítulo.

Metadatos a nivel de subtítulo:
· <i>id</i> : ID único para el subtítulo
· <i>number/position</i> : número del subtítulo en el texto, se empieza a contar desde 1
· <i>begin</i> : código de tiempo para la entrada del subtítulo
· <i>end</i> : código de tiempo para la salida del subtítulo
· <i>duration</i> : duración del subtítulo en pantalla
· <i>number of lines</i> : número de línea en el subtítulo
· <i>number of characters</i> : número de caracteres en el subtítulo
· <i>characters per second</i> : número de caracteres por segundo
· <i>pause</i> : tiempo transcurrido entre la salida de un subtítulo y la entrada del siguiente.

Tabla 26. Metadatos a nivel de línea.

Metadatos a nivel de línea:
· <i>id</i> : ID único para la línea
· <i>line number/position</i> : número de la línea en el subtítulo, se empieza a contar desde 1
· <i>number of characters</i> : número de caracteres por línea
· <i>text</i> : el texto que aparece en la línea

Conviene de nuevo incidir en el hecho de que la exportación de datos desde XML a Excel se produjo gracias al *script from_xml_to_excel.py*. En definitiva, después de la ejecución de este *script*, se obtuvieron dos archivos de Excel (para las dos lenguas del corpus EMPAC; véase anexo XXX para consultarlos ambos) con tres hojas de datos cada uno: hoja *texts* (con datos sobre los textos), hoja *subtitles* (con datos sobre los subtítulos) y hoja *lines* (con datos para las líneas).

La información recogida en estos documentos de Excel constituye la base del análisis de todas las variables de estudio, salvo la de “causas de segmentación inadecuada”, para la que utilizaremos CQPweb, como detallaremos más adelante en un apartado separado.

a) Información recuperada a nivel texto

En la hoja *texts*, la información se encuentra organizada de la siguiente forma: cada línea representa un texto y cada columna un campo que describe el texto. Los campos coinciden con los metadatos a nivel de texto. El campo ID aparece en los tres niveles por lo que este campo para cada nivel es: *text_id*, *subtitle_id* y *line_id*.

b) Información recuperada a nivel subtítulo

En el caso de la información de los subtítulos, la hoja *subtitles*, cada línea representa un subtítulo del corpus. En primer lugar, aparecen las columnas con la misma información que en la hoja de textos seguidas de las columnas que describen los datos específicos de los subtítulos. Para distinguir el nivel que contiene cada columna en caso de que el atributo sea compartido se utilizan los sufijos *text_t* and *subtitle_s*. Por ejemplo, la duración es un atributo tanto de texto como de subtítulo; para identificar el primer caso utilizaremos *duration_t*; para el segundo caso, *duration_s*. En la siguiente figura podemos ver la organización de la información a nivel de subtítulo.

	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	
	n_lines	t_line	l_iner	w_line	r_line	n_line	tokens_t	video_url	srt_url	description	subtitle_id	no	begin	end	duration_s	chars	cps	n_lines_s	tokens_s	pause
2	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da c2206d6c2	1	00:00:00,3	00:00:04,7	4,004	72	17,98	2	16	0			
3	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 144a9c4bf	2	00:00:04,8	00:00:09,2	4,994	76	15,22	2	15	0,001			
4	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 816eb1311	3	00:00:09,3	00:00:14,8	5,006	68	13,58	2	12	0,001			
5	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da bfefbba77	4	00:00:14,9	00:00:18,0	3,991	56	14,03	2	11	0,001			
6	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 775d271e8	5	00:00:18,1	00:00:19,5	1,004	22	21,91	1	5	0,001			
7	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 775d271e8	6	00:00:19,6	00:00:23,5	3,999	68	17	2	12	0,001			
8	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 1901a9594	7	00:00:23,6	00:00:27,4	3,999	63	15,75	2	16	0,001			
9	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 4a67fe55c	8	00:00:27,5	00:00:32,0	4,995	44	8,81	2	8	0,001			
10	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 7785eeb91	9	00:00:32,1	00:00:34,2	2,001	28	13,99	1	6	0,001			
11	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 14d16761c	10	00:00:34,3	00:00:36,6	2,002	38	18,98	1	8	0,001			
12	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da f4a352f7f	11	00:00:36,6	00:00:40,1	3,994	53	13,27	2	10	0,001			
13	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da dedb790b	12	00:00:40,2	00:00:43,3	3,002	37	12,33	1	8	0,001			
14	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da b7e93d461	13	00:00:43,4	00:00:48,2	4,998	74	14,81	2	14	0,001			
15	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 793196302	14	00:00:48,3	00:00:50,2	1,999	23	11,51	1	6	0,001			
16	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 091d95c72	15	00:00:50,3	00:00:53,6	3,003	51	16,98	2	13	0,001			
17	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 50a8e806e	16	00:00:53,7	00:00:56,5	2,998	32	10,67	1	7	0,001			
18	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 3e1c6905e	17	00:00:56,6	00:01:00,3	3,998	57	14,26	2	10	0,001			
19	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 43aa77e21	18	00:01:00,4	00:01:03,2	2,998	29	9,67	1	5	0,001			
20	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da f238d1494	19	00:01:03,2	00:01:08,4	5,001	57	11,4	2	12	0,001			
21	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 3973226f4	20	00:01:08,4	00:01:12,8	4,003	54	13,49	2	10	0,001			
22	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 7bbbab16	21	00:01:12,8	00:01:15,2	2,994	31	10,35	1	7	0,001			
23	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 5c713ba05	22	00:01:15,3	00:01:21,1	5,998	60	10	2	12	0,001			
24	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 4e909d621	23	00:01:21,2	00:01:23,4	2,003	18	8,99	1	6	0,001			
25	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 0190e1442	24	00:01:23,5	00:01:26,2	2,997	43	14,35	2	10	0,001			
26	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da ff3aed0a5	25	00:01:26,3	00:01:28,8	2,005	34	16,96	1	8	0,001			
27	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 1acc09fb7	26	00:01:28,8	00:01:35,0	6,991	65	9,3	2	14	0,001			
28	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da ba27769f1	27	00:01:35,0	00:01:39,0	3,999	65	16,25	2	16	0,001			
29	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 8e281a2d1	28	00:01:39,0	00:01:42,9	3,008	62	20,61	2	13	0,001			
30	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 690c735a9	29	00:01:43,0	00:01:45,5	2,006	37	18,44	1	9	0,991			
31	48	12	18	0	302	https://w	https://w	Dutch ECR MEP Peter van Da 7d7239402	30	00:01:45,6	00:01:47,8	2,002	10	5	1	3	0,001			
32	32	32	0	0	353	https://w	https://w	With Brexit, what kind of bo b08995254	1	00:00:08,8	00:00:12,1	3,993	58	14,53	1	11	0			

Figura 23. Excel de EMPAC_EN con la información recuperada a nivel de subtítulo.

c) Información recuperada a nivel línea

Y, por último, en la hoja *lines*, cada entrada representa una línea de subtítulo del corpus. Aquí se replica la distribución en las columnas de la hoja de textos y se agrega otra con datos específicos de los subtítulos y de las líneas. Para distinguir el nivel que contiene cada columna, en caso de que el atributo sea compartido se utilizan los sufijos *text_t* and *subtitle_s* y *line_l*. Por ejemplo, el número de caracteres por línea se representa (*chars_l*) mientras que el número de caracteres por subtítulo (*chars_s*)

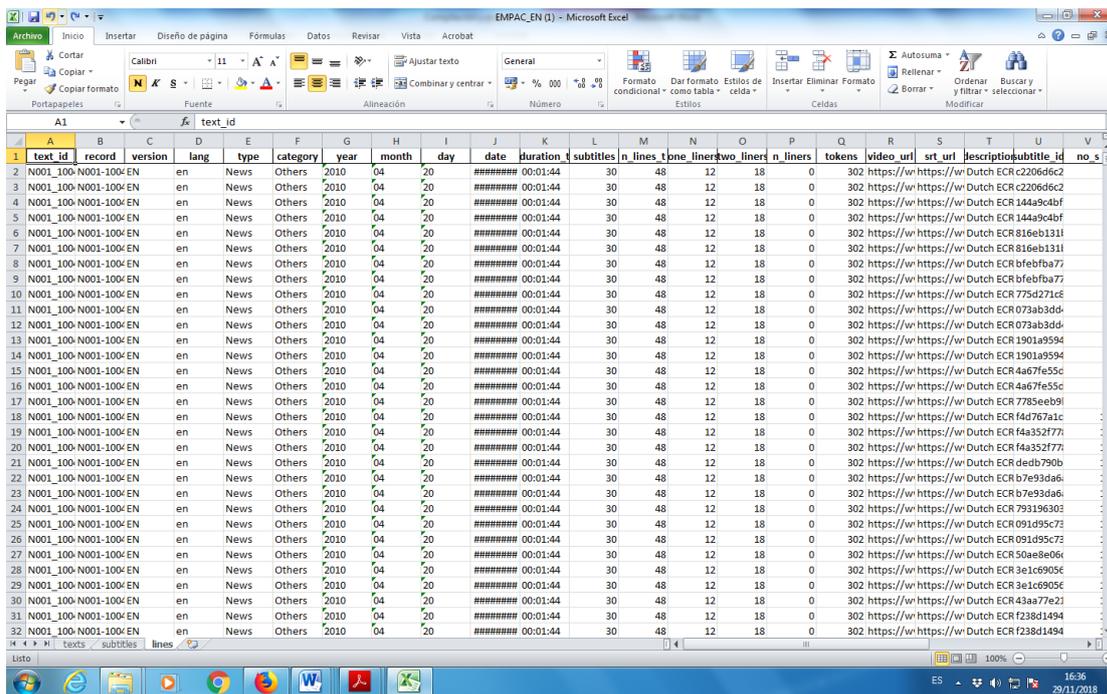


Figura 24. Excel con información recuperada a nivel de línea.

Como apuntamos anteriormente, toda la información, así recogida en estos dos documentos de Excel, que posteriormente se exportan a SPSS, son la base del análisis de todas las variables de estudio, salvo la de “causas de segmentación inadecuada”.

10.3.1. Metodología de análisis de los parámetros de estudio

A continuación y a modo de resumen, conviene tener presente de nuevo, las siguientes tablas con los parámetros, variables y valores de nuestro estudio ya explicadas con anterioridad (capítulo 7).

PARÁMETRO	VARIABLE	VALORES POSIBLES
Velocidad de lectura	Velocidad en caracteres por segundo	15- Norma UNE y DCMP 16-20 – Netflix ≥21
	Pausa	0 segundos ≤ 0,16 segundos ≥ 0,16 segundos
	Duración	> 1 segundo 1 y 6 segundos < 6 segundos

Parámetro formato y posicionamiento

PARÁMETRO	VARIABLE	VALORES POSIBLES
Parámetro formato y posicionamiento	Número de líneas	- 1 línea - 2 líneas - Más de dos líneas
	Número de caracteres por línea	- 32 - 37 Norma UNE - 42 Netflix ≥ 43
	Causas de segmentación inadecuada	- Preposición al final de línea - Determinante final de línea - Partición formas verbales compuestas - Conjunción al final de línea - Frase-Frase - Partición del sintagma - Partición nombre compuesto

Cabe destacar que hemos analizado los datos según las variables de cada parámetro desde diversos puntos de vista: de forma global, por años en un análisis más pormenorizado y por tipos de texto (o programas) por si detectáramos diferencias entre ellos.

Nótese que los valores de estas variables se han recodificado con el programa SPSS con el fin de poder agruparlos según resultara más oportuno para el análisis. Por ejemplo, para la variable de velocidad en caracteres por segundo y a partir de la bibliografía reseñada (véase capítulo 9), hemos agrupado los valores en tres grupos (15 cps, 16 a 20 cps y 21 cps o más) de modo que podremos identificar los porcentajes de cada grupo. Asimismo, puesto que nuestra intención es analizar estos datos no solo en el total del corpus y por versiones sino también por años y por tipo de texto, hemos realizado análisis de bivariantes. Con este fin, se crean tablas de contingencia, que permiten el análisis de los datos a partir de dos criterios de clasificación; por ejemplo, los cps y los años. De la misma forma, hemos aplicado la prueba del Chi Cuadrado, que nos permite establecer si existen diferencias significativas entre los valores. Adelantamos en este punto que en todos los casos hemos encontrado diferencias significativas.

A continuación, a modo de ejemplo, mostramos una tabla de contingencia (con la relación entre las variables de cps y años).

Tabla 27. Ejemplo de análisis bivariable cps y años.

			CPS_AÑOS			Total
			15 CPS	20 CPS	≤21 CPS	
año	2009	Recuento	16424	6010	686	23120
		% dentro de año	71,0%	26,0%	3,0%	100,0%
	2010	Recuento	31825	15205	1503	48533
		% dentro de año	65,6%	31,3%	3,1%	100,0%
	2011	Recuento	20232	12250	1572	34054
		% dentro de año	59,4%	36,0%	4,6%	100,0%
	2012	Recuento	19384	15313	2695	37392
		% dentro de año	51,8%	41,0%	7,2%	100,0%
	2013	Recuento	17126	14306	3272	34704
		% dentro de año	49,3%	41,2%	9,4%	100,0%
	2014	Recuento	10618	8749	2165	21532
		% dentro de año	49,3%	40,6%	10,1%	100,0%
	2015	Recuento	6821	5518	1600	13939
		% dentro de año	48,9%	39,6%	11,5%	100,0%
	2016	Recuento	3309	2775	910	6994
		% dentro de año	47,3%	39,7%	13,0%	100,0%
	2017	Recuento	1546	1571	648	3765
		% dentro de año	41,1%	41,7%	17,2%	100,0%
Total		Recuento	127285	81697	15051	224033
		% dentro de año	56,8%	36,5%	6,7%	100,0%

Como puede deducirse de la tabla de contingencia anterior, para todas las variables obtenemos la distribución de los datos por años y en total, además del porcentaje y el número exacto de casos.

A continuación, nos centramos en la metodología del análisis de cada parámetro de estudio (velocidad de lectura; formato y posicionamiento; y causas de segmentación) y, más en concreto, en la recodificación de cada una de sus respectivas variables.

10.3.1.1. Metodología del análisis del parámetro velocidad de lectura

En cuanto al parámetro de velocidad de lectura y a fin de facilitar el seguimiento de este apartado, recuperamos debajo la información que presentamos en el capítulo 7. Recordemos que, en el presente trabajo, este parámetro se examina a partir de tres variables: caracteres por segundo (cps), pausas y tiempo de permanencia en pantalla del subtítulo. Por lo que se refiere a los cps, nos interesa saber, por una parte, cuántos subtítulos respetan la norma UNE

o la guía del DCMP que indica que no se puede exceder los 15 cps. Por otra parte, buscamos identificar los subtítulos que se enmarcarían en velocidades más exigentes y que son reflejo de la realidad del mercado, es decir, hasta 20 cps (como señala Netflix en sus guías de estilo para traductores y como señalan los estudios realizados en este sentido, proyecto SURE). Por último, nos disponemos a constatar si hay *subtítulos imposibles* (véase apartado 2.1 del capítulo 7); es decir, aquellos con más de 20 cps.

Por su parte, para la variable pausa hemos establecido tres valores, un valor comúnmente aceptado (0,16 segundos), como explicamos en la revisión teórica de esta variable, y otros dos por debajo y por encima de dicho valor. De un lado, por tanto, contemplamos pausas 0, para aquellos subtítulos sin separación temporal que pueden provocar que el espectador no registre el cambio de un subtítulo a otro, por lo que es necesario que exista una pausa de entre 2 y 4 *frames* (0,16 segundos) entre subtítulos. De otro, registramos pausas de más 0,16 segundos que aglutinaría las pausas naturales que se producen en los vídeos o cuando nadie habla.

Por último, para la variable duración, analizaremos aquellos subtítulos que no permanezcan en pantalla como mínimo un segundo y aquellos que superen los seis, para determinar si EMPAC respeta la regla de los seis segundos. Asimismo, en esta variable, al margen del análisis que acabamos de mencionar, hemos considerado interesante analizar también los subtítulos que permanecen entre 0 y dos segundos en pantalla (cortos), los que permanecen entre dos y cuatro segundos (medios) y los que permanecen entre cuatro y seis segundos (largos) y cruzar estos datos con la variable velocidad en caracteres por segundo.

Los detalles concretos de estos estudios se explicarán en el capítulo de análisis para mayor claridad de lectura.

10.3.1.2. Metodología del análisis del parámetro formato y posicionamiento

En el estudio que lleva a cabo Tamayo (2015a) este parámetro incluye, como indicamos en el capítulo 7, los siguientes aspectos: segmentación en unidades de sentido, sincronización, número de líneas, distribución y forma en pantalla, número de caracteres por línea, posición y, color y fuente. Sin embargo, en nuestro estudio, este parámetro se concreta en tres variables: número de líneas, caracteres por línea (cpl) y segmentación de subtítulos. Las dos primeras se abordan en términos cuantitativos, con las herramientas que venimos revisando hasta ahora. La tercera se examina con CQPweb y ha supuesto un mayor reto.

Para realizar el análisis del número de líneas discriminamos entre subtítulos que tienen una o dos líneas e identificar, a su vez, si hay subtítulos con tres o más líneas. En cuanto a la variable de número de caracteres por línea, como vimos en el capítulo 7, las guías y estudios en este sentido apuntan tres valores de referencia: 32, 36, 42 caracteres por línea. Como ya avanzamos, nosotros hemos incorporado el valor >42 caracteres por línea. Las tablas de contingencia nos permitirán observar si hay (o no) evolución histórica o si se registran (o no) diferencias significativas entre tipos de texto.

Por último, destinamos, debajo, un apartado entero a la metodología de análisis de las “causas de segmentación inadecuada” por suponer esta un reto de especial complejidad. Antes de nada, parece obligado admitir que la segmentación de subtítulos no constituye una variable estrictamente técnica sino más bien lingüística, pero resulta innegable que se halla estrechamente vinculada al resto de parámetros ya que un subtítulo bien segmentado se procesa más rápidamente que aquel que no lo está. Es esta la razón por la que se analiza esta variable en la presente tesis (además de su innegable interés científico).

10.3.2. Metodología de análisis de las causas de segmentación inadecuada

La revisión teórica del capítulo 7 sobre los parámetros de estudio nos permitió la presentación de una taxonomía propia sobre las causas de segmentación inadecuada a partir de la cual se generan, como parte del análisis, búsquedas avanzadas en CQPweb. Con el objetivo de facilitar la lectura y antes de profundizar en estas búsquedas, recordamos a continuación, nuestra taxonomía.

Tabla para el análisis de las posibles causas de segmentación inadecuada.

CATEGORIAS GRAMATICALES	SEGMENTACIÓN ESPAÑOL	SEGMENTACIÓN INGLÉS
1. PREPOSICIÓN	1. <i>Prep</i> Preposición final de línea	1. <i>Prep</i> Preposición final de línea
2. DETERMINANTES	2. <i>Det_Indef</i> Artículos indefinidos + SN 3. <i>Det_Def</i> Artículos definidos + SN 4. <i>Det_Dem</i> Demostrativos + SN 5. <i>Det_Pos</i> Posesivos + SN	2. <i>Det_Indef</i> Indefinite article + NP 3. <i>Det_Def</i> Definite article + NP 4. <i>Det_Dem</i> Demonstratives + NP 5. <i>Det_Pos</i> Possessive + NP

(La tabla continúa en la página siguiente)

(La tabla continúa de la página anterior)

CATEGORIAS GRAMATICALES	SEGMENTACIÓN ESPAÑOL	SEGMENTACIÓN INGLÉS
3.FORMAS VERBALES COMPUESTAS	6. <i>Verb_Comp</i> Partición verbos compuestos 7. <i>Verb_Períf</i> Partición de perífrasis verbal 8. <i>Verb_PredNom</i> Partición del predicado nominal del atributo	6. <i>Verb_Aux</i> Partición auxiliary + lexical verb 7. <i>Verb_Inf</i> Partición <i>To + infinitive</i> 8. <i>Verb_Phrasal</i> Partición de un Pharsal verb
4. CONJUNCIÓN	9. <i>Conj</i> Conjunción al final	9. <i>Conj</i> Conjunción al final
5. FRASES	10. <i>Frase</i> Frase completa + Frase completa	10. <i>Frase</i> Frase completa + Frase completa
6. SINTAGMA	11. <i>Sintg_Adj</i> Partición del sintagma nominal 12. <i>Sintag_Adv</i> Partición del sintagma adverbial	11. <i>Sintg_Adj</i> Partición del sintagma nominal 12. <i>Sintag_Adv</i> Partición del sintagma adverbial
7. PARTICIÓN NOMBRE COMPUESTO	13. <i>Nom_Comp</i> Partición nombre compuesto	13. <i>Nom_Comp</i> Partición nombre compuesto

En definitiva, las causas de la segmentación que analizaremos quedan agrupadas en siete categorías para ambas lenguas y como vemos nuestra propuesta coincide con las de los teóricos presentados en el capítulo 7 pero con un nivel de concreción mayor de cada categoría.

10.3.2.1. *Diseño de búsquedas para las causas de segmentación inadecuada con CQPweb*

Como comentábamos anteriormente, en el capítulo 9, CQPweb proporciona una interfaz web que permite utilizar CQP desde un navegador y que, como apunta Molés-Cases (2016:167), destaca en los siguientes aspectos:

Sus principales ventajas son: la integración de un número ilimitado de etiquetas a nivel de palabra, metadatos y etiquetas estructurales (en forma de etiquetas XML de inicio y cierre) en sus búsquedas y la capacidad de realizar búsquedas muy generales en corpus grandes y gestionar de forma eficaz mucha información.

Esta versatilidad en el uso de etiquetas nos ha proporcionado la posibilidad de realizar, en EMPAC, búsquedas automatizadas de posibles segmentaciones erróneas, lo cual constituye una de las verdaderas innovaciones del presente trabajo, ya que, a nuestro entender, este examen resulta totalmente inédito.

Obsérvese que estas búsquedas conciernen tanto a la segmentación en el seno de subtítulos (es decir, entre las líneas que los conforman: segmentación intrasubtítulo) como a la que ocurre entre subtítulos (a la que nos referiremos como segmentación intersubtítulo). Nos ha parecido conveniente establecer esta diferenciación en consonancia con la bibliografía académica, que además considera que los errores en la segmentación intersubtítulo son más graves que los de segmentación intrasubtítulo.

A modo de ejemplo, reproducimos debajo la búsqueda en CQPWeb, para la versión en español, de “preposición al final de línea” entre líneas de un mismo subtítulo.

```
[pos="PREP"]</line><line_no="2">[pos="ADV|PREP"]*[pos="ART+|PPOIDM"]
*[pos="NC|NMEA|ORD|NMON|NP|ADJ"]+”
```

El resultado de esta búsqueda (que como queda claro se asienta sobre dosis de especialización notable) lo podemos ver en la siguiente figura:

Text	Solution 1 to 50
A002_90724_001_ES 04daab31d45ca446ac37d39c59bec334	Pensamos que la economía social de mercado , libre , pero con normas estrictas
A002_90724_003_ES e5bbf1465466179677aba4b615e183a8	Seguiremos luchando contra la discriminación y la desigualdad .
A002_90724_004_ES a361e230bdad4de4fb7d6c7b371153b9	sino que gracias al enorme apoyo de los movimientos verdes de toda Europa ,
A008_00006_ES f743f412013662a25a1875033b79c561	Los resultados aparecen en la página Web del Parlamento Europeo .
A008_00006_ES b5cea78224a4859ad5b0767ede34b6b1	y los detalles de las votaciones de los informes adoptados en la legislatura .

Figura 25. Búsqueda con CQPweb de preposición al final de línea.

De esta forma, hemos intensificado 2121 casos de segmentación inadecuada de esta categoría en la versión en español, de entre los que citamos 4 debajo:

1	Contribuirá a fijar estándares de comunicación en Europa y en el mundo.
2	con HERA, un proyecto de evaluación de riesgos para la salud y el medio ambiente.
3	En EE UU, la protección de los datos de ciudadanos de la EU no está garantizada.
4	para armonizar las decisiones de Bruselas en todos los Estados miembros .

A continuación, mostramos las cuatro tablas de búsquedas que diseñamos para identificar, en CQPWeb, la segmentación intersubtítulo e intrasubtítulo, inadecuada:

Tabla 28. Búsquedas de segmentación inadecuada intersubtítulo para inglés.

BÚSQUEDAS SEGMENTACIÓN INTER INADECUADA EN INGLÉS
Prep PREP [pos="IN"]</subtitle><subtitle>[pos="DT PP\$"]*[pos="J.+ RB N.+"]*[pos="N.+"];
Det_Indef DET [pos="DT" & lemma="a"]</subtitle><subtitle>[pos="J.+ N.+ RB"]*[pos="N.+"];
Det_Def DET [pos="DT" & lemma="the"]</subtitle><subtitle>[pos="J.+ N.+ RB"]*[pos="N.+"];
Det_Dem DET [pos="DT" & lemma="this that these those"]</subtitle><subtitle>[pos="J.+ N.+ RB"]*[pos="N.+"];
Det_Pos DET [pos="PP\$"]</subtitle><subtitle>[]
Verb_Aux VERB_COMP [pos="VB.+ VD.+ VH.+ MD.+"]</subtitle><subtitle>[pos="VV.+ VB.+ VD.+ VH.+"]
Verb_Inf VERB_COMP [pos="TO"]</subtitle><subtitle>[pos="VB VD VH VV"]
Verb_Phrasal VERB_COMP [pos="VV.* VB.* VH.* VD.*"]</subtitle><subtitle>[pos="RP+ IN+"]* [pos="RP+ IN+"]+;
Conj CONJ [pos="CC"]</subtitle><subtitle>
Frase FRASE []</s><s>[] within subtitle;
Sintag_Adj SINT [pos="PP\$ DT"] [pos="J.+"]</subtitle><subtitle>[pos="J.+ N.+ RB"]*[pos="J.+ N.+"];
Sintag_Adv SINT [pos="WRB RB"]</subtitle><subtitle>[pos="J."]*[pos="V.+ N."];
Nom_Comp NOM_COMP [pos="»NNS?!NPS?!NNS?» & word!="%"]+</subtitle><subtitle>[pos="NNS?!NPS?!NNS?"]+

Tabla 29. Búsquedas de segmentación inadecuada intersubtítulo para español.

BÚSQUEDAS SEGMENTACIÓN INTER INADECUADA EN ESPAÑOL
<p># Prep PREP</p> <pre>[pos="PREP"]</subtitle><subtitle>[pos="ADV PREP"]*[pos="ART+ PPOIDM"]* [pos="NCINMEAIORDINMONINPIADJ"]+;</pre> <p># Det_Indef DET</p> <pre>[pos="ART" & lemma="un"]</subtitle><subtitle>[pos="ADV NCINMEAI MONINPIADJ"]* [pos="NCINMEAIORDINMONINP"]+;</pre> <p># Det_Def DET</p> <pre>[pos="ART" & lemma="el"]</subtitle><subtitle>[pos="ADV NCINMEAI MONINPIADJ"]* [pos="NCINMEAIORDINMONINP"]+;</pre> <p># Det_Dem DET</p> <pre>[pos="DM"]</subtitle><subtitle>[pos="ADV NCINMEAI MONINPIADJ"]* [pos="NCINMEAIORDINMONINP"]+;</pre> <p># Det_Pos DET</p> <pre>[pos="PPO"]</subtitle><subtitle>[pos="ADV NCINMEAI MONINPIADJ"]* [pos="NCINMEAIORDINMONINP"]+;</pre> <p># Verb_Comp VERB_COMP</p> <pre>[pos="V[HS].+"]</subtitle><subtitle>[pos="V.+ " & pos!="V.+fin"];</pre> <p># Verb_Perif_1 VERB_COMP</p> <pre>[pos="V.+ " & pos!="V[HS].+"]</subtitle><subtitle>[pos="V.+ " & pos!="V.+fin"];</pre> <p># Verb_Perif_2 VERB_COMP</p> <pre>([pos="V.+ "][pos="PREP" word="que"]</subtitle>[pos="V.+ " & pos!="V.+fin"] [pos="V.+ "]</ subtitle> <subtitle>[pos="PREP" word="que"]</subtitle>[pos="V.+ " & pos!="V.+fin"]);</pre> <p># Verb_PredNom VERB_COMP</p> <pre>[pos="VS.+ VE.+ "][lemma="parecer"]+</subtitle><subtitle> [pos="ADV NCINMEAI MONIORDI PREP INPIADJ"]+;</pre> <p># Conj CONJ</p> <pre>[pos="CC.* CQUE CSUB.*"]</subtitle><subtitle>[];</pre> <p># Frase FRASE</p> <pre>[]</s><s>[] within subtitle;</pre> <p># Sintag_Adj SINT</p> <pre>[pos="ART+ PPOIDM"]</subtitle><subtitle>[pos="ADJ"]+;</pre> <p># Sintag_Adv SINT</p> <pre>[pos="V.+ " & _subtitle_n_lines="2" & _line_no="1"]</subtitle><subtitle>[pos="ADV"]</pre> <p># Nom_Compuesto NOM_COMP</p> <pre>[pos="NCINMEAI MONINP"]</subtitle><subtitle>[pos="NCINMEAI MONINP"]+;</pre>

Tabla 30. Búsquedas de segmentación inadecuada intrasubtítulo para inglés

BÚSQUEDAS SEGMENTACIÓN INTRA INADECUADA EN INGLÉS
Prep PREP
[pos="IN"]</line><line_no="2">[pos="DT PP\$"]*[pos="J.+ RB N.+"]*[pos="N.+"];
Det_Indef DET
[pos="DT" & lemma="a"]</line><line_no="2">[pos="J.+ N.+ RB"]*[pos="N.+"];
Det_Def DET
[pos="DT" & lemma="the"]</line><line_no="2">[pos="J.+ N.+ RB"]*[pos="N.+"];
Det_Dem DET
[pos="DT" & lemma="this that these those"]</line><line_no="2">[pos="J.+ N.+ RB"]*[pos="N.+"];
Det_Pos DET
[pos="PP\$"]</line><line_no="2">[]
Verb_Aux VERB_COMP
[pos="VB.+ VD.+ VH.+ MD.+"]</line><line_no="2">[pos="VV.+ VB.+ VD.+ VH.+"]
Verb_Inf VERB_COMP
[pos="TO"]</line><line_no="2">[pos="VB VD VH VV"]
Verb_Phrasal VERB_COMP
[pos="VV.* VB.* VH.* VD.*"] [pos="RP+ IN+"]</line><line_no="2">[pos="RP+ IN+"]*[pos="RP+ IN+"]+;
Conj CONJ
[pos="CC"]</line><line_no="2">
Frase FRASE
[]</s><s>[] within line;
Sintag_Adj SINT
[pos="PP\$ DT"] [pos="J.+"]</line><line_no="2">[pos="J.+ N.+ RB"]*[pos="J.+ N.+"];
Sintag_Adv SINT
[pos="WRB RB"]</line><line_no="2">[pos="J."]*[pos="V.+ N."];
Nom_Comp NOM_COMP
[pos="»NNS?!NPS?!NNS?» & word!="%"]+</line><line_no="2">[pos="NNS?!NPS?!NNS?"]+;

Tabla 31. Búsquedas de segmentación inadecuada intrasubtítulo para español.

BÚSQUEDAS SEGMENTACIÓN INTRA INADECUADA EN ESPAÑOL
<pre># Prep PREP [pos="PREP"]</line><line_ no="2">[pos="ADV PREP"]*[pos="ART+ PPO DM"]*[pos="NC NMEA ORD INMON NPI ADJ"]+; # Det_Indef DET [pos="ART" & lemma="un"]</ line><line_no="2">[pos="ADV NC NMEA INMON NPI ADJ"]*[pos="NC NMEA ORD INMON NP"]+; # Det_Def DET [pos="ART" & lemma="el"]</ line><line_no="2">[pos="ADV NC NMEA INMON NPI ADJ"]*[pos="NC NMEA ORD INMON NP"]+; # Det_Dem DET [pos="DM"]</ line><line_no="2">[pos="ADV NC NMEA INMON NPI ADJ"]*[pos="NC NMEA ORD INMON NP"]+; # Det_Pos DET [pos="PPO"]</ line><line_no="2">[pos="ADV NC NMEA INMON NPI ADJ"]*[pos="NC NMEA ORD INMON NP"]+; # Verb_Comp VERB_COMP [pos="V[HS].+" & _line_no="1" & _subtitle_n_lines="2"%c]</line><line_no="2">[pos="V.+" & pos!="V.+fin"]; # Verb_Perif_1 VERB_COMP [pos="V.+" & pos!="V[HS].+" & _line_no="1" & _subtitle_n_lines="2"%c]</line><line_ no="2">[pos="V.+" & pos!="V.+fin"]; # Verb_Perif_2 VERB_COMP ([pos="V.+" & _line_no="1" & _subtitle_n_lines="2"%c][pos="PREP" word="que"]</line><line_ no="2">[pos="V.+" & pos!="V.+fin"] [pos="V.+" & _line_no="1" & _subtitle_n_lines="2"%c]</ line>[pos="PREP" word="que"]</line>[pos="V.+" & pos!="V.+fin"]); # Verb_PredNom VERB_COMP [pos="VS.+ VE.+"]lemma="parecer"]+</ line><line_no="2">[pos="ADV NC NMEA INMON ORD IPREP NPI ADJ"]+; # Conj CONJ [pos="CC.* CQUE CSUB.*"]</line><line_no="2">[]; # Frase FRASE [</s><s>[] within line; # Sintag_Adj SINT [pos="ART+ PPO DM"]</line><line_no="2">[pos="ADJ"]+; # Sintag_Adv SINT [pos="V.+" & _subtitle_n_lines="2" & _line_no="1"]</line><line_no="2">[pos="ADV"] # Nom_Compuesto NOM_COMP [pos="NC NMEA INMON NP"]</line><line_no="2">[pos="NC NMEA INMON NP"]+;</pre>

Con estas búsquedas, generamos cuatro documentos en texto plano a los que invocará un *script* con el que automatizaremos la extracción de los datos en este sentido.

- *queries_wrong_segmentation_inter_en.txt*, para segmentación inter-subtítulo en inglés
- *queries_wrong_segmentation_inter_es.txt*, para segmentación inter-subtítulo en español
- *queries_wrong_segmentation_intra_en.txt*, para segmentación intra-subtítulo en inglés
- *queries_wrong_segmentation_intra_es.txt*, para segmentación intra-subtítulo en español

Posteriormente generamos otros dos documentos para normalizar los resultados tanto para el análisis de la segmentación intrasubtítulo como intersubtítulo. Normalizar en este caso es calcular la frecuencia relativa mediante la fórmula:

casos de segmentación inadecuada / valor de normalización × 1000,

siendo el valor de normalización para la segmentación intrasubtítulo el número total de subtítulos de dos líneas y para la segmentación intersubtítulo el número de frases que abarcan más de un subtítulo. Resulta evidente que la normalización es esencial para comparar los datos en ambas versiones.

10.3.2.2. Extracción automatizada de resultados de las causas de segmentación inadecuada

El *script* *get_wrong_segmentation.py* ejecuta las búsquedas en EMPAC-EN y EMPAC-ES y en sus respectivos subcorpus (año, categoría y tipo de texto) al tiempo que normaliza los resultados. El resultado de la acción lo obtenemos en distintos archivos Excel para cada idioma y según se analice la segmentación intrasubtítulo o intersubtítulo con las distribuciones de las segmentaciones inadecuadas por categoría de segmentación inadecuada (*class*) y por cada tipo de segmentación inadecuada (*feature*).

En la siguiente tabla, vemos un ejemplo de la visualización de los datos de segmentación intersubtítulos en inglés con la distribución de los datos totales por categoría y subcategoría:

Tabla 32. Ejemplo visualización de resultados de segmentación intersubtítulo.

Class	Feature	Freq	Total	Rel_freq
CONJ	Conj	35	15915	2,199183
DET	Det_Def	11	15915	0,691172
	Det_Dem	11	15915	0,691172
	Det_Indef	1	15915	0,062834
	Det_Pos	15	15915	0,942507
NOM_COMP	Nom_Comp	672	15915	42,22432
PREP	Prep	163	15915	10,24191
SINT	Sintag_Adj	21	15915	1,31951
	Sintag_Adv	146	15915	9,173735
VERB_COMP	Verb_Aux	7	15915	0,439837
	Verb_Inf	3	15915	0,188501
	Verb_Phrasal	1609	15915	101,0996

Como vemos, la tabla contiene los campos siguientes:

Feature: cada tipo de segmentación inadecuada que nosotros agrupamos por categorías en nuestra propuesta de clasificación de segmentación inadecuada.

Class: las seis categorías completas de nuestra propuesta: conjunción, determinante, nombre compuesto, preposición, sintagma nominal, verbo compuesto.

Freq: frecuencia absoluta, el número de hits encontrados en el corpus.

Total: el número total de subtítulos de dos líneas que utilizamos para normalizar la segmentación para esa muestra

Rel_freq: la frecuencia relativa (normalizada), es decir, la proporción de casos respecto al total. Lo hemos calculado en un tanto por 1000 para facilitar la lectura.

La información que hemos descrito en este capítulo nos permite destacar y explicar los aspectos fundamentales en los que hemos basado el análisis de nuestro corpus y nos permite dar paso ahora al análisis de los resultados obtenidos.

CAPÍTULO 11

ANÁLISIS DEL CORPUS

11.1. Análisis inicial del corpus EMPAC: accesibilidad al contenido audiovisual

11.1.1. Fase 1: Análisis para detectar si los vídeos incorporan subtítulos

11.1.2. Fase 2: Análisis para detectar si son subtítulos convencionales o subtítulos para personas sordas

11.1.3. Fase 3: Análisis para determinar si existen audiodescripciones

11.2. Análisis central del corpus EMPAC

11.2.1. Análisis del parámetro velocidad de lectura

11.2.1.1. Variable número de caracteres por segundo

11.2.1.1.1. Variable número de caracteres por segundo: análisis global

11.2.1.1.2. Variable número de caracteres por segundo: análisis por años

11.2.1.1.3. Variable número de caracteres por segundo: análisis por tipo de texto

11.2.1.2. Variable pausa entre subtítulos

11.2.1.2.1. Variable pausa entre subtítulos: análisis global

11.2.1.2.2. Variable pausa entre subtítulos: análisis por años

11.2.1.2.3. Variable pausa entre subtítulos: análisis por tipo de texto

11.2.1.3. Variable duración de los subtítulos

11.2.1.3.1. Variable duración de los subtítulos: análisis global

11.2.1.3.2. Variable duración de los subtítulos: análisis por años

11.2.1.3.3. Variable duración de los subtítulos: análisis por tipo de texto

11.2.1.3.4. Variable duración de los subtítulos: análisis global de subtítulos cortos, medios, largos

11.2.2. Análisis del parámetro formato y posicionamiento

11.2.2.1. Variable número de líneas

11.2.2.1.1. Variable número de líneas: análisis global

11.2.2.1.2. Variable número de líneas: análisis por años

11.2.2.1.3. Variable número de líneas: análisis por tipo de texto

11.2.2.2. Variable número de caracteres por línea

11.2.2.2.1. Variable número de caracteres por línea: análisis global

11.2.2.2.2. Variable número de caracteres por línea: análisis por años

11.2.2.2.3. Variable número de caracteres por línea: análisis por tipo de texto

11.2.2.3. Variable causas de segmentación inadecuada

CAPÍTULO 11. ANÁLISIS DEL CORPUS

Es este capítulo mostraremos, por una parte, el resultado del análisis de nuestro estudio inicial, en el que pretendemos determinar la accesibilidad al contenido audiovisual siguiendo la metodología expuesta en el capítulo 10 y, por otra parte, expondremos el análisis del estudio central de nuestro trabajo, en el que, a partir de los parámetros objeto de estudio, exploramos el corpus EMPAC y mostraremos los resultados obtenidos en la versión en inglés y en español (EMPAC_EN y EMPAC_ES).

11.1. Análisis inicial del corpus EMPAC: accesibilidad al contenido audiovisual

En el capítulo de metodología expusimos cómo hemos realizado el análisis de la evaluación de los criterios de éxito 1.2.2 (los vídeos deben incorporar subtítulos) y 1.2.3 (los vídeos deben incorporar audiodescripción) adaptando ligeramente la propuesta de Voces Merayo (2010). Con el objetivo específico de comprobar si se incorporaba la SPS y la AD en los vídeos de la web del EuroParlTV, recordemos que a día de hoy es el Multimedia Center, dividimos el análisis en tres fases para comprobar si los vídeos de la web de EuroParlTV son accesibles según las recomendaciones de las WCAG 2.0. La fase 1 del análisis determina si el vídeo alojado en la web incorpora subtítulos. La fase 2, en caso de que el vídeo incorpore subtítulos, estudia si son subtítulos para personas sordas o subtítulos convencionales y, por último, en la fase 3 analizamos si el vídeo va acompañado de audiodescripción.

Como ya apuntamos en el apartado de metodología, las fases del estudio previo nos permitieron la creación de una base de datos que contiene la siguiente información de cada texto de nuestro corpus: título del vídeo, tipo de programa, fecha de publicación en la plataforma, ID con el nombre del archivo, idioma del subtítulo, categoría del texto, y el campo (has_srt) que indica si el vídeo contiene subtítulo (VERDADERO) o no (FALSO). A partir de estos datos podemos evaluar la accesibilidad de los contenidos en las fases de nuestra metodología. (Una muestra de esta base de datos en formato Excel aparece en el capítulo 10 y la información al completo en los anexos).

11.1.1. Fase 1: Análisis para detectar si los vídeos incorporan subtítulos

Recordemos que la ficha de evaluación adaptada para esta fase (y que presentamos en el capítulo 10) es la siguiente:

Definición	Se evalúa si el contenido incorpora subtítulos.
Examen	Se comprueba si existe subtitulación del contenido audiovisual.
Procedimiento	Se ejecuta un <i>script</i> que descarga todos los vídeos y registra cuales disponen de un archivo SRT con la subtitulación.
Resultado	El significado de los valores es: SÍ NO

Al ejecutar el *script* conseguimos recabar, de forma automatizada, la información acerca del número de vídeos alojados en la web de EuroparlTV y cuántos de ellos tienen asociado un archivo SRT con la subtitulación del vídeo en inglés y/o español. En inglés logramos descargar los SRT para 3 878 vídeos únicos, mientras que 169 fueron los vídeos que no incorporaban subtítulos. Por su parte, en español conseguimos descargar los SRT para 3 974 vídeos únicos y en esta ocasión 71 fueron los vídeos sin subtítulos.

En la siguiente tabla podemos ver el total de vídeos agrupados por lengua y desglosados por el número de vídeos que disponían de archivo SRT (VERDADERO) o no (FALSO) y el porcentaje de vídeos que incorporan subtítulos.

Tabla 33. Datos vídeos sin subtítulo en Corpus EMPAC.

Idioma	Incorpora subtítulos	Vídeos	Porcentaje
en	FALSO	169	4,18
	VERDADERO	3878	95,82
es	FALSO	71	1,76
	VERDADERO	3974	98,24

Como podemos comprobar casi la totalidad de los vídeos incorporan subtítulos en los dos idiomas, en especial la versión en español, con un 98,24% de los vídeos con subtítulos.

Por otra parte, cabe mencionar en este punto la diferencia entre el número total de vídeos/textos que se publicaron en ese periodo de tiempo y el total de vídeos con subtítulos que conseguimos descargar y que conforman el corpus EMPAC. En inglés, finalmente descargamos 3 812 y en español 3 974 (tras las vicisitudes ocurridas en la descarga que ya se explicaron en el capítulo 9).

Tras un análisis global de la cantidad de vídeos alojados en la web con y sin subtítulos, consideramos que analizar los datos por años es otra forma especialmente reveladora de presentar los resultados. A continuación, mostramos los datos de los vídeos en inglés:

Tabla 34. Datos vídeos sin subtítulo por año en EMPAC_EN.

Idioma	Año	Subtítulos	Vídeos	Porcentaje
EN	2009	FALSO	110	23,35
		VERDADERO	361	76,65
	2010	FALSO	15	1,98
		VERDADERO	744	98,02
	2011	FALSO	6	1,05
		VERDADERO	563	98,95
	2012	FALSO	6	1
		VERDADERO	594	99
	2013	FALSO	5	1,02
		VERDADERO	485	98,98
	2014	FALSO	13	3,51
		VERDADERO	357	96,49
	2015	FALSO	2	0,55
		VERDADERO	361	99,45
	2016	FALSO	6	2,8
		VERDADERO	208	97,2
	2017	FALSO	6	2,84
		VERDADERO	205	97,16

Descubrimos que es en 2009 cuando el 23% de los vídeos no incorporan subtítulos en inglés mientras que el resto de años los porcentajes de vídeos con subtítulos está por encima del 96%. Si ahora lo comparamos con los resultados del análisis en español veremos que ambas versiones, excepto en 2009, presentan resultados muy afines.

Tabla 35. Datos de vídeos sin subtítulo por año en EMPAC_ES.

Idioma	Año	Subtítulos	Vídeos	Porcentaje
ES	2009	FALSO	6	1,27
		VERDADERO	465	98,73
	2010	FALSO	2	0,26
		VERDADERO	757	99,74
	2011	FALSO	2	0,35
		VERDADERO	567	99,65
2012	FALSO	3	0,5	
	VERDADERO	597	99,5	

(La tabla 35 continúa en la página siguiente)

(La tabla 35 continúa de la página anterior)

Idioma	Año	Subtítulos	Videos	Porcentaje
ES	2013	FALSO	6	1,23
		VERDADERO	483	98,77
	2014	FALSO	22	5,96
		VERDADERO	347	94,04
	2015	FALSO	5	1,38
		VERDADERO	358	98,62
	2016	FALSO	12	5,61
		VERDADERO	202	94,39
2017	FALSO	13	6,16	
	VERDADERO	198	93,84	

El análisis de las dos tablas nos permite ver cómo, efectivamente, en 2009, al contrario que en la versión en inglés (76,66%), casi la totalidad de los vídeos en español (98,73%) incorporan subtítulos. Es a partir de 2010 cuando se produjo el primer gran cambio en la web de EuroparlTV, ya que a partir de ese año los porcentajes en ambas lenguas están por encima del 93%. Podemos deducir que es en esa fecha cuando ya de forma estandarizada la subtitulación de todos los videos se hace a partir de los *templates* o *Master Files*. Se realiza un pautado y una transcripción en inglés de la información que aparece en la pista de audio y con ese *template* se genera el resto de subtítulos a las diferentes lenguas.

Por su parte, analizamos los datos según el tipo de texto con el objetivo de detectar diferencias significativas entre los programas. Podemos ver en la siguiente tabla que los datos arrojan los siguientes resultados en el caso de la versión en inglés y en español:

Tabla 36. Datos de vídeos sin subtítulos por tipo de texto en EMPAC_EN y EMPAC_ES.

Idioma	Año	Subtítulos	Videos	Porcentaje
EN	Background	FALSO	24	3,36
		VERDADERO	691	96,64
	Discovery	FALSO	10	2,48
		VERDADERO	394	97,52
	History	FALSO	16	12,4
		VERDADERO	113	87,6
	Interview	FALSO	24	5,57
		VERDADERO	407	94,43
News	FALSO	95	4,01	
	VERDADERO	2273	95,99	

(La tabla 36 continúa en la página siguiente)

(La tabla 36 continúa de la página anterior)

Idioma	Año	Subtítulos	Vídeos	Porcentaje
ES	Background	FALSO	20	2,8
		VERDADERO	695	97,2
	Discovery	FALSO	10	2,48
		VERDADERO	394	97,52
	History	FALSO	15	11,72
		VERDADERO	113	88,28
	Interview	FALSO	3	0,7
		VERDADERO	428	99,3
News	FALSO	23	0,97	
	VERDADERO	2344	99,03	

Observamos que, en la versión en inglés, es el programa News el que mayor número de vídeos sin subtítulo contiene (95 archivos), mientras que, en español, este tipo de texto no destaca sobre los demás y encontramos más homogeneidad entre los tipos de programa/texto.

Realizamos un análisis en el que cruzamos los años y el tipo de texto en la versión en inglés para comprobar en qué medida el programa News era el que menos subtítulos incorporaba. La tabla muestra los siguientes valores:

Tabla 37. Datos de vídeos sin subtítulos por tipo de texto en EMPAC_EN.

Lengua	Año	Tipo	Has_srt	Vídeos	Porcent.
EN	2009	Background	FALSO	15	13,04
			VERDADERO	100	86,96
		Discovery	VERDADERO	49	100
		History	FALSO	2	14,29
			VERDADERO	12	85,71
		Interview	FALSO	21	34,43
			VERDADERO	40	65,57
		News	FALSO	72	31,03
	VERDADERO		160	68,97	
	2010	Background	FALSO	5	3,82
			VERDADERO	126	96,18
		Discovery	FALSO	3	3,7
			VERDADERO	78	96,3
		History	VERDADERO	16	100
Interview		FALSO	1	1,75	
		VERDADERO	56	98,25	
News		FALSO	6	1,27	
	VERDADERO	468	98,73		

(La tabla 37 continúa en la página siguiente)

(La tabla 37 continúa de la página anterior)

Lengua	Año	Tipo	Has_srt	Vídeos	Porcent.	
EN	2011	Background	FALSO	1	0,99	
			VERDADERO	100	99,01	
		Discovery	FALSO	1	2,17	
			VERDADERO	45	97,83	
		History	FALSO	1	8,33	
			VERDADERO	11	91,67	
		Interview	VERDADERO	43	100	
		News	FALSO	3	0,82	
			VERDADERO	364	99,18	
		2012	Background	VERDADERO	135	100
				FALSO	1	1,92
			Discovery	VERDADERO	51	98,08
	FALSO			11	100	
	History		VERDADERO	11	100	
	Interview		VERDADERO	54	100	
	News		FALSO	5	1,44	
			VERDADERO	343	98,56	
	2013		Background	VERDADERO	107	100
				FALSO	3	6
		Discovery	VERDADERO	47	94	
			FALSO	2	18,18	
		History	VERDADERO	9	81,82	
			FALSO	43	100	
		Interview	VERDADERO	43	100	
		News	VERDADERO	279	100	
		2014	Background	VERDADERO	47	100
				FALSO	1	2,38
	Discovery		VERDADERO	41	97,62	
			FALSO	11	18,33	
	History		VERDADERO	49	81,67	
FALSO			63	100		
Interview	VERDADERO		63	100		
News	FALSO		1	0,63		
	VERDADERO		157	99,37		
2015	Background		FALSO	1	3,7	
		VERDADERO	26	96,3		
	Discovery	FALSO	1	2,04		
		VERDADERO	48	97,96		
	Interview	VERDADERO	62	100		
	News	VERDADERO	225	100		

(La tabla 37 continúa en la página siguiente)

(La tabla 37 continúa de la página anterior)

Lengua	Año	Tipo	Has_srt	Vídeos	Porcent.
EN	2016	Background	FALSO	2	11,11
			VERDADERO	16	88,89
		Discovery	VERDADERO	17	100
		Interview	VERDADERO	34	100
		News	FALSO	4	2,76
			VERDADERO	141	97,24
	2017	Background	VERDADERO	34	100
		Discovery	VERDADERO	18	100
		History	VERDADERO	5	100
		Interview	FALSO	2	14,29
			VERDADERO	12	85,71
		News	FALSO	4	2,86
VERDADERO	136		97,14		

Vemos cómo, efectivamente, es en 2009 en el programa News donde se ubican 72 de los 95 vídeos de esta sección que no contienen subtítulos en la versión en inglés, mientras que los 23 vídeos sin subtítulos restantes de esta sección se reparten de forma homogénea entre 2010 y 2017. Recordemos que en 2009 el 23,35% de los vídeos no incorporaban subtítulos, y por tipo de texto (programa) es News, seguido de Interview los dos que, precisamente en este año, contienen más vídeos sin subtítulos.

Por su parte, el análisis cruzado de vídeos con subtítulos por año y programa/tipo de texto de la versión en español visibiliza de nuevo que la distribución de los vídeos sin subtítulos es más homogénea tanto en años como en tipo de texto y que no se evidencia ninguna tendencia concreta.

Tabla 38. Datos de vídeos sin subtítulos por tipo de texto en EMPAC_ES.

Lengua	Año	Tipo	Has_srt	Vídeos	Porcent.
ES	2009	Background	FALSO	2	1,74
			VERDADERO	113	98,26
		Discovery	VERDADERO	49	100
		History	VERDADERO	14	100
		Interview	FALSO	1	1,64
			VERDADERO	60	98,36
		News	FALSO	3	1,29
			VERDADERO	229	98,71

(La tabla 38 continúa en la página siguiente)

(La tabla 38 continúa de la página anterior)

Lengua	Año	Tipo	Has_srt	Vídeos	Porcent.
ES	2010	Background	VERDADERO	131	100
		Discovery	FALSO	1	1,23
			VERDADERO	80	98,77
		History	VERDADERO	16	100
		Interview	VERDADERO	57	100
		News	FALSO	1	0,21
	VERDADERO		473	99,79	
	2011	Background	FALSO	1	0,99
			VERDADERO	100	99,01
		Discovery	VERDADERO	46	100
		History	FALSO	1	8,33
			VERDADERO	11	91,67
		Interview	VERDADERO	43	100
	News	VERDADERO	367	100	
	2012	Background	VERDADERO	135	100
		Discovery	FALSO	1	1,92
			VERDADERO	51	98,08
		History	VERDADERO	11	100
		Interview	VERDADERO	54	100
		News	FALSO	2	0,57
	VERDADERO		346	99,43	
	2013	Background	VERDADERO	107	100
		Discovery	FALSO	3	6
			VERDADERO	47	94
		History	FALSO	3	27,27
			VERDADERO	8	72,73
		Interview	VERDADERO	43	100
	News	VERDADERO	278	100	
	2014	Background	FALSO	10	21,28
			VERDADERO	37	78,72
Discovery		FALSO	1	2,38	
		VERDADERO	41	97,62	
History		FALSO	10	16,95	
		VERDADERO	49	83,05	
Interview		VERDADERO	63	100	
News		FALSO	1	0,63	
	VERDADERO	157	99,37		

(La tabla 38 continúa en la página siguiente)

(La tabla 38 continúa de la página anterior)

Lengua	Año	Tipo	Has_srt	Vídeos	Porcent.
ES	2015	Background	FALSO	3	11,11
			VERDADERO	24	88,89
		Discovery	FALSO	1	2,04
			VERDADERO	48	97,96
		Interview	VERDADERO	62	100
		News	FALSO	1	0,44
	VERDADERO		224	99,56	
	2016	Background	FALSO	2	11,11
			VERDADERO	16	88,89
		Discovery	VERDADERO	17	100
		Interview	FALSO	2	5,88
			VERDADERO	32	94,12
		News	FALSO	8	5,52
	VERDADERO		137	94,48	
	2017	Background	FALSO	2	5,88
			VERDADERO	32	94,12
		Discovery	FALSO	3	16,67
			VERDADERO	15	83,33
		History	FALSO	1	20
			VERDADERO	4	80
		Interview	VERDADERO	14	100
News		FALSO	7	5	
	VERDADERO	133	95		

Consideramos que en la versión en inglés el porcentaje de vídeos que incorporan subtítulos, exceptuando 2009, hace que podamos hablar de porcentajes muy altos de vídeos que incorporan subtítulos, mientras que, en la versión en español, el porcentaje es todavía más alentador puesto que el 98,24% de los vídeos incorporan subtítulos.

A continuación, pasamos a analizar el tipo de subtítulos que acompañan a los vídeos del EuroparlTV.

11.1.2. Fase 2: Análisis para detectar si son subtítulos convencionales o subtítulos para personas sordas

Entre los objetivos de nuestra tesis se encuentra el de evaluar en qué medida los vídeos alojados en la web del Parlamento Europeo son accesibles y determinar si hay traducción multilingüe, subtítulos para personas sordas y audiodescripciones. En esta fase, con el

objetivo de evaluar qué tipo de subtítulos (SPS o convencional) incorpora cada vídeo, como describimos en el capítulo de metodología, generamos la siguiente ficha de evaluación:

Definición	Se evalúa si los subtítulos son convencionales o SPS.
Examen	Se comprueba si los subtítulos incorporan información sonora relevante.
Procedimiento	Analizar todos los archivos SRT y examinar si hay información sonora relevante entre paréntesis o corchetes con CQPweb. examinar si hay información sonora relevante en mayúsculas con CQPweb.
Resultado	SUBTITULACIÓN CONVENCIONAL SPS El significado de los valores es: SUBTITULACIÓN CONVENCIONAL: si los subtítulos únicamente recogen la información textual. SPS: si los subtítulos incorporan (entre paréntesis o corchetes) información sonora relevante.

Como vimos en el capítulo 9, el primer paso es crear una búsqueda que nos permita explorar todo el corpus y encontrar si hay información contenida entre paréntesis. Recordemos que la forma en la que se muestra la información sonora relevante (diferencia fundamental entre la subtitulación convencional y la SPS) es, como indican las guías y recomendaciones de referencia de nuestro estudio, entre paréntesis o corchetes. Por tanto, con la herramienta CQPweb hemos creado una búsqueda que nos permite localizar todos los paréntesis y corchetes del corpus y la información contenida en los mismos de forma automática.

[word="\p{Ps}"]

El resultado de la búsqueda en la versión en inglés es de 21 paréntesis o corchetes en todo el corpus. En la siguiente tabla recogemos el resultado de la búsqueda:

No	Solution 1 to 21 Page 1 / 1
1	↓ 70s : CAP TOO SUCCESSFUL ?]
2	The Chinese firm is the world 's leading manufacturer of photovoltaic cells (PV) ,
3	The Chinese firm is the world 's leading manufacturer of photovoltaic cells (PV) ,
4	With benzo [a] pyrene the situation is even worse .
5	Almost 90% of Poland 's benzo [a] pyrene emissions come from low stacks .
6	Under guidance from Eurocontrol , (EU aviation security organisation)
7	Under guidance from Eurocontrol , (EU aviation security organisation)

8	with both Belarus and Russia (Kaliningrad) .
9	with both Belarus and Russia (Kaliningrad) .
10	It 's very technical . I 'm proposing to remove the removal of (d) of Annex IX
11	to trigger Article 42 (7) of the EU Treaty ,
12	I think that it 's bizarre that he has gone for 42 (7) .
13	I have heard it suggested that calling on 42 (7) will be a catalyst
14	Could article 42 (7) be the basis of future collective action ?
15	They 're afraid that this would concentrate more data power [
16	and ultimately , she ended up hiring [a trainee] , thus creating more employment incentive .
17	introduce the RDE (real-driving emissions) testing ,
18	BPM (Beats per Minute) by director Robin Campillo
19	Group of the European People 's Party (Christian Democrats)
20	The European People 's Party (Christian Democrats)
21	Group of the European People 's Party (Christian Democrats)

Podemos apreciar que en estos 21 casos no hay descripción alguna de la información sonora relevante. Los paréntesis contienen principalmente siglas y enumeraciones. En la entrada 15, en la que CQPweb solo nos muestra un corchete, de forma manual, constatamos que se trata de un error de tipografía.



Figura 26. Subtítulo en contexto en CQPweb.

Por su parte, en el análisis de la versión en español del corpus obtuvimos como resultado 48 paréntesis y unos corchetes en el total de los subtítulos, pero tras analizar todos los casos de forma manual constatamos que ninguno de ellos contenía información sonora relevante al igual que el caso de la versión en inglés. El tipo de información entre paréntesis coincide con la versión en inglés: siglas y enumeraciones. A continuación, mostramos una imagen con ejemplos del tipo de información que aparece entre paréntesis o corchetes:

No	Solution 1 to 45 Page 1 / 1
1	Cuestor (quaestor) es una palabra latina que designa a un funcionario encargado de las finanzas públicas .
2	de la escuela PHL de Hasselt (Bélgica) .
3	Se celebra en Estrasburgo (Francia) .
4	500 jóvenes han viajado a Metz (Francia) para averiguarlo .
5	(edad , dirección , preferencias) que se envían a las autoridades estadounidenses .
6	Este fabricante es líder mundial en células fotovoltaicas (PV) .
7	En el aeropuerto de Schiphol (Países Bajos) faltó poco para la tragedia .
8	cuando el avión se preparaba para el aterrizaje en Detroit (EE . UU .) .
9	pero tras dos reuniones productivas en Bonn y Tianjin (China) ,
10	Este fabricante es líder mundial en células fotovoltaicas (PV) .

La segunda parte del análisis que proponemos en nuestra ficha de evaluación supone determinar si aparece la información de los sonidos importantes representados en mayúsculas. Para extraer estos datos, de nuevo generamos una búsqueda que nos permite explorar todo el corpus de forma automática y analizar la información en mayúscula.

`(?longest)[word="\p{Lu}+"]\{2,\} within line;`

En la versión en inglés, obtenemos 347 casos de texto en mayúscula. Utilizamos la herramienta de CQPweb *Frequency Breakdown* que agrupa los casos por tipos y facilita la interpretación de los resultados y el acceso a la información. De este modo, obtuvimos una lista con todos los casos de mayúsculas en el que se juntan los casos repetidos. El primer caso que aparece en la lista es que más ocurrencias tiene. Podemos encontrar la lista completa de los resultados obtenidos en los anexos. A continuación, mostramos los diez primeros resultados:

(La tabla continúa en la página siguiente)

(La tabla continúa de la página anterior)

Número	Resultado de la búsqueda	Número ocurrencias	Porcentaje
1	HOW IT WORKS	17	4,91
2	SIS II	14	4,05
3	EU GDP	12	3,47
4	CHALLENGING MOMENT FOR YOU	9	2,6
5	EUROPEAN ELECTIONS	9	2,6
6	IN THE UPCOMING ELECTIONS	9	2,6
7	IN THIS PARLIAMENTARY TERM	9	2,6
8	WHY SHOULD PEOPLE VOTE	9	2,6
9	WHAT HAS BEEN THE MOST	7	2,02
10	THE NEWS IN BRIEF	6	1,73

Como podemos ver, el caso de *HOW IT WORKS* es el que más veces aparece por lo que nos servirá como ejemplo del análisis manual que realizamos de los casos obtenidos.

Puesto que EMPAC (y su consulta en CQPweb) permite el acceso a los enlaces del vídeo y el SRT (además de recoger información referente al nombre del archivo, idioma, duración, etc.) una comprobación manual corrobora el hecho de que *HOW IT WORKS* es el título de un programa.

Metadata for text A008_00020_EN	
Text identification code	A008_00020_EN
Structure ``text_category``	Others
Structure ``text_date``	2009-10-06 14:50:00
Structure ``text_day``	06
Structure ``text_description``	It's potentially all change at the EU, but will the facelift really affect citizens, or is it purely cosmetic?
Structure ``text_duration``	00:02:05
Structure ``text_lang``	en
Structure ``text_month``	10
Structure ``text_n_lines``	
Structure ``text_n_lines``	38
Structure ``text_one_lines``	4
Structure ``text_record``	A008-00020
Structure ``text_srt_url``	https://www.europartv.europa.eu/programme-files/924/srt/A008-00020-EN.srt
Structure ``text_subtitles``	21
Structure ``text_title``	How It Works: The EU goes under the knife
Structure ``text_tokens``	244
Structure ``text_two_lines``	17
Structure ``text_type``	Discovery
Structure ``text_version``	EN
Structure ``text_video_url``	https://www.europartv.europa.eu/programme/others/how-it-works-the-eu-goes-under-the-knife
Structure ``text_year``	2009
No. words in text	244

Figura 27. Metadatos a nivel de texto en CQPweb.

De esta forma, hemos accedido al vídeo y al SRT y comprobado que efectivamente, en todos los casos, es el nombre del programa. En el caso de SIS II, las siglas de *Schengen Information System*. En definitiva, después de minuciosos análisis manuales y automatizados confirmamos que las mayúsculas en la versión en inglés corresponden a siglas, acrónimos, nombres de programas o de secciones y a los agradecimientos que mayoritariamente aparecen al final del vídeo.

Por su parte, en el caso de la versión en español, la cantidad de casos encontrados de información en mayúsculas es más elevado y asciende a 2.188 ocurrencias. De nuevo, a partir de la opción *Frequency Breakdown*, accedimos a los casos de la versión en español. La lista completa con todos los casos de información en mayúscula se encuentra en los anexos, seguidamente mostramos los primeros 10 casos.

Número	Resultado de la búsqueda	Número ocurrencias	Porcentaje
1	EE UU	460	21,05
2	SÍ O NO	177	8,1
3	DÍAS EN EUROPA	91	4,16
4	PASO A PASO	77	3,52
5	EN PORTADA	63	2,88
6	LA ENTREVISTA	34	1,56
7	ENTREVISTA DEL DÍA	30	1,37
8	DÍAS ACTUALIDAD	23	1,05
9	DÍAS REPORTAJE	23	1,05
10	CERCA DE TI	22	1,01

Para empezar, llama la atención que el programa *PASO A PASO (How it Works* en inglés) aquí aparece en segundo lugar, mientras que SIS II, que en inglés aparecía en segundo lugar, en español ocupa el puesto 17. Esta diferencia entre las lenguas llamó nuestra atención y quisimos indagar en ella. A modo de ejemplo, describiremos el análisis manual de las dos primeras entradas de los resultados, que son los que mayor número de casos reúnen, 460 y 177 respectivamente.

La alineación de nuestro corpus resultó fundamental para este examen ya que ilustró de forma inmediata las razones de las diferencias. Por ejemplo, en el caso de las siglas EE UU, quedó claro que, en la versión en inglés, no se utilizan sino que se prefiere otras opciones como: *United States, American* o *US*.

Your query "(?longest)[word="\p{Lu}+"]{2,} within line;" returned 2,188 matches in 969 different texts (in 2,475,690 words [3,922 texts]; frequency: 883.79 instances per million words), reduced to results where query node matches word: **EE UU** (463 hits) [0.012 seconds - retrieved from cache]

No	Text	Solution 1 to 50	Page 1 / 10
1	A010_00013_ES_6cd6572704a0be6bc22ae2f43e9de73b	En EE UU , la protección de los datos de ciudadanos de la EU no está garantizada .	
		In the United states , data protection regarding EU citizens is n't guaranteed .	
2	A010_00013_ES_53e03c30e3a952e87ca659085698c396	Cuando nuestros datos van a EE UU , no están protegidos .	
		But when our data go to the United States , there is no protection .	
3	A010_00013_ES_0cf2a09b0a02e8882f012b9748aae223	Martine Roure , el mensaje del nuevo presidente de EE UU , Obama ,	
		Martine Roure , the message of the new American president . Obama ,	
4	A010_00016_ES_abf4902937982cb507e151430025ef45	No es un fenómeno exclusivamente europeo , ocurre también en EE UU .	
		It 's not the case in Europe only but also elsewhere , notably in the USA .	
5	A010_00018_ES_f743bae6f51ee0134ba38a3b0c61ea32	Estamos en EE UU en 2008. Al igual que otras miles , esta casa está en venta ,	
		We 're in the US in 2008. Like thousands of others this house is up for sale .	

Figura 28. Subtítulos alineados en CQPweb (ejemplo siglas EEUU).

Por parte, para el caso de SÍ O NO, el subtítulo en inglés no contiene esta información dado que corresponde a un mensaje que aparece sobreimpreso en pantalla y que, por tanto, no hay que traducir, como queda patente en la captura que sigue:



Figura 29. Vídeo de EuroparlTV con información textual en pantalla.

El contraste es evidente con el mismo vídeo de la versión en español, que requiere, ahora sí, subtitulación del mensaje sobreimpreso en pantalla en inglés.



Figura 30. Vídeo de EuroParlTV con información textual en pantalla subtitulada.

Sin embargo, de nuevo al analizar de forma manual estos casos observamos que, de nuevo, la información en mayúscula no recoge los sonidos relevantes para las personas con discapacidad auditiva.

Por tanto, de un análisis minucioso (tanto manual como automatizado) observamos que la totalidad de los subtítulos (en inglés y en español) son convencionales puesto que ninguno recoge la información sonora relevante; es decir, no hemos detectado SPS en ninguno de los archivos de EMPAC.

11.1.3. Fase 3: Análisis para determinar si existen audiodescripciones

Cómo ya se avanzó en el capítulo metodológico con el objetivo de determinar si los vídeos incorporaban audiodescripción, consultamos, en primer lugar, las webs donde se alojaban dichos vídeos. En ellas no encontramos ninguna información para activar la audiodescripción correspondiente. Acto seguido, nos adentramos en el código fuente de dichas páginas y tampoco localizamos ninguna etiqueta ni información adicional sobre recursos de audiodescripción. Constatamos por duplicado que, en los HTML asociados a EuroParlTV, no hay ningún tipo de información (ni visible ni incrustada en código) en forma de enlaces o elementos de audio (por ejemplo) audiodescripción.

Tras realizar la evaluación de la accesibilidad en las tres fases en las que dividimos el análisis, podemos afirmar que, si bien casi todos los vídeos incorporan subtitulación,

especialmente para el caso de español, ninguno incorpora la información sonora relevante que exige el cumplimiento del criterio de éxito, ni audiodescripción para personas ciegas y con discapacidad visual. Sin embargo, consideramos que es un punto de partida privilegiado hacia la accesibilidad de los contenidos y no debíamos restarle importancia.

En el siguiente apartado llevamos a cabo el estudio central de nuestro análisis que estudia y describe los parámetros objeto de nuestro estudio que presentamos en el capítulo 7.

11.2. Estudio central del corpus EMPAC

Como se vio en el capítulo de metodología, estudiaremos aquí el parámetro velocidad de lectura y el parámetro formato y posicionamiento. Estos parámetros los analizaremos desde diversos puntos de vista: de forma global, por años y por tipo de texto. Todas las variables, excepto una (segmentación) se han analizado con programas estadísticos (Excel y SPSS).

En este mismo apartado, primero recordamos los datos globales del Corpus EMPAC para después, en el siguiente apartado reproducir los parámetros y las variables que vamos a analizar. Finalmente, mostraremos, los resultados del análisis en las dos versiones del corpus

De forma global, por tanto, el número de subtítulos del corpus EMPAC asciende a 467 694 que aglutinan 4 912 716 palabras, es decir, casi cinco millones de palabras en medio millón de subtítulos. En el capítulo 7 describimos de forma exhaustiva nuestro corpus de estudio, por lo que la tabla que sigue sirve de recordatorio de los datos totales por versiones.

Tabla 39. Datos de EMPAC_EN y EMPAC_ES con número de textos, subtítulos, palabras y duración.

Idioma	Núm. de textos	Núm. de subtítulos	Tokens	Duración
EN	3812	224036	2448772	274 horas y 12 min
ES	3922	243660	2463944	283 horas y 2 min

Como podemos observar, la cantidad de textos y subtítulos de nuestro corpus supone un reto a la hora de analizar los datos, por lo que una metodología de corpus inspirada y refinada en el equipo de investigación European Comparable and Parallel Corpus (ECPC) de la Universitat Jaume I, como indicamos en el capítulo 9, nos parece la más adecuada para este tipo de estudio.

11.2.1. Análisis del parámetro velocidad de lectura

Sirva este apartado para plasmar uno de los objetivos de nuestro estudio que no es otro que el describir en qué medida los parámetros técnicos de la subtitulación (velocidad de lectura y formato y posicionamiento, incluida la segmentación) están presentes en EMPAC y dar muestra de la evolución (estudio diacrónico) en los valores asignados a cada parámetro técnico.

Las variables y valores que planteamos aquí nos permiten describir si los subtítulos de nuestro corpus respetan las normas técnicas de la subtitulación, descritas en el capítulo 7, teniendo en cuenta distintas realidades (académicas y profesionales) de ahí que analicemos más de un valor para cada variable. Nuestra intención no es otra que describir qué está pasando y en qué espejo nos podemos mirar. Recordemos que, en el caso de la velocidad de lectura, las variables que investigamos son:

PARÁMETRO	VARIABLE	VALORES POSIBLES
Velocidad de lectura	Velocidad en caracteres por segundo	15- Norma UNE y DCMP 16-20 – Netflix ≥21
	Pausa	0 segundos ≤ 0,16 segundos ≥ 0,16 segundos
	Duración	> 1 segundo 1 y 6 segundos < 6 segundos

Las tres variables de este parámetro tienen asignados unos valores que expusimos en los capítulos teóricos y que aparecen reflejados en la tabla anterior, y con los que pretendemos desvelar qué medida nuestro corpus se acerca más a unos valores o a otros. El análisis estadístico de los datos con el programa SPSS nos ha permitido hacer estas agrupaciones y cruzar los datos con las siguientes variables: años y tipo de texto.

11.2.1.1. Variable número de caracteres por segundo

11.2.1.1.1. Variable número de caracteres por segundo: análisis global

En el capítulo de metodología de análisis explicamos que, para agrupar los datos en torno a los valores que hemos establecido para cada variable, debemos recodificarlos con el programa estadístico SPSS. En el caso de los cps, hemos recodificado la variable en tres grupos que responden al número de caracteres establecidos por la norma UNE (15cps); a

una horquilla entre 16 y 20 cps, que según los últimos estudios de percepción esta es la franja en la que la mayor parte de las agencias se mueve; y por último a resultados iguales o superiores a 21 cps, que según los expertos corresponden a subtítulos excesivamente exigentes para un espectador medio.

A continuación, y al igual que haremos con el resto de variables, comenzamos el análisis con la exposición de los datos globales de la versión en español y en inglés, lo cual nos permite mostrar una fotografía panorámica de datos y porcentajes registrados.

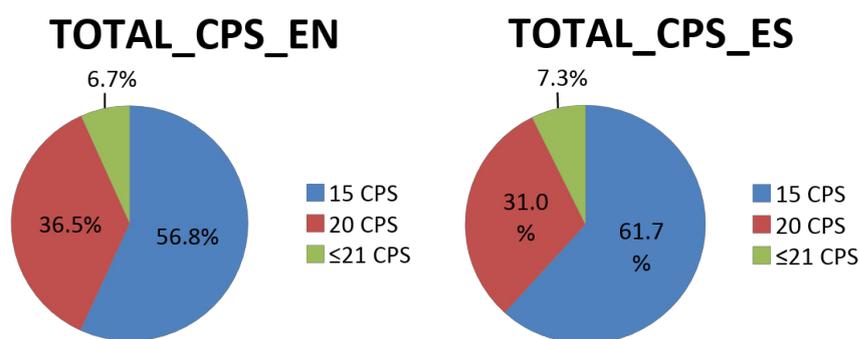


Figura 31. Resultado de análisis cps en EMPAC_EN Y EMPAC_ES.

Como podemos comprobar, del total de subtítulos de nuestro corpus en la versión en inglés (224036), los datos señalan que el 56,8% de los subtítulos no superan los 15 cps, es decir, siguen la pautas establecida en la norma UNE y DCMP. Sin embargo, el 36,5% se mostraría a velocidades de lectura por encima de los 15 cps e incluso encontramos un 6,7% con velocidades de lectura de 21 cps y más. Por lo tanto, de forma global podemos afirmar que prácticamente la mitad de los subtítulos supera los 15 cps.

En la versión española (con 243660 subtítulos) de nuestro corpus los datos que arroja nuestro análisis muestran que el 61,7% de los subtítulos son acordes a la norma, porcentaje ligeramente superior al de la versión en inglés, lo que reduce el porcentaje de subtítulos del grupo 2 que implicaría velocidades de lectura muy altas, mientras que el 7% corresponde a subtítulos de 21 cps o más que implican velocidades de lectura inasumibles por el espectador y que implicaría una traducción *verbatim*.

11.2.1.1.2. Variable número de caracteres por segundo: análisis por años

Con el objetivo de realizar un estudio pormenorizado y obtener una aproximación más realista sobre la evolución de la velocidad de lectura en los subtítulos que componen

nuestro corpus, realizamos un análisis bivariante de años y cps. De esta forma podemos comparar según las diferentes agrupaciones que hemos hecho, la cantidad de subtítulos con ese número de cps y rastrear su comportamiento a lo largo del tiempo.

En la siguiente gráfica, con los datos de los cps por años, el color azul 15 cps el rojo 20 cps y el verde 21 cps en la versión en inglés.

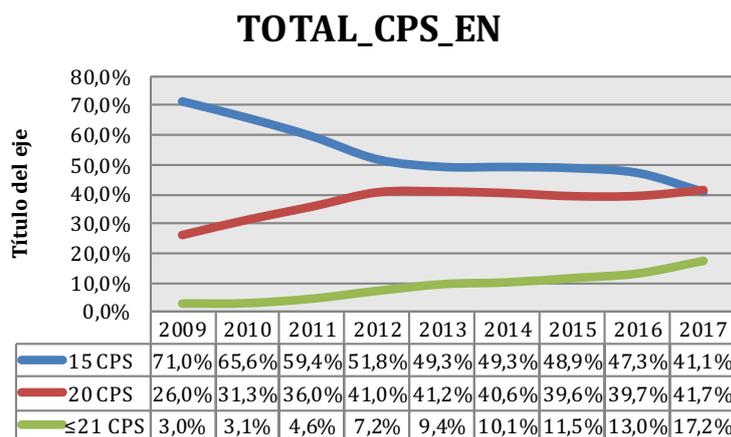


Figura 32. Resultado análisis cps por años en EMPAC_EN.

Se puede apreciar con claridad que, para la versión en inglés, el grupo de 15 cps es descendente, mientras que el de 21 cps es ascendente, y aunque el valor 20 cps aumente hasta 2012 en los siguientes años se mantiene relativamente estable.

Asimismo, los datos muestran cómo en 2009 el 71% de los subtítulos cumplía con la norma de 15 cps; sin embargo, a partir de ese punto se produce un descenso paulatino en el número de subtítulos que se encuentran en este grupo hasta alcanzar un 41,1% en 2017. En el grupo de 21 cps o más vemos la misma tendencia pero a la inversa aunque el cambio no es tan acusado y pasamos del 3% al 17,2%, sin olvidar que en un corpus de las dimensiones que hemos descrito, el 17,2% no es una cantidad desdeñable. Y por último, el grupo de 20 cps no sigue este mismo tipo de evolución ascendente/descendente. Si bien durante los primeros años, hasta 2012, sí se produce un incremento, en los siguientes años la cifra se mantiene relativamente estable. Es importante destacar que el 40% de los subtítulos de cada año tienen una velocidad de entre 16 y 20 cps.

En el caso de la versión en español la gráfica muestra la siguiente evolución en la velocidad de lectura exigida a los espectadores históricamente (el código de colores sigue siendo el mismo que para la gráfica anterior).

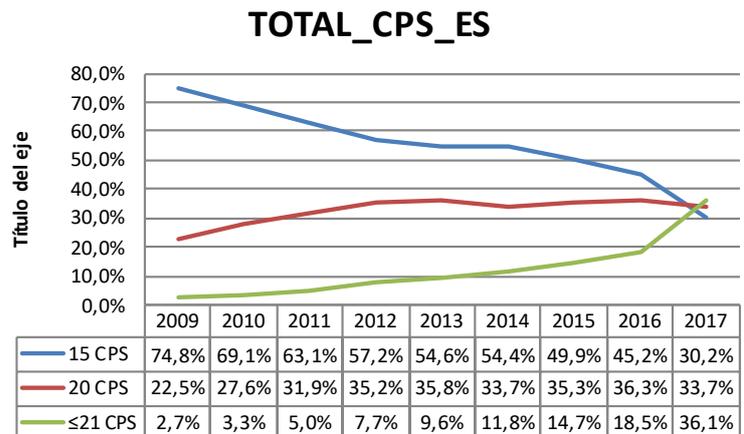


Figura 33. Resultado análisis cps por años en EMPAC_ES.

Al igual que en el caso anterior, en la versión en español, se observa una tendencia al aumento de la velocidad de lectura con el paso de los años. Sin embargo, mientras que el número de subtítulos con velocidades de lectura de hasta 15 cps en 2009 era del 75%, en 2017 solo es del 30%. Con todo, quizás el dato más significativo es que, en 2017, el número de subtítulos con velocidades de lectura de 21 cps o más es el más abultado de entre los tres valores. Es más, el grupo de hasta 15 cps, es el que ahora ocupa la última posición, por lo que, en realidad, tan solo el 30% de los subtítulos, en 2017, respetaría la norma en cuanto a la velocidad de lectura.

Por otra parte, si comparamos los datos de los subtítulos en inglés y en español de 2017, vemos claramente que el aumento de la velocidad en los subtítulos en español es mucho más elevado que en inglés.

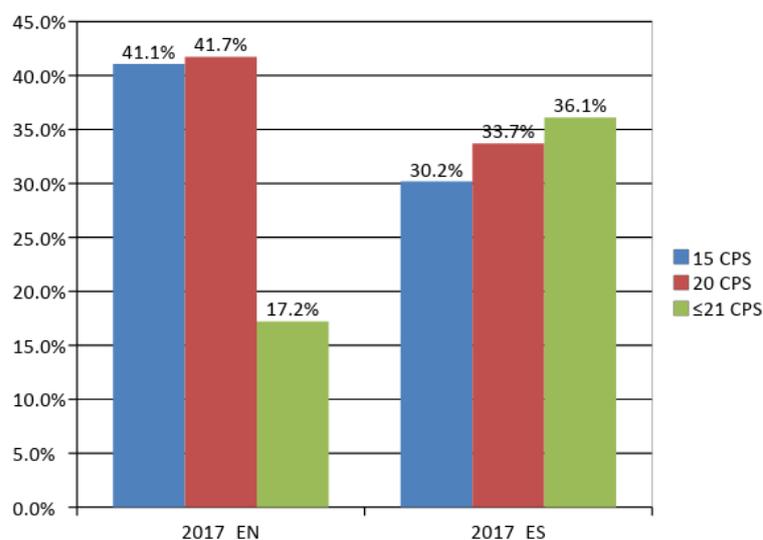


Figura 34. Resultados análisis cps en 2017 en EMPAC_EN y EMPAC_ES.

Esto nos lleva a pensar que es factible que, en 2017, exista la tendencia de generar subtítulos cada vez menos editados y más difíciles de leer.

11.2.1.1.3. Variable número de caracteres por segundo: análisis por tipo de texto

Por último, mostramos el resultado del análisis de la velocidad de lectura teniendo en cuenta los cinco tipos de textos (lo denominamos tipo de texto porque es la forma en la que está etiquetado en HTML, pero somos conscientes de que hablamos de géneros) en los que está dividido el corpus en consonancia con el etiquetado origen (*Background, Discovery, History, Interview, News*)

Tabla 40. Resultados cps por tipo de texto en EMPAC_EN y EMPAC_ES.

Tipo de texto	INGLÉS			ESPAÑOL		
	15 CPS	20 CPS	≤21 CPS	15 CPS	20 CPS	≤21 CPS
<i>Background</i>	60,2%	34,4%	5,4%	66,2%	28,9%	4,9%
<i>Discovery</i>	51,5%	40,7%	7,8%	55,8%	35,8%	8,5%
<i>History</i>	56,8%	36,2%	7,0%	60,1%	32,2%	7,7%
<i>Interview</i>	61,2%	32,5%	6,4%	67,2%	26,7%	6,1%
<i>News</i>	51,6%	40,1%	8,3%	54,5%	34,6%	10,9%

Aquí los datos son más homogéneos (en comparación que los datos obtenidos en el análisis anterior), tanto entre el tipo de textos como entre idiomas, si bien parece que *News* es el tipo de texto con los subtítulos de mayor número de caracteres por segundo; los porcentajes son por lo demás similares en todas las variables.

11.2.1.2. Variable pausa entre subtítulos

Esta variable es especialmente importante puesto que los estudios que mencionábamos en el capítulo 7 ponen de relieve que, para que el cerebro registre el cambio de un subtítulo a otro, es conveniente dejar un espacio de tiempo determinado entre ambos con el fin de que el cerebro pueda registrar el cambio de subtítulo. La pausa mínima entre subtítulos es de 0,16 segundos, y será este el valor que tomaremos como central para la presente variable. Sin embargo, nos interesa también detectar si hay subtítulos que no tienen pausa alguna entre ellos, lo que aumenta la velocidad de lectura requerida para el espectador, y si existen subtítulos con pausas superiores a los 0,16 segundos. La recodificación de la variable en este

caso es de tres grupos: subtítulos sin pausa, subtítulos con la pausa mínima más habitual y subtítulos por encima de los 0,16 segundos requeridos para asegurar la lectura del subtítulo.

11.2.1.2.1. Variable pausa entre subtítulos: análisis global

En el análisis del total del corpus en su versión en inglés (EMPAC_EN) observamos que mientras que el 78,6% del total de subtítulos proporcionan una pausa según indica la norma, el 18,3% realiza una pausa superior, lo que apunta a que hay un porcentaje muy bajo de pausas en el parlamento de las personas que participan en el vídeo. En una película, por ejemplo, los actores no se pasan todo el tiempo hablando; sin embargo, parece ser que probablemente es una de las características que definen los contenidos del EuroParlTV, es precisamente la brevedad y la densidad lingüística del contenido. Por último, encontramos un porcentaje relativamente bajo de subtítulos que no dejan ningún tipo de espacio entre ellos, el 3,1%. La diferencia entre las versiones en inglés y en español es mínima, lo que se explica por el uso de Master Files para la generación de los subtítulos.

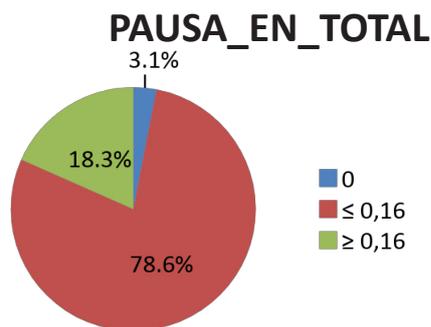


Figura 35. Resultado pausa entre subtítulos en EMPAC_EN.

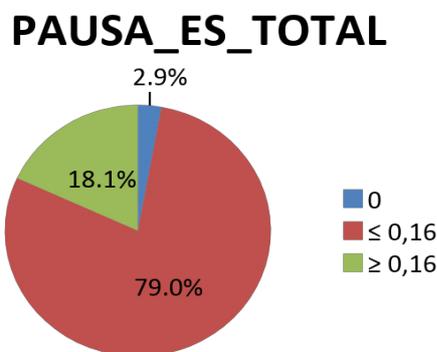


Figura 36. Resultado pausa entre subtítulos en EMPAC_ES.

11.2.1.2.2. Variable pausa entre subtítulos: análisis por años

El análisis por años nos revela que entre, 2009 y 2015, el número de subtítulos sin pausa era extraordinariamente bajo; sin embargo, en 2016 se observa un incremento significativo respecto a los años anteriores. Pero es en 2017 donde se produce un cambio radical en la tendencia en ambas versiones y más de la mitad de los subtítulos tienen Pausa 0, es decir, no hay pausa alguna entre ellos.

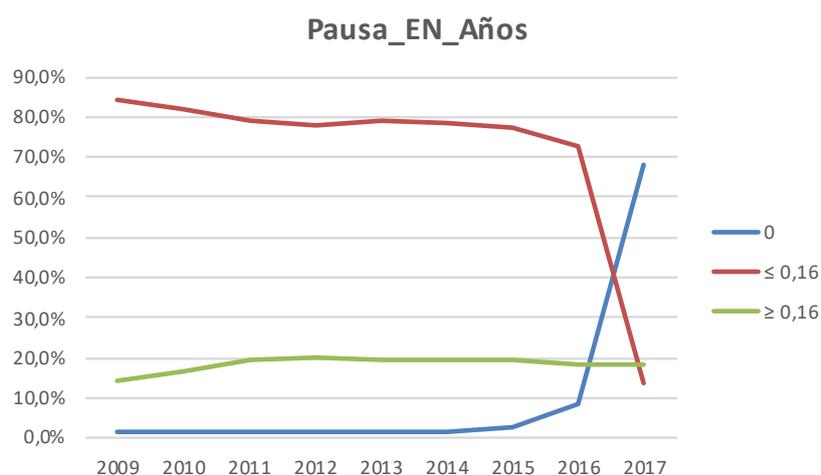


Figura 37. Evolución de la pausa entre subtítulos entre 2009 y 2017 en EMPAC_EN.

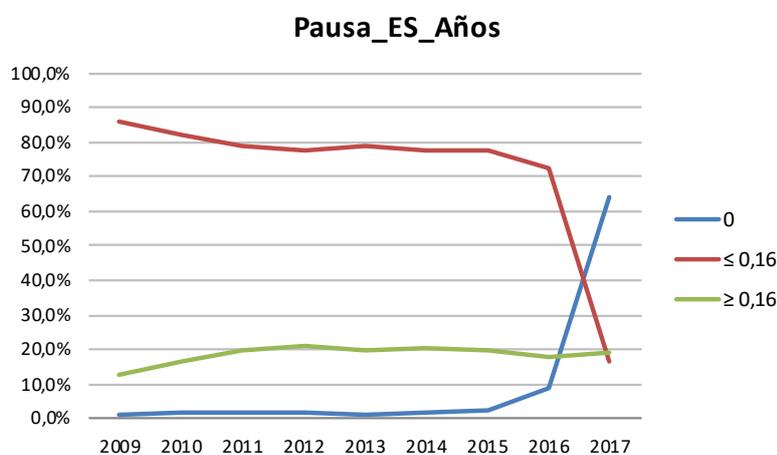


Figura 38. Evolución de la pausa entre subtítulos entre 2009 y 2017 en EMPAC_ES.

En la tabla de datos que mostramos a continuación, podemos observar cómo en la versión en inglés pasamos de un 1,7% de subtítulos sin pausa en 2009 a un 68,1% en 2017, mientras el resto de años se mantiene en la misma tónica excepto en 2016 que se registra un aumento al 9,1% para dispararse en 2017. El grupo de más de 16 milésimas (línea verde) es homogéneo posiblemente por el tipo de textos audiovisuales donde se registra en mayor

medida: textos institucionales que, como apuntábamos anteriormente, no son obras de ficción, sino obras de no ficción del género de los informativos.

En la tabla que mostramos a continuación se recoge la información en detalle de los porcentajes por años y por versión del corpus.

Tabla 41. Resultados pausa entre subtítulos por años en EMPAC_EN y EMPAC_ES.

EN	PAUSA 0	≤ 0,16	≥ 0,16	ES	PAUSA 0	≤ 0,16	≥ 0,16
2009	1,7%	84,1%	14,2%	2009	1,2%	85,7%	13,0%
2010	1,8%	81,8%	16,4%	2010	1,7%	81,9%	16,4%
2011	1,7%	79,0%	19,3%	2011	1,6%	78,6%	19,8%
2012	1,6%	78,0%	20,4%	2012	1,6%	77,4%	21,1%
2013	1,4%	79,2%	19,3%	2013	1,4%	78,8%	19,8%
2014	1,7%	78,6%	19,6%	2014	1,9%	77,9%	20,2%
2015	2,6%	77,6%	19,8%	2015	2,6%	77,6%	19,8%
2016	8,8%	72,9%	18,3%	2016	9,1%	72,7%	18,1%
2017	68,1%	13,8%	18,1%	2017	64,3%	16,5%	19,2%

Como podemos apreciar en esta visualización minuciosa de los datos, los porcentajes son muy similares en ambas versiones.

11.2.1.2.3. Variable pausa entre subtítulos: análisis por tipo de texto

En este caso podemos ver, en la tabla que ofrecemos seguidamente, que el tipo de texto *News* tiene en ambas versiones el porcentaje más alto de subtítulos sin pausa y, en *Discovery* e *History*, encontramos los porcentajes de pausas por encima de 0,16 milésimas ($\geq 0,16$) más altos, lo que probablemente indique que, los reportajes y documentales (géneros que aparecen en *Discovery* e *History*) permiten más pausas entre parlamentos que las noticias. Entre los programas con más pausas, encontramos también la serie *Blink*, un formato para niños y jóvenes en el que se les explica la idea de Europa.

Por otra parte, parece lógico que sea en las entrevistas donde menos pausas largas se den puesto que el formato no lo permite. No es lo mismo un documental, en el que son más habituales los tiempos en los que las imágenes no van acompañadas de parlamento, que una entrevista, donde por lo general la imagen no va más allá de un plano medio del entrevistador y del entrevistado, que preguntan y contestan respectivamente.

Tabla 42. Resultados pausa entre subtítulos por tipo de texto en EMPAC_EN y EMPAC_ES.

EN	PAUSA 0	≤ 0,16	≥ 0,16	ES	PAUSA 0	≤ 0,16	≥ 0,16
Background	1,1%	80,5%	18,4%	Background	1,1%	81,1%	17,8%
Discovery	2,5%	74,5%	23,0%	Discovery	2,4%	74,2%	23,3%
History	2,7%	76,1%	21,2%	History	2,6%	76,1%	21,3%
Interview	2,5%	82,3%	15,2%	Interview	2,6%	82,5%	14,9%
News	6,0%	75,6%	18,4%	News	5,5%	75,9%	18,6%

11.2.1.3. Variable duración de los subtítulos

Como señalamos en el capítulo 7, los subtítulos, según la bibliografía académica y la mayoría de las guías y recomendaciones generadas desde el sector profesional, establecen que la duración mínima de un subtítulo es de un segundo y de seis segundos la máxima; por debajo de un segundo se puede producir el efecto *flash* y por encima de seis la relectura. Por tanto, la recodificación de esta variable se ha traducido en tres agrupaciones: aquellos subtítulos con una duración inferior a un segundo; aquellos que están en pantalla entre un segundo y seis y, por último, aquellos que superan ese tiempo máximo de permanencia en pantalla, es decir, más de seis segundos. Por otra parte, nos ha parecido interesante averiguar la duración más habitual de los subtítulos en nuestro corpus y su relación con la cantidad de caracteres por segundo por lo que añadiremos este análisis en esta sección.

11.2.1.3.1. Variable duración de los subtítulos: análisis global

En la tabla que mostramos a continuación podemos observar cómo prácticamente la totalidad de los subtítulos tienen una duración entre uno y hasta seis segundos, cumpliendo así con la norma vigente, mientras que el porcentaje de subtítulos con una duración inferior a un segundo es prácticamente inexistente. Tampoco es reseñable el porcentaje de subtítulos de más de 6 segundos.

Tabla 43. Resultados duración de los subtítulos en EMPAC_EN y EMPAC_ES.

DURACIÓN	> 1 sg	1 y 6 sg	< 6 sg
EN	0,1%	93,3%	6,6%
ES	0,1%	93,2%	6,6%

11.2.1.3.2. Variable duración de los subtítulos: análisis por años

Un análisis pormenorizado por años revela que ambas versiones muestran un repunte en el número de subtítulos de menos de un segundo en 2017 y que el 90% de los subtítulos permanece en pantalla entre 1 y 6 segundos, como indica la norma.

Tabla 44. Resultados variable duración por años en EMPAC_EN y EMPAC_ES.

EN	> 1 sg	1 y 6 sg	< 6 sg	ES	>1 sg	1 y 6 sg	< 6 sg
2009	0,0%	89,6%	10,3%	2009	0,0%	90,3%	9,6%
2010	0,0%	92,6%	7,4%	2010	0,0%	92,6%	7,4%
2011	0,1%	93,7%	6,2%	2011	0,1%	93,8%	6,1%
2012	0,2%	94,4%	5,4%	2012	0,2%	94,6%	5,2%
2013	0,3%	94,7%	5,0%	2013	0,3%	94,8%	4,8%
2014	0,0%	93,3%	6,7%	2014	0,1%	93,5%	6,4%
2015	0,1%	93,4%	6,5%	2015	0,1%	93,4%	6,5%
2016	0,1%	93,2%	6,7%	2016	0,1%	93,6%	6,3%
2017	1,2%	95,9%	2,9%	2017	1,1%	96,2%	2,8%

El análisis de la versión en inglés de esta variable por años muestra, por una parte, que en 2009 el 10,3% de los subtítulos están en pantalla más de seis segundos, mientras que no hay subtítulos por debajo del segundo; sin embargo, en 2017, el número de subtítulos de menos de un segundo asciende hasta un 1,2% y los de más de 6 segundos se reducen hasta un 2,9%. Los datos de los subtítulos en español son muy similares a los de la versión en inglés y de nuevo aquí aumenta, en los mismos periodos, el porcentaje de subtítulos con menos de un segundo de permanencia en pantalla y disminuye, también escasamente, los subtítulos de más de seis segundos de duración.

Por su parte, el análisis de los datos cruzando la duración de los subtítulos por grupos y el tipo de texto, no aporta información muy clarificadora y los datos en la versión en inglés y en español son iguales porque lógicamente todos los subtítulos se generan a partir del *Master File*.

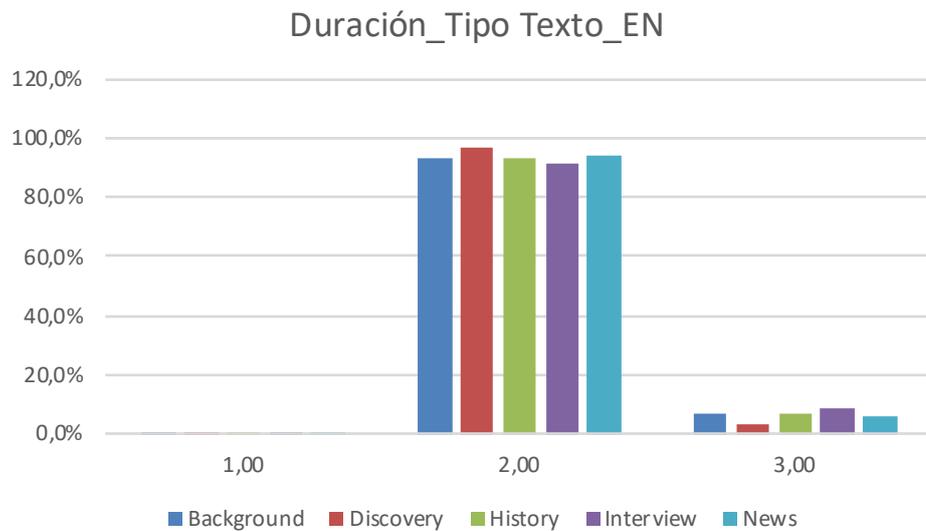


Figura 39. Resultados de variable duración por tipo de texto en EMPAC_EN.

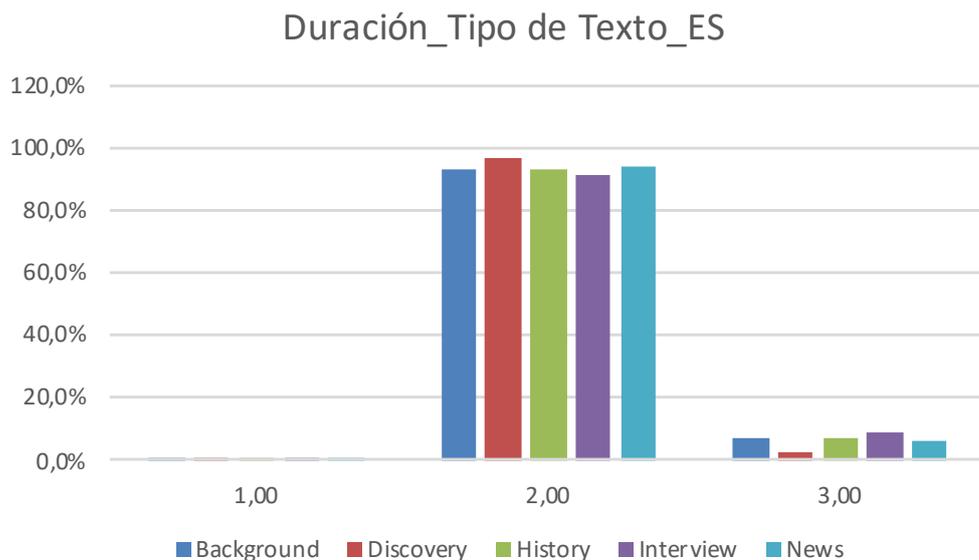


Figura 40. Resultados de variable duración por tipo de texto en EMPAC_ES.

11.2.1.3.4. Variable duración de los subtítulos: análisis global de subtítulos cortos, medios, largos

En este apartado, realizamos un análisis con el que, como indicamos en el capítulo de metodología, pretendemos desvelar la duración más habitual de los subtítulos de nuestro corpus. Por este motivo, recodificamos la variable duración del subtítulo en cuatro franjas de valores: a) grupo 1, de 0 a 2 segundos (cortos), b) grupo 2, de 2 a 4 segundos (medios), c)

grupo 3, de 4 a 6 segundos (largos) y d) grupo 4, de más de 6 segundos. En la siguiente tabla podemos ver la distribución de los datos globales, en inglés y en español:

DURACIÓN_TOTAL_GRUPOS_EN

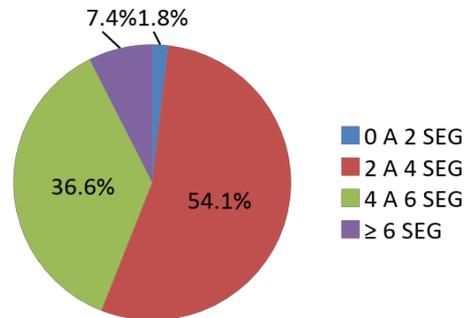


Figura 41. Resultados duración por grupos en EMPAC_EN.

DURACIÓN_TOTAL_GRUPOS_E S

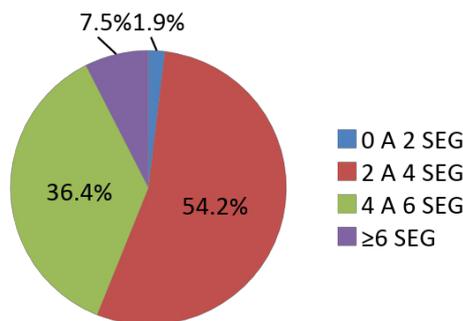


Figura 42. Resultados duración por grupos en EMPAC_ES.

Los datos nos descubren que más de la mitad de los subtítulos del corpus, el 54,1% para ser más precisos, tienen una duración de entre 2 y 4 segundos, seguido del grupo c, de 4 a 6 segundos, con un 36,6% de los subtítulos. Obviamente la diferencia entre las versiones es escasísima puesto que se trabaja con un Master File.

Este mismo análisis, ahora por años, revela que en 2017 se produce un marcado aumento en el porcentaje de subtítulos en el grupo 1 (pasa del 2% en años anteriores a casi el 13%) y un ligero aumento en el grupo 2; de nuevo, las versiones en inglés y en español son coincidentes.

Tabla 45. Resultados duración por grupos y años en EMPAC_EN y EMPAC_ES.

	DURACIÓN_EN				DURACIÓN_ES			
	0 A 2	2 A 4	4 A 6	≥6	0 a 2	2 A 4	4 A 6	≥6
2009	1,7%	46,5%	40,3%	11,5%	1,4%	48,2%	39,6%	10,7%
2010	1,6%	51,4%	38,7%	8,3%	1,6%	51,4%	38,7%	8,3%
2011	1,2%	53,3%	38,4%	7,1%	1,3%	53,6%	38,0%	7,0%
2012	1,5%	56,7%	35,6%	6,2%	1,7%	57,5%	34,8%	5,9%
2013	2,1%	58,5%	33,7%	5,7%	2,3%	59,0%	33,1%	5,5%
2014	1,4%	56,1%	35,2%	7,4%	1,8%	56,8%	34,3%	7,1%
2015	2,0%	55,4%	35,2%	7,3%	2,1%	55,6%	35,1%	7,2%
2016	2,9%	55,2%	34,3%	7,5%	3,0%	56,0%	33,7%	7,2%
2017	12,8%	59,4%	24,5%	3,3%	12,8%	59,6%	24,4%	3,1%

2.1.3.4 Variable duración de los subtítulos: análisis de subtítulos cortos, medios, largos y cps

Por último, nos parece interesante vincular la duración de los subtítulos y su relación con la variable caracteres por segundo por lo que realizamos un análisis de estas dos variables, como mostramos en la siguiente gráfica.

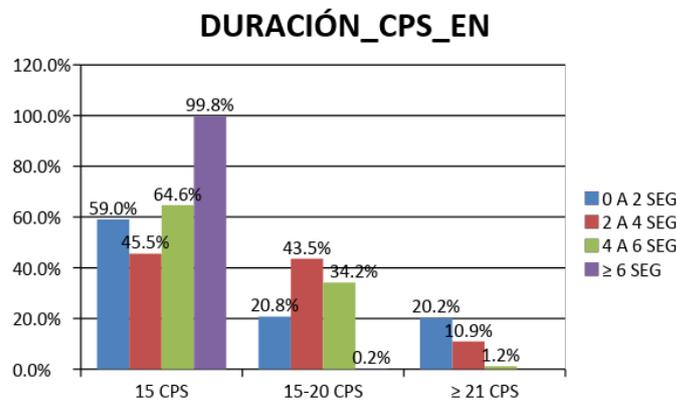


Figura 43. Resultados de duración por grupos y cps en EMPAC_EN.

Observamos que el grupo 4 (más 6 segundos) cumple con la norma de 15 cps, seguido del grupo 3 (de 4 a 6 segundos) con casi un 65% de los subtítulos expuestos a 15 cps. El grupo de 2 a 4 segundos (precisamente el más numeroso del corpus) es el más exigente en cuanto a velocidades de lectura, con un 54,5% de los subtítulos por encima de los 15 cps. El 59% de subtítulos del grupo 1 se proyecta a velocidades de hasta 15. El porcentaje de subtítulos restante se expone, por igual, entre las velocidades de 15 a 20 y 21 cps. De hecho, es en este grupo donde el porcentaje de subtítulos con 21 cps o más es más abultado y por tanto más exigente.

Por su parte, el análisis de estas dos variables en español desvela datos muy similares como podemos ver en la siguiente gráfica:

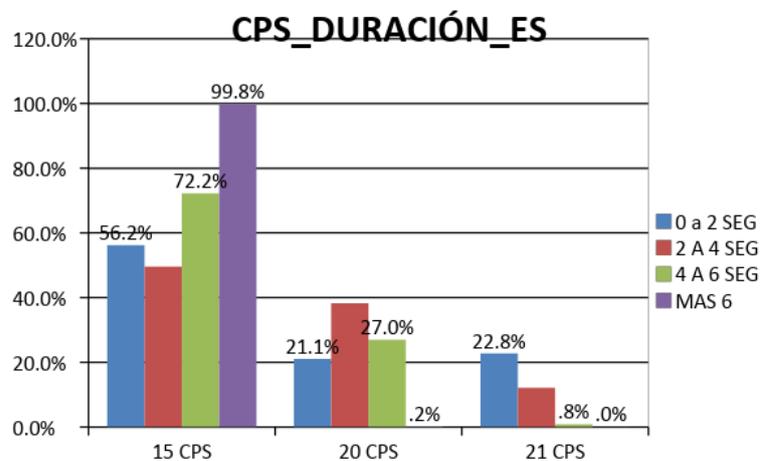


Figura 44. Resultados de duración por grupos y cps en EMPAC_ES.

11.2.2. Análisis del parámetro formato y posicionamiento

En nuestro estudio, el parámetro formato y posicionamiento comprende las variables y valores que reunimos en la siguiente tabla (véase también capítulo 7):

PARÁMETRO	VARIABLE	VALORES POSIBLES
Parámetro formato y posicionamiento	Número de líneas	- 1 línea - 2 líneas - Más de dos líneas
	Número de caracteres por línea	- 32 - 37 Norma UNE - 42 Netflix ≥ 43
	Causas de segmentación inadecuada ⁹⁷	- Preposición al final de línea - Determinante final de línea - Partición formas verbales compuestas - Conjunción al final de línea - Frase-Frase - Partición del sintagma - Partición nombre compuesto

Para las dos primeras variables (el número de líneas por subtítulo y los caracteres por línea) hemos realizado el análisis con las mismas herramientas estadísticas y siguiendo la misma estructura que hemos venido aplicando en el análisis de las variables del parámetro

⁹⁷ En la tabla 7 definimos las causas de segmentación inadecuada por lenguas

velocidad de lectura. En cuanto a la tercera variable, causas de segmentación inadecuada, ésta ha supuesto un reto significativo por la relativa complejidad del lenguaje de búsquedas de CQPweb y la cantidad de tiempo que hemos empleado en familiarizarnos con la herramienta. Una variable que, aunque no encaja estrictamente en las normas técnicas por ser un aspecto lingüístico de los subtítulos, sí está estrechamente vinculada a los parámetros técnicos y al impacto en la velocidad de lectura; un subtítulo bien segmentado se tarda menos en procesar que uno con una segmentación inadecuada por lo que se puede leer y comprender con mayor agilidad. Por esta razón, como Tamayo (2015a), la incluimos en este parámetro y realizamos lo que, a nuestro entender, es el estudio más pormenorizado hasta la fecha en este sentido.

11.2.2.1 Variable número de líneas

Es una norma aceptada que los subtítulos pueden ser de una o dos líneas (a excepción de algunos países con dos lenguas oficiales donde se utilizan subtítulos bilingües) y, de forma excepcional, se pueden componer de hasta tres, en los subtítulos para personas sordas. Por otra parte, según la bibliografía mencionada en el capítulo 7, algunos autores indican que debe haber un equilibrio entre el número de subtítulos de una línea y el número de subtítulos de dos líneas. Sin embargo, opinamos que esto no es del todo acertado porque, a nuestro entender, este criterio se halla sujeto al tipo de texto y al estilo del traductor. Veamos a continuación las características de nuestro material en este sentido.

11.2.2.1.1. Variable número de líneas: análisis global

Los datos muestran que la cantidad total de subtítulos de una línea o dos líneas en ambas versiones es muy similar.

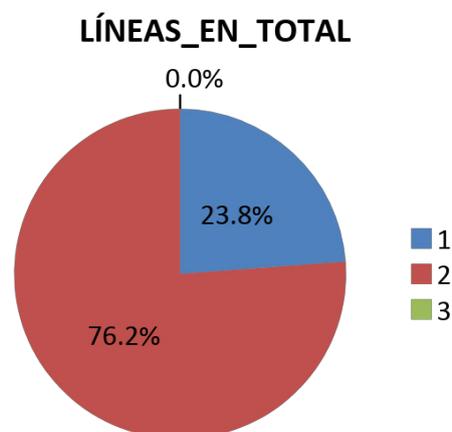


Figura 45. Resultados de la variable número de líneas en EMPAC_EN.

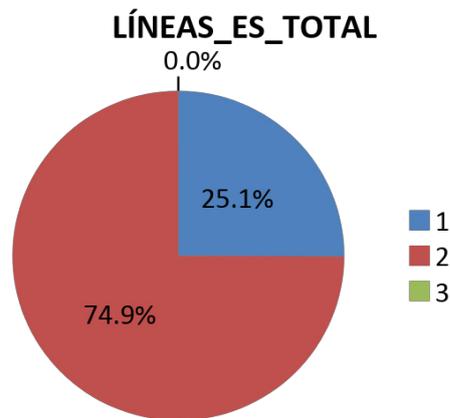


Figura 46. Resultados de la variable número de líneas en EMPAC_ES.

Quizás, lo más revelador del resultado de nuestro análisis es el porcentaje muy superior de subtítulos de dos líneas frente a los de una línea, además del porcentaje cero de subtítulos de más de dos líneas.

11.2.2.1.2. Variable número de líneas: análisis por años

En el análisis de esta variable por años observamos en ambas versiones un cambio drástico en el porcentaje del número de líneas que resulta muy llamativo.

En la siguiente gráfica podemos apreciar cómo, entre 2009 y 2016, la tendencia del número de líneas es prácticamente igual en EMPAC_EN y EMPAC_ES, donde más del 70% de los subtítulos son de dos líneas. Sin embargo, en 2016 la tendencia se invierte por completo y en 2017 hasta el 80% de los subtítulos son de una línea. Podemos apreciar la evolución en las siguientes gráficas, en las que los subtítulos de una línea se marcan en azul y los de dos en rojo.

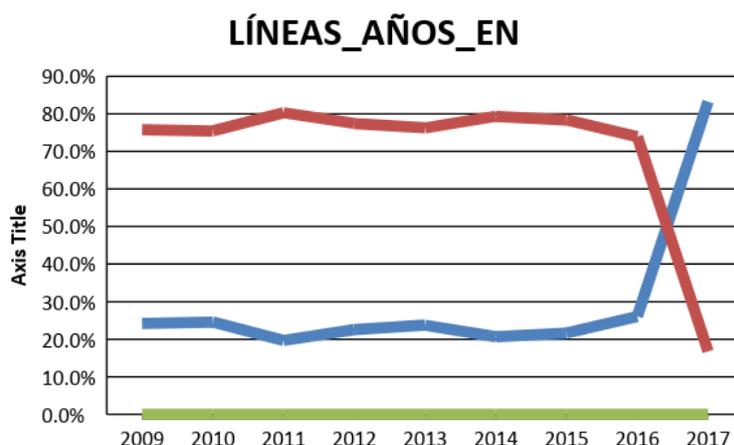


Figura 47. Evolución de la variable número de líneas en EMPAC_EN.

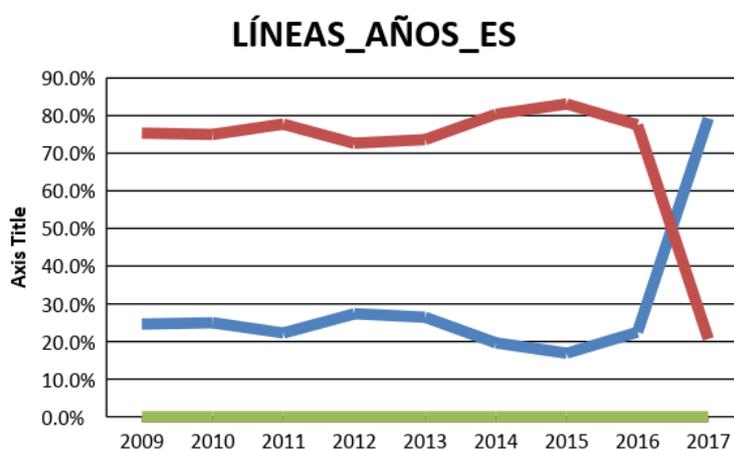


Figura 48. Evolución de la variable número de líneas en EMPAC_ES.

11.2.2.1.3 Variable número de líneas: análisis por tipo de texto

En cuanto al análisis de esta variable según el tipo de texto exponemos los resultados en la siguiente tabla:

Tabla 46. Resultados variable número de líneas por tipo de texto en EMPAC_EN y EMPAC_ES.

Núm Líneas	LÍNEAS_EN			LÍNEAS_ES		
	1	2	3	1	2	3
<i>Background</i>	13,7%	86,3%	0,0%	15,5%	84,5%	0,0%
<i>Discovery</i>	14,3%	85,7%	0,0%	14,8%	85,2%	0,0%
<i>History</i>	11,2%	88,7%	0,1%	11,6%	88,4%	0,0%
<i>Interview</i>	13,3%	86,7%	0,0%	13,9%	86,0%	0,0%
<i>News</i>	13,3%	86,7%	0,0%	13,0%	87,0%	0,0%

No detectamos diferencia destacable atendiendo a la variable tipo de texto en la versión en inglés ni en la versión en español.

11.2.2.2. Variable número de caracteres por línea

El número de caracteres por línea ha ido en aumento en los últimos años y en nuestro estudio analizamos cuatro valores para esta variable a partir de las recomendaciones de las guías, manuales y las aportaciones de los académicos expertos en subtitulación (véase capítulo 7). Para realizar el análisis recodificamos la variable caracteres por línea en cuatro valores que recogen todo el espectro de caracteres por línea del corpus: 32, 37, 42 y más de 42 (43) caracteres por línea (cpl). Los dos primeros son los que recomienda tanto la academia como la normativa vigente (Norma UNE y DCMP) y los dos siguientes los más habituales en el sector profesional. En el capítulo 7 expusimos que el valor 42 representa lo que plataformas de *streaming* como Netflix marcan en sus guías de estilo, mientras que por encima de esa cifra, recogeríamos todos los subtítulos que están por encima de los 42 cpl y normativas y guías entienden como inadecuados por dificultar la lectura de los subtítulos.

11.2.2.2.1. Variable número de caracteres por línea: análisis global

El análisis global nos permite apreciar que los grupos de 32 y 37 caracteres por línea (los que apunta la normativa vigente) suman casi el 80% del corpus.

CPL_EN_TOTAL

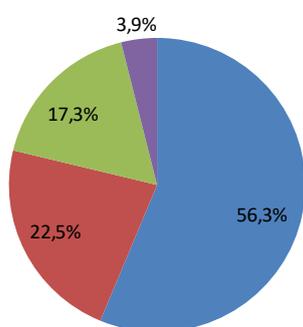


Figura 49. Resultado de análisis de cpl en EMPAC_EN.

CPL_ES_TOTAL

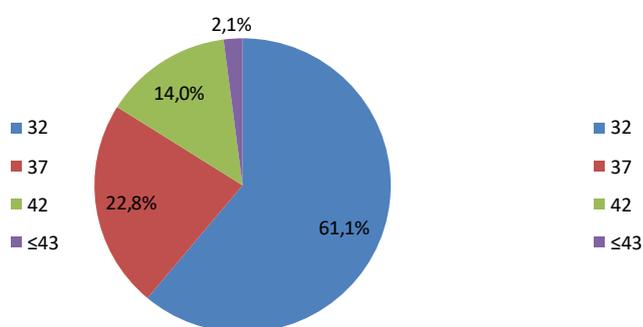


Figura 50. Resultado de análisis de cpl en EMPAC_ES.

Por otra parte, nos resulta interesante que sea la versión en inglés la que tenga más subtítulos en el grupo de 42 y 43 cpl, y no parece descabellado plantear la hipótesis de que esto se deba a un proceso de transcripción automatizada seguido de una escasa edición posterior. Esta interpretación de los datos requiere, claro está, confirmación futura.

11.2.2.2.2. Variable número de caracteres por línea: análisis por años

El análisis de esta variable por años nos procura información acerca de la evolución del número de caracteres por línea y permite ubicaciones cronológicas en cuanto a posibles desviaciones de los valores que se entienden adecuados (hasta 37 caracteres por línea).

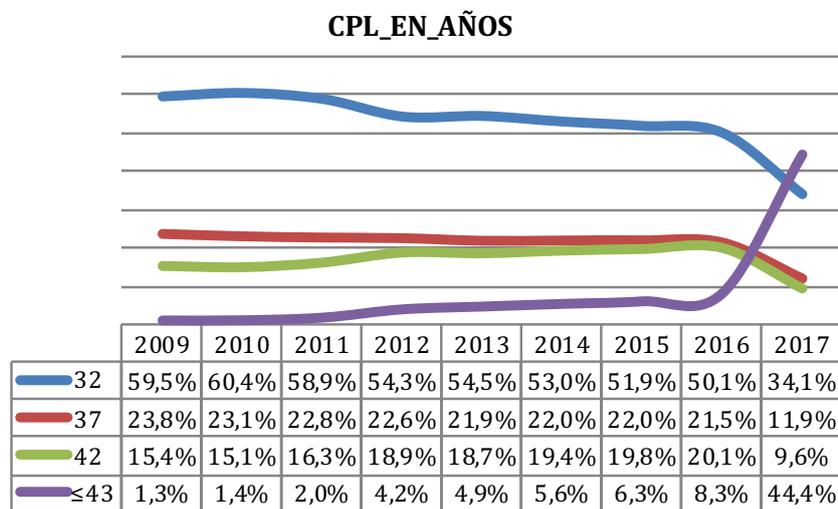


Figura 51. Evolución de cpl en EMPAC_EN.

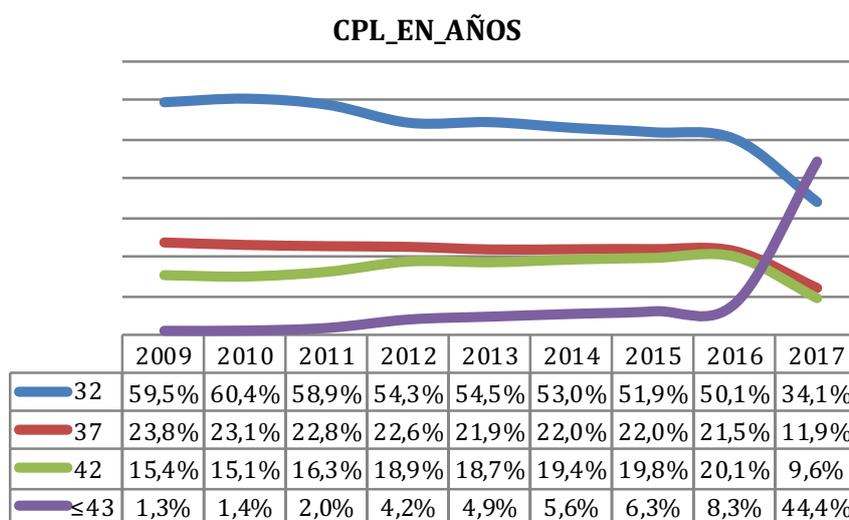


Figura 52. Evolución de cpl en EMPAC_ES.

Observamos que, tanto en EMPAC_EN como en EMPAC_ES, desde 2016 hasta 2017 se produce un cambio en la tendencia de líneas: descienden las de 32, 37 y 42 caracteres mientras que ascienden las de 43 cpl o más caracteres.

Por su parte, en EMPAC_EN vemos que, efectivamente pasamos del 59,5% de las líneas con 32 caracteres de 2009 al 34% de estas en 2017, mientras que el grupo de ≤ 43 cpl registra el 1,3% en 2009 y se dispara hasta 44,4% en 2017.

Apreciamos un comportamiento muy similar en EMPAC_ES; sin embargo, los cambios son sensiblemente más bruscos, si cabe, en los subtítulos en español.

11.2.2.2.3. Variable número de caracteres por línea: análisis por tipo de texto

Por último, según el tipo de texto no apreciamos diferencias significativas, siendo en todos ellos los 32 y 37 caracteres por línea el número más utilizado. Sin embargo, apreciamos que en la versión en inglés existe un porcentaje superior de líneas con el valor 42 y 43 cpl que en la versión en español.

Tabla 47. Resultados análisis cpl por tipo de texto en EMPAC_EN y EMPAC_ES.

cpl	CPL_EN				CPL_ES			
	32	37	42	43	32	37	42	43
<i>Background</i>	56,9%	22,7%	17,0%	3,3%	62,4%	22,2%	13,6%	1,8%
<i>Discovery</i>	60,9%	21,3%	15,1%	2,7%	64,3%	21,5%	12,3%	1,9%
<i>History</i>	57,7%	22,2%	16,9%	3,1%	62,7%	21,5%	13,9%	1,9%
<i>Interview</i>	55,5%	22,7%	17,7%	4,0%	62,5%	22,7%	13,1%	1,7%
<i>News</i>	54,6%	22,4%	18,0%	5,0%	57,9%	24,2%	15,3%	2,6%

Los datos contrastados de EMPAC_EN y EMPAC_ES muestran que en todos los tipos de texto siempre hay más subtítulos con el valor 32 cpl en español; por su parte, el valor 37 cpl es muy parejo en español e inglés. Sin embargo, el valor 42 cpl en inglés siempre tiene un porcentaje superior al español. A su vez, en EMPAC_EN el valor 43 cpl se da en mayor medida que en EMPAC_ES. Es decir que, mientras que en lo referente al tipo de textos, el español parece más conservador (con más 32 que el inglés), el inglés tiende en mayor medida al uso de 42 y de 43 cpl.

11.2.2.3. Variable causas de segmentación inadecuada

Como anunciamos anteriormente, el estudio y análisis de la segmentación en nuestro corpus merece atención especial por lo que dedicaremos un extenso apartado a describir el análisis de las causas de segmentación inadecuada. Para ello, generamos en primer lugar, a partir de la literatura existente y de estudios de recepción realizados sobre este punto (explicados en el capítulo 7), una clasificación de categorías y subcategorías que nos permite estudiar y comparar los resultados en ambas lenguas. A continuación, recordamos nuestra propuesta:

CATEGORIAS GRAMATICALES	SEGMENTACIÓN ESPAÑOL	SEGMENTACIÓN INGLÉS
1. PREPOSICIÓN	1. <i>Prep</i> Preposición final de línea	1. <i>Prep</i> Preposición final de línea
2. DETERMINANTES	2. <i>Det_Indef</i> Artículo indefinido + SN 3. <i>Det_Def</i> Artículo definido + SN 4. <i>Det_Dem</i> Demostrativos + SN 5. <i>Det_Pos</i> Posesivos + SN	2. <i>Det_Indef</i> Artículo indefinido + NP 3. <i>Det_Def</i> Artículo definido + NP 4. <i>Det_Dem</i> Demostrativos + NP 5. <i>Det_Pos</i> Posesivos + NP
3. FORMAS VERBALES COMPUESTAS	6. <i>Verb_Comp</i> Partición verbos compuestos 7. <i>Verb_Perif</i> Partición de perífrasis verbal 8. <i>Verb_PredNom</i> Partición del predicado nominal del atributo	6. <i>Verb_Aux</i> Partición auxiliar + verbo 7. <i>Verb_Inf</i> Partición <i>To</i> + infinitivo 8. <i>Verbo frasal</i> Partición de un verbo frasal
4. CONJUNCIÓN	9. <i>Conj</i> Conjunción al final	9. <i>Conj</i> Conjunción al final
5. FRASES	10. <i>Frase</i> Frase completa + Frase completa	10. <i>Frase</i> Frase completa + Frase completa
6. SINTAGMA	11. <i>Sintg_Adj</i> Partición del sintagma nominal 12. <i>Sintag_Adv</i> Partición del sintagma adverbial	11. <i>Sintg_Adj</i> Partición del sintagma nominal 12. <i>Sintag_Adv</i> Partición del sintagma adverbial
7. PARTICIÓN NOMBRE COMPUESTO	13. <i>Nom_Comp</i> Partición nombre compuesto	13. <i>Nom_Comp</i> Partición nombre compuesto

A partir de esta clasificación, en la que establecemos las posibles causas de segmentación inadecuada que queremos analizar, generamos las búsquedas de las unidades lingüísticas que entendemos no se deben fragmentar y que analizaremos con la herramienta CQPweb.

Como ya avanzamos en el capítulo 10, nuestro análisis de la segmentación requiere de búsquedas diferentes para cada idioma, que distingan por lo demás, entre la segmentación *intrasubtítulo* segmentación *intersubtítulo*. Por este motivo, generamos cuatro grupos de búsquedas que recogimos en sendas tablas para las 13 causas de segmentación inadecuada. En total generamos 52 búsquedas que expusimos en el capítulo 10.

La herramienta CQPweb no sólo nos permite crear las búsquedas de las unidades lingüísticas que entendemos no se debían fragmentar sino que también facilita el cómputo de los casos en todo el corpus y el examen de estos por año y por tipo de texto, como en el resto de variables de cada parámetro.

Antes de mostrar los resultados de nuestro análisis explicamos a continuación cómo ha sido el proceso del análisis en sí. Hemos preferido realizar esta explicación más en detalle en este apartado en lugar de en el capítulo de metodología de análisis por la extensión y la relevancia de la misma.

Como ejemplo, mostramos a continuación la búsqueda de la categoría preposición a final de línea en inglés:

The screenshot shows the CQPweb interface. On the left is a purple sidebar menu with the following items: Menu, Corpus queries (Standard query, Restricted query, Word lookup, Frequency lists, Keywords, Analyse corpus, Export corpus), Saved query data (Query history, Saved queries, Categorized queries, Upload a query, Create/edit subcorpora). The main area is titled 'EMPAC English: powered by CQPweb' and 'Standard Query'. It contains a text input field with the query: `[pos="IN"]</line><line_no="2">[pos="DT|PP$"]*[pos="J.+|RB|N.+"]*[pos="N.+"];`. Below the input field are several settings: Query mode: CQP syntax (with a link to Simple query language syntax); Number of hits per page: 50; Display alignment: Do not show aligned text in parallel corpus; Restriction: None (search whole corpus). At the bottom are 'Start Query' and 'Reset Query' buttons.

Figura 53. Interfaz de CQPweb y el menú de búsquedas.

El resultado de la búsqueda lo podemos ver en la siguiente imagen. En la parte superior de esta vemos que tenemos 6.971 casos de segmentación inadecuada por acabar las líneas en preposición.

Your query "[pos="IN"]</line><line_no="2">[pos="DT|PP\$"]*[pos="J,+|RB|N,+"]*[pos="N,+"]; returned 6,971 matches in 2,479 different texts (in 2,466,812 words [3,817 texts]; frequency: 2,825.91 instances per million words), ordered randomly (0.009 seconds - retrieved from cache)

Showing parallel corpus EMPAC Spanish

No	Text	Solution 1 to 50	Page 1 / 140
1	N003_0106_EN 098db00430eda66d55aa0940f6586a2b	A majority <u>without</u> <u>the Socialists</u> and the Greens ?	
		Por mayoría , pero sin los socialistas ni los verdes .	
2	M001_0054_EN 41ff602eb1a74b6c85d455a0db1a8838	In the event of massive imports <u>of</u> <u>Korean products</u> ...	
		en caso de importaciones masivas de productos coreanos .	

Figura 54. Visualización de los resultados de la búsqueda con CQPweb.

Por su parte, si queremos ver los resultados según la frecuencia de las entradas agrupadas por las etiquetas de POS podemos seleccionar la opción *Frequency Breakdown* en el menú superior de CQPweb. De este modo, en lugar de ver todos los casos detectados, vemos los casos agrupados por categorías y ordenados según el número de entradas de mayor a menor.

Your query "[pos="IN"]</line><line_no="2">[pos="DT|PP\$"]*[pos="J,+|RB|N,+"]*[pos="N,+"]; returned 6,971 matches in 2,479 different texts (in 2,466,812 words [3,817 texts]; frequency: 2,825.91 instances per million words), ordered randomly

Showing frequency breakdown of annotation in this query, at the query node; there are 59 different types and 6,971 tokens at this concordance position.

Breakdown position: Node

No.	Query result	No. of occurrences	Percent
1	IN_DT_NN	1433	20.56%
2	IN_DT_NP	990	14.2%
3	IN_DT_JJ_NN	785	11.26%
4	IN_NN	764	10.96%
5	IN_NP	749	10.74%

Figura 55. Visualización de resultados al filtrar con el *frequency breakdown*.

Seguidamente para comprobar si había diferencias significativas en las causas de segmentación inadecuada entre años hicimos búsquedas restringidas en CQPweb para así obtener los datos de segmentación por años a lo que añadimos la restricción del tipo de texto en consonancia con el estudio de las otras variables. La interfaz de CQPweb nos permite este tipo de búsquedas como mostramos en la imagen que viene a continuación:

The screenshot shows the EMPAC English interface. On the left is a navigation menu with sections: 'Menu', 'Corpus queries' (containing Standard query, Restricted query, Word lookup, Frequency lists, Keywords, Analyse corpus, Export corpus), and 'Saved query data' (containing Query history, Saved queries, Categorised queries). The main area is titled 'EMPAC English: powered by CQPweb' and 'Restricted Query'. It features a large text input field, a 'Query mode' dropdown set to 'CQP syntax' with a link to 'Simple query language syntax', a 'Number of hits per page' dropdown set to '50', and a 'Display alignment' dropdown set to 'Do not show aligned text in parallel corpus'. There are 'Start Query' and 'Reset Query' buttons. Below this is a section titled 'Select the text-type restrictions for your query:' with three columns of checkboxes:

Structure ``text_category``	Structure ``text_type``	Structure ``text_year``
<input type="checkbox"/> EU_affairs <input type="checkbox"/> Economy <input type="checkbox"/> Others <input type="checkbox"/> Security <input type="checkbox"/> Society <input type="checkbox"/> World	<input type="checkbox"/> Background <input type="checkbox"/> Discovery <input type="checkbox"/> History <input type="checkbox"/> Interview <input type="checkbox"/> News	<input type="checkbox"/> 2009 <input type="checkbox"/> 2010 <input type="checkbox"/> 2011 <input type="checkbox"/> 2012 <input type="checkbox"/> 2013 <input type="checkbox"/> 2014 <input type="checkbox"/> 2015 <input type="checkbox"/> 2016 <input type="checkbox"/> 2017

Figura 56. Búsqueda restringida por categoría, tipo de texto, año en CQPweb.

A continuación, pasamos a mostrar los resultados obtenidos de las búsquedas que hemos generado para los distintos tipos de segmentación inadecuada para EMPAC_EN y EMPAC_ES.

11.2.2.3.1. Análisis de la segmentación en inglés entre líneas de subtítulo: segmentación intrasubtítulo

Seguidamente recogemos, en primer lugar, una tabla con el número de tipos (y porcentajes) de segmentación intrasubtítulo e intersubtítulo inadecuada que ilustraremos posteriormente con ejemplos. En este sentido, los ejemplos que presentamos de cada categoría son siempre los diez primeros que aparecen, alineados, en orden aleatorio en CQPweb. Con esta práctica perseguimos alejarnos de búsquedas intencionadas que pueden ser tachadas de parciales o subjetivas.

Primero, como apuntamos en la introducción de este apartado, mostramos en la siguiente tabla el resultado del análisis de la segmentación *intrasubtítulos* en inglés, es decir, la segmentación en EMPAC_EN:

Tabla 48. Resultado de análisis de la segmentación intrasubtítulo en inglés.

class	feature	freq	total	rel_freq
CONJ	Conj	1337	173287	7,715524
DET	Det_Def	1667	173287	9,619879
	Det_Dem	324	173287	1,869731
	Det_Indef	224	173287	1,292653
	Det_Pos	353	173287	2,037083
FRASE	Frase	6427	173287	37,08876
NOM_COMP	Nom_Comp	3946	173287	22,77147
PREP	Prep	6971	173287	40,22806
SINT	Sintag_Adj	2326	173287	13,42282
	Sintag_Adv	2105	173287	12,14748
VERB_COMP	Verb_Aux	949	173287	5,476464
	Verb_Inf	587	173287	3,387444
	Verb_Phrasal	8770	173287	50,60968

Observamos, que en EMPAC_EN, atendiendo a los resultados globales, parece ser que la categoría VERBO COMPUESTO, en concreto *Phrasal Verb*, es la que más casos de segmentación inadecuada recoge, seguida de preposición al final de línea y la categoría FRASE, es decir, cuando tenemos dos frases y en lugar de que cada frase ocupe una línea, estas se reparten entre las dos líneas.

En el análisis en detalle de cada categoría veremos si efectivamente estos resultados son válidos o si, por el contrario, nuestra búsqueda ha reconocido como error aquello que en realidad no lo era.

Seguidamente, exponemos en la siguiente tabla el resultado del análisis de la segmentación *intersubtítulos* en EMPAC_EN:

Tabla 49. Resultados de segmentación intersubtítulo en EMPAC_EN.

class	feature	freq	total	rel_freq
CONJ	Conj	4	85643	0,046706
DET	Det_Dem	6	85643	0,070058
	Det_Pos	5	85643	0,058382
NOM_COMP	Nom_Comp	532	85643	6,211833
PREP	Prep	172	85643	2,008337
SINT	Sintag_Adj	9	85643	0,105087
	Sintag_Adv	136	85643	1,587987
VERB_COMP	Verb_Aux	6	85643	0,070058
	Verb_Inf	3	85643	0,035029
	Verb_Phrasal	2039	85643	23,80813

En el caso de la segmentación *intersubtítulo*, lógicamente no hemos tenido en cuenta la categoría frase y parece ser que la es también la categoría VERBO COMPUESTO, en concreto *Phrasal Verb*, la que más casos de segmentación inadecuada recoge. Sin embargo, el segundo lugar, en esta ocasión lo ocupa la categoría NOMBRE COMPUESTO, seguida de preposición.

Pasamos ahora a analizar en detalle cada una de estas categorías comenzando con la segmentación intrasubtítulo.

a. CATEGORÍA CONJUNCIÓN

Como apuntamos en el capítulo 7, todas las guías recogen la importancia de intentar evitar que la línea de un subtítulo termine con una conjunción. A continuación, mostramos el resultado de este tipo de segmentación en nuestro corpus en los subtítulos de dos líneas y, como apuntamos anteriormente, mostramos la alineación de los diez primeros ejemplos que, de forma aleatoria, nos muestra la aplicación.

No	Solution 1 to 50 Page 1 / 27
1	Thus looking for dialogue <u>and</u> applying pressure is the best solution .
	El diálogo y la presión son la mejor solución .
2	work together <u>and</u> we can achieve this goal .
	necesitamos trabajar juntos y alcanzar esa meta .
3	I 'm very much in favour of democracy , <u>but</u> I do n't see this working well democratically .
	Estoy muy a favor de la democracia , pero dudo que esto funcione democráticamente .
4	solidarity , responsibility <u>and</u> for EU states to support each other .
	solidaridad , responsabilidad y colaboración entre los Estados .
5	the European Central Bank <u>and</u> the International Monetary Fund .
	del BCE y del FMI .
6	that have always existed <u>and</u> of which there is an increasing number .
	que siempre han existido y de los que cada vez hay más .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 27
7	There is the law <u>and</u> there is the interpretation of the law .
	por un lado está la ley y , por el otro , la interpretación de la ley .
8	The states have to agree , <u>and</u> then there 's a public hearing at the EP .
	Cuando los Estados se ponen de acuerdo celebramos una audiencia en el PE .
9	We talk a lot about the casino economy , <u>but</u> there are differences between EU states .
	Se habla mucho de la economía de casino , pero existen diferencias entre los países .
10	which will allow SMEs to flourish <u>and</u> to develop quality products and services .
	que permitan a las pymes prosperar y desarrollar productos y servicios de calidad .

Al filtrar los resultados de la búsqueda con el *frequency breakdown de CQPweb*, se genera una tabla de frecuencias que muestra que hay un total de 11 conjunciones a final de línea en mismo un subtítulo. El *frequency breakdown* se puede hacer por anotación (POS) o por palabra. En este caso lo hemos realizado por palabra.

No.	Query result	No. of occurrences	Percent
1	<u>and</u>	1131	84.59%
2	<u>but</u>	105	7.85%
3	<u>or</u>	65	4.86%
4	<u>both</u>	19	1.42%
5	<u>either</u>	7	0.52%
6	<u>nor</u>	4	0.3%
7	<u>versus</u>	2	0.15%
8	<u>&</u>	1	0.07%
9	<u>less</u>	1	0.07%
10	<u>plus</u>	1	0.07%
11	<u>yet</u>	1	0.07%

Ahora podemos examinar, por ejemplo, las 4 primeras entradas de segmentación inadecuada de “and”, que es la conjunción que más veces aparece en la tabla. Además la alineación de nuestro corpus nos permite monitorizar si una segmentación inadecuada en inglés tiene algún impacto en la segmentación en español. Mostramos las primeras diez entradas con el subtítulo alineado en español.

No	Solution 1 to 50 Page 1 / 23
1	Thus looking for dialogue <u>and</u> applying pressure is the best solution .
	El diálogo y la presión son la mejor solución .
2	work together <u>and</u> we can achieve this goal .
	necesitamos trabajar juntos y alcanzar esa meta .
3	solidarity , responsibility <u>and</u> for EU states to support each other .
	solidaridad , responsabilidad y colaboración entre los Estados .
4	the European Central Bank <u>and</u> the International Monetary Fund .
	del BCE y del FMI .
5	that have always existed <u>and</u> of which there is an increasing number .
	que siempre han existido y de los que cada vez hay más .
6	There is the law <u>and</u> there is the interpretation of the law .
	por un lado está la ley y , por el otro , la interpretación de la ley .
7	The states have to agree , <u>and</u> then there 's a public hearing at the EP .
	Cuando los Estados se ponen de acuerdo celebramos una audiencia en el PE .
8	which will allow SMEs to flourish <u>and</u> to develop quality products and services .
	que permitan a las pymes prosperar y desarrollar productos y servicios de calidad .
9	We need support for economic <u>and</u> social reforms in these countries ,
	Debemos apoyar las reformas económicas y sociales en estos países ,
10	I 'm from France <u>and</u> I 'm third camera assistant .
	Yo vengo de Francia y soy el tercer ayudante de cámara .

Observamos que la segmentación inadecuada en inglés no se replica necesariamente en español; de hecho, cuando la conjunción se ha quedado al final de la línea en español, como en el ejemplo número 8, es previsiblemente por falta de espacio. Por su parte, la revisión manual del resto de entradas (tanto de *and* como de las otras conjunciones) nos permite

constatar que, efectivamente, el impacto de una segmentación inadecuada del inglés en la versión española es testimonial.

b. CATEGORÍA DETERMINANTE

En la categoría determinante incluimos 4 subcategorías que recogen los casos en los que se ha separado el artículo (definido, indefinido), demostrativo y posesivo del sintagma nominal.

b.1. ARTÍCULO INDEFINIDO

En esta primera tabla podemos ver los diez primeros casos que nos muestra CQPweb al realizar la búsqueda de esta subcategoría.

No	Solution 1 to 50 Page 1 / 7
1	This is the case of ‘ Class Enemy ‘ , by a Slovenian director whom EuroparITV met .
	Es el caso de “ Class enemy “ , de un director esloveno al que hemos podido entrevistar .
2	You said governments have a softly-softly approach to organised crime .
	Dice que los Gobiernos son muy blandos con la delincuencia organizada .
3	The vote can also be taken using an electronic box , either to get a clear result ,
	También se hace con urna electrónica para que el resultado esté más claro
4	Palestine have already submitted a roadmap for lasting peace .
	Palestina ya ha ofrecido una hoja de ruta para una paz duradera .
5	Jaka Jakopič was a professional footballer .
	Jaka Jakopič era futbolista profesional ,
6	and the Commission backed a guideline daily amount label .
	y la Comisión quiere indicar el consumo máximo diario .
7	We do n’t have any tools and a European Monetary Fund is the answer .
	No hay instrumentos , y uno sería el Fondo Monetario Europeo .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 7
8	It is hoped there may be a safety agreement on oil and gas
	El acuerdo sobre la seguridad del petróleo y del gas
9	Tehran reacted by threatening to block a key oil shipping lane , the Strait of Hormuz .
	Teherán amenaza con bloquear el estrecho de Ormuz , por donde circula el petróleo .
10	You are an MEP from Luxembourg , a member of the European People 's Party .
	Eurodiputada por Luxemburgo y miembro del Partido Popular Europeo .

Por su parte, el *frequency breakdown*, en este caso del POS, muestra que, los artículos indefinidos van seguidos por sustantivos o adjetivos en el mayor número de casos y que la unidad de sentido que más se ve afectada por una segmentación inadecuada es la de artículo indefinido seguido de nombre común.

No.	Query result	No. of occurrences	Percent
1	DT NN	102	45.54%
2	DT JJ NN	83	37.05%
3	DT JJ JJ NN	9	4.02%
4	DT NP	8	3.57%
5	DT RB JJ NN	8	3.57%
6	DT JJR NN	7	3.13%
7	DT JJ NNS	2	0.89%
8	DT JJ NP	2	0.89%
9	DT RB JJ JJ NN	2	0.89%
10	DT JJ JJ JJ NN	1	0.45%

b.2. ARTÍCULO DEFINIDO

En la subcategoría de artículo definido es en la que hay un mayor número de casos de segmentación inadecuada de entre las cuatro subcategorías mencionadas arriba. En esta primera tabla podemos ver los diez primeros casos de segmentación inadecuada que nos muestra CQPweb para esta subcategoría.

No	Solution 1 to 50 Page 1 / 34
1	The regulation hopes to develop the internal market and crisis measures
	La normativa espera desarrollar el mercado interior y medidas anticrisis
2	climate change , the financial crisis the Echelon electronic eavesdropping system ,
	el cambio climático , la crisis financiera , el sistema de espionaje Echelon ,
3	All MEPs will meet in the hemicycle , the new MEPs officially taking up their seats .
	Todos los eurodiputados viajarán a Estrasburgo para tomar posesión .
4	The majority of the Council want the system of private investment to remain .
	La mayoría del Consejo quiere mantener el sistema de inversión privada ,
5	take unbelievable steps , like the surveillance of everyone ‘s individual life .
	dar paso a cosas increíbles , como la vigilancia de la vida privada .
6	Pyotr Gusenkov received the authorisation to dock
	El Pyotr Gusenkov recibe autorización para atracar
7	take part in the European Voluntary Service .
	participan en el servicio voluntario europeo .
8	Will you be active as Chairman of the Foreign Affairs Committee ?
	¿ Tendrá un papel activo como presidente de la Comisión de Asuntos Exteriores ?
9	He said that Blair will be President of the European Council “ over my dead body “ .
	Ha dicho que Blair será presidente del Consejo “ por encima de su cadáver “ .
10	You say there ‘s been a change in the mindset here but regards transparency ,
	Afirma que se ha producido un cambio en la mentalidad , respecto a la transparencia ,

Por una parte, observamos que nuevamente la segmentación inadecuada en inglés no influye en la segmentación en español y, por otra parte, el *frequency breakdown* del POS muestra que los artículos definidos van seguidos por sustantivos o adjetivos en el mayor

número de casos y que las unidades de sentido que más se ven afectadas por una segmentación inadecuada son dos: la de artículo definido seguido de nombre común y nombre propio.

No.	Query result	No. of occurrences	Percent
1	DT NN	549	32.93%
2	DT NP	479	28.73%
3	DT JJ NN	255	15.3%
4	DT NNS	200	12%
5	DT JJ NNS	80	4.8%
6	DT JJ NP	35	2.1%
7	DT JJ JJ NN	24	1.44%
8	DT JJS NN	12	0.72%
9	DT NPS	10	0.6%
10	DT JJS NNS	8	0.48%

En las siguientes tablas podemos ver las diez primeras entradas de las dos entradas que aglutinan el mayor número de casos: la de artículo definido seguido de nombre común y nombre propio (DT NN y DT NP).

No	Solution 1 to 50 Page 1 / 11
1	The majority of the Council want the system of private investment to remain .
	La mayoría del Consejo quiere mantener el sistema de inversión privada ,
2	take unbelievable steps , like the surveillance of everyone 's individual life .
	dar paso a cosas increíbles , como la vigilancia de la vida privada .
3	Pyotr Gusev received the authorisation to dock
	El Pyotr Gusev recibe autorización para atracar
4	You say there 's been a change in the mindset here but regards transparency ,
	Afirma que se ha producido un cambio en la mentalidad , respecto a la transparencia ,
5	it 's the same rules , no matter who 's the owner , who 's the share holder ,
	tenga las mismas normas , sin importar quiénes sean sus propietarios o accionistas :
6	Meetings with the ECB and Eurogroup
	de reuniones con los presidentes del BCE y el Eurogrupo ,

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 11
7	than the fiscal problems of the eurozone or of the European Union .
	que los de la eurozona o la Unión Europea .
8	The problem is that the 27 postponed the question of the eurozone 's new governance .
	El problema es que los 27 han pospuesto el asunto de la nueva gobernanza .
9	The rapes are extremely violent , with the insertion of foreign objects into the vagina ,
	Estas violaciones son muy agresivas , les introducen objetos en la vagina ,
10	We will have to phase them out once the rationale for these measures fades away
	Tienen que desaparecer progresivamente a medida que dejen de hacer falta

A continuación, la tabla con los diez primeros ejemplo de casos de artículo definido seguido de nombre propio:

No	Solution 1 to 50 Page 1 / 10
1	climate change , the financial crisis the Echelon electronic eavesdropping system ,
	el cambio climático , la crisis financiera , el sistema de espionaje Echelon ,
2	take part in the European Voluntary Service .
	participan en el servicio voluntario europeo .
3	Will you be active as Chairman of the Foreign Affairs Committee ?
	¿ Tendrá un papel activo como presidente de la Comisión de Asuntos Exteriores ?
4	for an allied military offensive against the Islamic State of Iraq and the Levant or ISIL .
	a una ofensiva militar aliada contra el Estado Islámico de Irak y el Levante .
5	It is not enough . But when the European Commission made this proposal
	No es suficiente , pero cuando nos presentó la Comisión esta propuesta

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 10
6	appeared before MEPs on Monday as the United Nations Special Envoy for the Sahel .
	compareció ante los eurodiputados como enviado especial de la ONU para el Sahel .
7	Last unknown factor : the date of the MEPs ' vote on the new commissioners .
	Otra incógnita : la fecha en que la cámara votará el nuevo colegio de comisarios .
8	from the European Trade Union Confederation .
	de la Confederación europea de sindicatos .
9	In political or legislative matters , the Ombudsman will direct you to Parliament ,
	En cuestiones políticas o legislativas , el Defensor le derivará al Parlamento ,
10	Since the EPP were allied with the UK Conservatives in the last legislature ,
	Ya que los conservadores británicos formaban parte del PPE hasta ahora

De nuevo la segmentación inadecuada en inglés no influye en la segmentación en español.

En esta ocasión, nos ha parecido interesante mostrar el conjunto léxico (ARTÍCULO + SUSTANTIVO) que más veces se ha separado de forma incorrecta, para lo que mostramos la *frequency breakdown* de las palabras:

No.	Query result	No. of occurrences	Percent
1	the European	74	4.44%
2	the EU	52	3.12%
3	the Commission	51	3.06%
4	the Parliament	35	2.1%
5	the Member	17	1.02%
6	the environment	15	0.9%
7	the EP	14	0.84%
8	the eurozone	13	0.78%
9	the Council	10	0.6%
10	the government	9	0.54%

El artículo y sustantivo *the European* son la combinación de palabras que más veces se han segmentado de forma inadecuada. Algo por otra parte bastante previsible ya que *Europa* y todos sus derivados es una de las palabras más frecuentes en nuestro corpus.

No	Solution 1 to 50 Page 1 / 2
1	take part in the European Voluntary Service . participan en el servicio voluntario europeo .
2	It is not enough . But when the European Commission made this proposal No es suficiente , pero cuando nos presentó la Comisión esta propuesta
3	from the European Trade Union Confederation . de la Confederación europea de sindicatos .
4	the Chair of the Confederal Group of the European United Left / Nordic Green Left . presidente del GUE / NGL .
5	as one of the European Capitals of Culture . como una de las capitales de la cultura .
6	She took part in the European Voluntary Service , Participa en el servicio voluntario europeo ,
7	When more people are aware of how the European Parliament or the Commission Si más gente conociera cómo el Parlamento Europeo o la Comisión
8	Via the ‘ Tweet your MEP ‘ initiative , the European Parliament has built a platform La plataforma Tweet your MEP del Parlamento Europeo
9	I wish to acknowledge that the European Union had severely condemned Quiero agradecer a la Unión Europea que haya condenado duramente
10	according to a new report by the European Agency for Fundamental Rights , según un nuevo informe de la Agencia Europea de Derechos Fundamentales ,

Tanto esta entrada, como el resto de las que analizamos a partir del *frequency breakdown*, muestran cómo la segmentación en una lengua no afecta a la otra.

b.3. SUBCATEGORÍA DEMOSTRATIVOS

En esta primera tabla podemos ver los diez primeros casos que nos muestra CQPweb al realizar la búsqueda de esta subcategoría.

No	Solution 1 to 50 Page 1 / 7
1	During an anti-occupation march , this MEP from Morphou , in the north ,
	Durante una marcha antiocupación , esta eurodiputada de Morfou , en el norte ,
2	between the European Union and those countries which are further away from us .
	entre la Unión Europea y otros países que están más alejados de nosotros .
3	Mr Souvira , how do these criminal networks work ?
	Sr. Souvira , ¿ cómo operan estas redes criminales ?
4	Because we must n't forget that these people have not committed any crime .
	No debemos olvidar que estas personas no han cometido ningún delito .
5	In other words , it 's disingenuous to say that interconnection is an issue for renewables .
	Es falso afirmar que la interconectividad no es un problema para las renovables .
6	because they are forming this Constituent Assembly . That 's the first point .
	porque se trata de formar esa asamblea constituyente . Es el primer punto .
7	At first glance , a 24g portion of this product will give us 98 kilocalories -
	Una porción de 24g nos aporta 98 kilocalorías ,
8	A bad idea , according to this representative of European farmers .
	Mala idea , según este representante de los agricultores europeos .
9	The EP sends about a dozen of these delegations around the world each year
	El Parlamento envía una docena de delegaciones por el mundo cada año
10	Last summer scenes like these hit the headlines .
	El verano pasado , estos casos fueron noticia .

En cuanto al posible impacto de la segmentación inadecuada en la versión en inglés observamos que no condiciona la segmentación en español.

b.4. SUBCATEGORÍA POSESIVOS

De nuevo, en la siguiente tabla podemos ver los diez primeros casos que nos muestra CQPweb al realizar la búsqueda de esta subcategoría.

No	Solution 1 to 50 Page 1 / 8
1	the 6 countries decide to extend their partnership to cover the whole economy .
	los 6 países deciden ampliar el acuerdo al resto de la economía .
2	came to defend his 2010 action plan before MEPs .
	defendió su plan de acción 2010 ante la Eurocámara .
3	MEPs have now stepped up their responsibilities , voting for a directive
	Ahora , el PE asume su responsabilidad aprobando una directiva
4	this gynaecologist was awarded for his combat in his hospital in eastern DR Congo .
	Este ginecólogo obtuvo el galardón por su lucha en un hospital del este del Congo .
5	ring up the train company , give our information and buy the ticket that way .
	a la compañía ferroviaria para dar nuestra información y comprar el billete .
6	Women were forced to defer to their husbands for authorisation
	(no alignment found)
7	A full-time job , in a country which on its way to becoming 28th member of the EU .
	Trabaja a tiempo completo en un país que se convertirá en el 28º país de la UE .
8	He constantly trains his employees on new technologies .
	Da formación continua a sus empleados en nuevas tecnologías .
9	They also want to strengthen their relationship with the countries of the South ,
	Ellos también quieren reforzar su relación con los países del sur ,
10	They were nervous about entrusting their economy to a European Central Bank .
	No querían confiar su economía a un Banco Central Europeo .

Por su parte, el *frequency breakdown* del POS muestra 12 posibles combinaciones en las que destaca aquellas en las que los posesivos van seguidos por sustantivos o adjetivos.

No.	Query result	No. of occurrences	Percent
1	PP\$ NN	166	47.03%
2	PP\$ NNS	87	24.65%
3	PP\$ JJ	72	20.4%
4	PP\$ NP	7	1.98%
5	PP\$ CD	5	1.42%
6	PP\$ RB	5	1.42%
7	PP\$ JJR	3	0.85%
8	PP\$ JJS	3	0.85%
9	PP\$ VVN	2	0.57%
10	PP\$ CC	1	0.28%
11	PP\$ RBS	1	0.28%
12	PP\$ VBZ	1	0.28%

Mostramos en la siguiente tabla, los ejemplos de la primera entrada, es decir, posesivo seguido de nombre común por ser la que más ocurrencias recoge.

No	Solution 1 to 50 Page 1 / 4
1	the 6 countries decide to extend their partnership to cover the whole economy .
	los 6 países deciden ampliar el acuerdo al resto de la economía .
2	this gynaecologist was awarded for his combat in his hospital in eastern DR Congo .
	Este ginecólogo obtuvo el galardón por su lucha en un hospital del este del Congo .
3	ring up the train company , give our information and buy the ticket that way .
	a la compañía ferroviaria para dar nuestra información y comprar el billete .
4	A full-time job , in a country which on its way to becoming 28th member of the EU .
	Trabaja a tiempo completo en un país que se convertirá en el 28º país de la UE .
5	They also want to strengthen their relationship with the countries of the South ,
	Ellos también quieren reforzar su relación con los países del sur ,

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 4
6	They were nervous about entrusting <u>their</u> <u>economy</u> to a European Central Bank .
	No querían confiar su economía a un Banco Central Europeo .
7	had no qualms expressing <u>their</u> <u>hostility</u> towards the SWIFT deal .
	expresaron sin reparos su hostilidad al pacto SWIFT .
8	What would we find if we put <u>our</u> <u>shopping</u> baskets under the microscope ?
	¿ Qué encontraríamos si pusiéramos la bolsa de la compra bajo el microscopio ?
9	We want free countries in <u>our</u> <u>neighbourhood</u> . That 's a huge challenge .
	Queremos países vecinos libres . Ese es un gran reto .
10	The biggest priority , shared by <u>our</u> <u>student</u> association , is integrating Europe
	La prioridad de nuestra asociación de estudiantes es integrar Europa

Asimismo, mostramos ejemplos de la tercera entrada con más casos, es decir, posesivo seguido de adjetivo.

No	Solution 1 to 50 Page 1 / 2
1	Why does n't nuclear count towards <u>our</u> <u>renewable</u> targets , if it 's low carbon ?
	La energía nuclear tiene bajas emisiones .
2	and have been elected because of <u>their</u> <u>different</u> views on how to create policies .
	y han sido elegidos por sus ideas políticas innovadoras .
3	Foreign Policy Chief presented <u>her</u> <u>new</u> right arm :
	La alta representante de Exteriores presenta su nuevo brazo derecho :
4	Several countries have already frozen <u>their</u> <u>preparatory</u> work to ratify the agreement .
	Varios países ya han congelado los preparativos para ratificar el acuerdo .
5	The EU 's most pressing topics are <u>its</u> <u>new</u> regime of financial supervision
	Problemas acuciantes para la UE son su nuevo régimen de supervisión financiera

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 2
6	for using the experience of our European partners and their support .
	para aprovechar la experiencia de nuestros socios europeos y su respaldo .
7	Every year , the MEPs vote for their favourite European film for the LUX Prize .
	Cada año , los eurodiputados eligen su filme europeo preferido para el Premio LUX .
8	could Tunisia , like Egypt , see its democratic flowering stamped out ?
	¿ Túnez , al igual que Egipto , fracasará en su transición democrática ?
9	It 's evening and Stella , not her real name , searches for friends online .
	Por la noche , Stella , nombre ficticio , busca amigos en la red .
10	Crisis-struck Portugal has seen its public debt rating regularly downgraded
	Con la crisis , la calificación de la deuda soberana portuguesa ha ido bajando

En cuanto al posible impacto de la segmentación inadecuada en la versión en inglés observamos que tampoco en este caso condiciona la segmentación en español.

c. CATEGORÍA FRASE

Como apunta, entre otros, Díaz Cintas y Remael (2007:176) “If a subtitle consists of two or more sentences, put one sentence on each line”. El resultado de nuestro análisis, muestra un elevado número de casos comparado con el resto de categoría: 6427 casos. Veamos en qué circunstancias se dan y si realmente pueden considerarse segmentaciones inadecuadas.

No	Solution 1 to 50 Page 1 / 129
1	at the Heads of State and Government Council in Brussels this week . But it won't .
	en el Consejo Europeo de esta semana en Bruselas , pero no será así .
2	a species from the Caspian Sea , which is destroying the fauna and flora . Let 's see .
	una especie del mar Caspio que está destruyendo la fauna y la flora . Veámoslo .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 129
3	One minute with two possible answers . It 's Yes or No. Off we go .
	Tiene un minuto con dos respuestas posibles . Sí o no . Allá vamos .
4	The slogan ' Act . React . Impact . '
	El eslogan es " Acción . Reacción . Decisión . "
5	You should get to choose . You might choose to be naked or public online ,
	Deberíamos poder elegir . Puedes elegir estar desnudo en público , en Internet ,
6	That 's not important . It 's important that he 's an important whistle-blower .
	Eso no es lo importante , sino que su denuncia es muy relevante .
7	OK , sent . Now I 'll just have to wait .
	Enviado . Ahora , a esperar .
8	Hello . Only just back from Strasbourg ,
	Hola . A su regreso de Estrasburgo ,
9	but how many countries will make up the EU of tomorrow ? Zoom in on the map !
	pero ¿ cuántos países integrarán la UE del mañana ? ; Veamos el mapa !
10	It is not enough . But when the European Commission made this proposal
	No es suficiente , pero cuando nos presentó la Comisión esta propuesta

En este caso, el *Frequency breakdown* del POS no puede ofrecer información relevante por la cantidad de casos que se dan, en concreto 157; sin embargo, en la siguiente tabla, podemos observar las primeras 20 combinaciones recogidas:

No.	Query result	No. of occurrences	Percent
1	SENT PP	1624	25,27%
2	SENT DT	784	12,2%
3	SENT IN	363	5,65%
4	SENT NN	354	5,51%
5	SENT NP	336	5,23%
6	SENT RB	318	4,95%
7	SENT VV	317	4,93%
8	SENT JJ	227	3,53%
9	SENT WP	188	2,93%

(Continúa en la página siguiente)

(Continúa de la página anterior)

No.	Query result	No. of occurrences	Percent
10	SENT CC	177	2,75%
11	SENT NNS	156	2,43%
12	SENT WRB	144	2,24%
13	SENT UH	129	2,01%
14	SENT EX	127	1,98%
15	SENT VBZ	103	1,6%
16	SENT CD	67	1,04%
17	‘ NP	63	0,98%
18	SENT MD	61	0,95%
19	SENT PPS	59	0,92%
20	: PP	54	0,84%

Al analizar cada entrada de forma manual nos dimos cuenta de que los casos que van precedidos de comillas seguidos de la etiqueta del POS, por ejemplo: ([‘ NP](#)) no son casos de segmentación inadecuada pues son los entrecomillados que aparecen en el corpus:

No	Solution 1 to 50 Page 1 / 2
1	‘ Again ‘ , because in the 90s the picture was much less rosy .
	‘ De nuevo ‘ , porque en la década de 1990 , el panorama era menos favorable .
2	It ‘s an expression we do n’t use . ‘ Aiuto ‘ , a request of help maybe .
	No utilizamos esa expresión , decimos más bien “ aiuto “ .
3	‘ Europe ? Give me a name and a phone number , please . ‘
	‘ ¿ Europa ? Deme un nombre y un número de teléfono , por favor ‘ .
4	“ Europe must go further in terms of solidarity . ‘
	Pero también pide más solidaridad en la UE
5	‘ Smoking kills ‘ is well-known . It ‘s on every pack of cigarettes with shocking pictures .
	Es bien sabido que fumar mata . Lo pone en todos los paquetes de tabaco .
6	‘ Warhol 1 ‘ by Slovenian artist Marko Blažo
	‘ Warhol 1 ‘ , del artista esloveno Marko Blažo ,

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 2
7	' Sparpolitik ' is a very positive idea .
	' Sparpolitik ' es una idea muy positiva .
8	' Buy-local ' entrepreneurs have partnered up
	Los emprendedores con el lema ' compra local ' se han asociado
9	' United in diversity . ' That 's the motto of the European Union .
	" Unidos en la diversidad " . Es el lema de la Unión Europea .
10	' Austerity plan ' does n't rhyme with ' recovery ' at the moment . How do you see the future ?
	' Plan de austeridad ' no rima con ' recuperación ' . ¿ Cómo ve el futuro ?

Queremos señalar (a modo de curiosidad) que el alineado con la versión en español nos permite ver cómo en ocasiones se utilizan las comillas simples mientras que en otros casos son las dobles las escogidas. Este tipo de aspectos bien merecerán nuestra atención en estudios posteriores.

Por su parte, los casos que van precedidos de dos puntos seguidos de una etiqueta de POS, por ejemplo: (: [PP](#)) tampoco serían casos de segmentación inadecuada. Mediante un análisis manual hemos constatado que son puntos suspensivos que indican duda, interrupción voluntaria del discurso o énfasis expresivo:

No	Solution 1 to 50 Page 1 / 2
1	On the other hand ... I think it 's important for European taxpayers to know this .
	Por otro lado , es importante que el contribuyente europeo sepa
2	Even now ... They used to talk about the creative class in Russia .
	Incluso ahora ... Antes se hablaba de la clase creativa en Rusia .
3	When you listen to Labwani ... He did n't say it there . I 've known him for 10 years .
	Cuando escuchamos a Labwani ... Ahí no lo dice . Le conozco desde hace 10 años .
4	but based on ... I do n't know ... the financial transaction tax
	pero basándose en ... en el impuesto sobre transacciones financieras

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 2
5	shot or injured ... They indiscriminately fire into demonstrations ,
	Disparan de manera indiscriminada en las manifestaciones ,
6	They say ... They leak in advance , they speak of a federal Europe ,
	Dicen ... Adelantan información , hablan de una Europa federal ,
7	No ... It was quite honestly confronted with such a shockwave
	No ... Tuvo que enfrentarse a tal seísmo
8	So if it happens ... It will probably not be a Greek situation in Spain .
	Si ocurriera ... La situación en España no será como la griega ,
9	Irreproachable elections ... I do n't think anyone 's ever seen one anywhere .
	Elecciones irreprochables . Dudo que se hayan visto antes .
10	we forget ... It 's like that story .
	olvidamos ... Como en el cuento .

En cuanto a la forma en la que la segmentación en inglés pudiera influir en la segmentación del subtítulo en español, consideramos que si bien hay ocasiones en la que la segmentación inadecuada es coincidente, también son muchos los casos en los que la segmentación en español, cuando el espacio lo permite, separa las frases en líneas.

d. CATEGORÍA NOMBRE COMPUESTO

Los nombres compuestos son unidades de sentido que hemos de evitar separar puesto que ralentizan la lectura de la información textual que aparece en el subtítulo. Hemos de tener en cuenta que en el caso de la lengua inglesa son muy habituales, en especial si lo comparamos con la española. Mostramos a continuación las primeras diez entradas que obtenemos como resultado de la búsqueda de subtítulos que segmentan los nombres compuestos:

No	Solution 1 to 50 Page 1 / 79
1	as it does n't look like the General Affairs summit is going to sort this ?
	No parece que el Consejo de Asuntos Generales vaya a resolverlo .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 79
2	The Environment Committee votes on the draft report on Thursday .
	La Comisión de Medio Ambiente votará el proyecto de informe el jueves .
3	I and the Danish government support that ,
	El Gobierno danés y yo apoyamos esa postura ,
4	The aim is to form a European Voluntary Humanitarian Aid Corps .
	El objetivo es formar un Cuerpo Voluntario Europeo de Ayuda Humanitaria .
5	I 'd like to thank Commissioner Connie Hedegaard and our MEPS ,
	Gracias a la comisaria Connie Hedegaard y a nuestros eurodiputados ,
6	and they have n't got off-balance-sheet things in Member States .
	y los Estados miembros no puedan omitir nada en sus balances .
7	Find out when Agriculture Commissioner Phil Hogan stops by on Wednesday .
	Lo sabremos en la comparecencia del comisario de Agricultura el miércoles .
8	and the EU High Representative for Foreign Affairs and Security , Federica Mogherini .
	y la Alta Representante de Exteriores y Seguridad , Federica Mogherini .
9	Because there was an emergency summit of European heads of state .
	Se celebraba una cumbre de emergencia de jefes de estado .
10	It 's hoped the Industry and Environment Committees will consider the opinion .
	Se espera que las comisiones de Industria y Medio Ambiente la consideren .

A partir del *frequency breakdown* vemos en la siguiente tabla las etiquetas de POS con un mayor número de entradas:

No.	Query result	No. of occurrences	Percent
1	NP NP	898	22.76%
2	NN NN	783	19.84%

(Continúa en la página siguiente)

(Continúa de la página anterior)

No.	Query result	No. of occurrences	Percent
3	NN NNS	558	14.14%
4	NP NP NP	297	7.53%
5	NP NN	239	6.06%
6	NN NP	164	4.16%
7	NP NNS	125	3.17%
8	NP NPS	110	2.79%
9	NNS NN	84	2.13%
10	NP NP NP NP	81	2.05%
11	NN NN NN	75	1.9%
12	NNS NNS	72	1.82%
13	NP NPS NP	57	1.44%
14	NN NN NNS	56	1.42%
15	NP NP NN	34	0.86%

Encontramos en la siguiente tabla los diez primeros casos de la primera entrada con más casos (NP NP), es decir, nombre propio seguido de nombre propio:

No	Solution 1 to 50 Page 1 / 18
1	The aim is to form a European Voluntary Humanitarian Aid Corps .
	El objetivo es formar un Cuerpo Voluntario Europeo de Ayuda Humanitaria .
2	I 'd like to thank Commissioner Connie Hedegaard and our MEPS ,
	Gracias a la comisaria Connie Hedegaard y a nuestros eurodiputados ,
3	The key , says a unit head at the European Employment Service , is mobility .
	La clave , según el director del Servicio Europeo de Empleo , es la movilidad .
4	No more bill shock , says European People 's Party MEP Pilar del Castillo
	Se acabaron las facturas abultadas , según la eurodiputada del PPE Pilar del Castillo ,
5	The newly elected French President François Hollande
	El nuevo presidente francés , François Hollande ,
6	about who is standing to be Commission President , what they stand for ,
	sobre quién se presenta a presidente de la CE , qué ideas defiende

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 18
7	15 MEPs did take part in the European Union 's election observation mission .
	15 eurodiputados participaron en la misión de observación electoral de la UE
8	It was n't until 1989 that Mikhail Gorbachev , then President of Russia ,
	Hubo que esperar a 1989 a que Mijaíl Gorbachov , entonces presidente ,
9	To discuss all this , we have Mercedes Bresso from the Socialists and Democrats .
	(no alignment found)
10	Should that group split up , would Liberal Graham Watson stand a chance ?
	Si se dividiera el grupo , ¿ tendría posibilidades el liberal Graham Watson ?

Y seguidamente, la segunda entrada, con los casos de nombre común seguido de otro nombre común (NN NN):

No	Solution 1 to 50 Page 1 / 16
1	I and the Danish government support that ,
	El Gobierno danés y yo apoyamos esa postura ,
2	Because there was an emergency summit of European heads of state .
	Se celebraba una cumbre de emergencia de jefes de estado.
3	saying , 'This seems to be a sub-standard product . 'But nothing was done.
	que indicaron : 'Parece ser un producto de baja calidad ' . Pero no se hizo nada.
4	MEPs want an international regulation agency for raw materials sales.
	La Cámara pide una agencia reguladora sobre la venta de materias primas.
5	should n't be seen as this huge event right before the elections.
	no debería ser visto como un gran evento antes de las elecciones.

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 16
6	Is positive discrimination, for example recruitment quotas, a solution?
	¿La discriminación positiva, por ejemplo, las cuotas de contratación, es la solución?
7	By combining this in one external action service we can save a lot of money.
	Al combinarse en un servicio único, podemos ahorrar mucho dinero.
8	This is the aim of a new free trade deal that MEPs agreed on Thursday.
	en virtud de un nuevo acuerdo de libre comercio recién aprobado por el PE .
9	but refused to discuss raising the rescue fund 's € 0.5 billion ceiling until March.
	pero rechaza debatir el aumento del techo del fondo de rescate, de 500 millones.
10	The trainers in the youth centre really helped me,
	Los entrenadores del centro juvenil me ayudaron mucho

Por último, al filtrar la búsqueda de esta categoría con el *frequency breakdown* por palabras que aparecen con mayor frecuencia estás son las diez palabras compuestas que más veces aparecen segmentadas de manera inadecuada en nuestro corpus:

No.	Query result	No. of occurrences	Percent
1	European Parliament	76	1.93%
2	Member States	58	1.47%
3	European Commission	55	1.39%
4	European Union	17	0.43%
5	Commission President	14	0.35%
6	European Globalisation Adjustment	14	0.35%
7	billion euros	12	0.3%
8	climate change	11	0.28%
9	European People	11	0.28%
10	Monetary Affairs	11	0.28%

Las primeras seis entradas muestran la segmentación del subtítulo en nombres de instituciones o cargos, como el caso de *European Commission* que vemos a continuación:

No	Solution 1 to 50 Page 1 / 2
1	In March 2011, the European Commission proposed for the first time
	En marzo de 2011, la Comisión Europea propuso por primera vez
2	the humanitarian office of the European Commission , is in a crisis meeting.
	la oficina humanitaria de la Comisión, convocaba una reunión de crisis.
3	on the condition that the European Commission is informed in advance.
	a condición de que se informe previamente a la Comisión Europea.
4	If they fail to comply, the European Commission will act once again.
	Si no la cumplen, la Comisión Europea interviene de nuevo.
5	They knew the system at the European Commission very well, the procedure,
	Conocían bien el sistema de la Comisión Europea, el procedimiento,
6	The President of the European Commission , Mr Barroso.
	El presidente de la Comisión Europea, el Sr. Barroso.
7	on the condition that the European Commission is informed in advance.
	a condición de que se informe previamente a la Comisión Europea.
8	The European Commission said
	La Comisión Europea se ha justificado :
9	when he was President of the European Commission from 1985 to 1995.
	como presidente de la Comisión Europea entre 1985 y 1995 ,
10	The President of the European Commission , José Manuel Barroso .
	El presidente de la Comisión Europea , José Manuel Durão Barroso .

Por otra parte, en la séptima y octava entrada del *frequency breakdown* por palabras es donde encontramos los casos de nombres comunes compuestos. A modo de ejemplo, ilustramos a continuación el caso de *climate change*:

No	Solution 1 to 11 Page 1 / 1
1	A key component of Europe 's climate change policy is energy efficiency,
	Algo crucial en la política climática de la UE es la eficiencia energética,
2	Find more tricks to combat climate change soon on A-Z.
	Descubre más trucos para combatir el cambio climático en A-Z.
3	And we stay on the subject of climate change with today 's guest.
	Seguimos con el cambio climático con nuestra invitada de hoy.
4	Tackling the economic crisis and climate change will be crucial for Stockholm.
	La crisis económica y el cambio climático son las prioridades para Estocolmo.
5	Has the world lost interest in climate change over the last two decades?
	¿Se ha perdido el interés por el cambio climático en las últimas décadas?
6	There are challenges ahead: climate change , immigration, the economic crisis.
	Nos esperan desafíos como el cambio climático, inmigración, crisis económica ...
7	We 'll come back to the question of climate change and CO2 emissions in a moment.
	Después hablaremos del cambio climático y las emisiones de CO2.
8	And finally, a leading expert on climate change comes to the EP on Tuesday.
	Y por último, un experto en cambio climático visita el PE el martes .
9	We 're joined by Maeve McLynn , climate change and development policy officer
	Nos acompaña Maeve McLynn , encargada del cambio climático y el desarrollo
10	The probable priorities include climate change , business competition
	Las probables prioridades incluyen el cambio climático , la competitividad
11	And 2 out of 3 believe fighting climate change will be good for the economy.
	y 2 de cada 3 creen que combatirlo sería bueno para la economía .

En esta categoría, observamos que los nombres compuestos que han sido segmentados en inglés de forma inadecuada (segmentación que puede o no atribuirse a cuestiones espaciales), en la versión subtitulada al español la segmentación se ha gestionado de otra forma en la mayor parte de los casos.

e. CATEGORÍA PREPOSICIÓN

Otra de las unidades de sentido que es recomendable no separar cuando segmentamos un subtítulo es una preposición del sintagma nominal que le sigue; de hecho, es una de las unidades de sentido de mayor relevancia. En la siguiente tabla mostramos las diez primeras entradas que obtenemos como resultados de nuestra búsqueda. En el total de categorías que hemos diseñado para nuestro estudio este es el segundo tipo de segmentación inadecuada que más casos acumula, solo por detrás de la separación de dos elementos de una forma verbal o de una perífrasis.

No	Solution 1 to 50 Page 1 / 140
1	A majority <u>without the Socialists</u> and the Greens?
	Por mayoría, pero sin los socialistas ni los verdes.
2	In the event of massive imports <u>of Korean products</u> ...
	en caso de importaciones masivas de productos coreanos .
3	The fifth edition <u>of the European</u> Health Award.
	Quinta edición del Premio Europeo de la Salud ,
4	We have those metals <u>in everything</u> we touch in our day-to-day life ,
	Esos metales están presentes en todo lo que tocamos cada día ,
5	According to some MEPs, a slim rise <u>in this sum</u> may be in store in 2011.
	Según los eurodiputados, esta suma podría aumentar un poco en 2011.
6	The main conclusion is <u>that politics</u> is much more complicated
	La conclusión principal es que la política es mucho más complicada
7	One thing which causes problems <u>for students</u> when it comes to recognition
	Uno de los problemas de los estudiantes con las homologaciones

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 140
8	I 'm asking <u>for</u> <u>a mechanism</u> at the level of the Union
	Lo que pedimos es un mecanismo a nivel comunitario
9	and a member <u>of</u> <u>the non-attached members.</u>
	y un miembro de los no alineados,
10	Lots of conflicts arose <u>because</u> <u>nobody</u> was able to decide.
	Surgieron muchos conflictos porque nadie era capaz de decidir.

Por su parte, el *frequency breakdown* muestra 55 combinaciones de preposición seguida de otras categorías. Las 20 primeras y las que recogen más del 90% de los casos son las que podemos apreciar en la siguiente tabla:

No.	Query result	No. of occurrences	Percent
1	<u>IN DT NN</u>	1433	20.56%
2	<u>IN DT NP</u>	990	14.2%
3	<u>IN DT JJ NN</u>	785	11.26%
4	<u>IN NN</u>	764	10.96%
5	<u>IN NP</u>	749	10.74%
6	<u>IN NNS</u>	452	6.48%
7	<u>IN JJ NNS</u>	395	5.67%
8	<u>IN DT NNS</u>	393	5.64%
9	<u>IN JJ NN</u>	314	4.5%
10	<u>IN DT JJ NNS</u>	187	2.68%
11	<u>IN DT JJ JJ NN</u>	96	1.38%
12	<u>IN DT JJ NP</u>	82	1.18%
13	<u>IN JJ JJ NNS</u>	45	0.65%
14	<u>IN DT NPS</u>	36	0.52%
15	<u>IN DT JJ JJ NNS</u>	32	0.46%
16	<u>IN JJR NN</u>	25	0.36%
17	<u>IN DT RB JJ NN</u>	19	0.27%
18	<u>IN JJ JJ NN</u>	18	0.26%
19	<u>IN JJ NP</u>	16	0.23%
20	<u>IN NPS</u>	14	0.2%

La primera entrada, con 1433 casos, es la de preposición + determinante + nombre común. En la siguiente tabla recogemos las primeras diez entradas con el subtítulo alineado en español:

No	Solution 1 to 50 Page 1 / 29
1	According to some MEPs, a slim rise <u>in this sum</u> may be in store in 2011.
	Según los eurodiputados, esta suma podría aumentar un poco en 2011.
2	I 'm asking <u>for a mechanism</u> at the level of the Union
	Lo que pedimos es un mecanismo a nivel comunitario
3	it 's considered <u>as the charter</u> of rights for seafarers.
	considerado como una carta de derechos de los marineros.
4	Cloning is the reproduction <u>of an individual</u> without fertilisation.
	La clonación consiste en reproducir un individuo sin fertilización.
5	and does her homework sitting <u>on the toilet</u> , with no table to work on.
	y hace los deberes sentada en el retrete, ya que no tiene mesa.
6	Eloisa is very happy <u>with this volunteer</u> initiative.
	Eloisa está muy contenta con el programa de voluntariado.
7	Whether he 'll be able to will depend <u>on the outcome</u> of this European debate.
	Que lo consiga dependerá del resultado de este debate europeo.
8	But the tools must be there so <u>that the day</u> they 're needed it can be done.
	Pero deben existir herramientas para que puedan utilizarse.
9	Standing in for the boss. The job <u>of an assistant</u> needs a jack of all trades:
	Siempre a disposición de su jefe. El trabajo del asistente es variopinto:
10	11 March is also the European Day <u>for the Victims</u> of Terrorism
	El 11 de marzo es el Día Europeo de las Víctimas del Terrorismo

Y en esta ocasión, en la segunda entrada, las categorías gramaticales se distribuyen de la siguiente forma: preposición seguida de determinante y de nombre propio. Mostramos las diez primeras entradas de esta distribución en la siguiente tabla:

No	Solution 1 to 50 Page 1 / 20
1	The fifth edition <u>of</u> <u>the European</u> Health Award.
	Quinta edición del Premio Europeo de la Salud,
2	to proposed revisions <u>of</u> <u>the EU</u> Firearms Directive,
	a las revisiones de la Directiva de Armas de Fuego,
3	But not all the creases <u>between</u> <u>the EU</u> and Moscow were ironed out.
	La cumbre no ha solucionado todos los roces entre la UE y Moscú.
4	Romano Prodi, former President <u>of</u> <u>the European</u> Commission, will meet MEPs
	Romano Prodi, expresidente de la Comisión Europea, se reunirá con los eurodiputados
5	The Parliament was n't happy <u>with</u> <u>the Bulgarian</u> Rumiana Jeleva
	La búlgara Rumiana Jeleva no convenció a los parlamentarios
6	We have to make sure <u>that</u> <u>the Treaty</u> is not questioned anymore ,
	Hay que asegurar que el Tratado no vuelve a ponerse en duda ,
7	that we catch by methods other <u>than</u> <u>the Corruption</u> Perceptions Index ,
	que no recoge el Índice de Percepción de la Corrupción , pero sí otros ,
8	MEPs will decide <u>upon</u> <u>the EU</u> budget 2011 in October.
	La Eurocámara se pronunciará en octubre sobre el presupuesto 2011.
9	and Gorbachev was Chair <u>of</u> <u>the Supreme</u> Council of the USSR, etc.
	y Gorbachov, el Consejo Superior de la URSS, etc.
10	which is not endorsed <u>by</u> <u>the Commission</u> itself.
	no refrendado por la propia Comisión.

A continuación, damos cuenta del caso de preposición + sustantivo, que es el que presenta mayor número de incidencias:

No	Solution 1 to 50 Page 1 / 16
1	We have those metals in everything we touch in our day-to-day life ,
	Esos metales están presentes en todo lo que tocamos cada día,
2	The main conclusion is that politics is much more complicated
	La conclusión principal es que la política es mucho más complicada
3	Lots of conflicts arose because nobody was able to decide.
	Surgieron muchos conflictos porque nadie era capaz de decidir.
4	and walks his dog, a symbol of friendship for many US presidents.
	y saca a su perro, una muestra de amistad para muchos presidentes.
5	It distinguishes between two categories of spending : agricultural and other types.
	Distingue entre dos categorías de gastos: el agrícola y los demás tipos.
6	there are problems with hotel or flight reservations.
	se dan muchos problemas con las reservas de hoteles y de aviones .
7	Europe may need to wait for energy independence day.
	la independencia energética tal vez esté lejos.
8	the part of universities , the part of business , the part of I do n't know what.
	en las universidades, las empresas, o lo que sea.
9	Let 's try working on insulation of our buildings.
	Intentemos mejorar el aislamiento de los edificios.
10	Has some momentum for dealing with climate change has been lost?
	¿Se ha perdido el ímpetu para abordar el cambio climático ?

Podemos apreciar que, en los ejemplos que hemos mostrado de esta categoría, la subtitulación en español, generalmente, respeta las normas de segmentación y no rompe la unidad de sentido como sí ocurre en la versión en inglés.

f. PARTICIÓN DEL SINTAGMA

f.1. SINTAGMA ADJETIVAL

El sintagma adjetival o adjetivo va acompañado de los complementos del adjetivo y debemos evitar romper esta unidad de sentido. A continuación, exponemos el resultado de la búsqueda de este tipo de segmentación potencialmente inadecuada con las primeras diez entradas y su subtítulo alineado:

No	Solution 1 to 50 Page 1 / 47
1	It 's the first time a permanent president chairs EU summits,
	Por primera vez hay un presidente permanente en las cumbres europeas
2	under Lisbon it should have an equal say with the Council .
	Según Lisboa, debería estar en pie de igualdad con el Consejo.
3	The Orange Revolution was a one-month show with one million people on the square.
	En la Revolución Naranja, un millón de personas ocupó la plaza durante un mes.
4	while around 85% of the European social fabric is composed of SMES.
	mientras que el 85% del tejido social europeo se compone de pymes.
5	Excluding cash , cards are the preferred method of payment in the European Union
	Después del efectivo , la tarjeta es nuestro modo de pago preferido ,
6	In the committee stage, any individual member can submit amendments .
	En las comisiones, cualquier miembro puede presentar enmiendas.
7	Maybe because a consistent EU approach in Africa was lacking.
	Debido a la ausencia de una política europea coherente en África.

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 47
8	The project could jump start the biggest energy alliance the EU has formed
	Este proyecto es la piedra angular de la mayor alianza energética
9	Martin Schulz, S , would be in a good position for the next President.
	Martin Schulz, S , estaría en una buena posición para la Presidencia.
10	as in 2001 the cross-border programme IZOM began.
	porque en 2001 nació el programa transfronterizo IZOM.

Las diez primeras entradas del *frequency breakdown* del POS recogen más del 90% de los casos las exponemos en la siguiente tabla:

No.	Query result	No. of occurrences	Percent
1	DT JJ NN	1295	55.67%
2	DT JJ JJ	326	14.02%
3	DT JJ NNS	297	12.77%
4	DT JJ NP	244	10.49%
5	DT JJS NN	51	2.19%
6	DT JJS JJ	29	1.25%
7	DT JJS NNS	28	1.2%
8	DT JJR NN	27	1.16%
9	DT JJS NP	7	0.3%
10	DT JJR JJ	6	0.26%

La primera entrada en la que la ruptura del sintagma se produce entre alguno de estos tres elementos determinante, adjetivo y nombre común, nos sirve de ejemplo para ilustrar este tipo de segmentación inadecuada:

No	Solution 1 to 50 Page 1 / 26
1	It 's the first time a permanent president chairs EU summits,
	Por primera vez hay un presidente permanente en las cumbres europeas
2	under Lisbon it should have an equal say with the Council.
	Según Lisboa, debería estar en pie de igualdad con el Consejo .

(Continúa en la página siguiente)

(Continúa de la página anterior)

3	The Orange Revolution was a one-month show with one million people on the square.
	En la Revolución Naranja, un millón de personas ocupó la plaza durante un mes.
4	Excluding cash, cards are the preferred method of payment in the European Union
	Después del efectivo , la tarjeta es nuestro modo de pago preferido,
5	In the committee stage, any individual member can submit amendments.
	En las comisiones, cualquier miembro puede presentar enmiendas.
6	Martin Schulz, S, would be in a good position for the next Presidency.
	Martin Schulz , S , estaría en una buena posición para la Presidencia.
7	as in 2001 the cross-border programme IZOM began .
	porque en 2001 nació el programa transfronterizo IZOM .
8	No, because this is not an exclusive power for the EU which it should be.
	No, porque no es competencia exclusiva de la UE, aunque debería serlo.
9	There was also an important leadership change :
	Han dado un salto cualitativo en liderazgo.
10	Let 's move on to the second part of our video sequence,
	Pasemos a la segunda parte de nuestro vídeo

Por su parte, en esta tabla mostramos ejemplos de la segunda entrada, es decir, la combinación de determinante + adjetivo + adjetivo (DT JJ JJ):

No	Solution 1 to 50 Page 1 / 7
1	Final question: from now until the first constitutive session what will happen ?
	Última pregunta. ¿Qué pasará de aquí a la sesión constitutiva?
2	In the view of a continuing European unification
	En el contexto de la unificación europea,

(Continúa en la página siguiente)

(Continúa de la página anterior)

3	due to the lack of a homogenous European citizenship concept,
	debido a la ausencia del concepto de una ciudadanía europea homogénea,
4	to project it into a certain supranational European construction.
	por proyectar una construcción europea supranacional.
5	and a disaster for the whole European rail organisation.
	y un desastre para toda la organización ferroviaria europea.
6	Let 's move on to a related second question.
	Escuchemos la segunda pregunta, que viene que ni pintada.
7	It has got itself into an extreme economic mess at the moment.
	En la actualidad, tiene un serio problema económico .
8	She 's an MEP with the Liberal Democratic group , the ALDE .
	Es eurodiputada del grupo liberal y democrático ALDE.
9	A question to all of you. What is the real added value of the Citizens ' Initiative?
	Una pregunta para todos. ¿Cuál es el valor añadido de la Iniciativa Ciudadana?
10	are you a proper democratic representative?

En las tablas que acabamos de mostrar damos cuenta de la ruptura de sintagmas adjetivales en la versión en inglés y hemos observado, por una parte, que la cantidad de información que cada subtítulo contiene es muy elevada, pudiendo ser esta precisamente una de las causas de segmentación inadecuada. Por su parte, la versión en español no parece copiar los errores de segmentación si bien, cuando estos se producen, son por falta de espacio. El número de caracteres por línea y la cantidad de tiempo que el subtítulo permanece en pantalla son determinantes a la hora de decidir cómo segmentar un subtítulo.

f.2. PARTICIÓN SINTAGMA ADVERBIAL

En el caso del sintagma adverbial nos encontramos aproximadamente con el mismo número de ocurrencias que en el de la partición del sintagma adjetivo. Vemos el resultado de la búsqueda en la siguiente tabla.

No	Solution 1 to 50 Page 1 / 43
1	But their small size can <u>easily</u> <u>allow</u> them to pass into the human body.
	Sin embargo, su pequeño tamaño les permite pasar al cuerpo humano.
2	If this Parliament can't <u>finally</u> <u>make</u> the Council see reason,
	Si este Parlamento no puede hacer entrar en razón al Consejo,
3	a team of EU experts will <u>also</u> <u>be</u> able to intervene in the process.
	un equipo de expertos de la UE podrá intervenir en el proceso.
4	Europe has to assert itself and it can <u>only</u> <u>do</u> so in a world with a global economy
	Europa debe hacerse valer y sólo lo logrará en una economía globalizada
5	Very <u>often</u> <u>electronic waste</u> leaves Europe
	Muchos residuos electrónicos salen de Europa
6	when they 're <u>deliberately</u> <u>flouting</u> the rules is wrong.
	cuando desacatan las normas deliberadamente es un error.
7	Many ships like this one end <u>up</u> <u>broken</u> up for recycling
	Muchos barcos como este acaban desmontados para ser reciclados
8	Jacques Chirac made a <u>very</u> <u>anti-European speech</u> in 1978.
	Jacques Chirac dio un discurso muy antieuropeo en 1978.
9	The next phase in this saga is <u>behind</u> <u>closed</u> doors, talks at the Council.
	La próxima fase del proceso es a puerta cerrada: las conversaciones en el Consejo.
10	I do n't agree. GMOs were <u>never</u> <u>introduced</u> to produce more.
	No estoy de acuerdo. No utilizan OMG para producir más.

El *frequency breakdown* muestra el tipo de categorías gramaticales que encontramos en el sintagma adverbial, así como las combinaciones en las que más número de casos de segmentación inadecuada se ha producido.

No.	Query result	No. of occurrences	Percent
1	RB VV	485	23.04%
2	RB VVN	377	17.91%
3	RB VVG	236	11.21%
4	RB JJ NN	109	5.18%
5	RB VVP	103	4.89%
6	RB VB	97	4.61%
7	RB VH	97	4.61%
8	RB VVZ	88	4.18%
9	RB VBZ	87	4.13%
10	RB VVD	80	3.8%
11	RB NN	55	2.61%
12	RB VHZ	52	2.47%
13	RB VBN	44	2.09%
14	RB NP	39	1.85%
15	RB VBP	36	1.71%
16	RB VHP	22	1.05%
17	WRB NP	20	0.95%
18	WRB NN	14	0.67%
19	RB VBG	11	0.52%
20	RB JJ JJ NN	10	0.48%

Seguidamente, recogemos en la siguiente tabla ejemplos de adverbio + infinitivo sin *to* (RB VV) que con 23,04% es la que más casos documenta:

No	Solution 1 to 50 Page 1 / 10
1	<p>But their small size can easily allow them to pass into the human body.</p> <p>Sin embargo, su pequeño tamaño les permite pasar al cuerpo humano.</p>
2	<p>If this Parliament ca n't finally make the Council see reason,</p> <p>Si este Parlamento no puede hacer entrar en razón al Consejo,</p>

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 10
3	Europe has to assert itself and it can only do so in a world with a global economy
	Europa debe hacerse valer y sólo lo logrará en una economía globalizada
4	that this Parliament will finally show conviction and pride
	de que el PE demuestre por fin convicción y orgullo
5	that we need to really break our dependence on oil,
	para abandonar el petróleo de golpe
6	It may be punished financially if it does n't obey the decision at the end of the trial.
	Podrá castigársele económicamente si no acata la decisión de los jueces.
7	The list of problems is endless. We ca n't count how many problems Somalia has.
	La lista de problemas es interminable. No podemos contarlos.
8	He said that the Serbs there will never accept Kosovo as an independent state.
	Dice que los serbios allí jamás aceptarán el estado independiente de Kosovo.
9	Aside from that, I 'd much rather look to the future than the past.
	Aparte de eso, prefiero mirar al futuro que al pasado.
10	But those that work here know that crisis is far from being resolved.
	Pero los que trabajan aquí saben que la crisis está lejos de estar resuelta.

Por último, presentamos una tabla que muestra ejemplos de subtítulos que separan adverbio del participio pasado del verbo:

No	Solution 1 to 50 Page 1 / 8
1	Many ships like this one end up broken up for recycling
	Muchos barcos como este acaban desmontados para ser reciclados

(Continúa en la página siguiente)

No	Solution 1 to 50 Page 1 / 8
2	The next phase in this saga is <u>behind closed</u> doors, talks at the Council .
	La próxima fase del proceso es a puerta cerrada: las conversaciones en el Consejo.
3	I do n't agree. GMOs were <u>never introduced</u> to produce more.
	No estoy de acuerdo. No utilizan OMG para producir más.
4	You ca n't change it. You have <u>n't changed</u> it in the last 50 years.
	No se puede cambiar. No ha cambiado en los últimos 50 años.
5	as long as the British government has <u>not decided</u> to leave the EU or notified the EU.
	mientras el Gobierno británico no decida abandonar la UE ni lo notifique a la UE.
6	where an ingredient is <u>specifically developed</u> for cosmetic purposes
	cuando un ingrediente se ha desarrollado específicamente con fines cosméticos
7	The Hungarian Presidency is <u>eagerly awaited</u> in the Parliament on Wednesday.
	El Parlamento recibirá a la presidencia húngara el miércoles.
8	Here the EP 's role has n't been so <u>clearly strengthened</u> , as our expert explains.
	El Parlamento no podrá bloquearlo, como explica este experto .
9	All Berlusconi 's media have <u>always opposed</u> Prodi and the left.
	Los medios de Berlusconi siempre se opusieron a Prodi y a la izquierda.
10	There was a collapse and, as was <u>rightly pointed</u> out, ordinary people are suffering .
	El sistema se derrumbó y como dicen, la gente es la que sufre.

g. CATEGORÍA VERBO COMPUESTO

En la categoría formas verbales compuestas hemos generado tres subcategorías (verbo auxiliar, *to* infinitivo y *phrasal verb*) la suma de estas se convierten en la categoría con mayor número de casos de segmentación inadecuada. Recordemos, que no se deben separar los elementos de una forma verbal o de una perífrasis. Comenzamos con los casos que

separan El auxiliar del verbo principal; continuaremos con la separación del *to* del verbo en infinitivo; y concluiremos con la tercera categoría de los verbos compuestos o *phrasal verbs*.

g.1. VERBO AUXILIAR

En nuestro corpus hemos encontrado 949 casos en los que se ha separado el auxiliar del verbo principal. En la siguiente tabla ofrecemos los diez primero casos de nuestra búsqueda:

No	Solution 1 to 50 Page 1 / 19
1	But do you understand , Europeans who <u>are fighting</u> against tax fraud and tax evasion, Pero entenderá que los europeos que luchan contra el fraude y la evasión fiscal
2	and the fact that we now <u>have developed</u> and developing countries y el hecho de que las naciones desarrolladas y en desarrollo
3	This gives rights to a product that <u>is produced</u> somewhere else in the world Esto otorga derechos sobre un producto que se fabrica en otro lugar del mundo,
4	Finland , other Member States , they ' <u>re going</u> to look for a lighter licensing regime en Finlandia y otros Estados , van a buscar un régimen de licencias más flexible
5	For 10 years an emergency service <u>has helped</u> the homeless in Brussels Este servicio de emergencias de Bruselas proporciona a los sin techo
6	which do have good governance and <u>are making</u> major progress in governance. donde hay un buen gobierno y progresan de forma importante en este sentido.
7	If , as soon as an infected leave <u>is discovered</u> , you cut the branches on top Cuando se descubre una hoja infectada , si se cortan las ramas superiores
8	that is in danger of <u>being overshadowed</u> by roaming , que podría ser eclipsado por la itinerancia .
9	when they are <u>being reported</u> to the Member States. cuando se informa a los Estados miembros.

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 19
10	I can tell you that what 's <u>being done</u> in Greece right now - Les aseguro que es lo que se está haciendo ahora en Grecia,

El *frequency breakdown* por POS muestra una amplia gama de posibilidades que recogemos en la siguiente tabla:

No.	Query result	No. of occurrences	Percent
1	<u>VBN VVN</u>	159	16.75%
2	<u>VHP VVN</u>	109	11.49%
3	<u>VBD VVN</u>	105	11.06%
4	<u>VBP VVG</u>	102	10.75%
5	<u>VHZ VVN</u>	83	8.75%
6	<u>VBP VVN</u>	73	7.69%
7	<u>VBZ VVG</u>	64	6.74%
8	<u>VBN VVG</u>	58	6.11%
9	<u>VBG VVN</u>	40	4.21%
10	<u>VBZ VVN</u>	39	4.11%
11	<u>VBD VVG</u>	26	2.74%
12	<u>VHP VBN</u>	26	2.74%
13	<u>VHZ VBN</u>	25	2.63%
14	<u>VHD VVN</u>	10	1.05%
15	<u>VBZ VBG</u>	7	0.74%
16	<u>VHG VVN</u>	6	0.63%
17	<u>VBP VBG</u>	5	0.53%
18	<u>VBZ VBZ</u>	3	0.32%
19	<u>VHP VBZ</u>	2	0.21%
20	<u>VHP VHD</u>	2	0.21%

Ejemplos de las dos primeras entradas los podemos ver a continuación. El primer lugar lo ocupan los siguientes dos elementos: el participio pasado del verbo *to be* seguido del participio pasado del verbo principal:

No	Solution 1 to 50 Page 1 / 4
1	And my region has <u>been under-represented</u> for that time. Mi región ha estado infrarrepresentada durante este periodo.

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 4
2	Do you think it will be decisive, as has been said , for a future coalition with the CDU?
	¿Cree que será decisivo, como se afirma, para la futura coalición con el CDU ?
3	There 's already a position that 's been decided by the majority of governments .
	La mayoría de los gobiernos ya han adoptado una postura.
4	is that Ireland has been held responsible and accountable for that ,
	es que se ha hecho responsable a Irlanda y ha pagado,
5	The Sakharov Prize has been awarded by the EP each year since 1988
	El Premio Sájarov ha sido otorgado por el PE cada año desde 1988
6	This was very brave. Had she been caught , she would 've ended up in prison.
	Fue muy valiente, podría haber acabado en la cárcel si la hubiesen descubierto .
7	where the most recent application has been lodged , that should examine the case,
	el que tiene la solicitud de asilo más reciente , debe examinar el caso,
8	The decision has been made in the Council -
	Se ha tomado la decisión en el Consejo
9	I think that peace has been consolidated . That 's the first phase.
	La paz está consolidada. Es el final de la primera etapa.
10	to present the conclusions that have been adopted . He 's then grilled by MEPs.
	para presentar las conclusiones adoptadas. Los eurodiputados lo interrogan.

En segundo lugar, la forma verbal que más se separa es el participio pasado de *have* seguido del participio pasado del verbo principal:

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 3
1	and the fact that we now <u>have developed</u> and developing countries
	y el hecho de que las naciones desarrolladas y en desarrollo
2	where they <u>have put</u> solar panels in place
	en las que se han instalado paneles solares
3	Eastern European countries <u>have contributed</u> the least digitised content.
	Europa del Este es la que menos ha aportado al contenido digital.
4	The 27 ministers of agriculture <u>have begun</u> working on the reform of the CAP.
	Los 27 ministros de agricultura han empezado la reforma de la PAC.
5	and now Europe. Certain media <u>have renamed</u> him the new Forrest Gump.
	y ahora Europa. Algunos medios lo han bautizado como el nuevo Forrest Gump.
6	The cogs driving Europe 's economy <u>have seized</u> up, slowing it down significantly .
	La maquinaria que impulsa la economía europea ha aminorado su marcha.
7	you 've had a British general - who <u>'ve headed</u> up the so-called EU Military Staff.
	hemos tenido un general británico que ahora lidera el llamado Estado Mayor.
8	They have to give you the money you <u>'ve built</u> up. They will pay you this pension.
	Le darán el dinero que haya cotizado. Le pagarán la pensión.
9	so the Greeks <u>have put</u> up a physical barrier.
	y los griegos han optado por levantar un muro.
10	They are not the only ones who <u>have extended</u> their stay beyond three months.
	No son los únicos que han prolongado su estancia más allá de los tres meses.

Como se puede observar la versión alineada en español no copia la forma de segmentar los subtítulos en la versión en inglés.

g.2. SEPARACIÓN DE TO + INFINITIVE

En esta subcategoría encontramos un número relativamente bajo de casos y por lo que veremos seguidamente, por la densidad del subtítulo, exceptuando algunos casos, parece una tarea especialmente complicada decidir por dónde segmentar el subtítulo.

No	Solution 1 to 50 Page 1 / 12
1	I hope companies are starting <u>to own</u> up more to their responsibility.
	Espero que las empresas empiecen a admitir sus responsabilidades.
2	We have failed to group together and <u>to be</u> commercially aggressive. That 's true.
	Nos ha faltado agruparnos y ser más agresivos en la comercialización.
3	It 's misguided . There 's no reason <u>to look</u> for someone to blame at EU level.
	Está equivocad . No hay motivos para buscar culpables a nivel de la UE.
4	But next year voters will also be helping <u>to choose</u> the next EU Commission President,
	Pero en 2014 , los votantes también ayudarán a elegir el presidente de la CE
5	Many consumers want <u>to avoid</u> long animal transport.
	Muchos consumidores quieren evitar transportes largos.
6	Consumer tsar — the Parliament wants <u>to defend</u> consumers by creating a new role.
	El zar del consumidor: un nuevo cargo para defender sus derechos .
7	And if the companies continue <u>to put</u> on the breaks despite all our efforts ,
	Y si las empresas siguen poniendo obstáculos a pesar de todo ,
8	Germany and Greece — you are about <u>to visit</u> Greece before the end of the year.
	Alemania y Grecia. Usted visitará Grecia antes de que acabe el año.
9	who are working on a project <u>to harmonise</u> the quality of care in Europe .
	que tienen un proyecto para armonizar la calidad de la atención en Europa .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 12
10	Slovenia managed to push it back in 10 days . Eslovenia logró rechazarlo en 10 días .

El *frequency breakdown* por su parte, nos muestra que *to* seguido de verbo es el que mayor casos tiene:

No.	Query result	No. of occurrences	Percent
1	TO VV	547	93.19%
2	TO VB	22	3.75%
3	TO VH	18	3.07%

Vemos a continuación las diez primeras entradas del caso de separación de *to* más infinitivo:

No	Solution 1 to 50 Page 1 / 11
1	I hope companies are starting to own up more to their responsibility. Espero que las empresas empiecen a admitir sus responsabilidades.
2	It 's misguided . There 's no reason to look for someone to blame at EU level . Está equivocada. No hay motivos para buscar culpables a nivel de la UE.
3	But next year voters will also be helping to choose the next EU Commission President , Pero en 2014 , los votantes también ayudarán a elegir el presidente de la CE
4	Many consumers want to avoid long animal transport. Muchos consumidores quieren evitar transportes largos.
5	Consumer tsar — the Parliament wants to defend consumers by creating a new role. El zar del consumidor: un nuevo cargo para defender sus derechos .
6	And if the companies continue to put on the breaks despite all our efforts, Y si las empresas siguen poniendo obstáculos a pesar de todo,

(Continúa en la página siguiente)

(Continúa de la página anterior)

7	Germany and Greece — you are about <u>to visit</u> Greece before the end of the year.
	Alemania y Grecia. Usted visitará Grecia antes de que acabe el año.
8	who are working on a project <u>to harmonise</u> the quality of care in Europe.
	que tienen un proyecto para armonizar la calidad de la atención en Europa.
9	Slovenia managed <u>to push</u> it back in 10 days.
	Eslovenia logró rechazarlo en 10 días.
10	police forces apply all measures <u>to dissuade</u> drivers from breaking the law.
	Son las medidas que garantizan el respeto de las normas de tráfico.

g.3. SUBCATEGORÍA DE VERBO COMPUESTO: PHRASAL VERB

Por su parte, en esta subcategoría en la que incluimos tanto los casos de verbos con preposición como de *Phrasal Verb*, la búsqueda nos da como resultado un número de casos muy elevado, siendo esta la categoría con más ocurrencias. En la siguiente tabla podemos ver las diez primeras entradas de la búsqueda:

No	Solution 1 to 50 Page 1 / 176
1	Hello. Next week the MEPs will <u>be in</u> Strasbourg for two plenaries.
	Hola. La próxima semana los eurodiputados se reunirán en Estrasburgo en dos plenos.
2	all breast implants would be <u>tested in</u> unannounced spot checks ,
	obligan a analizar todos los implantes en controles sorpresa
3	because until we <u>accept that</u> Mediterranean affairs
	porque hasta que no reconozcamos que los asuntos mediterráneos
4	with university degrees being <u>recognised in</u> all countries .
	para el reconocimiento de títulos universitarios en todos los países.
5	They do n't <u>realise that in</u> this world nothing is easy.
	No se dan cuenta de que no hay nada fácil en este mundo.

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 176
6	On the programme is the life and <u>works</u> <u>of</u> the film-maker ,
	(no alignment found)
7	but MEPs <u>proved</u> <u>that</u> even after 100 years
	pero tras 100 años de celebraciones ,
8	and France only grows and <u>progresses</u> <u>through</u> this openness to the world
	y Francia solo puede crecer y progresar abriéndose al mundo,
9	this packaging has been carefully <u>planned</u> <u>at</u> the opposite end of the food chain.
	la industria alimentaria diseñó concienzudamente este envase.
10	to use up food surpluses <u>produced</u> <u>under</u> the Common Agricultural Policy.
	para utilizar los excedentes producidos según la Política Agraria Común.

Por su parte el *frequency breakdown* por POS muestra 52 combinaciones de tiempo verbal con preposición o adverbio. Las primeras 20 entradas las recogemos en la tabla que aparece a continuación:

No.	Query result	No. of occurrences	Percent
1	<u>VVN IN</u>	3343	38.12%
2	<u>VV IN</u>	1868	21.3%
3	<u>VVG IN</u>	875	9.98%
4	<u>VVP IN</u>	618	7.05%
5	<u>VVD IN</u>	433	4.94%
6	<u>VVZ IN</u>	402	4.58%
7	<u>VVN RP IN</u>	145	1.65%
8	<u>VB IN</u>	118	1.35%
9	<u>VV IN IN</u>	104	1.19%
10	<u>VBZ IN</u>	101	1.15%
11	<u>VV RP IN</u>	97	1.11%
12	<u>VVN IN IN</u>	76	0.87%
13	<u>VBP IN</u>	55	0.63%
14	<u>VBD IN</u>	52	0.59%
15	<u>VVD RP IN</u>	46	0.52%
16	<u>VVP IN IN</u>	43	0.49%
17	<u>VV RP</u>	38	0.43%

(Continúa en la página siguiente)

(Continúa de la página anterior)

No.	Query result	No. of occurrences	Percent
18	VHP IN	33	0.38%
19	VBZ IN IN	32	0.36%
20	VVG RP IN	31	0.35%

El verbo principal (en los distintos tiempos verbales) seguido de preposición es la segmentación inadecuada con mayor número de casos. Las seis primeras concordancias recogen 7539 casos de los 8770 casos de segmentación inadecuada. En séptimo lugar, es cuando aparece la estructura: verbo + adverbio+ preposición.

A continuación, mostramos dos tablas: la primera con los diez subtítulos que aparecen en la primera combinación verbo en pasado más preposición y la segunda con los ejemplos de la separación de alguna de los elementos del *phrasal verb* que encontramos en la séptima entrada y que corresponden a [VVN RP IN](#).

No	Solution 1 to 50 Page 1 / 67
1	all breast implants would be tested in unannounced spot checks , obligan a analizar todos los implantes en controles sorpresa
2	with university degrees being recognised in all countries . para el reconocimiento de títulos universitarios en todos los países .
3	this packaging has been carefully planned at the opposite end of the food chain . la industria alimentaria diseñó concienzudamente este envase .
4	to use up food surpluses produced under the Common Agricultural Policy . para utilizar los excedentes producidos según la Política Agraria Común .
5	We are absolutely united that Assad has to go Estamos todos de acuerdo en que Assad debe irse
6	in the report that will be published in a few days , is the creation of a mediator . en el informe que se publicará dentro de unos días , es la creación de un mediador .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 67
7	This Fukushima scenario is not <u>covered</u> <u>in</u> our catastrophe plans. Lo ocurrido en Fukushima no se contempla en nuestros planes de emergencia.
8	I 've been rather <u>encouraged</u> <u>by</u> all the meetings I 've had . He salido reconfortado de todas las reuniones.
9	Are you saying that Italy has <u>acted</u> <u>while</u> waiting for a European answer? ¿Que mientras se espera una respuesta europea Italia toma medidas ?
10	I think it should be <u>completed</u> <u>with</u> an initiative for growth. Habría que completarlo con una iniciativa en favor del crecimiento.

No	Solution 1 to 50 Page 1 / 3
1	As a result, it 's <u>marked out</u> <u>by</u> its dependence on external energy . Por lo tanto, depende de la energía proveniente de fuentes exteriores .
2	is <u>swallowed up</u> <u>by</u> the restless waters beneath it ? sea absorbida por las aguas incansables que la rodean ?
3	But negotiations have <u>dragged on</u> <u>since</u> 2001. Pero las negociaciones duran desde 2001.
4	The environment of trust <u>built up</u> <u>by</u> all these measures brings results. (no alignment found)
5	that people are really <u>fed up</u> <u>with</u> this austerity programme la población está harta de este programa de austeridad
6	These are tests that are <u>carried out</u> <u>on</u> suspicious samples. Estas pruebas se realizan en muestras sospechosas.

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 3
7	because wine growing is <u>made up of</u> so many micro-problems
	porque la viticultura tiene tantos problemas
8	We cannot allow that this is all <u>taken over by</u> the Member States ,
	No podemos permitir que los Estados miembros lo monopolicen todo ,
9	In Europe, financing is mainly <u>carried out by</u> the banks .
	En Europa, la financiación la realizan principalmente los bancos .
10	is a text that was <u>drawn up by</u> the ILO in 2006
	es un texto que redactó la OMT en 2006

Por una parte, la primera tabla sí recoge varios ejemplos de verbos con preposición que no habrían de ser segmentados. Sin embargo, la segunda tabla, recoge muchos casos de verbos con preposición seguidos de preposición o partícula que realmente sí están bien segmentados. Por ejemplo, en la última entrada, la segmentación se produce detrás de *drawn up*, pero como le sigue la preposición *by*, el patrón de búsqueda que hemos generado entiende que la segmentación es incorrecta.

Si aplicamos una búsqueda en la que en lugar de incluir la categoría preposición solo incluimos la categoría que *Treetagger* denomina partícula, obtenemos un resultado mucho más exacto.

```
[pos="VV.*|VB.*|VH.*|VD.*"]</line><line_no="2">[pos="RP+"]*[pos="RP+"]+;
```

Los casos que encontramos en esta ocasión son 109:

No	Solution 1 to 50 Page 1 / 3
1	The fact that the Turkish plane was <u>shot down</u> , Turkey being a member of NATO ,
	El hecho de que derribaran un avión turco y Turquía sea miembro de la ONU ,
2	but the governments want to <u>allow up</u> to 600% of the monthly salary .
	pero los gobiernos quieren permitir hasta 600% del sueldo mensual .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 3
3	Smoking-related diseases are said to kill up to 700,000 people a year in the EU
	Las enfermedades asociadas al tabaco se cobran hasta 700000 vidas al año en la UE
4	and another 500,000 will be waiting out of the million we have .
	y nos faltarán otros 500000 por colocar del millón que tenemos ahora .
5	I 'm currently touring around the countries of Europe ,
	(no alignment found)
6	but a common home which we are building up together , Mr Viktor Orbán .
	(no alignment found)
7	Deirdre, I 'm conscious we 're running out of time , but I want to bring in Marian .
	(no alignment found)
8	Another issue: do we ask farmers to take out insurance to ensure their income?
	Otro asunto: ¿pedimos al agricultor que garantice sus ingresos con un seguro?
9	It 's clearly a useful candidacy for calming down the British Euro-sceptics .
	Sería una candidatura ideal para calmar a los euroescépticos británicos.
10	We know what these countries went through , politically and economically ,
	Sabemos lo que estos países sufrieron a nivel económico y político

Como podemos observar, la revisión manual nos permite depurar posibles errores que se producen al automatizar el análisis de los datos y nos impulsa, en próximos estudios, a redefinir algunas búsquedas o a utilizar métodos de análisis paralelos.

11.2.2.3.2. Análisis de la segmentación en inglés entre subtítulos: segmentación intersubtítulo.

En este apartado, queremos dar cuenta de la segmentación entre subtítulos; a pesar de que, como indican Díaz Cintas y Remael (2007), en este tipo de segmentación se aplican las mismas reglas que en la segmentación intrasubtítulo, lo cierto es que el número de casos que

revela nuestro análisis es inferior y parece ser que de alguna manera la segmentación aquí está más cuidada.

Recordemos el resultado del análisis de los datos de este tipo de segmentación que aportábamos al principio de este apartado.

Tabla 50. Resultado del análisis de segmentación intersubtítulo en inglés.

class	feature	freq	total	rel_freq
CONJ	Conj	4	85643	0,046706
DET	Det_Dem	6	85643	0,070058
	Det_Pos	5	85643	0,058382
NOM_COMP	Nom_Comp	532	85643	6,211833
PREP	Prep	172	85643	2,008337
SINT	Sintag_Adj	9	85643	0,105087
	Sintag_Adv	136	85643	1,587987
VERB_COMP	Verb_Aux	6	85643	0,070058
	Verb_Inf	3	85643	0,035029
	Verb_Phrasal	2039	85643	23,80813

Vemos cómo el porcentaje de segmentación inadecuada es significativamente menor respecto a la segmentación intrasubtítulos, pero coinciden las categorías con más casos en ambos tipos de segmentación. Como podemos observar, la categoría de *phrasal verb* es la que más casos presenta en ambos tipos, pero dado que hemos detectado errores en la formulación de la búsqueda de esta categoría, revisaremos concienzudamente este apartado.

A continuación, y en consonancia con la descripción de los resultados de nuestro análisis para la segmentación intrasubtítulo, mostraremos ejemplos de segmentación inadecuada en nuestro corpus de cada una de las categorías. Recordemos también que mostraremos los resultados de las búsquedas por orden aleatorio de aparición (siempre las diez primeras) y alineadas con el subtítulo en español para poder analizar si la segmentación inadecuada en la lengua pivote (inglés) influye también en la forma de segmentar en español.

a. CATEGORÍA CONJUNCIÓN

Los resultados de la búsqueda desvelan que son únicamente cuatro los casos encontrados. Los mostramos y analizamos seguidamente:

No	Solution 1 to 4 Page 1 / 1
1	All you need is health insurance , and Tan solo necesita un seguro médico

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 4 Page 1 / 1
2	In large quantities , it sticks to artery walls <u>and</u>
	En grandes cantidades , se adhiere a la pared de las arterias
3	Meeting the targets set out in the treaty will mean an additional € 6 billion <u>plus</u>
	Cumplir los objetivos fijados en el tratado requerirá 6000 millones más
4	I think it 's got to be a mixture of <u>both</u>
	Creo que debe favorecer a ambos ,

De los cuatro casos que aparecen *plus* y *both* no serían incorrectos puesto que *plus* actúa como adjetivo y *both* aquí no actúa como conjunción sino como determinante. Por lo tanto se reduciría a dos el número de casos en el total del corpus.

Comprobamos, asimismo, la segmentación de esos subtítulos en EMPAC_EN y observamos que en español la segmentación no mantiene coloca la conjunción en el siguiente subtítulo.

b. CATEGORÍA DETERMINANTE

Antes de comenzar con la presentación de los resultados, recordemos que la categoría determinante contiene cuatro subcategorías (artículos indefinidos, artículos definidos, demostrativos y posesivos). En las subcategorías artículos indefinidos y artículos definidos no hay casos de segmentación intersubtítulos inadecuada. En el caso de las otras dos subcategorías (demostrativos y posesivos), tan solo hay 6 casos de demostrativos y 5 casos de posesivos que pasamos mostramos y analizamos en las siguientes tablas:

No	Solution 1 to 6 Page 1 / 1
1	An agreement has been signed by the Coalition parties saying <u>that</u> <u>money</u> from the Regional Funds will make up for the difference in CAP.
	Los partidos de la coalición han firmado un acuerdo que estipula que el dinero de los Fondos Regionales compensará la diferencia de la PAC .
2	What you do n't want is the net effect of all <u>this</u> <u>being</u> a reduction in the attractiveness of the industry
	No es deseable que el efecto final de todo esto sea una reducción del atractivo del sector

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 6 Page 1 / 1
3	What you do n't want is the net effect of all this being a reduction in the attractiveness of the industry
	No nos conviene que todo esto haga perder atractivo al sector
4	Harmonisation is better for business . But while some countries oppose this MEPs ' proposal will remain in cold storage and paper work will pile up .
	La armonización facilita el comercio . Pero mientras se opongán los estados , la propuesta del PE seguirá en punto muerto y aumentará la burocracia .
5	This MEP thinks that despite all this Hungary 's priorities reflect continuity with the Belgian presidency .
	Este eurodiputado belga considera que las prioridades de Hungría dan continuidad a la presidencia belga .
6	This MEP thinks that despite all this Hungary 's priorities reflect continuity with the Belgian presidency .
	Este eurodiputado belga considera que las prioridades de Hungría dan continuidad a la presidencia belga .

Apreciamos que, en el primer caso, *that* actúa como pronombre relativo en lugar de determinante. La segunda y tercera entradas están repetidas y en este caso *this* actúa como pronombre, no como determinante y tampoco se considera causa de segmentación inadecuada. Ocurre lo mismo en la cuarta, quinta y sexta entrada, que también están repetidas. Por tanto, nuestro corpus no contiene segmentación inadecuada en esta categoría.

En el caso de los adjetivos posesivos, cuando estos van seguidos de pronombre relativo o adverbio o son pronombre de objeto, no se puede considerar casos de segmentación inadecuada. Veamos los ejemplos concretos:

(Continúa en la página siguiente)

No	Solution 1 to 5 Page 1 / 1
1	<p>This Nigerian woman found the courage to report those that exploited her who will be brought before the French courts to face heavy sentences .</p> <p>Esta mujer nigeriana tuvo el valor de denunciar a sus explotadores , que se enfrentan a penas muy duras impuestas por la justicia francesa .</p>
2	<p>But we ‘ve made it so that she takes the right to protection with her wherever she goes . So Europe is moving .</p> <p>Lo hemos cambiado de forma que mantiene el derecho de protección a donde quiera que vaya . Europa avanza .</p>
3	<p>that with a high representative and a service that is loyal to her Europe will speak with a single voice when it comes to diplomacy .</p> <p>que con una alta representante y un servicio que le sea leal Europa hable con una sola voz en cuestiones diplomáticas .</p>
4	<p>This Nigerian woman found the courage to report those that exploited her who will be be brought before the French courts to face heavy sentences .</p> <p>Esta mujer nigeriana tuvo el valor de denunciar a sus explotadores , que se enfrentan a penas muy duras impuestas por la justicia francesa .</p>
5	<p>that with a high representative and a service that is loyal to her Europe will speak with a single voice when it comes to diplomacy .</p> <p>que con una alta representante y un servicio que le sea leal Europa hable con una sola voz en cuestiones diplomáticas .</p>

Por tanto, como evidencian los casos concretos de la tabla anterior, no consideramos que estos sean casos de segmentación inadecuada.

c. CATEGORÍA NOMBRE COMPUESTO

En este caso, y a diferencia de las categorías anteriores, por cuestiones de espacio hemos decidido no alinear el texto con la versión en español y centrarnos en explicar el resultado de la búsqueda.

No	Solution 1 to 50 Page 1 / 11
1	Some MEPs have raised gender equality issues others want the nominations to be geographically balanced .
2	of this law in the Member States lead to a problem with regard to the internet environment ,
3	to award the European Parliament 's Sakharov Prize to Ms Malala Yousafzai Mr President . Honourable Members .
4	The report by MEPs recommends that foods involving nano-technology face strict testing .
5	But some of the people that were in the clip talk about the difficulty of accessing finance for their business .
6	IRAQ INSTABILITY Iraq is in the depths of its own severe and bloody crisis .
7	so obviously we 're also expecting business expansion in the EU thanks to the growth of car sales .
8	THE NEWS IN BRIEF : GAZA AIR STRIKES Israeli missiles continue to rain down on the Palestinian Territories in Gaza .
9	Over the past four decades South Korea has witnessed one of the fastest-ever economic growths .
10	Our answer is a definite no. Jean-Marie Devos , from industry 's point of view ,

Los resultados de esta búsqueda son elevadamente inexactos pues, como podemos comprobar, en las diez primeras entradas encontramos casos que realmente están bien segmentados; sin embargo, aunque técnicamente es correcto, con la información lingüística de la que disponemos, no podemos afinar más esta búsqueda. Necesitaríamos *constituent parsing* o *dependency parsing* para poder precisar de una forma más exacta esos casos;

esto podría constituir un objeto de estudios diferenciado en trabajos futuros pero ahora nos tenemos que limitar a presentarlo de forma extremadamente sintética.

Los resultados de la búsqueda que generamos para esta categoría ofrecen ejemplos en los que la omisión de puntuación (comas al final del subtítulo), como es el caso de la tercera y novena entrada, hace que CQPweb interprete que la segmentación es inadecuada cuando en realidad no lo es. El caso de la segunda, cuarta y quinta entrada en la que una misma palabra puede ser verbo y sustantivo como es el caso de *face* y *talk*, también encontramos que tiene difícil solución. Por su parte, la sexta y octava entrada, son ejemplos de intertítulos que aparecen en pantalla y que tampoco serían casos de segmentación inadecuada, así como la novena entrada, que también es adecuada. El caso que aparece en la entrada número 10, es un caso que encontramos repetidas veces en esta búsqueda y que, obviamente no es un caso de segmentación inadecuada, sino un error de interpretación puesto que *TreeTagger* se ha confundido y ha pensado que *no.* es una abreviatura de *number* en lugar de *no* y *punto final*. En total encontramos casi 50 casos de *no.*

Hemos realizado un minucioso análisis manual de esta búsqueda y, si bien se dan algunos casos de segmentación inadecuada, un porcentaje muy elevado de las 536 no debería considerarse como tal. Por tanto, entendemos que en el futuro habremos de redefinir la búsqueda para reducir sustancialmente los casos detectados de forma incorrecta, como el caso de *no.* y analizar cada caso utilizando *constituent parsing* o *dependency parsing*. Sin embargo, esta es una limitación del presente trabajo por motivos de extensión del mismo y lo dejamos para estudios futuros. Por este motivo, y para no distorsionar los resultados finales, no tendremos en cuenta los datos cuantitativos de esta categoría.

d. CATEGORÍA PREPOSICIÓN

Como apuntamos en la revisión teórica (capítulo 7) y como también consta en las guías de referencia que hemos utilizado en nuestro estudio, es conveniente evitar separar la preposición del sintagma que le precede. Veamos un ejemplo de los 152 casos que encontramos en nuestro corpus:

No	Solution 1 to 14 Page 1 / 1
1	and the migration policy because for countries like Malta , Cyprus or Greece , the burden is too heavy .
	y la política de migración , ya que algunos países como Malta , Chipre o Grecia , deben cargar con todo el peso .
2	Greece was n't an issue then , as attention was focused on Morocco , Algeria and the Canaries ,
	Grecia no era prioritaria en aquel momento , se centraron más bien en Marruecos , Argelia y las islas Canarias ,
3	But Kim Clijsters ' opponent Justine Henin was replaced by Serena Williams , US and world number 1.
	Al final , Kim Clijsters no jugó contra Justine Henin , lesionada , sino contra Serena Williams , número uno del mundo .
4	It 's very important for the EU that Hungary is defending the strengthening of the single currency .
	Es muy importante para la UE que Hungría defienda el refuerzo de la moneda única .
5	It 's very important for the EU that Hungary is defending the strengthening of the single currency .
	Es muy importante para la UE que Hungría defienda el refuerzo de la moneda única .
6	Since 1993 the CE logo has guaranteed that certain imports from China and elsewhere conform to EU health and safety standards .
	Desde 1993 , el logo " CE " garantiza que los productos importados de China y otros países respetan las normas de seguridad de la UE .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 14 Page 1 / 1
7	Since 1993 the CE logo has guaranteed that certain imports from China and elsewhere conform to EU health and safety standards .
	Desde 1993 , el logo “ CE “ garantiza que los productos importados de China y otros países respetan las normas de seguridad de la UE .
8	In the meeting we just had at the Parliament , you did mention that Europol is setting up a counter-terrorism unit to help Member States .
	En la audiencia de hoy en el Parlamento , usted ha afirmado que Europol está creando una unidad antiterrorista para ayudar a los Estados .
9	What is not said often enough is that Greece has hosted a million immigrants for several years already .
	Lo que no suele decirse es que Grecia ha acogido a un millón de inmigrantes desde hace varios años .
10	To understand for example how it could be that Panama was blacklisted by the OECD and now it 's not ,
	Entender , por ejemplo , cómo puede ser que Panamá estuviera en la lista negra de la OCDE y ahora no ,

TreeTagger interpreta que *that* es una preposición y no un pronombre de relativo como es el caso en los ejemplos que mostramos en la tabla, pero en cualquier caso no es conveniente separar el pronombre relativo de la frase que introduce. Por otra parte, la subtitulación al español es bastante más cuidadosa con los pronombres relativos ya que, por lo general, no los coloca al final de la frase. Asimismo, entendemos que hay casos que, por cuestiones de espacio, no se pueden segmentar de otra forma.

e. PARTICIÓN DEL SINTAGMA

Esta categoría, formada por dos subcategorías, partición del sintagma nominal y partición del sintagma adverbial, presenta un número reducido de ocurrencias; 9 en el caso del sintagma adjetival y un poco más, 136 en el caso del sintagma adverbial.

e.1. PARTICIÓN DEL SINTAGMA ADJETIVAL

Veamos a continuación los 9 casos de la primera subcategoría:

No	Solution 1 to 9 Page 1 / 1
1	Since 1 October 2010 , anyone selling a cigarette to a minor risks a prison sentence of six months . Is n't that going a bit too far ?
	Desde el 1 de octubre de 2010 , quien venda cigarrillos a un menor se arriesga a una pena de prisión de seis meses . ¿ No es excesiva ?
2	If you look back , you can see that in the majority of governments in the past Social Democrats and other left-wing groups were members of government
	Mirando atrás , se ve que en la mayoría de los Gobiernos anteriores , los socialdemócratas y otros grupos de izquierdas estaban en el Gobierno
3	That 's another no. Thanks for playing the game .
	No . Gracias por participar en el juego .
4	That 's a no. Returning to the single market , I had several questions left ,
	Entonces , no . Tenía más preguntas sobre el Mercado Único
5	Good news : despite the economic and monetary tempest rocking the eurozone European growth should reach 1% this year , predicts the Commission .
	Buenas noticias : a pesar de la tormenta económica y monetaria en la eurozona el crecimiento en Europa debería ser del 1% este año , según la Comisión .
6	In Copenhagen the negotiations have come to a standstill representatives from Tuvalu compared it to the Titanic .
	En Copenhague , las negociaciones se han paralizado y los representantes de Tuvalu las han comparado con el Titanic .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 9 Page 1 / 1
7	60 days is the maximum ? - That 's <u>the cap.</u> <u>Private</u> businesses may exceed these 60 days with justified exceptions .
	¿ 60 días es el máximo ? - Ese es el límite . Las empresas podrán superar los 60 días de forma excepcional ,
8	for example giving <u>a conservative</u> <u>responsibilities</u> relating to financial services .
	por ejemplo , otorgar a un conservador responsabilidades referentes a los servicios financieros .
9	The revolutionary idea of <u>a supranational</u> <u>common</u> agricultural market , independently financed and controlled ,
	La idea de un mercado agrícola común supranacional , con financiación y control independiente ,

De nuevo, vemos errores que se producen en este tipo de metodologías cuantitativas y que han de ser combinadas con metodologías manuales. En algunos casos *TreeTagger* otorga una categoría gramatical que en ocasiones se confunde con otra, como se refleja en el ejemplo 8.

e.2. PARTICIÓN DEL SINTAGMA ADVERBIAL

Nuestro corpus presenta los siguientes ejemplos de partición del sintagma adverbial:

No	Solution 1 to 50 Page 1 / 3
1	The financial framework for seven years <u>2007-2013</u> <u>was</u> fixed during times of plenty , and now we 're in times of scarcity .
	El marco económico para el septenio 2007-2013 se decidió en época de abundancia , y ahora es tiempo de escasez .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 3
2	which means that the only thing we can do right now is financially support them .
	y lo único que podemos hacer ahora es apoyarlos económicamente .
3	The issue of prior authorisation by a Member State for treatment abroad is particularly tricky where rare diseases are concerned .
	La cuestión de la autorización previa de los tratamientos en el extranjero se complica en el caso de las enfermedades raras .
4	The challenge right now is that we have the structure and the mechanism ;
	Ahora mismo el reto es que tenemos la estructura y el mecanismo ;
5	To discuss this issue I have here Michele LeVoy , director of PICUM ,
	Para ello tenemos con nosotros a Michele LeVoy , directora de PICUM ,
6	This should get the economy moving again provided that financial players re-inject the excess liquidity
	Con eso la economía debería reactivarse . Siempre y cuando los actores financieros reinyecten el exceso de liquidez ,
7	Brussels wants to show that the fall of the Wall 20 years ago influenced the history of Europe and changed the lives of all European citizens .
	Bruselas quiere recordar que la caída del muro ha marcado la historia de Europa y cambiado la vida de los europeos .
8	Turkey 's refusal to recognise the genocide of Armenians 100 years ago was also raised , as was the Cyprus question .
	La negativa de Turquía a reconocer el genocidio armenio de hace 100 años ha salido a relucir , así como el problema chipriota .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 3
9	In my view , the first thing we have to deal with <u>now</u> <u>is</u> not to coordinate sanctions
	En mi opinión , lo primero que debemos hacer ahora es no coordinar sanciones ,
10	Chair , the question of whether it 's a milestone or <u>not</u> <u>has</u> to be discussed in detail .
	Sr. presidente , la cuestión de si se trata de un hito o no debe debatirse a fondo .

f. CATEGORÍA VERBO COMPUESTO

f.1 SUBCATEGORÍA VERBO AUXILIAR

En este caso solo encontramos las siguientes seis muestras:

No	Solution 1 to 6 Page 1 / 1
1	As I said , this 7th EDD with the discussions we 've already <u>had</u> <u>has</u> let us draw some conclusions . The first is , we must be honest ,
	En estas 7ª Jornadas , los debates que ya hemos tenido nos han permitido sacar conclusiones . La primera , seamos sinceros ,
2	I think the problems journalists covering European affairs <u>have</u> <u>are</u> the same in all countries .
	(no alignment found)
3	The sort profiling we have to do <u>is</u> <u>profiling</u> based on intelligence , based upon a random nature
	El tipo de perfiles que tenemos que hacer debe basarse en los datos de inteligencia y ser aleatorios ,
4	As you said , the difficulty we <u>have</u> <u>is</u> that the EU budget is financed by national contributions ,
	Como ha señalado , el problema es que el presupuesto de la UE se financia con contribuciones nacionales

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 6 Page 1 / 1
5	What you see now as the result of the vote we had is that 80% of the money will still go to the 20 biggest farmers .
	La votación dará como resultado que el 80% del dinero se destine a los 20 mayores agricultores .
6	For me the main tasks of the period ahead us are reinforcing internal market integration
	Para mí , los principales objetivos durante esta legislatura son reforzar la integración del mercado interior

Observamos de nuevo que, en la mayor parte de los casos, los casos de segmentación que hemos detectado con la búsqueda no son casos de segmentación inadecuada.

f.2. SUBCATEGORÍA SEPARACIÓN DE TO+INFINITIVO

Tan solo detectamos tres casos, pero los dos primeros están repetidos por lo que reduciremos el número de ocurrencias a dos:

No	Solution 1 to 3 Page 1 / 1
1	Now you do n't have to go to the exchange bureau at the airport , you can exchange currency in the shop .
	Ahora ya no es necesario ir a la oficina de cambio del aeropuerto , puedes cambiar en las tiendas .
2	Now you do n't have to go to the exchange bureau at the airport , you can exchange currency in the shop .
	Ahora ya no es necesario ir a la oficina de cambio del aeropuerto , puedes cambiar en las tiendas .
3	I found him little inclined to develop renewable energies and energy efficiency , which would be essential .
	No le veo muy dispuesto a fomentar la eficiencia energética y las renovables , y eso es crucial .

e.3. SUBCATEGORÍA VERBO CON PREPOSICIÓN Y PHRASAL VERB

Encontramos que en la búsqueda de este tipo de segmentación se dan, lógicamente, los mismos errores que en la segmentación intrasubtítulo. Veamos la siguiente tabla para ilustrarlo:

No	Solution 1 to 50 Page 1 / 41
1	A big problem in Europe <u>is that</u> many people suffer from poverty although they have jobs .
	Un gran problema en Europa es que muchos trabajadores se encuentran en situación de pobreza .
2	This is very rare but in 2009 , no agreement could be <u>reached on</u> the working time directive .
	Esto no ocurre con frecuencia , pero en 2009 no se llegó a un acuerdo sobre la directiva de ordenación del tiempo de trabajo .
3	As a result , the current directive is still <u>enforced until</u> the Commission makes another proposal to reopen the debate .
	Como resultado , se seguirá aplicando la directiva actual hasta que la Comisión presente otra propuesta que reabra el debate .
4	According to the new status , travel expenses will from now on be <u>reimbursed on</u> the basis of actual expenses .
	Según el nuevo estatuto , los gastos de desplazamiento se reembolsarán de acuerdo con la cantidad pagada .
5	As our expert explains , the national influence <u>disappears after</u> the approval by the European Council and the EP .
	Como nos explica este experto , los parlamentos pierden su derecho una vez el Consejo y el Parlamento aprueban la ley .
6	The president 's power <u>grew after</u> the introduction of universal suffrage to appoint the European Parliament ,
	Su poder se ha ampliado desde que se introdujo el sufragio universal para el Parlamento

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 41
7	The Chinese authorities were n't too pleased with the Sakharov prize going to imprisoned Chinese dissident Hu-Jia .
	A las autoridades chinas tampoco les sentó bien que se concediera el premio Sájarov al disidente chino Hu-Jia .
8	We as the European Union did n't contribute by sending soldiers under European command .
	La Unión Europea no participó con un envío de soldados bajo mando europeo .
9	Cooperation is key and must be encouraged at Justice level , with Eurojust , and at police level , with Europol .
	Debemos fomentar la colaboración a nivel judicial , con Eurojust , y a nivel policial , con Europol .
10	The movement against genetic engineering in Europe has ensured that the only genetically modified corn let it was stopped by many countries .
	La oposición a la manipulación genética en Europa garantizó que muchos países prohibieran el único cultivo de maíz modificado permitido .

Si cambiamos ligeramente el patrón de la búsqueda y lo dejamos como en el caso de la segmentación intrasubtítulo obtenemos tan solo cuatro entradas:

No	Solution 1 to 4 Page 1 / 1
1	So if you want to be in the lead , you must have something to show up until the end .
	Para liderar , hay que tener algo que demostrar hasta el final .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 4 Page 1 / 1
2	Flavours like menthol or vanilla will be banned along with such claims as ‘ mild ‘ and ‘ low tar ‘ ,
	Los sabores como mentol y vainilla estarán prohibidos y también distintivos como ‘ suave ‘ y ‘ bajo en alquitrán ‘ ,
3	for speaking mainly in English , was also questioned on how and what the plan was for making research competitive .
	por hablar siempre en inglés , también le han preguntado qué plan tiene para aumentar la competitividad de la investigación .
4	saying they ‘ll hold their own elections , leaving the Parliament 27 members shy out of a total of 450 seats available .
	afirman que celebrarán sus propias elecciones , por lo que en el Parlamento faltarán 27 escaños de 450.

Sin embargo, hay casos de verbo con preposición que, aunque no recogemos en esta búsqueda, si suponen casos de segmentación inadecuada, pero dejaremos esta categoría para un estudio en profundidad en futuros trabajos de investigación.

11.2.2.3.3. Análisis de la segmentación en español entre líneas de subtítulo: segmentación intrasubtítulo

Para la versión en español, mostramos en primer lugar, dos tablas con el número de casos y el porcentaje correspondiente de segmentación inadecuada en los casos de segmentación intrasubtítulos e intersubtítulos en la versión en español según las categorías propuestas para este análisis. Y seguidamente, exponemos ejemplos de los subtítulos con segmentación inadecuada por categorías que hemos localizado en nuestro corpus.

Los resultados obtenidos del análisis de la segmentación intrasubtítulo en español son los siguientes:

Tabla 51. Resultados análisis segmentación intrasubtítulo en español.

class	feature	freq	total	rel_freq
CONJ	Conj	2534	184485	13,73553
DET	Det_Def	940	184485	5,095265
	Det_Dem	208	184485	1,127463
	Det_Indef	244	184485	1,322601
	Det_Pos	192	184485	1,040735
FRASE	Frase	7159	184485	38,80532
NOM_COMP	Nom_Compuesto	1620	184485	8,781202
PREP	Prep	2142	184485	11,6107
SINT	Sintag_Adj	3490	184485	18,91753
	Sintag_Adv	1849	184485	10,0225
VERB_COMP	Verb_Comp	404	184485	2,18988
	Verb_Períf_1	3270	184485	17,72502
	Verb_Períf_2	1181	184485	6,401604
	Verb_PredNom	826	184485	4,477329

Por otra parte, los resultados obtenidos análisis de la segmentación intersubtítulo en español quedan de la siguiente forma:

Tabla 52. Resultados análisis segmentación intersubtítulo en español.

class	feature	freq	total	rel_freq
CONJ	Conj	83	95514	0,868983
DET	Det_Dem	2	95514	0,020939
	Det_Pos	1	95514	0,01047
NOM_COMP	Nom_Compuesto	1549	95514	16,21752
PREP	Prep	18	95514	0,188454
SINT	Sintag_Adj	96	95514	1,005088
VERB_COMP	Verb_Comp	15	95514	0,157045
	Verb_Períf_1	294	95514	3,078083
	Verb_Períf_2	179	95514	1,874071
	Verb_PredNom	20	95514	0,209393

Procedemos ahora al análisis de cada una de las categorías propuestas en nuestra investigación.

a. CATEGORÍA CONJUNCIÓN

En la siguiente tabla, y siguiendo el mismo patrón que el utilizado en la visualización de los resultados en inglés, mostramos las diez primeras entradas que recoge CQPweb cuando buscamos este tipo de segmentación inadecuada.

No	Solution 1 to 50 Page 1 / 51
1	<p>Hay unas negociaciones a las <u>que</u> <u>no</u> acudirán los rebeldes , y lo entiendo .</p> <p>There are supposedly talks going on which the rebels won't attend , and I understand .</p>
2	<p>Sigue en pie porque lo <u>que</u> <u>pretendo</u> antes que nada</p> <p>I 'm running because I have a strong desire and ambition</p>
3	<p>En el futuro , los Estados no tendrán <u>que</u> <u>actuar</u> solos para rescatar a los bancos .</p> <p>In future , States should no longer have to act alone to bail out the financial sector .</p>
4	<p>La Unión hace lo <u>que</u> <u>estos</u> 16 Estados deciden .</p> <p>The Bund actually does what all these 16 countries decide .</p>
5	<p>es que las reglas que tenían y <u>que</u> <u>nosotros</u> tenemos no están a la altura .</p> <p>is that the rules they had and we currently have are n't fit for purpose .</p>
6	<p>pero sabemos perfectamente <u>que</u> <u>necesitamos</u> un año , más o menos ,</p> <p>as time goes on , knowing full well it 'll still take up to a year or thereabouts</p>
7	<p>Estamos casi terminando , así <u>que</u> <u>brevemente</u> , expongan sus conclusiones .</p> <p>It 's nearly the end so I 'll ask you to conclude briefly .</p>
8	<p>Una última pregunta : ¿ Cree <u>que</u> <u>la</u> izquierda perderá frente a la derecha</p> <p>One last question . Do you think the left will lose out to the radical right</p>
9	<p>hay eurodiputados que temen <u>que</u> <u>muchos</u> proyectos no puedan terminarse .</p> <p>MEPs fear , however , that many projects cannot be brought to an end .</p>
10	<p>Si no , no habrá crecimiento , <u>aunque</u> <u>hubiera</u> un espejismo de crecimiento .</p> <p>Without that there will be no growth , although there 'd be an illusion of growth .</p>

El *frequency breakdown* nos permite ver la locución conjuntiva *que* (CQUE) y las conjunciones que introducen una subordinada, son las que más casos acumulan:

Nº	Resultado búsquedas	Número ocurrencias	Porcentaje
1	COUE ART	392	15,47%
2	COUE VLfin	322	12,71%
3	CSUBI VLinf	216	8,52%
4	CSUBX ART	177	6,99%
5	COUE SE	120	4,74%
6	CSUBX VLfin	92	3,63%
7	COUE NEG	77	3,04%
8	COUE VHfin	64	2,53%
9	COUE PREP	53	2,09%
10	COUE VLinf	48	1,89%

Sirva de ejemplo de este *frequency breakdown* los casos de *que* seguido de artículo que recoge la primera entrada de la tabla anterior:

No	Solution 1 to 50 Page 1 / 8
1	Una última pregunta : ¿ Cree que la izquierda perderá frente a la derecha
	One last question. Do you think the left will lose out to the radical right
2	Según Eurobarómetro , el 73% opina que las instituciones europeas son corruptas .
	Eurobarometer says 73% of Europeans think the European institutions are corrupt .
3	pero no debemos pensar que las ventas por Internet son coser y cantar .
	but it would be a mistake to think that sales 2.0 is child 's play .
4	Todo el mundo está de acuerdo en que la eficiencia es lo más importante
	Everybody could agree that energy efficiency is most important
5	Bélgica es pequeña , por lo que la gente se va a trabajar a otro lugar .
	Belgium is small so people go and work somewhere else .
6	La Convención de Dublín , por la que el país de llegada del inmigrante
	The Dublin Convention , under which the country in which a migrant arrives
7	gano en una semana lo que un trabajador gana al mes en Ucrania .
	I earn as much in a week as a Ukrainian worker on average earns in a month .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 8
8	El Parlamento insistió en <u>que</u> <u>el</u> presidente de la Comisión Europea
	The Parliament insisted on the fact that the President of the European Commission
9	es que permitiría , por ejemplo , <u>que</u> <u>un</u> piloto tuviera ocho horas de descanso
	is that it would allow , for example , a pilot to be on standby for eight hours
10	siempre por cada 100g o 100ml , para <u>que</u> <u>el</u> consumidor pueda comparar productos .
	always per 100g or 100ml , so the consumer can compare the products .

Como podemos observar la forma de segmentar en la versión en inglés no influye en la versión en español.

b. CATEGORÍA DETERMINANTE

En esta categoría distinguimos cuatro tipos de determinantes: artículo definido, artículo indefinido, demostrativo y posesivo.

b.1. ARTÍCULO INDEFINIDO

Las diez primeras ocurrencias de este tipo de segmentación inadecuada son las que siguen

No	Solution 1 to 50 Page 1 / 5
1	Los eurodiputados también quieren <u>una</u> <u>separación</u> entre las autoridades públicas
	MEPs also want a separation between public authorities
2	La caída del Muro de Berlín fue <u>una</u> <u>revolución</u> pacífica de unión de pueblos .
	The fall of the Berlin Wall was a peaceful revolution that united East and West .
3	Los negociadores del PE esperan <u>un</u> <u>acuerdo</u> con los Estados este año .
	EP negotiators hope to reach an agreement with the Member States by the year 's end .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 5
4	Facilitará nuestra tarea , pero es una directiva para la gente , para salvar vidas .
	It will greatly facilitate our work but it 's a directive for the people , to save lives ,
5	La inestabilidad en el Sahel es una amenaza para la zona , como en Mali ,
	Clearly , instability in the Sahel is a threat to countries in the region , like Mali recently ,
6	La información en tiempo real la envía un centro de coordinación de trasplantes .
	The information is relayed in real time by a transplant coordination centre .
7	¿ Han oído hablar de él ? Se trata de un proyecto faraónico de fusión nuclear .
	Have you heard of it? It 's a massive nuclear fusion project .
8	el millón de firmas le da derecho a una reunión de alto nivel en la Comisión .
	your one million signatures will get you a top-level meeting at the Commission .
9	porque su moneda tiene una circulación sumamente limitada
	because their currency is one with an extremely limited circulation
10	Por tanto , lo que vemos es una infraestructura de la revolución industrial
	So what we 're seeing is an entire industrial revolution infrastructure

Por su parte, el *frequency breakdown* por categorías muestras que el artículo al final de subtítulo va seguido de las siguientes categorías gramaticales

Nº	Resultado búsqueda	Número ocurrencias	Porcentaje
1	ART NC	205	84,02%
2	ART ADJ NC	28	11,48%
3	ART NP	8	3,28%
4	ART ORD	2	0,82%
5	ART ADJ ADJ NC	1	0,41%

Ejemplos de artículo indefinido seguido de nombre común es el que más casos recoge. Ilustramos en el siguiente ejemplo las diez primeras entradas con subtítulos que han separado el artículo del sustantivo que le sigue:

No	Solution 1 to 50 Page 1 / 5
1	Los eurodiputados también quieren una separación entre las autoridades públicas MEPs also want a separation between public authorities
2	La caída del Muro de Berlín fue una revolución pacífica de unión de pueblos . The fall of the Berlin Wall was a peaceful revolution that united East and West .
3	Los negociadores del PE esperan un acuerdo con los Estados este año . EP negotiators hope to reach an agreement with the Member States by the year 's end .
4	Facilitará nuestra tarea , pero es una directiva para la gente , para salvar vidas . It will greatly facilitate our work but it 's a directive for the people , to save lives ,
5	La inestabilidad en el Sahel es una amenaza para la zona , como en Mali , Clearly , instability in the Sahel is a threat to countries in the region , like Mali recently ,
6	La información en tiempo real la envía un centro de coordinación de transplantes . The information is relayed in real time by a transplant coordination centre .
7	¿ Han oído hablar de él ? Se trata de un proyecto faraónico de fusión nuclear . Have you heard of it ? It 's a massive nuclear fusion project .
8	el millón de firmas le da derecho a una reunión de alto nivel en la Comisión . your one million signatures will get you a top-level meeting at the Commission .
9	porque su moneda tiene una circulación sumamente limitada because their currency is one with an extremely limited circulation
10	Por tanto , lo que vemos es una infraestructura de la revolución industrial So what we 're seeing is an entire industrial revolution infrastructure

Aunque el porcentaje es reducido, estimamos que son también destacables los casos de artículo indefinido seguido de un adjetivo que precede al sustantivo. Podemos verlos en la siguiente tabla:

No	Solution 1 to 28 Page 1 / 1
1	3000 millones de euros y la apertura de un nuevo capítulo en su proceso de adhesión :
	3 billion euros and the opening of a new chapter in its bid to access the EU ,
2	Pese a su diminuto tamaño , tienen una gran reactividad y pueden ser tóxicas .
	Through their tiny size , they have a great surface reactivity , so they can be toxic .
3	Con este informe , queremos lograr una mayor utilización del transporte inteligente
	What we want to achieve with this report is to get more use of intelligent transport
4	Hemos comprendido que somos una sola moneda y debemos trabajar juntos
	We ‘ve just realised we ‘ve a common currency and must now work together
5	va a presentar los deportes en una importante cadena codificada de Francia ?
	is going to be a sports presenter on a major encrypted channel in France ?
6	¿ No sería mejor hablar ahora de una posible adhesión para solucionar la crisis ?
	Would n’t it now be better to talk about a possible membership to solve the problem ?
7	Cuando el premio LUX coincide con un gran festival , los finalistas no dan abasto .
	As the LUX prize coincides with a big festival , the finalists ‘ schedule is packed .
8	Necesitas contactos y encontrar una buena idea con la que estés a gusto .
	You must develop your network and find the right project for yourself .
9	3000 millones de euros y la apertura de un nuevo capítulo en su proceso de adhesión :
	3 billion euros and the opening of a new chapter in its bid to access the EU ,
10	Hay vastas extensiones dedicadas a una única planta . Las abejas no tienen cabida .
	Huge fields contain a single plant variety . The bees ca n’t cope .

De nuevo, el texto alineado en inglés nos permite constatar que la segmentación en español no toma como ejemplo la segmentación en inglés.

b.2. ARTÍCULO DEFINIDO

El número de casos que encontramos en esta subcategoría es relativamente elevado, un total de 924 casos, que, aunque en el cómputo total no representan más que el 5 por mil de los casos, consideremos es un porcentaje elevado:

No	Solution 1 to 50 Page 1 / 19
1	Para engrasar los engranajes , el Banco Central Europeo tiene un programa To grease the gears and get things moving , the EU Central Bank has a programme
2	No todos están convencidos de que la austeridad devuelva el crecimiento . Not everyone is convinced an austerity straitjacket will restore Europe to growth .
3	Necesitamos más transparencia , que las empresas declaren su estrategia fiscal . We must create more transparency , making companies declare aggressive tax planning .
4	que estalló pocos años después que la burbuja estadounidense de las sub-prime . that burst a few years after the US subprime mortgage bubble burst .
5	preside el Consejo , pero ¿ es el fin de las presidencias rotatorias ? Lo veremos . chairs the Council is this the end of rotating presidencies ? Answers later on .
6	y también debemos pensar en las condiciones de los otros refugiados . and we also have to consider the conditions of the refugees who are not in the camps .
7	En Bélgica solo se habló de las elecciones regionales . In Belgium they only spoke about the regionals held at the same time .
8	Los países europeos por donde pase el Gasoducto Transadriático saldrán ganando . The EU countries the future trans-Adriatic pipeline will cross will benefit .
9	¿ No cree que el problema es que los gitanos no se comprometen en política ? Is part of the problem that the Roma do n't get politically engaged ,

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 19
10	La idea es tener una base común para las grandes empresas , y eso se ha bloqueado .
	The idea is to have a common base for large companies , and that 's blocked .

Por su parte, el *frequency breakdown* por POS arroja los siguientes resultados:

Nº	Resultados búsqueda	Número ocurrencias	Porcentaje
1	ART NC	801	85.21%
2	ART NP	90	9.57%
3	ART ADJ NC	40	4.26%
4	ART ORD	9	0.96%

Por tanto, el artículo definido, al igual que el indefinido, está seguido en una amplia mayoría de casos por un nombre común, en menos ocasiones de un nombre propio y a veces de un adjetivo. En la primera tabla mostramos las diez primeras entradas de artículo seguido de sustantivo (ART NC) y en la segunda, los casos de adjetivo antes del sustantivo.

No	Solution 1 to 50 Page 1 / 17
1	No todos están convencidos de que la austeridad devuelva el crecimiento .
	Not everyone is convinced an austerity straitjacket will restore Europe to growth .
2	Necesitamos más transparencia , que las empresas declaren su estrategia fiscal .
	We must create more transparency , making companies declare aggressive tax planning .
3	que estalló pocos años después que la burbuja estadounidense de las sub-prime .
	that burst a few years after the US subprime mortgage bubble burst .
4	preside el Consejo , pero ¿ es el fin de las presidencias rotatorias ? Lo veremos .
	chairs the Council is this the end of rotating presidencies ? Answers later on .
5	y también debemos pensar en las condiciones de los otros refugiados .
	and we also have to consider the conditions of the refugees who are not in the camps .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 17
6	En Bélgica solo se habló de las elecciones regionales .
	In Belgium they only spoke about the regionals held at the same time .
7	Los países europeos por donde pase el Gasoducto Transadriático saldrán ganando .
	The EU countries the future trans-Adriatic pipeline will cross will benefit .
8	¿ No cree que el problema es que los gitanos no se comprometen en política ?
	Is part of the problem that the Roma do n't get politically engaged ,
9	sobre todo accidentes de coche . En las mujeres , es una muerte de cada diez .
	especially in connection to car accidents . In women , alcohol causes 1 in 10 deaths .
10	No obstante , parece que tanto la eurozona como el FMI ofrecerán ayuda ,
	However , it seems that both the eurozone and the IMF will provide assistance ,

Y seguidamente los casos de artículo + adjetivo + sustantivo (ART ADJ NC):

No	Solution 1 to 40 Page 1 / 1
1	La idea es tener una base común para las grandes empresas , y eso se ha bloqueado .
	The idea is to have a common base for large companies , and that 's blocked .
2	Ahora tienen un año para examinar las presuntas violaciones de la ley europea :
	Now they have one year to examine alleged breaches of EU law .
3	No se puede responder al ascenso de la extrema derecha , izquierda o lo que sea
	(no alignment found)
4	Lady Ashton ha intentado dar las mismas oportunidades a todos los países ,
	Lady Ashton has tried to give all the countries the same opportunities
5	Los conductores de Uber cumplen las mismas normas que los chóferes privados .
	In London the Uber drivers have to comply with the same rules as private hire drivers .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 40 Page 1 / 1
6	en un gran hangar de Lampedusa , con los pequeños féretros blancos en primera fila .
	in a vast hangar in Lampedusa , with tiny white coffins in the first line .
7	ha premiado obras que se centran en las mayores preocupaciones de los europeos .
	has recognised cinematographic works that focus on European citizens ‘ main concerns .
8	Es de suponer que los asesinatos y la posible implicación de su servicio secreto
	Presumably , assassinations and possible links to his secret service
9	Es importante dar mayor alcance a las futuras posibilidades del presupuesto .
	It ‘s important to give further scope for the budget and future possibilities .
10	se quiere obligar a los proveedores de los nuevos combustibles a ser transparentes .
	MEPs want to force the suppliers of these new fuels to be more transparent .

Por similitud con los ejemplos que hemos tomado para la versión en inglés, añadimos aquí la combinación de palabras (artículo + sustantivo) que más veces se ha separado en la versión en español. El *frequency breakdown* por palabra muestra los siguientes resultados:

Nº	Resultado búsquedas	Número ocurrencias	Porcentaje
1	la Comisión	28	3,03%
2	los Estados	16	1,73%
3	los eurodiputados	15	1,62%
4	los derechos	13	1,41%
5	las elecciones	9	0,97%
6	las instituciones	9	0,97%
7	las negociaciones	9	0,97%
8	el Parlamento	8	0,87%
9	la Delegación	8	0,87%
10	los países	7	0,76%

Recordemos que en la versión en inglés *the* seguido de *European* es la que más veces se segmenta inadecuadamente. En el caso del español es *la Comisión*. Veamos las diez primeras entradas:

No	Solution 1 to 28 Page 1 / 1
1	<p>con invitados como el presidente de la Comisión , el Defensor del Pueblo Europeo ,</p> <p>with guests like the President of the Commission , the European Ombudsman ,</p>
2	<p>Europa se ha quedado sin batería y la Comisión ofrece los cables de arranque .</p> <p>Europe needs a kick-start , and today the Commission is applying the jump cables .</p>
3	<p>Queremos que el programa de la Comisión incluya nuestras propuestas .</p> <p>We want to see a lot of our points in the Commission 's programme .</p>
4	<p>que Jean-Claude Juncker , presidente de la Comisión , proponga un plan de inversión ,</p> <p>the Commission President Jean-Claude Juncker rolling out an investment plan</p>
5	<p>Será indispensable la ayuda de la Comisión para reestructurar el sector .</p> <p>Commission aid to allow us to restructure the fishing sector will be indispensable .</p>
6	<p>Tras su adopción por la Comisión de Libertades Civiles ,</p> <p>After its adoption in the Civil Liberties Committee ,</p>
7	<p>y ha optado por la Comisión de Transportes .</p> <p>If Isabelle Durant could choose , she would opt for the Transport Committee .</p>
8	<p>son algunas de las campañas de la comisión del PE dedicada a estos asuntos .</p> <p>some of many campaigns being fought by members of the EP 's dedicated committee .</p>
9	<p>Pero la semana pasada , pese a que la Comisión de Comercio dio su visto bueno ,</p> <p>But last week , despite the Trade Committee giving the green light</p>
10	<p>Cuando Trichet se reunió con la Comisión de Asuntos Económicos</p> <p>When Mr Trichet met the economic affairs committee ,</p>

Podemos apreciar que, de nuevo, la segmentación en inglés no parece tener impacto en la subtitulación al español.

b.3. SUBCATEGORÍA DEMOSTRATIVOS

En esta subcategoría el número de casos es menor que en la anterior y, por lo que hemos podido apreciar en los casos que hemos detectado, se producen generalmente por falta de espacio, aunque no sería el caso de las entradas 4 y 8:

No	Solution 1 to 50 Page 1 / 5
1	Seguro que ha escuchado mucho estos argumentos estos cuatro o cinco años .
	I ‘m sure you ‘ve heard many of those arguments over the past four or five years .
2	Dice que necesitamos esos coordinadores en todos los países .
	You say we need these new coordinators in every country .
3	Es un motor de crecimiento , pero este mercado debe reforzarse y armonizarse .
	It ‘s a driver for growth but a market that needs to be strengthened and harmonised .
4	Espero que hagan estas cosas más a menudo ,
	(no alignment found)
5	La UE anima a los 27 a financiar esta política , aunque no a cualquier precio .
	The EU is encouraging Member States to finance this policy , but not at all costs .
6	Los eurodiputados traducirán esta experiencia en un marco legal
	MEPs will then translate this shared knowledge into a legal framework
7	si queremos resolver este gran problema al que nos enfrentamos .
	Otherwise we will not solve this immense problem that we have .
8	Y supongo que eso forma parte del problema ,
	And I guess that ‘s part of the problem ,
9	Para algunos países , el coste de estas nuevas normas es desproporcionado .
	For some countries , the costs of moving to these new norms are disproportionate .
10	Los socialistas participarán en esta manifestación contra la austeridad
	The S & D will join them in this demonstration against austerity

En cuanto al *frequency breakdown* por POS este nos muestra que, en su mayoría, los determinantes van seguidos de sustantivos como podemos apreciar en la tabla anterior.

Nº	Resultado búsquedas	Número ocurrencias	Porcentaje
1	DM NC	196	94,23%
2	DM ADJ NC	9	4,33%
3	DM NP	2	0,96%
4	DM ORD	1	0,48%

En la siguiente tabla mostramos los escasos casos de segmentación inadecuada de determinante seguido de adjetivo precediendo al sustantivo:

No.	Solution 1 to 9 Page 1 / 1
1	<p>si queremos resolver este gran problema al que nos enfrentamos .</p> <p>Otherwise we will not solve this immense problem that we have .</p>
2	<p>Para algunos países , el coste de estas nuevas normas es desproporcionado .</p> <p>For some countries , the costs of moving to these new norms are disproportionate .</p>
3	<p>Esperamos ganar la votación sobre este importante recurso para el presupuesto .</p> <p>We hope to win the vote on this important own resource for the EU budget .</p>
4	<p>que se han producido en Siria estos últimos días ? Pónganos al día .</p> <p>that have taken place in Syria in the past few days ? Just so we can set the scene .</p>
5	<p>Algunos pasajeros suben a estos modernos trenes cofinanciados por la UE .</p> <p>Some lucky passengers take these modern trains co-financed by EU regional aid .</p>
6	<p>Veremos , con mis invitados , si esta nueva reforma será realmente positiva .</p> <p>We 'll see with my guests if this new reform will really have a positive impact .</p>
7	<p>Estoy segura de que si introducimos estos nuevos tacógrafos se pueden salvar vidas .</p> <p>I 'm confident that the introduction of the new digital tachograph could save lives .</p>
8	<p>Ningún país está a salvo de este peligroso crimen organizado</p> <p>No country is safe from this very serious organised crime</p>

(Continúa en la página siguiente)

(Continúa de la página anterior)

No.	Solution 1 to 9 Page 1 / 1
9	Algunos pasajeros suben a estos modernos trenes cofinanciados por la UE . Some lucky passengers take these modern trains co-financed by EU regional aid .

Aquí sí que la constricción del espacio es determinante en la segmentación del subtítulo, así como de nuevo evidenciamos que la segmentación en inglés no condiciona la segmentación en la versión en español.

b.4. SUBCATEGORÍA POSESIVOS

En primera instancia mostramos las diez primeras entradas de esta subcategoría de segmentación inadecuada:

No	Solution 1 to 50 Page 1 / 4
1	No estamos de acuerdo con sus argumentos contra la reestructuración . He added a new argument against restructuring , which is a problem for us .
2	Intentaremos convencer a nuestros colegas que no apoyen una Comisión (no alignment found)
3	usted defenderá los intereses de su cliente e intentará actuar lo antes posible . you 'll act in the interests of your client , trying to move as quickly as possible .
4	usted defenderá los intereses de su cliente e intentará actuar lo antes posible . you 'll act in the interests of your client , trying to move as quickly as possible .
5	Pedir prestado para pagar a sus proveedores y al personal era el límite . Borrowing to pay for one 's supplies and employees was the limit .
6	Pero Europa , primero , debe solucionar sus problemas sobre la base de este Tratado . But Europe should first solve its problems on the basis of the current Treaty .
7	Problemas acuciantes para la UE son su nuevo régimen de supervisión financiera The EU 's most pressing topics are its new regime of financial supervision

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 4
8	Pero primero quiero que <u>nuestros invitados</u> nos den una primera impresión .
	But first I want a brief feel of the situation as it is on the ground , from our guests .
9	y pide una mayor participación de <u>su presidente</u> en las cumbres europeas
	It wants the president to have more access to European Council meetings
10	Los jóvenes deben quedarse en <u>sus países</u> y cooperar con las instituciones
	Young people must stay in their region and cooperate with the institutions ,

Esta subcategoría es de las cuatro que conforman la categoría determinante la que menos casos recoge. Como hemos visto en la tabla anterior llama mucho la atención cuando se segmenta así un subtítulo. Por otra parte, el *frequency breakdown* muestra cuatro categorías de las que va seguido, en este caso, el posesivo:

Nº	Resultado búsquedas	Número de ocurrencias	Porcentaje
1	<u>PPO NC</u>	176	92,15%
2	<u>PPO ADJ NC</u>	12	6,28%
3	<u>PPO ORD</u>	2	1,05%
4	<u>PPO ADJ ORD</u>	1	0,52%

Los ejemplos de la primera concordancia (PPO NC) quedan recogidos en la primera tabla de esta categoría, y a continuación, ofrecemos ejemplo de segmentación inadecuada de la segunda entrada, es decir, (PPO ADJ NC):

No	Solution 1 to 12 Page 1 / 1
1	Problemas acuciantes para la UE son <u>su nuevo régimen</u> de supervisión financiera
	The EU 's most pressing topics are its new regime of financial supervision
2	Internet se ha convertido en <u>nuestro último espacio</u> de libertad en Argelia
	The internet has really become our single last space of freedom in Algeria
3	pronunciará un discurso inaugural . <u>Su homólogo nigeriano</u> , Muhammadu Buhari ,
	will be making a keynote address , with the President of Nigeria , Muhammadu Buhari ,

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 12 Page 1 / 1
4	Pasamos a nuestro próximo apartado .
	We 'll go to the next part of the programme .
5	Fue muy astuta y todo se debe a sus excelentes habilidades de negociación .
	She was very clever then and it was all down to her excellent negotiation skills .
6	Nos conviene tener nuestro propio sistema europeo .
	It 's good to have our own European system .
7	Pasamos a nuestro siguiente apartado .
	We now come to a different point in the programme .
8	Nos conviene tener nuestro propio sistema europeo .
	(no alignment found)
9	hasta el 15% de nuestra principal fuente de energía .
	to 15% in our primary energy consumption .
10	para hablar de nuestras propias debilidades ,
	(no alignment found)
11	pero antes debo someterla a nuestro habitual minuto de preguntas Sí o No .
	but first I must subject you to our usual minute of Yes or No questions .
12	Llegamos a nuestro último apartado ,
	We come to our last section

Aquí observamos, por una parte, que el problema se agrava a medida que las unidades de sentido están formadas por más palabras, pero es interesante observar el fenómeno en este sentido. En cuanto a la segmentación en inglés y su influencia en la versión española, podemos decir que es, al igual que en el resto de casos, insignificante o nula.

c. CATEGORÍA FRASE

En la versión en inglés de nuestro corpus esta categoría tiene un porcentaje relativamente alto de casos de segmentación inadecuada. En el caso de la versión en español también se da esta circunstancia con 7159 casos. En la siguiente tabla mostramos las diez primeras ocurrencias y posteriormente, a partir del *frequency breakdown*, realizaremos un análisis minucioso de los casos.

No	Solution 1 to 50 Page 1 / 164
1	prosperidad para todos , no para unos pocos . .Es nuestra tarea principal .
	prosperity for all , not great riches for the few , this is our main duty .
2	EDF . .Líderes : Nigel Farage , Francesco Enrico Speroni .
	(no alignment found)
3	EFD ha llegado . .Somos críticos con Europa , somos euroescépticos .
	(no alignment found)
4	CRE . .Líder : Michał Tomasz Kamiński . Miembros : 54 / Países : 8
	(no alignment found)
5	La mayoría proceden de los ministerios . .¿ Por qué ?
	The majority stems from the government ministries . Why is that ?
6	Habrá menos de los 785 que tenemos hoy . .Se elegirán 736.
	We will have less than/IN the 785 MEPs we have today , 736 will be elected .
7	A menudo los eurodiputados votan en la primera lectura . .¿ Por qué ?
	MEPs often vote at first reading . Why does it happen ?
8	Hola . .Soy Petr , de República Checa .
	Hello . I am Petr from the Czech Republic .
9	Hola . .Me llamo Jonathan y vengo de Dinamarca .
	Hi . My name is Jonathan , and I 'm from Denmark .
10	No . .Hay una gran diferencia ...
	(no alignment found)

A partir del *frequency breakdown* por palabras y por POS observamos que había muchos casos que no deberían ser considerados casos de segmentación inadecuada. En la siguiente tabla mostramos los resultados para interpretarlos seguidamente:

Nº	Resultado búsquedas	Número de ocurrencias	Porcentaje
1	.UU_FS NP	502	6,16%
2	.¿_FS FS	333	4,08%
3	.Es_FS VSfin	260	3,19%
4	.No_FS NEG	224	2,75%
5	.Adiós_FS NP	221	2,71%
6	?¿_FS FS	212	2,6%
7	.La_FS ART	157	1,93%
8	.El_FS ART	150	1,84%
9	.¿_FS INT	137	1,68%
10	.._FS CM	124	1,52%
11	?¿_FS INT	108	1,32%
12	.y_FS CC	102	1,25%
13	.Los_FS ART	100	1,23%
14	.Pero_FS CCAD	85	1,04%
15	.Se_FS SE	81	0,99%
16	.En_FS PREP	75	0,92%
17	.Hola_FS NP	74	0,91%
18	.¿_FS FS	74	0,91%
19	.Gracias_FS NC	69	0,85%
20	.Eso_FS DM	67	0,82%

La primera entrada, con 502 casos, es el caso de *EE.UU.* en el que la herramienta entiende que el punto separa dos frases, por lo que estos casos no son casos de segmentación inadecuada.

No	Solution 1 to 50 Page 1 / 11
1	en torno al acuerdo entre la Unión Europea y EE .UU .
	on the agreement between the European Union and the USA
2	el Consejo y los EE .UU . durante un breve periodo de tiempo .
	the Council 's and the US 's court for a very short period of time .
3	entre la Unión Europea , Australia , Canadá , Corea , EE .UU . y otros países .
	between the European Union , Australia , Canada , Korea , the US and other countries .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 11
4	con un banco de inversión de EE .UU . , Grecia ocultó su deuda
	with a US investment bank Greece hid the size of its debt
5	y no solo en Europa , ya que la crisis comenzó en EE .UU .
	not just in Europe as the financial crisis started in the US ,

Así como la décima entrada en la que la herramienta también interpreta que el segundo punto del acrónimo es final de frase y comienzo de otra cuando este va seguido de una coma. En esta ocasión hablamos de 124 casos:

No	Solution 1 to 50 Page 1 / 3
1	con un banco de inversión de EE . UU . Grecia ocultó su deuda
	with a US investment bank Greece hid the size of its debt
2	Durante generaciones , los irlandeses emigraron a EE.UU . Canadá o Australia .
	For generations , the Irish were migrating to the US , Canada , Australia and so on .
3	El problema es que los políticos ... Desde luego en EE . UU .
	Part of the problem is that/IN politicians are to an extent , certainly in the US ,
4	pero no para EE . UU . que importa el 5% del volumen de su demanda de petróleo .
	but not the United States . The US imports only 5% of the volume of their oil needs .
5	si , por ejemplo , Israel decide seguir adelante sin EE . UU .
	if , for instance , Israel decided to go it alone without America ,

Tanto en el caso de la segunda entrada como en el resto, el motor de búsquedas sí que ha detectado correctamente los casos de segmentación inadecuada. Veamos algunos ejemplos a continuación:

No	Solution 1 to 50 Page 1 / 7
1	La mayoría proceden de los ministerios .¿? Por qué ?
	The majority stems from the government ministries . Why is that ?
2	A menudo los eurodiputados votan en la primera lectura .¿? Por qué ?
	MEPs often vote at first reading . Why does it happen ?
3	la directiva no resistirá los argumentos jurídicos .¿? Qué opina ?
	this directive may not withstand the legal arguments . What do you think ?
4	Última pregunta .¿? Qué pasará de aquí a la sesión constitutiva ?
	Final question : from now until the first constitutive session what will happen ?
5	No ha respondido a la pregunta . Barroso no es su candidato .¿? Quién es ?
	You 're not answering my question . Mr Barroso is not your candidate . Who is ?
6	y que mermó la capacidad de la UE para exigir reformas .¿? Sr. Szymański ?
	(no alignment found)
7	Doris Pack .¿? Cerramos las puertas ?
	(no alignment found)
8	Sr. El Midaoui , usted conoce bien la realidad .¿? Está de acuerdo ?
	Mr El Midaoui , do you see that ? Do you have the same impression ?
9	Solemos ver países en desarrollo en los últimos puestos .¿? Por qué ?
	We tend to find the developing countries at the bottom of the list .
10	Nicolas Veron .¿? Europa ha tomado decisiones acertadas ?
	Nicolas Veron . Has Europe taken the right decisions ?

Por una parte, observamos de nuevo que la segmentación en inglés no tiene por qué ser la causa de la segmentación inadecuada en español y por otra parte, la riqueza de datos que hemos encontrado en los resultados de esta búsqueda nos alienta a profundizar en este punto en estudios que realizaremos en el futuro.

d. CATEGORÍA NOMBRE COMPUESTO

Si bien los nombres compuestos no son tan habituales en español como en inglés hemos detectado un número considerable de casos, un total de 1620. Veamos algunos ejemplos en la siguiente tabla:

No	Solution 1 to 50 Page 1 / 33
1	Ni el gobierno ni el partido chino podrán detener este avance .
	It 's impossible for the Chinese government or the Party to stop it .
2	La adhesión de la Antigua República Yugoslava de Macedonia se hará esperar
	The Former Yugoslavian Republic of Macedonia aspires EU-accession ,
3	del Foro Europeo de Juventud Musulmana y otras asociaciones .
	from the Forum of European Muslim Youth and Student Organisations .
4	el ginecólogo congoleño Denis Mukwege ?
	Congolese gynaecologist Dr Denis Mukwege ?
5	2013 AÑO EUROPEO DE LOS CIUDADANOS
	(no alignment found)
6	No es que quieran ser el estado número 51 de la Unión ,
	(no alignment found)
7	Costará olvidar el desastre de Deepwater Horizon en el golfo de México .
	The Deepwater Horizon disaster is still preying on people 's minds .
8	y su viceprimer ministro proeuropeo Nick Clegg dejó perplejos a muchos .
	and his pro-Europe deputy Nick Clegg left many bemused .
9	y las constantes dificultades económicas del sur .
	and constant economic difficulties of the South .
10	como el de los estados miembros del Consejo
	(no alignment found)

Podemos observar como el caso 4 y 8 no son nombres compuestos, sino un nombre común seguido de un nombre propio, NC NP, pero sin que el nombre común esté en mayúsculas y forme parte del nombre de un proyecto o iniciativa.

En cuanto al análisis a partir del *frequency breakdown* por POS obtenemos las siguientes concordancias de categorías:

Nº	Resultado búsquedas	Número ocurrencias	Porcentaje
1	NP NP	645	39,69%
2	NC NP	428	26,34%
3	NC NC	407	25,05%
4	NP NC	144	8,86%
5	NMON NP	1	0,06%

La primera entrada, NP NP, es clara en cuanto a la segmentación inadecuada excepto de nuevo en los casos en los que toda la información aparece en mayúsculas como en la entrada número 6:

No	Solution 1 to 50 Page 1 / 13
1	Croacia , Antigua República Yugoslava de Macedonia y Turquía . (no alignment found)
2	Líder : Joseph Daul Miembros : 265 / Países : 26 (no alignment found)
3	Tras las elecciones al Parlamento Europeo de 2009 , habrá 736. After 2009 and the European Parliament election , they will be 736.
4	En las últimas elecciones , el Partido Popular Europeo obtuvo 265 escaños The results of the last election gave 265 seats to the European People 's Party
5	Se creará el puesto de Alto Representante de Exteriores de la UE . A post of High Representative for Foreign Affairs and Security Policy will be created .
6	NECESITAMOS VIGILANCIA INTRUSIVA , Struan Stevenson . (no alignment found)
7	SIN GUERRAS NO HABRÍA TERRORISMO (no alignment found)

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 13
8	CONTRIBUÍMOS AL ESTADO DE DERECHO , Karl von WOGAU (no alignment found)
9	CRISIS DEL MODELO SOCIAL EUROPEO , Francis Wurtz . (no alignment found)
10	LA COMISIÓN DEBE ACEPTAR EL RECHAZO A LOS OMG

En el *frequency breakdown* de esta entrada observamos que las diez primeras entradas recogen casos de mayúsculas que no son casos de segmentación inadecuada:

Nº	Resultado búsquedas	Número ocurrencias	Porcentaje
1	Parlamento Europeo	41	6,36%
2	MÁS IMPORTANTE	11	1,71%
3	VOTAR EN	11	1,71%
4	Reino Unido	10	1,55%
5	RETO PARA	10	1,55%
6	CONTIN-UE-RÁ EN	9	1,40%
7	República Yugoslava	9	1,40%
8	Consejo Europeo	8	1,24%
9	Izquierda Unitaria	8	1,24%
10	Iniciativa Ciudadana	7	1,09%

Por su parte, la segunda entrada, NC NP, nos permite detectar los casos que ya mencionamos tras analizar la primera entrada de esta categoría:

No	Solution 1 to 50 Page 1 / 9
1	del Centro Europeo de Inteligencia Estratégica y Seguridad . from the European Strategic Intelligence and Security Centre .
2	Me alegra mucho ver a mi amigo Hans-Gert Pöttering I 'm so delighted to see my friend Hans-Gert Pöttering
3	¿ Qué país será el 28º Estado miembro de la UE ? (no alignment found)

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 9
4	Hoy me acompañan en el estudio Paul Rübzig , miembro austriaco del PPE ;
	With me in the studio I have Paul Rübzig , an Austrian member of the EPP ;
5	También a Libor Rouček , de la Alianza Progresista de Socialistas y Demócratas
	We have Libor Rouček of the Progressive Alliance of Socialists and Democrats ,
6	Gracias a mis invitados Jean-Paul Gauzès , Philippe Lamberts ,
	My thanks to my studio guests Jean-Paul Gauzès , Philippe Lamberts ,
7	pero la Carta de los Derechos Fundamentales es bastante reciente .
	but the Charter of Fundamental Rights is a relatively recent thing .
8	En la última sesión debatimos el informe Vălean sobre la ciudadanía europea ,
	We had in the last session the Vălean report on EU citizenship ,
9	Empecemos por el profesor Pierre Piccinin .
	Let 's start with Professor Pierre Piccinin .
10	a mi izquierda , la eurodiputada danesa Anne Jensen del Grupo de los Liberales ;
	On my left , Danish MEP , Anne Jensen of the Liberals Group . Welcome .

e. CATEGORÍA PREPOSICIÓN

La categoría preposición recoge un total de 2142 entradas y es, al igual que en el caso de la versión en inglés, la segunda causa de segmentación inadecuada con mayor número de entradas.

No	Solution 1 to 50 Page 1 / 43
1	Visitó el PE para participar en una audiencia con Sir Paul McCartney
	He visited the EP to take part in a hearing with Sir Paul McCartney ,

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 43
2	y yo le acompañé <u>durante</u> <u>una exposición</u> sobre este tema . I joined him around an exhibition on this topic .
3	sino que crea inseguridad <u>para</u> <u>las personas</u> que poseen armas legales . but creates insecurity for those concerned who legally own arms .
4	Porque existe una lista de motivos <u>para</u> <u>la detención</u> de los solicitantes de asilo , Because there 's still a series of grounds for the detention of asylum seekers ,
5	ya que su economía ha crecido <u>durante</u> <u>la crisis</u> y espero que lo siga haciendo . as its economy has grown throughout the crisis and I hope it will continue to .
6	para acabar con los ensayos <u>con</u> <u>animales</u> y desarrollar nuevos métodos . that aim to stop animal testing and develop new methods .
7	Diana Wallis , eurodiputada reelecta <u>de</u> <u>la ALDE</u> , LIB / DEM de Yorkshire , en el RU . (no alignment found)
8	Incluso en Francia , donde el 70% <u>de</u> <u>la energía</u> procede de centrales nucleares , Even in France , where 70% of the power is generated by nuclear plants ,
9	lanzando más de 80 proyectos <u>de</u> <u>desarrollo</u> y reconstrucción en Somalia . launching more than 80 development and reconstruction projects in Somalia .
10	los eurodiputados votarán <u>sobre</u> <u>el acceso</u> a créditos más económicos MEPs will vote on access to cheaper loans

Como podemos ver en el *frequency breakdown* por etiquetas del POS, se dan 23 posibles combinaciones. Las tres primeras entradas, que recogen el porcentaje más significativo, son las de preposición + artículo + nombre común, y preposición + nombre común + nombre propio. Mostramos a continuación ejemplo de estas entradas y de la quinta entrada, por incluir un adjetivo después de la preposición:

Nº	Resultado búsquedas	Número de ocurrencias	Porcentaje
1	PREPART NC	1031	48,13%
2	PREP NC	509	23,76%
3	PREP NP	167	7,8%
4	PREPART NP	120	5,6%
5	PREPART ADJ	113	5,28%
6	PREP PPO NC	55	2,57%
7	PREPADJ	45	2,1%
8	PREP PREPART NC	28	1,31%
9	PREP DM NC	23	1,07%
10	PREP DM ADJ	6	0,28%

En el caso de preposición seguida de artículo y nombre común, las diez primeras entradas sirven como ejemplo de la segmentación de la preposición y el sintagma que le sigue:

No	Solution 1 to 50 Page 1 / 21
1	Seguiremos luchando contra la discriminación y la desigualdad . (no alignment found)
2	sino que gracias al enorme apoyo de los movimientos verdes de toda Europa , but thanks to the great support of green movements from all over Europe
3	Los resultados aparecen en la página Web del Parlamento Europeo . In this case , take a look at the European Parliament 's website .
4	y los detalles de las votaciones de los informes adoptados en la legislatura . and all the details of the votes on the reports adopted in the legislature .
5	Está claramente estipulado en el reglamento interno del Parlamento : It 's clearly written in the internal regulation of the Parliament :
6	Las leyes contra la discriminación se ampliaron (no alignment found)
7	que los productos no son nocivos para los seres humanos y el medio ambiente . whether products are harmless for humans and the environment .

(Continúa en la página siguiente)

No	Solution 1 to 50 Page 1 / 21
8	Necesitamos el Tratado de Lisboa <u>para el proceso</u> de codecisión automática .
	We need the Lisbon Treaty so that joint decision becomes automatic .
9	La gente piensa que <u>según la configuración</u> del nuevo parlamento
	People are now saying that in consideration of the new face of the EP
10	hay que repartir puestos <u>entre los socios</u> de la coalición .
	then you have to find a place for each of your partners in that coalition .

En el caso de preposición seguida de nombre común sin artículo entre ellos, tenemos 509 casos:

No	Solution 1 to 50 Page 1 / 11
1	para acabar con los ensayos <u>con animales</u> y desarrollar nuevos métodos .
	that aim to stop animal testing and develop new methods .
2	lanzando más de 80 proyectos <u>de desarrollo</u> y reconstrucción en Somalia .
	launching more than 80 development and reconstruction projects in Somalia .
3	Perdemos dinero <u>para servicios</u> sociales , para sanidad .
	This is the money we lose for social services , for health services .
4	No había muchos ánimos en la reunión <u>de eurodiputados</u> ante la cumbre de la UE .
	There was not much cheer as senior MEPs met ahead of the EU summit .
5	La decisión llega <u>tras meses</u> de debates ,
	The decision came after months of heated debate ,
6	A principios de julio , la Oficina <u>de Información</u> del Parlamento en Suecia
	At the beginning of July , the European Parliament Information Office in Sweden

No	Solution 1 to 50 Page 1 / 11
7	pero el sistema informático de control <u>de personas</u> y mercancías aún no funciona .
	But the new IT system to control people and goods is still not operational .
8	Tenemos necesidades <u>de infraestructura</u> , investigación , clima ,
	We have needs in infrastructure , research , climate ,
9	Y repito , tenemos a alguien <u>con experiencia</u> en todo tipo de coaliciones .
	(no alignment found)
10	Los casos más frecuentes son los <u>de gente</u> que pide obtener documentos .

En la siguiente tabla, mostramos los diez primeros casos de la categoría preposición en los que la preposición va seguida también de un adjetivo que modifica al sustantivo. Estos son también casos de segmentación inadecuada, según las recomendaciones que mencionamos en capítulos anteriores, pero pone de manifiesto que si hubiéramos hecho una búsqueda más sencilla no los hubiéramos detectado:

No	Solution 1 to 50 Page 1 / 3
1	Tras la visita , ¿ votará Laurent <u>en las próximas</u> elecciones europeas en 2014 ?
	After this visit , will Laurent vote in the next European election in 2014 ?
2	sigue en el aire , pese <u>a las nuevas</u> leyes respaldadas por el PE
	is still up in the air , despite new laws backed by the EP
3	Vamos a debatir este paquete <u>durante los próximos</u> meses y años .
	Of course we will be discussing this package in the coming months and years .
4	para poder luchar <u>contra la mala</u> educación religiosa
	in order to fight the lack of a religious education
5	en ponerse de acuerdo para salvar <u>a un Estado</u> miembro y la moneda común .
	agreed to save a Member State and our common currency .

No	Solution 1 to 50 Page 1 / 3
6	y nos dejaron <u>sin una verdadera</u> elección , (no alignment found)
7	Debemos intentar convencerles <u>con una nueva</u> visión , sin austeridad . We have to try to convince them with a real post-austerity vision , not an austerity one .
8	La perspectiva de que continúe <u>en los próximos</u> años atormenta a los afganos . The prospect of more in the next years haunts Afghans today .
9	La idea es que la UE hable <u>con una sola</u> voz en la escena internacional . The idea is for the EU to speak with a sole voice on the international stage .
10	Quiero pedirle sus impresiones <u>sobre los últimos</u> acontecimientos de la UE , I intend to ask him about his impressions of what recently happened in the EU ,

f. PARTICIÓN DEL SINTAGMA (*ADJETIVALY ADVERBIAL*)

f.1. *SINTAGMA ADJETIVAL*

El número de entradas de esta categoría es relativamente elevado (3490 subtítulos) en comparación con otras categorías, sin embargo, a medida que analizamos los subtítulos nos damos cuenta que en muchas ocasiones no se podía segmentar de otra manera, dado que el número de subtítulos es igual para ambas lenguas por el uso de *Master Files*. A continuación, mostramos las diez primeras entradas:

No	Solution 1 to 50 Page 1 / 70
1	en una gran reunión con <u>los presidentes parlamentarios</u> de los estados involucrados . at a big meeting of the presidents of the relevant states ' parliaments .
2	El martes , el presidente de <u>la Rada ucraniana</u> asistirá a la Comisión On Tuesday , the Ukrainian Rada Chair comes to the Foreign Affairs Committee .

No	Solution 1 to 50 Page 1 / 70
3	¿ Nos falta una unión fiscal y de mercado laboral ? Fiscal and labour-market union . Do you see a gap ?
4	¿ Tiene cabida la agricultura ecológica en la reforma ? What place for organic farming in the future reform ?
5	Un político ruso visita el Parlamento Europeo para comunicar la postura rusa . a senior Russian politician arrived at the EU Parliament to put Moscow 's position .
6	Les digo que la austeridad , las reformas fiscales y laborales , están muy bien , What I 'm saying now is that austerity , fiscal market and labour reforms , fine , I 'm for that
7	que cumple con los objetivos medioambientales y turísticos . which meets targets for the environment and tourism .
8	Pero debo decir que estos países emergentes son rara vez But I have to say that these emerging countries are not or are rarely
9	sobre la veracidad de la vigilancia masiva practicada por la NSA about whether there is any truth to mass surveillance by the NSA ,
10	Los pesticidas son útiles , pero su uso excesivo contamina el suelo y el agua Pesticides are useful , but too much of them pollutes the soil and water .

Por su parte, el *frequency breakdown* muestra cinco posibles combinaciones de categorías y vemos como la entrada de artículo seguido de nombre común y adjetivo es la que más casos recoge:

Nº	Resultado búsquedas	Número ocurrencias	Porcentaje
1	ART NC ADJ	3256	93,3%
2	PPO NC ADJ	101	2,89%
3	DM NC ADJ	73	2,09%
4	ART NP ADJ	56	1,6%
5	ART NMEA ADJ	4	0,11%

En la siguiente tabla mostramos los diez primeros ejemplos de (artículo + nombre común + adjetivo):

No	Solution 1 to 50 Page 1 / 66
1	que acabó con <u>el régimen comunista</u> en Polonia . (no alignment found)
2	La regulación de <u>los mercados financieros</u> es su prioridad legislativa The regulation of financial markets is a legislative priority
3	Han dado <u>un salto cualitativo</u> en liderazgo . There was also an important leadership change :
4	que tiene <u>unos objetivos europeístas</u> bien definidos . (no alignment found)
5	En comparación con <u>los países europeos</u> , China , Rusia o India (no alignment found)
6	Una de <u>las causas principales</u> de la crisis (no alignment found)
7	Queremos <u>una industria europea</u> sostenible We want to make European industry sustainable
8	Queremos mantener <u>el modelo estatal</u> europeo de bienestar We want to preserve the European welfare state model ,
9	forman <u>el triángulo institucional</u> de la UE . form the institutional triangle of the EU .
10	pero el Parlamento tiene <u>un papel clave</u> en la elaboración de estas leyes . but the EP has an essential role in the expansion of these laws .

Los ejemplos que vemos a continuación, los hemos extraído de la segunda entrada que en lugar de un artículo es un posesivo el que va seguido de sustantivo y adjetivo:

No	Solution 1 to 50 Page 1 / 3
1	y favorecer <u>su integración efectiva</u> en Francia .
	and help them to really integrate in France .
2	Pero es un país grande y <u>su déficit presupuestario</u> es del 5,3% .
	But it 's a big country and its budget deficit is just 5.3% .
3	ni para Europa ni para <u>nuestros socios asiáticos</u> o de otros lugares ,
	Neither for Europe nor our partners from Asia or elsewhere .
4	Es también capaz de cerrar <u>su sector público</u> a las empresas europeas ,
	It 's also capable of closing its public sector contracts to European companies ,
5	pese a <u>nuestras diferencias políticas</u> e ideológicas ,
	despite our political and ideological differences ,
6	e ignorar <u>sus aspectos humanos</u> y humanitarios .
	while turning a blind eye on the humanitarian and human aspects of it all .
7	serán capaces de reducir <u>su deuda soberana</u> de forma significativa ?
	will be able to significantly reduce their sovereign debt ?
8	porque perdería <u>su liderazgo político</u> aquí .
	because Britain would lose its leading role in making policies work here .
9	y mejorar <u>su formación periodística</u> .
	It 's a chance to expand their journalistic knowledge .
10	independientemente de <u>su situación económica</u> y su clase social .
	(no alignment found)

Por último, nos parece interesante añadir en este caso, que al analizar el *frequency breakdown* por palabras, vemos que los siguientes sintagmas son los que más veces se segmentan inadecuadamente:

Nº	Resultado búsquedas	Número ocurrencias	Porcent.
1	las instituciones europeas	23	0,66%
2	los derechos humanos	23	0,66%
3	los mercados financieros	15	0,43%
4	los parlamentos nacionales	15	0,43%
5	la Comisión Europea	14	0,4%
6	los ciudadanos europeos	14	0,4%
7	el cambio climático	12	0,34%
8	la crisis económica	11	0,32%
9	las elecciones europeas	11	0,32%
10	la eficiencia energética	9	0,26%

A continuación, mostramos ejemplos de *las instituciones europeas* y *los derechos humanos* que posiblemente sean los términos que más veces aparecen en el discurso parlamentario. En la primera de las tablas, tenemos los ejemplos de *las instituciones europeas*:

No	Solution 1 to 23 Page 1 / 1
1	<p>es el mismo que las instituciones europeas han enfatizado en 2015 :</p> <p>was also one that the EU institutions dedicated 2015 to :</p>
2	<p>2012 : pasan los meses y las instituciones europeas no se ponen de acuerdo</p> <p>2012. Months went by and the European institutions were unable to agree</p>
3	<p>Cuanto más se esfuerzen las instituciones europeas por facilitarnos la vida</p> <p>Yeah , I think whatever the European institutions can do to make life easier ,</p>
4	<p>Ahora , los Estados y las instituciones europeas trabajan de manera más estrecha</p> <p>The Member States and EU institutions now work closely together</p>
5	<p>Otra explicación es que las instituciones europeas están muy alejadas .</p> <p>Another explanation is that seen from Malacky , EU institutions seem far away .</p>
6	<p>Tal vez usted crea que las instituciones europeas están muy lejos .</p> <p>Maybe you feel that European institutions are far removed from your everyday life .</p>

No	Solution 1 to 23 Page 1 / 1
7	las instituciones europeas no pueden funcionar .
	that must exist otherwise we cannot deliver as a European institution .
8	que el Parlamento y las instituciones europeas sirven para algo .
	that the EP and other European institutions are not pointless .
9	Otra explicación es que las instituciones europeas están muy alejadas .
	Another explanation is that seen from Malacky , EU institutions seem far away .
10	empezando por las instituciones europeas . ¡Ya lo creo!
	starting with the European institutions. You bet!

Y, a continuación, los ejemplos de *los derechos humanos* en la siguiente tabla:

No	Solution 1 to 23 Page 1 / 1
1	Se hizo por el respeto a los derechos humanos , los derechos fundamentales
	It was for the respect of human rights , fundamental rights
2	Existen abusos de los derechos humanos en estos países
	Human rights abuses exist in all post-soviet areas .
3	se basa en esos valores : los derechos humanos , el respeto a los demás .
	is these values , the values of human rights , respect for others .
4	Sé que el respeto de los derechos humanos es un criterio importante
	I know that the respect for human rights is one very important criterion
5	¿Cuál es la situación de los derechos humanos en estas repúblicas ?
	How would you summarise the human rights situation in ex-Soviet Republics ?
6	En cuanto a los derechos humanos en Rusia ,
	With regards to human rights in Russia ,

No	Solution 1 to 23 Page 1 / 1
7	Su objetivo es mejorar los derechos humanos en todo el mundo .
	It 's designed to advance human rights across the world .
8	y el respeto por los derechos humanos y civiles .
	and more respect for human and civil rights .
9	mientras la situación de los derechos humanos se deteriora .
	when human rights are deteriorating there .
10	y otras violaciones de los derechos humanos de los sospechosos .
	and other breaches of suspects ' human rights .

Observamos que, en ninguno de los dos casos, la versión en inglés segmenta estas unidades de sentido, mientras que en el caso de la versión española sí lo hace.

f.2. PARTICIÓN SINTAGMA ADVERBIAL

En la siguiente tabla mostramos las diez primeras entradas de la búsqueda que hemos generado para esta categoría.

No	Solution 1 to 50 Page 1 / 37
1	El Tratado de Lisboa concedió más poderes al Parlamento ,
	As the Lisbon Treaty enhanced the EP 's lawmaking powers
2	El Parlamento siempre ha sido más ambicioso que los Estados .
	Parliament 's always been more ambitious than the Member States .
3	pero además entender cómo funciona el mundo .
	(no alignment found)
4	pero los funcionarios europeos están tranquilamente en sus despachos
	but in the practical situation the European bureaucrats sit in their nice offices

No	Solution 1 to 50 Page 1 / 37
5	Eso depende . Quizá haya que <u>gastar más</u> en la agricultura en el futuro .
	That depends . There may be spending on agriculture in the future .
6	Debemos admitir que esta crisis <u>es seguramente</u> la peor que hemos visto .
	We must recognize that this crisis is probably the worst we have ever seen .
7	Y siguiendo , cada vez <u>circulan más</u> camioneros en Europa .
	Sticking with transport , more and more truckers deliver goods in Europe .
8	está empezando a <u>desarrollarse incluso</u> el comercio transfronterizo .
	even crossborder e-commerce is starting to develop .
9	La ayuda presupuestaria se <u>controla mucho</u> mejor que la ayuda por proyectos .
	Budget support is more strictly controlled than project aid .
10	Maroun , ¿ qué le ha <u>enfurecido últimamente</u> ?
	Maroun , what annoyed you or what moved you recently ?

Podemos apreciar que hay cuatro entradas con el adverbio *más*, pero en estos casos está modificando a un grupo nominal y no al verbo, por lo que no debemos considerar estos casos como segmentación inadecuada. Por otra parte, *cómo* es adverbio interrogativo en una oración interrogativa indirecta, por lo que entendemos que en este caso también la segmentación es adecuada. Hemos realizado una búsqueda que nos permite detectar los casos de *más* y *cómo* y estos son los resultados obtenidos:

La búsqueda que hemos generado para el caso de *más*:

```
[pos="V.+" & _subtitle_n_lines="2" & _line_no="1"]
</line><line_no="2">[word="más"]
```

El resultado es de 492 casos, de los que mostramos, los diez primeros que aparecen de forma aleatoria en el corpus:

No	Solution 1 to 50 Page 1 / 10
1	eligiera a los comisarios y <u>tuviera</u> <u>más</u> poderes compartidos con el Consejo . (no alignment found)
2	y esto hará que <u>tenga</u> <u>más</u> visibilidad . (no alignment found)
3	Con el Tratado de Niza , el Consejo <u>tenía</u> <u>más</u> peso en el proceso legislativo . Under the Nice Treaty the Council carried the most weight in legislative procedures .
4	que las personas que beben café <u>viven</u> <u>más</u> tiempo . No tengo pruebas de ello . (no alignment found)
5	Marian Dony , ¿ qué le ha <u>sorprendido</u> <u>más</u> , la abstención ? Marian Dony what struck you the most about election night ?
6	porque <u>tienen</u> <u>más</u> experiencia política en el PE ? because they behave in a more politically experienced way in the EP ?
7	Unos países <u>tendrán</u> <u>más</u> eurodiputados y otros menos . Some countries will get more MEPs , some less .
8	Significa que probablemente <u>estará</u> <u>más</u> o menos equilibrado en el futuro , This means it will probably be more or less balanced in future ,
9	Por lo tanto , tenemos que <u>establecer</u> <u>más</u> medidas de protección de las fronteras . Therefore , we need to work on setting up additional border protection .
10	Eslovenia <u>estaba</u> <u>más</u> que preparada para entrar . (no alignment found)

En el caso de *cómo* obtenemos 111 casos, y la búsqueda ha sido la siguiente:

```
[pos="V.+" & _subtitle_n_lines="2" & _line_no="1"]</line>
<line_no="2">[word="cómo"]
```

No	Solution 1 to 50 Page 1 / 3
1	Esta persona no debería <u>decirnos</u> <u>cómo</u> debemos comportarnos . That person should n't be telling us how to behave .
2	Deben <u>explicar</u> <u>cómo</u> se utilizan sus productos (no alignment found)
3	Esta semana <u>veremos</u> <u>cómo</u> se reorganiza . This week , we 'll see how to reorganise ,
4	y no <u>saben</u> <u>cómo</u> son aquí las cosas . and they do n't know how things are here .
5	pero nos fuerzan a <u>pensar</u> <u>cómo</u> dar a las personas but they force you to think about how people can be given
6	creo que debe <u>preocuparnos</u> <u>cómo</u> van los donantes tradicionales , I think we should worry about how our traditional donors ,
7	Muchos no <u>saben</u> <u>cómo</u> se toman estas decisiones . Many do n't know how these decisions are being made .
8	Se está <u>investigando</u> <u>cómo</u> modernizarse sin excluir . So research is being done on for example how to modernise without excluding .
9	Hacen falta varios años para <u>ver</u> <u>cómo</u> las medidas ... We need a number of years now to see how these measures are going to ...
10	Nosotros debemos <u>explicar</u> <u>cómo</u> nos afectan a todos . So let 's see how this legislation is applied to you and me .

Hemos añadido el caso de *menos* con la misma función que *más*:

```
[pos="V.+" & _subtitle_n_lines="2" & _line_no="1"]</line>
<line_no="2">[word="menos"]
```

Y el total de casos que hemos encontrado es de 56. En la siguiente tabla mostramos las diez primeras entradas:

No	Solution 1 to 50 Page 1 / 2
1	los verdes pensamos que no se <u>necesita</u> <u>menos</u> política europea , sino más .
	we Greens think that we do n't need less but more of European politics .
2	el Parlamento <u>tendrá</u> <u>menos</u> peso que el Consejo .
	the EP will not have the same say as the Council .
3	pero las mujeres <u>tienen</u> <u>menos</u> derechos que los hombres .
	but women are in a situation of less rights than men .
4	Crearía una situación en la que <u>hay</u> <u>menos</u> variedad o prácticamente ninguna
	It would bring about a situation where you have far less choice or almost no choice
5	Si se discrimina a una mujer y se le <u>paga</u> <u>menos</u> en su lugar de trabajo ,
	If a woman is discriminated against and underpaid in a workplace
6	También es difícil porque <u>supone</u> <u>menos</u> ingresos para los Estados .
	But that is also difficult because it means less revenue for the states .
7	Nos gustaría que <u>hubiera</u> <u>menos</u> polacos en el extranjero .
	We 'd like fewer Poles to work abroad .
8	Y para ello no <u>necesitamos</u> <u>menos</u> integración europea ,
	And for that we need not less European integration ,
9	y estas medidas nos ayudarían a <u>ser</u> <u>menos</u> dependientes .
	and all these new measures could help us to be less dependent .
10	pero tienen que <u>consumir</u> <u>menos</u> energía para ello .
	but their development must consume a lot less energy .

g. CATEGORÍA VERBO COMPUESTO

Esta es la categoría que más casos de segmentación inadecuada contiene. Si presentamos los datos de forma conjunta se aprecia claramente:

Tabla 53. Resultados de segmentación intrasubtítulo inadecuada por categoría en EMPAC_ES.

class	freq	total	rel_freq
CONJ	2534	184485	13,73553
DET	1584	184485	8,586064
FRASE	7159	184485	38,80532
NOM_COMP	1620	184485	8,781202
PREP	2142	184485	11,6107
SINT	5339	184485	28,94002
VERB_COMP	5681	184485	30,79383

Veamos los resultados desglosados en las tres subcategorías que lo componen (partición verbos compuestos, partición de perífrasis verbal y partición del predicado nominal)

g.1. PARTICIÓN DE VERBOS COMPUESTOS

Esta es, de las tres subcategorías, en la que menos casos se dan, representando el 2,1 por mil del total, lo que se traduce en 404 casos en 184 485 subtítulos de dos líneas:

No	Solution 1 to 50 Page 1 / 9
1	A usted le parece positivo que se <u>hayan</u> <u>excluido</u> porque son perjudiciales . (no alignment found)
2	En Francia los expertos <u>han</u> <u>instrumentalizado</u> las elecciones . In France experts gain the upperhand in the elections .
3	Pero los socialistas no <u>han</u> <u>propuesto</u> una solución conjunta . But the Socialists did n't have a single , viable response ...
4	También están los Verdes que <u>han</u> <u>logrado</u> un magnifico resultado There are the Greens who having won in spectacular fashion ,

No	Solution 1 to 50 Page 1 / 9
5	Los presidentes americanos a veces son elegidos con un 40% del voto popular .
	American presidents are often elected with 40% of popular votes .
6	Gracias por la cooperación que hemos obtenido de ustedes , los eurodiputados .
	Thanks to the cooperation we 've received from you , the members of this parliament ,
7	Como el Acuerdo no ha sido aceptado por el Parlamento ,
	As the Agreement has not been accepted by the Parliament ,
8	El Parlamento Europeo es elegido directamente por los ciudadanos de la UE .
	The European Parliament is directly elected by EU citizens .
9	una posible respuesta sería otorgar a este tipo de robots
	a possible answer might be to give that category of robots
10	En el futuro , el reto será llegar a nuevos países
	Tomorrow , the challenge will be to move into new countries

El *frequency breakdown* por POS nos muestra que el verbo *haber* finito seguido de un verbo principal en participio pasado:

Nº	Resultado búsquedas	Número ocurrencias	Porcentaje
1	VHfin VLadj	247	61,14%
2	VSfin VLinf	40	9,9%
3	VSadj VLadj	27	6,68%
4	VSfin VLadj	22	5,45%
5	VHinf VLadj	20	4,95%
6	VHfin VMadj	8	1,98%
7	VSinf VLadj	8	1,98%
8	VHfin VSadj	7	1,73%
9	VSfin VCLinf	5	1,24%
10	VSinf VLinf	5	1,24%

Exponemos, seguidamente, ejemplos de la primera entrada, dado que es la que más ocurrencias tiene:

No	Solution 1 to 50 Page 1 / 5
1	También están los Verdes que han logrado un magnífico resultado
	There are the Greens who having won in spectacular fashion ,
2	no significa que la Comisión haya dejado de interesarse por esos temas .
	does not really mean the Commission has stopped following these issues .
3	con las prácticas clientelistas que hemos visto en los grandes partidos griegos .
	with the clientelist practices we ‘ve seen with the two main Greek parties .
4	No se aplicó , por lo que Grecia ha presionado a la UE para que tome cartas ,
	It was n’t implemented , so Greece had been pushing the EU to get involved
5	Financia proyectos de jóvenes que hayan pasado 3 años en el extranjero .
	It finances the research projects of young people with 3 years ‘ experience abroad .
6	Eso sigue una lógica económica. Hemos creado un mercado único en Europa .
	There is an economic logic here. We ‘ve created a unique European market .
7	La Conferencia de Presidentes del PE ha decidido que el ganador del Premio Sájarov
	The EP Conference of Presidents decided to award the 2014 Sakharov Prize
8	y según fuentes sindicales se han perdido cerca de 50000 empleos ,
	(no alignment found)
9	hasta ahora , los jefes de Estado se han reunido en el edificio Justus Lipsius .
	up till now the heads of state met here in the Justus Lipsius building .
10	Exportamos menos , pero lo hemos compensado con las ventas nacionales .
	We ‘re exporting less but the national market has made up for that .

Y la segunda entrada, la del verbo *ser* seguido de un verbo en infinitivo:

No	Solution 1 to 40 Page 1 / 1
1	El objetivo <u>era</u> <u>dar</u> un nuevo impulso a Europa ,
	It was the goal to give Europe a new boost ,
2	Creo que lo mejor <u>será</u> <u>esperar</u> hasta el otoño
	(no alignment found)
3	más difícil <u>será</u> <u>garantizar</u> la integración .
	the more difficult it becomes to guarantee integration .
4	una posible respuesta <u>sería</u> <u>otorgar</u> a este tipo de robots
	a possible answer might be to give that category of robots
5	una de las cosas que pedía <u>es</u> <u>reforzar</u> el énfasis en la repatriación :
	(no alignment found)
6	Para algunos , pasar de Energía a Digital <u>es</u> <u>salir</u> de Málaga para meterse en Malagón .
	For some , the journey from energy to digital is a journey from one side of hell to the other .
7	La primera cosa que tenemos que hacer <u>es</u> <u>pedir</u> una cumbre con la ONU y la OCDE ,
	The first thing we must do is to call for a summit with the UN and OECD
8	En el futuro , el reto <u>será</u> <u>llegar</u> a nuevos países
	Tomorrow , the challenge will be to move into new countries
9	que el objetivo de esta política <u>es</u> <u>salvaguardar</u> nuestra posición en el mundo .
	is that it 's a policy which aims to safeguard our position in the world .
10	La situación ideal <u>sería</u> <u>acabar</u> con la guerra en Siria
	The ideal situation would be to stop the war in Syria

Por último, ofrecemos la tercera entrada que muestra la segmentación de verbos conjugados en voz pasiva:

No	Solution 1 to 27 Page 1 / 1
1	las tres películas finalistas han <u>sido visionadas</u> . ¿Cuál será la ganadora ?
	screenings have been taking place for all three films , which one wins ?
2	Como el Acuerdo no ha <u>sido aceptado</u> por el Parlamento ,
	As the Agreement has not been accepted by the Parliament ,
3	Sin duda , para un hombre que ha <u>sido separado</u> de sus hijos durante cuatro años
	Of course , for a man who has been away from his children for four years
4	cientos de miles de tibetanos han <u>sido asesinados</u> , torturados y encarcelados .
	hundreds of thousands of Tibetans have been killed , tortured and imprisoned .
5	Pero la libre competencia ha <u>sido rechazada</u> por los operadores afectados ,
	But free competition has been countered by incumbent operators
6	Más de 500 enmiendas han <u>sido añadidas</u> a su informe del Parlamento .
	More than 500 amendments have been submitted to her report to the EP .
7	Conoce muy bien la UE tras haber <u>sido eurodiputada</u> durante siete años .
	She knows the EU very well having been an MEP for seven years .
8	La Agenda del Consumidor ha <u>sido examinada</u> en una comparecencia en el PE .
	The Consumer Agenda came under scrutiny at a public hearing in the EP .
9	Las estructuras de madera , que han <u>sido fabricadas</u> por una empresa del lugar .
	The wood structures , made by a local company .
10	Aquí, en Europa, Edward Snowden ha <u>sido nominado</u> al Premio Sájarov de 2013.
	Back in Europe Edward Snowden has been nominated for this year 's Sakharov Prize .

g.2 CATEGORÍA VERBO COMPUESTO PERÍFRASIS_1

Aquí buscamos la segmentación de perífrasis compuestas por dos verbos. Mostramos aquí el resultado de la búsqueda con las diez primeras entradas:

No	Solution 1 to 50 Page 1 / 66
1	sin que esto <u>suponga</u> <u>perder</u> de vista (no alignment found)
2	Un total de 18 eurodiputados <u>podrían</u> <u>entrar</u> en el Parlamento más adelante , So potentially , a total of 18 MEPs will enter Parliament at a later stage ,
3	Como resultado , se <u>seguirá</u> <u>aplicando</u> la directiva actual As a result , the current directive is still enforced
4	¿ Cómo se controla si se <u>están</u> <u>ganando</u> el sueldo con su trabajo ? How is it controlled that they 're paid according to the work they 've done ?
5	En casos extremos <u>puede</u> <u>haber</u> sospechas de fraude It 's very rare , but when an MEP is suspected of fraud ,
6	de todo el mundo <u>están</u> <u>destinados</u> a Bruselas . from all over the world are in Brussels .
7	El Parlamento Europeo está <u>intentando</u> <u>lograr</u> la colaboración directa The European Parliament is trying to reach a more intensive direct cooperation
8	Sin embargo , las dietas se <u>siguen</u> <u>pagando</u> según un baremo . But other expenses will be on a flat-rate basis .
9	Algunos de estos cambios <u>pretenden</u> <u>acercar</u> la UE a los ciudadanos . Many of these changes may bring the EU closer to its citizens .
10	El Tratado de Lisboa <u>permitirá</u> <u>engrasar</u> esta maquinaria . The Lisbon Treaty will revamp this enormous machine .

Por su parte, el *frequency breakdown*, muestra que un verbo conjugado seguido de verbo en infinitivo es el caso que más ocurrencias tiene:

No	Solution 1 to 50 Page 1 / 22
1	la nueva ley de prensa no <u>parece</u> <u>afectar</u> el trabajo de todos . (no alignment found)
2	Pero espero que <u>podamos</u> <u>demostrar</u> nuestra buena fe But I hope obviously that we are demonstrating our good faith
3	También <u>decidieron</u> <u>crear</u> un banco central . They envisaged the creation of a central bank .
4	Para que así sea , <u>debemos</u> <u>acordar</u> una ruta en Cancún For that to happen we need to focus on a roadmap that we can agree to in Cancun ,
5	Los aumentamos o los <u>hacemos</u> <u>bajar</u> de intensidad . We can increase or lower the intensity of them .
6	El régimen soviético ya no <u>podía</u> <u>mantener</u> el control sobre la zona . The USSR could no longer rule with an iron fist ,
7	estimular la producción <u>significa</u> <u>permitir</u> el acceso a aguas limpias , boosting output means the industry competing for clean water ,
8	Todo esto <u>aconseja</u> <u>introducir</u> la baja por paternidad . That all tells us that it 's necessary to include paternity leave .
9	A la comunidad le <u>gusta</u> <u>dar</u> muestras de su lealtad , (no alignment found)
10	ya que el presidente Poroshenko <u>logró</u> <u>establecer</u> , durante más de un mes , when President Poroshenko had managed to establish , for more than a month ,

En cuanto a la segunda entrada, esta recoge las perífrasis compuestas por verbo modal seguido de un verbo finito conjugado. Mostramos, a continuación, los diez primeros ejemplos:

No	Solution 1 to 50 Page 1 / 14
1	Las víctimas y allegados <u>pueden llamar</u> a un número de emergencia . that it is forbidden in our society to act in this way ,
2	Sin pagos , la empresa no <u>puede hacer</u> frente a sus facturas y salarios . Without payments , the company cannot honour its bills or salaries .
3	23 millones de alemanes no <u>pueden recordar</u> la caída del Muro de Berlín , About 23 million Germans ca n't remember the fall of the Berlin Wall
4	Suena muy bien , pero ¿ cómo se <u>puede llevar</u> a la práctica ? ¿ Creen que es difícil ? It sounds good , but how can it be put into practice ? Is it difficult ?
5	En los juguetes , los ftalatos <u>pueden provocar</u> trastornos renales y testiculares . In toys , phthalates can upset hormones and cause kidney and testicular problems .
6	El PE insiste en que se <u>debe utilizar</u> tecnología punta . The EP insists it must use the latest technology available .
7	mientras que el pleno solo <u>puede aprobar</u> o rechazar el resultado . and the plenary can merely say yes or no to the result .
8	ni control , y <u>deberían competir</u> unos con otros . and no control , and they should compete with each other .
9	En la farmacia se <u>podrá escanear</u> el producto In the pharmacy you can scan the product
10	No duran todo el día y <u>pueden programarse</u> en función de los partidos , They do n't happen all day long and are scheduled around upcoming games ,

g.2.1. SUBCATEGORÍA PERÍFRASIS (CON PREPOSICIÓN)

En esta subcategoría encontramos 1181 casos de los que mostramos las diez primeras entradas que aparecen en CQPweb:

No	Solution 1 to 50 Page 1 / 24
1	En 1962 la Asamblea <u>pasa</u> <u>a llamarse</u> Parlamento Europeo . (no alignment found)
2	Y tenemos que <u>elegir entre</u> <u>estar</u> divididos con poco peso (no alignment found)
3	Si el PE <u>insiste</u> <u>en defender</u> sus enmiendas , When MEPs insist on defending their amendments ,
4	Cada país <u>tendrá</u> <u>que determinar</u> sus objetivos (no alignment found)
5	El desarrollo debe <u>contribuir</u> <u>a reducir</u> la emigración forzosa , Development must contribute to reducing forced emigration ,
6	cuando las personas se ven <u>obligadas</u> <u>a abandonar</u> sus países o regiones , when people are forced to leave their country or their region ,
7	pero se ha hecho siempre <u>pensando en</u> <u>proteger</u> los derechos fundamentales . but it was always done with a view to protect our fundamental rights .
8	Nos <u>esforzamos</u> <u>por lograr</u> decisiones sensatas . So we endeavoured to make reasonable decisions .
9	El Parlamento no será capaz de <u>dar</u> <u>a conocer</u> el proyecto político The EP will be incapable of indicating its wider political project
10	debemos resistirnos porque <u>hay</u> <u>que inyectar</u> dinero en la economía . (no alignment found)

g.3. VERBO SEGUIDO DE PREDICADO NOMINAL

Ofrecemos los datos que el corpus nos ofrece cuando seleccionamos los verbos *ser*, *estar* o *parecer*:

No	Solution 1 to 50 Page 1 / 17
1	El Parlamento siempre ha <u>sido</u> <u>más</u> ambicioso que los Estados .
	Parliament 's always been more ambitious than the Member States .
2	El 82,6% de las aguas <u>son</u> <u>aptas</u> para el baño y de alta calidad .
	82.6% of waters open to swimming are high quality .
3	pero los funcionarios europeos <u>están</u> <u>tranquilamente</u> en sus despachos
	but in the practical situation the European bureaucrats sit in their nice offices
4	Debemos admitir que esta crisis <u>es</u> <u>seguramente</u> la peor que hemos visto .
	We must recognize that this crisis is probably the worst we have ever seen .
5	Pero los controles sanitarios <u>serán</u> <u>más</u> estrictos . Eso es bueno , ¿ no ?
	But the health checks will be stricter . That 's a good thing , is n't it ?
6	la pobreza sigue <u>estando</u> <u>muy</u> extendida en Europa .
	real poverty is still widespread in Europe .
7	a los trabajadores que han <u>sido</u> <u>víctimas</u> de grandes cambios
	to those who fall victim to major changes
8	Sin duda , eso habría <u>sido</u> <u>doloroso</u> para los grandes bancos ,
	Of course , it would have been painful for the big banks ,
9	Los trabajadores desplazados <u>estarán</u> <u>en</u> la agenda del Consejo el 6 de diciembre .
	Posted workers will be on the Council 's agenda on 6 December .
10	La salud es importante y debemos <u>ser</u> <u>prudentes</u> al comprar medicamentos .
	Our health is precious and we need to be careful when buying medicines .

Por su parte, el *frequency breakdown* desvela las siguientes concordancias de POS:

Nº	Resultado búsquedas	Número ocurrencias	Porcentaje
1	VEfin PREP	194	23,49%
2	VSfin ADJ	124	15,01%
3	VSfin ADV	85	10,29%
4	VEfin ADV	63	7,63%
5	VSfin NC	50	6,05%
6	VSfin PREP	42	5,08%
7	VEfin ADJ	40	4,84%
8	VSinf ADV	29	3,51%
9	VSinf ADJ	24	2,91%
10	VSadj ADV	23	2,78%

A continuación, mostramos ejemplos de la entrada con más ocurrencias, 194, que corresponde al verbo *estar* seguido de preposición:

No	Solution 1 to 50 Page 1 / 4
1	Los trabajadores desplazados estarán en la agenda del Consejo el 6 de diciembre . Posted workers will be on the Council 's agenda on 6 December .
2	Muchos empresarios alemanes estarían en desacuerdo con usted , Sr. Hübner , Many German business people would disagree with you , Mr Hübner ,
3	y que no estamos en condiciones de afrontar and that we 're unable to deal with the challenges
4	Además , Galileo no está en manos de militares . And Galileo is n't controlled by the military .
5	En ese sentido , estamos entre los más euroentusiastas . From that perspective , we are among the greatest Euro-enthusiasts .
6	Verano de 1985 , estamos en el río Mosela en Luxemburgo . It 's early summer in 1985 on the River Moselle in Luxembourg .
7	El Sr. Treiner dice que no está en contra de la fusión nuclear Mr Treiner said that he 's not against nuclear fusion

No	Solution 1 to 50 Page 1 / 4
8	más adelante cuando <u>esté</u> <u>en</u> pleno funcionamiento .
	(no alignment found)
9	Debe construirse sobre pilares porque <u>está</u> <u>en</u> una zona con riesgo de terremotos .
	It has to be built on piles because it 's in a seismic zone , an earthquake zone .
10	¿ Los nuevos comisarios <u>estarán</u> <u>a</u> la altura de los anteriores ?
	Will the new Commissioners be able to fill the shoes of the old ?

Por su parte, la siguiente concordancia en número de casos, es la del verbo *ser* seguido de adjetivo. A continuación, vemos la selección de los diez primeros ejemplos:

No	Solution 1 to 50 Page 1 / 3
1	El 82,6% de las aguas <u>son</u> <u>aptas</u> para el baño y de alta calidad .
	82.6% of waters open to swimming are high quality .
2	Creo que el papel de la mujer <u>es</u> <u>importante</u> en los países islámicos
	(no alignment found)
3	Para algunas medidas , aunque <u>sean</u> <u>modestas</u> , el progreso podría ser muy lento .
	For some measures , modest as they may seem , progress may be difficult and slow .
4	A mayor confianza, mayores ventas. ¿ <u>Será</u> <u>bueno</u> para el consumidor y la industria ?
	Consumer confidence means more sales. Will both consumers and industry benefit?
5	porque creo que China también <u>es</u> <u>consciente</u> de la flagrante corrupción .
	because I think the Chinese are also aware of the raging corruption .
6	Puede que cambie , pero <u>serán</u> <u>escépticos</u> hasta el 1 de enero de 2014.
	Perhaps it will change , but it will be pretty sceptical until 1 January next year .
7	Países como Suecia estiman que <u>es</u> <u>necesario</u> aplicar reducciones aún mayores .
	But some countries like Sweden think it 's necessary to impose larger reductions .

No	Solution 1 to 50 Page 1 / 3
8	Otra pregunta es si los medios <u>son responsables</u> de la baja participación .
	Another issue is the responsibility of the media in the low turnout in this election .
9	¿ Por qué cree que <u>será diferente</u> esta vez ?
	What is different this time, do you feel?
10	se atribuyen el trabajo, algo que <u>es cierto</u> en parte , pero sólo en parte .
	claim it to be their work , which is partly true , but only partly .

Se observa que los casos de segmentación inadecuada no se producen como respuesta a una traslación de la segmentación en inglés y entendemos que, presumiblemente, la mayor parte de los casos es por falta de espacio.

11.2.2.3.4. Análisis de la segmentación en español entre subtítulos: segmentación intersubtítulo.

Recordemos los resultados obtenidos del análisis de la segmentación antes de mostrar ejemplos de cada una de las categorías:

Tabla 54. Resultados segmentación intersubtítulo inadecuada en EMPAC_ES.

class	feature	freq	total	rel_freq
CONJ	Conj	83	95514	0,868983
DET	Det_Dem	2	95514	0,020939
	Det_Pos	1	95514	0,01047
NOM_COMP	Nom_Compuesto	1549	95514	16,21752
PREP	Prep	18	95514	0,188454
SINT	Sintag_Adj	96	95514	1,005088
VERB_COMP	Verb_Comp	15	95514	0,157045
	Verb_Períf_1	294	95514	3,078083
	Verb_Períf_2	179	95514	1,874071
	Verb_PredNom	20	95514	0,209393

Vemos cómo, al igual que sucede en EMPAC_EN, la segmentación intersubtítulo cuenta con un número muy inferior de casos al de la segmentación intrasubtítulo en EMPAC_ES.

a. CATEGORÍA CONJUNCIÓN

Nos hallamos ante 83 casos de segmentación inadecuada de los que mostramos las diez primeras entradas. Teniendo en cuenta que la conjunción *que* es la más usada en castellano y que tiene múltiples valores (copulativo, condicional, causal, final), observamos que el resultado en nuestro corpus muestra que en la mayor parte de los casos está bien segmentado; por motivos de espacio no podemos profundizar ahora en todos estos aspectos, pero sin duda se trata de objetos de estudios para investigaciones futuras en las que podemos analizar con más detalle estas cuestiones lingüísticas.

No	Solution 1 to 50 Page 1 / 2
1	El gobierno que pretenda impedir la entrada de bienes tendrá <u>que probar</u> con documentos que por motivos de seguridad (no alignment found)
2	Entonces , a ver si lo entiendo , ¿ usted dice <u>que cuando</u> los precios caen , So you 're saying that - to get this straight - when prices drop too low
3	y todo el sistema está diseñado para <u>que sea</u> más fácil que reciban esa información , and the whole system is designed to make it easier for that information to get to them ,
4	Uzbekistán sigue siendo una férrea dictadura , y es cierto <u>que no</u> estamos de acuerdo con ello , pero al mismo tiempo tenemos que entablar un diálogo , Uzbekistan is still a harsh dictatorship , and we can very well say we do n't agree with that , but at the same time we need to engage in dialogue ,
5	La Comisión y el Parlamento lo rechazan <u>porque si</u> el riesgo es alto , debe trazarse un perímetro The Commission and the Parliament are refusing and saying , ' If there 's a high risk , there must be a safety zone
6	Tenemos que comprometernos más para hacer <u>que los</u> países vecinos y la Unión Africana demuestren su valía . We need to be more engaged in order to make sure that , neighbouring African countries and the African Union prove themselves .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 2
7	con gran satisfacción es <u>que</u> <u>el</u> Consejo , el Parlamento y la Comisión
	with a great sense of achievement is that the Council , Parliament and Commission
8	Alrededor de un 50% de los accidentes se deben al hecho de <u>que</u> <u>los</u> coches no están lo bastante atentos .
	Around 50% percent of accidents are caused by the fact that there is not enough recognition by car drivers .
9	Si son normas europeas , está claro <u>que</u> <u>la</u> industria europea tomará ventaja .
	If they are European standards , it 's clear that European industries will be at an advantage .
10	Las queremos porque consideramos <u>que</u> <u>dispondremos</u> de mejores herramientas a nivel europeo
	We want them because we think that we will have better instruments on the European level

b. CATEGORÍA DETERMINANTE

De las cuatro subcategorías que analizamos, no hay casos de segmentación intersubtítulos inadecuada del artículo definido e indefinido como ocurre en EMPAC_EN. De las otras dos subcategorías (demostrativos y posesivos), tan solo hay 2 casos de demostrativos y 1 de posesivos que pasamos mostramos en las siguientes tablas

b.1. SUBCATEGORÍA DEMOSTRATIVOS

No	Solution 1 to 2 Page 1 / 1
1	Demuestra que hoy en día uno no puede ganarse la vida con <u>esto</u> <u>si</u> no es haciendo triquiñuelas con los subsidijs y con el sistema .
2	Todos deberíamos defender <u>esto</u> <u>si</u> de verdad queremos una Europa de los ciudadanos

Ninguno de los casos detectados se considera segmentación.

b.2. SUBCATEGORÍA POSESIVOS

Encontramos un ejemplo en nuestro corpus sobre segmentación en esta categoría:

No	Solution 1 to 1 Page 1 / 1
1	<p>Pero todavía no se ha hecho nada . -También incluyen una de nuestras prioridades , que obtuvimos a nivel político : la creación de un cuerpo policial</p> <p>Everything is still to be done. -They even point out one of our priorities , that we obtained on a political level : the creation of a police force</p>

En este caso sí que se considera segmentación inadecuada y como podemos ver la segmentación de una lengua no influye en la otra.

c. CATEGORÍA NOMBRE COMPUESTO

En esta entrada, que en principio mostraba un número relativamente de entradas, hemos constatado que, en el caso de la segmentación intersubtítulo, todos los casos que nuestra búsqueda detecta son incorrectos.

Mostramos el *frequency breakdown* por palabras de las primeras 25 entradas que constatan que casi toda la información forma parte de títulos de programas en mayúsculas o etc. seguido de principio de frase en mayúscula:

Nº	Resultado búsquedas	Número ocurrencias	Porcent.
1	EUROPA Hola	49	2,88%
2	REPORTAJE Hola	29	1,7%
3	etc. Pero	19	1,12%
4	EUROPA Bienvenidos	18	1,06%
5	EUROPEO HANGOUT	14	0,82%
6	etc. El	12	0,71%
7	NO La	12	0,71%
8	etc. La	11	0,65%
9	EUROPEO GRUPO	11	0,65%
10	FÓRUM Hola	11	0,65%
11	NOTICIAS Hola	11	0,65%
12	ACTUALIDAD La	10	0,59%
13	REPORTERO La	10	0,59%
14	ESTUDIO Hola	9	0,53%

(Continúa en la página siguiente)

(Continúa de la página anterior)

Nº	Resultado búsquedas	Número ocurrencias	Porcent.
15	etc. No	9	0,53%
16	ACCIÓN Hola	8	0,47%
17	MAYO Vicepresidente	8	0,47%
18	REPORTAJE Están	8	0,47%
19	PORTADA EI	7	0,41%
20	etc. Si	6	0,35%
21	NOTICIAS Bienvenidos	6	0,35%
22	TITULARES PRESUPUESTO	6	0,35%
23	etc. Hay	5	0,29%
24	etc. Nos	5	0,29%
25	etc. Son	5	0,29%

Por lo tanto, los resultados de esta categoría no serían válidos.

d. CATEGORÍA PREPOSICIÓN

En esta categoría tan solo encontramos 18 casos que mostramos a continuación:

No	Solution 1 to 18 Page 1 / 1
1	Los Estados miembros siempre han tenido miedo a una reforma profunda de la Unión . Member States have always feared an in-depth reform of the Union .
2	el etiquetado energético , los efectos de la cafeína y Europa se prepara para la gran batalla del fútbol . energy labelling , the effects of caffeine , and Europe gears up to do battle in football .
3	El Sr. Papandreou participó activamente en el proceso que culminó con la Declaración de Salónica y el programa para integrar a los Balcanes Occidentales (no alignment found)
4	El Parlamento Europeo trabaja estrechamente con la Comisión Europea , que elabora proyectos de ley , The European Parliament works closely with : the European Commission , which prepares draft laws ,

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 18 Page 1 / 1
5	No se trata de que cuando tomamos una decisión <u>sobre una directiva</u> sobre energía o sobre CO2 , por ejemplo ,
	It is not the case that when we decide on , say , an energy directive or the CO2 Directive ,
6	Después de que 5 comisiones votaran en <u>contra de este acuerdo</u> internacional para combatir la falsificación ,
	After five committee votes against ACTA , an international agreement to fight counterfeit trade ,
7	Pero este polémico plan ha planteado muchas dudas <u>entre los ciudadanos</u> europeos , que publicaron sus preocupaciones en Facebook .
	But the controversial plan has raised many questions amongst EU citizens , who posted their concerns on Facebook .
8	de 15 Estados miembros diferentes para plantear preguntas a un grupo <u>de expertos</u> destacados de la UE sobre los derechos de los niños .
	came to the European Parliament to pose questions to a panel of some of the EU 's leading experts on children 's rights .
9	23 de los 27 estados miembros están en <u>contra de la retirada</u> de la moratoria y la cláusula de salvaguardia .
	23 out of 27 Member States are against the lifting the moratorium and the safeguard clause .
10	Pero fuera de Europa , ¿ qué actitud existe <u>hacia el comercio</u> de derechos de emisión , tan importante en la política europea ?
	(no alignment found)

e. CATEGORÍA PARTICIÓN DEL SINTAGMA

Tan solo encontramos casos de sintagma adjetival, pero ninguno del sintagma adverbial.

e.1. SUBCATEGORÍA PARTICIÓN DEL SINTAGMA ADJETIVAL

No	Solution 1 to 50 Page 1 / 2
1	<p>Estamos intentando imponer una nueva cultura de una accesibilidad auténtica , cercana a la realidad de las personas con movilidad reducida .</p> <p>We ‘re trying to impose a new culture of accessibility , real accessibility that is close to the reality of citizens with reduced mobility .</p>
2	<p>que les permitan vivir y de aspirar a un futuro similar al que disfruta el mundo occidental ?</p> <p>that will allow them to live and to look forward to the future in the same way that the Western world can ?</p>
3	<p>Ahora hemos descubierto que el sistema solo funciona en inglés y las traducciones no son buenas .</p> <p>We ‘ve now discovered that the system only works in one language , English , and the translations are no good .</p>
4	<p>Frédéric Dufour consiguió hacer realidad su sueño de ampliar su empresa especializada en transportar turbinas eólicas .</p> <p>Frédéric Dufour was able to realise his dream of expanding his company specialising in transporting wind turbines .</p>
5	<p>Una prueba de estrés para el Tratado de Lisboa , es lo que propuso el ministro italiano Giulio Tremonti en su visita al PE .</p> <p>The Italian Finance Minister proposed a stress test on the Lisbon Treaty . Giulio Tremonti , who was visiting the EP ,</p>
6	<p>Desde las elecciones , en el Parlamento Europeo y en los medios solo se habla de este asunto : ¿ Hay una crisis política en Europa ?</p> <p>Here in the European Parliament and in the media , since the elections , all focus is on the question , are we seeing a political crisis in Europe ?</p>

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 2
7	<p>Los políticos no crean empleo , pero crean <u>el clima</u> <u>necesario</u> para que la gente que puede y quiere invertir lo haga ,</p> <p>Politicians do n't create jobs , but they create the climate in which people who can and want to invest will do it ,</p>
8	<p>En la UE , la Carta de los Derechos Fundamentales protege a <u>las personas</u> <u>junto</u> con otras leyes antidiscriminatorias ,</p> <p>In the EU , individuals are protected by the Charter of Fundamental Rights and anti-discrimination laws .</p>
9	<p>No debió hacerlo. He participado en un procedimiento contra <u>el BCE</u> <u>junto</u> con un profesor alemán , el Sr. Kerber ,</p> <p>It never should have done it. I 've been part of a procedure against the ECB together with a German professor , Mr Kerber ,</p>
10	<p>La práctica de negar un servicio a un cliente o de hacerle pagar más por <u>un producto</u> <u>solo</u> por vivir en un país diferente de la Unión Europea</p> <p>Refusing to provide services to a consumer , and making him pay more for goods because he lives in a different country in the European Union ,</p>

Las entradas en las que aparece *solo* con la función de adverbio, obviamente no son casos de segmentación inadecuada. Sin embargo, para CQPweb *solo* es siempre un adjetivo, de ahí el error y de ahí la importancia de, en futuros trabajos, ir más allá del etiquetado morfosintáctico, por los errores que detectamos, y también la importancia de profundizar en los resultados que hemos obtenido mediante un análisis manual especialmente detallado y laborioso.

f. CATEGORÍA VERBO COMPUESTO

f.1. SUBCATEGORÍA VERBO AUXILIAR

Nuestro corpus presenta un total de 15 casos que mostramos en la siguiente tabla:

No	Solution 1 to 15 Page 1 / 1
1	<p>Está claro que su intención no <u>es</u> <u>menoscabar</u> el pluralismo ni el estado de derecho .</p> <p>It is clear that they have no intention of undermining pluralism and the rule of law .</p>
2	<p>Tenemos que asumir que las prioridades deben <u>ser</u> <u>fomentar</u> el crecimiento y el empleo .</p> <p>We have to assume that the priorities should be promoting growth and promoting jobs .</p>
3	<p>Estamos lo más preparados posible. Lo que tenemos que hacer <u>es</u> <u>adaptar</u> los planes a la información que llegue sobre el virus H1N1 .</p> <p>(no alignment found)</p>
4	<p>En julio se producirá una cumbre de la OTAN , pero su secretario general no <u>ha</u> <u>querido</u> esperar , y visitará el Parlamento el martes para hablar sobre estrategias .</p> <p>A NATO Summit is set for July but that 's not putting off the Secretary General who will be at the Parliament to talk strategy on Tuesday .</p>
5	<p>Nuestra responsabilidad <u>es</u> <u>buscar</u> y pedir más transparencia en la democracia europea ,</p> <p>Our responsibility is to look for , to ask for more transparency in European democracy ,</p>
6	<p>También se dice que la verdadera solución <u>sería</u> <u>intentar</u> acabar con la pobreza en el Sahel y las regiones más pobres de África</p> <p>At the same time , some say the real solution would be to try to end poverty in the Sahel , in the poor areas of Africa ,</p>
7	<p>del Parlamento Europeo. ¿ Su postura seguirá <u>siendo</u> <u>mantener</u> el diálogo con las autoridades iraníes ?</p> <p>(no alignment found)</p>

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 15 Page 1 / 1
8	Mi objetivo principal en política exterior es establecer buenas relaciones de vecindad entre Kosovo y Serbia
	My most important goal in foreign policy is to establish good neighbourhood relations between Kosovo and Serbia
9	El objetivo de Naciones Unidas en 1990 era reducir a la mitad la malnutrición en el mundo en 25 años .
	The target of the United Nations in 1990 was to halve the proportion of undernourished people around the world in 25 years .
10	No se puede responder al ascenso de la extrema derecha , izquierda o lo que sea modificando las reglas del juego .
	(no alignment found)

f.2. SUBCATEGORÍA VERBO COMPUESTO PERÍFRASIS_1

Sin embargo, la siguiente subcategoría es, de las cuatro analizadas, en la que más casos hemos detectado:

No	Solution 1 to 50 Page 1 / 6
1	Hace 15 años se le ocurrió crear un museo muy particular sobre la cultura judía .
	It was 15 years ago that the idea came to her for a museum of Jewish Culture , but not just any old museum .
2	Debo decir que el Banco Central Europeo procura mantener una relación lo más estrecha posible con el PE .
	(no alignment found)
3	Sí. Los socialistas siempre han exigido este impuesto llamado el impuesto Tobin sobre las transacciones financieras .
	Yes. The Socialists have long been calling for this tax , this so-called Tobin tax . A tax on financial transactions.

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 6
4	Ahora contamos con una plataforma que nos permitirá hacer frente a uno de los retos más acuciantes para la humanidad :
	We now have the setup which will allow us to deal with one of the most urgent challenges for humanity ,
5	Los diputados quieren reformar la Unión Europea aprovechando que este año se celebra el 60º aniversario del Tratado de Roma.
	MEPs want to reform the European Union as the Treaty of Rome turns 60 this year .
6	Después de esta experiencia , por tanto , debemos clarificar y actualizar la definición de abuso de mercado
	Following this experience it 's therefore right for us to clarify and update the definition of market abuse
7	¿ Por qué esa gente prospera llevando a personas de una orilla a otra del Mediterráneo a cambio de dinero ,
	Why are these people prospering by getting paid to take people across the Mediterranean ,
8	pero cuando veo esto , sé que hay alguien a quien le encantaría tener un autógrafo de Cory Doctorow en un manuscrito original .
	but when I see this , I know there is someone who would love a Cory Doctorow signature on an original manuscript .
9	Interesa a la seguridad de la Unión Europea asociarse con EE UU
	It is in the European Union 's security interests to have a partnership with the United states
10	Al final , mientras la Comisión Europea proponía ampliar el periodo mínimo de permiso de maternidad a 18 semanas ,
	In the end , whereas the European Commission was proposing to extend the minimum period for maternity leave to 18 weeks ,

f.2.1. SUBCATEGORÍA PERÍFRASIS CON PREPOSICIÓN

En esta búsqueda, de nuevo, encontramos un número significativo de entradas que no serían casos de segmentación inadecuada. Por ejemplo, la primera y décima entrada de la siguiente tabla:

No	Solution 1 to 50 Page 1 / 4
1	¿ Rusia ha tenido éxito en su lucha con la Unión Europea por controlar el destino y el futuro de Ucrania ?
	Has Russia succeeded in ambushing the European Union for , in effect , control over the destiny or the future of Ukraine ?
2	Hay tanta tienda cerrada que el ayuntamiento anima a decorar las persianas .
	So many shops are closed that the council encourages decorating the shutters .
3	La Comisión de Control Presupuestario está interesada en saber cómo se utiliza ese dinero .
	The Budgetary Control Committee is naturally interested in knowing how this money is spent .
4	Es decir , en lugar de obsesionarse con supervisar y sancionar a los Estados , debería unirlos .
	I mean that , rather than being obsessed with supervising and punishing Member States , it should unify them .
5	Creemos que , además de solucionar la crisis económica , hay que centrarse en hablar con la gente de las posibilidades y oportunidades
	Our view is that , as well as solving our financial crisis , we have to focus on talking to people about the potential and opportunities
6	Al exsecretario de Estado de Portugal , que fue criticado por hablar siempre en inglés , también le han preguntado
	The former Portuguese Secretary of State who at one point was criticised for speaking mainly in English , was also questioned

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 50 Page 1 / 4
7	¿ Deben ayudarnos los países industrializados a asumir esta responsabilidad ? Los países en desarrollo necesitan ayuda .
	Do you as industrialised nations feel you should give us this aid , so that we can take responsibility ? The developing countries need help .
8	Es despreciable. Tenemos el deber hacia las personas que nos eligen de actuar en beneficio del interés público con transparencia y responsabilidad .
	(no alignment found)
9	un país del Sur al que la UE debería ayudar a democratizar y no entorpecer .
	a southern country which the EU should really assist in order to help democratisation , not to cause a hindrance .
10	y obligará a la Comisión Europea a presentar propuestas legislativas basadas en un millón de firmas ?
	and will force the European Commission to make legislative proposals based on a million signatures ?

f.3. SUBCATEGORÍA VERBO SEGUIDO DE PREDICADO NOMINAL

En esta subcategoría encontramos un total de 20 casos y damos cuenta de los diez primeros ejemplos:

No	Solution 1 to 20 Page 1 / 1
1	¿ Cree que es el comienzo de una nueva era de cooperación entre el Congreso y el Parlamento Europeo ?
	(no alignment found)
2	es decir , sustancias cuya estructura es parecida a las sustancias psicoactivas conocidas .
	i.e. substances whose structure is relatively close to known psychoactive substances .

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 20 Page 1 / 1
3	porque es evidente que los intereses nacionales vuelven a <u>ser mucho</u> más importantes para todos los Estados miembros .
	because we all see that national interests have again become much more important for all the Member States .
4	Es una sustancia que , por sí misma , puede ser o no <u>ser positiva</u> en su uso o positiva en su aplicación .
	It 's a substance which on its own may or may not be positive in its use or positive in its application .
5	Durante esta semana , los eurodiputados <u>están en</u> las circunscripciones de los ciudadanos a los que representan .
	During this week , MEPs can be found on the ground in the constituencies of the citizens they represent .
6	es decir , sustancias cuya estructura <u>es parecida a</u> las sustancias psicoactivas conocidas .
	i.e. substances whose structure is relatively close to known psychoactive substances .
7	En cualquier caso , los alemanes están cambiando de <u>parecer sobre</u> la cuestión del gobierno económico de la UE ,
	At any rate , the Germans are acting on the issue of economic governance in the EU -
8	pero la cuestión <u>es cómo</u> tener en cuenta también aquí
	but the question is , to what extent should we take into account

(Continúa en la página siguiente)

(Continúa de la página anterior)

No	Solution 1 to 20 Page 1 / 1
9	El problema para la Comisión <u>es</u> <u>cómo</u> combinar 28 modelos diferentes de protección de la maternidad en una directiva
	The problem of the Commission was how to combine 28 different models concerning maternity protection in one directive
10	porque las normas que aparezcan pueden <u>ser</u> <u>normas</u> americanas , pueden ser normas internacionales ...
	because those that will emerge , they might become American standards , they might be international standards .

Observamos, que, en estos casos, también hay ejemplos de segmentación, que no serían tales, y que se producen por errores de etiquetado del POS, por ejemplo:

“¿ Cree que es el comienzo
de una nueva era
de cooperación entre el Congreso
y el Parlamento Europeo ?”

En este caso, *Era* es un sustantivo si bien ha sido etiquetado como verbo, de ahí el error en la detección. Por tanto, como apuntamos en el análisis de los resultados de la segmentación intersubtítulo de EMPAC_EN, los resultados que hemos obtenido en este tipo de segmentación habrían de ser recalculados; sin embargo, dejaremos esta tarea para futuros trabajos y nos quedaremos con el dato revelador de la insignificante existencia de casos de segmentación inadecuada entre subtítulos.

11.2.2.3.5. Análisis de la segmentación en inglés entre líneas de subtítulo: segmentación intrasubtítulo por años.

Por último, en este apartado, dado que ya ha quedado perfectamente ejemplificado las posibles segmentaciones en los apartados anteriores, mostraremos únicamente los resultados por año y categoría. Mostramos en color verde el año en el que se la frecuencia relativa sea más alta.

Tabla 55. Resultados de análisis de causas de segmentación inadecuada por años en EMPAC_EN.

class	year	freq	total	rel_freq
CONJ	2009	146	17930	8,142777
	2010	353	38573	9,151479
	2011	195	27347	7,130581
	2012	216	28932	7,465782
	2013	176	26464	6,650544
	2014	93	17101	5,438278
	2015	93	10922	8,514924
	2016	53	5174	10,24353
	2017	12	844	14,21801
DET	2009	359	17930	20,02231
	2010	774	38573	20,06585
	2011	336	27347	12,28654
	2012	286	28932	9,885248
	2013	257	26464	9,711306
	2014	231	17101	13,50798
	2015	183	10922	16,75517
	2016	107	5174	20,68032
	2017	35	844	41,46919
FRASE	2009	614	17930	34,24428
	2010	1037	38573	26,88409
	2011	837	27347	30,60665
	2012	1544	28932	53,36651
	2013	1415	26464	53,46886
	2014	533	17101	31,16777
	2015	286	10922	26,18568
	2016	105	5174	20,29378
	2017	56	844	66,35071
NOM_COMP	2009	389	17930	21,69548
	2010	946	38573	24,52493
	2011	568	27347	20,7701
	2012	642	28932	22,18996
	2013	568	26464	21,46312
	2014	385	17101	22,5133
	2015	282	10922	25,81945
	2016	137	5174	26,47855
	2017	29	844	34,36019

(Tabla 55 continúa en la página siguiente)

(Tabla 55 continúa de la página anterior)

class	year	freq	total	rel_freq
PREP	2009	628	17930	35,0251
	2010	2130	38573	55,21997
	2011	1288	27347	47,0984
	2012	1076	28932	37,19065
	2013	788	26464	29,7763
	2014	405	17101	23,68283
	2015	418	10922	38,27138
	2016	202	5174	39,04136
	2017	36	844	42,65403
SINT	2009	566	17930	31,56721
	2010	963	38573	24,96565
	2011	698	27347	25,52382
	2012	703	28932	24,29835
	2013	638	26464	24,10822
	2014	440	17101	25,72949
	2015	242	10922	22,15711
	2016	144	5174	27,83147
	2017	37	844	43,83886
VERB_COMP	2009	1103	17930	61,51701
	2010	2110	38573	54,70148
	2011	1642	27347	60,04315
	2012	1748	28932	60,41753
	2013	1609	26464	60,79958
	2014	1027	17101	60,05497
	2015	669	10922	61,25252
	2016	332	5174	64,16699
	2017	66	844	78,19905

Parece evidente que, en términos cuantitativos, 2017 es el año en el que mayor casos de segmentación inadecuada se dan.

11.2.2.3.6. Análisis de la segmentación en español entre líneas de subtítulo: segmentación intrasubtítulo por años.

Asimismo, en este apartado, dado que ya ha quedado perfectamente ejemplificado las posibles segmentaciones en los apartados anteriores, mostraremos únicamente los resultados por año y categoría. Mostramos en color verde el año en el que se la frecuencia relativa sea más alta:

Tabla 56. Resultados de análisis de causas de segmentación inadecuada por años en EMPAC_ES.

class	year	freq	total	rel_freq
CONJ	2009	440	31590	13,92846
	2010	431	39289	10,96999
	2011	282	26811	10,51807
	2012	408	27888	14,62995
	2013	372	25969	14,32477
	2014	265	15318	17,29991
	2015	210	11566	18,15667
	2016	113	5204	21,71407
	2017	13	850	15,29412
DET	2009	294	31590	9,306743
	2010	424	39289	10,79182
	2011	242	26811	9,026146
	2012	208	27888	7,458405
	2013	123	25969	4,736416
	2014	105	15318	6,854681
	2015	95	11566	8,21373
	2016	82	5204	15,75711
	2017	11	850	12,94118
FRASE	2009	972	31590	30,76923
	2010	1402	39289	35,68429
	2011	897	26811	33,45642
	2012	1409	27888	50,52352
	2013	1444	25969	55,60476
	2014	619	15318	40,40998
	2015	250	11566	21,61508
	2016	108	5204	20,75327
	2017	58	850	68,23529
NOM_COMP	2009	326	31590	10,31972
	2010	361	39289	9,188322
	2011	219	26811	8,168289
	2012	266	27888	9,538153
	2013	171	25969	6,584774
	2014	143	15318	9,335422
	2015	79	11566	6,830365
	2016	45	5204	8,647194
	2017	10	850	11,76471

(Tabla 56 continúa en la página siguiente)

(Tabla 56 continúa de la página anterior)

class	year	freq	total	rel_freq
PREP	2009	358	31590	11,3327
	2010	458	39289	11,65721
	2011	197	26811	7,34773
	2012	348	27888	12,47849
	2013	315	25969	12,12985
	2014	176	15318	11,48975
	2015	184	11566	15,9087
	2016	88	5204	16,91007
	2017	18	850	21,17647
SINT	2009	1011	31590	32,0038
	2010	1273	39289	32,40093
	2011	929	26811	34,64996
	2012	649	27888	23,27166
	2013	603	25969	23,21999
	2014	394	15318	25,72137
	2015	301	11566	26,02455
	2016	161	5204	30,93774
	2017	18	850	21,17647
VERB_COMP	2009	1151	31590	36,43558
	2010	1334	39289	33,95352
	2011	902	26811	33,64291
	2012	699	27888	25,06454
	2013	604	25969	23,2585
	2014	415	15318	27,09231
	2015	388	11566	33,5466
	2016	163	5204	31,32206
	2017	25	850	29,41176

En el caso de EMPAC_ES no podemos afirmar con tanta rotundidad que es el año en el que más incidencia de causas de segmentación inadecuada se da. No obstante, dejamos para estudios futuros, análisis minuciosos en este sentido.

Concluye aquí el capítulo de análisis del corpus a partir del cual hemos podido extraer las primeras conclusiones. El capítulo que viene a continuación, el de las conclusiones, nos servirá para recapitular e interpretar los datos para así corroborar o refutar nuestra hipótesis de salida.

CAPÍTULO 12

CONCLUSIONES

CONCLUSIONES

La presente investigación nace de la inquietud por el impacto de la digitalización en la vida de las personas; sobre todo, en la de las personas con algún tipo de discapacidad sensorial. Sin embargo, no queremos dejar de señalar que esa inquietud y preocupación se extiende también a todas aquellas personas que no tienen conocimientos lingüísticos en otra lengua, en particular, de la lengua inglesa (hegemónica y globalizadora). Por tanto, en este trabajo hemos querido poner de relieve la importancia y el impacto de los avances tecnológicos en la sociedad actual, una sociedad en la que aquella persona que no tenga acceso a la información corre serio riesgo de quedar excluida y sin derecho a participar en ella de forma plena.

Los avances tecnológicos e Internet y, en especial, el aumento de la velocidad de transferencia de datos y almacenamiento en este medio, favorecen el hecho de que la información sea cada vez más audiovisual y menos textual. Por tanto, parece lógico recordar que, para que el mayor número de usuarios pueda disfrutar de estos contenidos, la traducción (entendida como forma de hacer accesible el material publicado en la web) es fundamental. Así pues, nos encontramos ante un escenario en el que la posición de la traducción audiovisual y de las nuevas modalidades, presentadas en el capítulo 2, es efectivamente central, y así lo demuestran los numerosos estudios que profundizan en esta materia.

Nosotros, con la intención de impulsar los estudios en el ámbito de la accesibilidad audiovisual en la web, hemos analizado una serie de material audiovisual alojado en la web del Parlamento Europeo, en concreto, en EuroparlTV; ello nos ha permitido, por una parte, estudiar la traducción audiovisual en el contexto institucional y por otra, analizar el cumplimiento de los requisitos de accesibilidad de los contenidos audiovisuales establecidos por las Pautas de Accesibilidad al Contenido Web (WCAG) 2.0.

La accesibilidad web es, obviamente, relevante en esta tesis, y hemos querido presentarla como un nuevo lugar en el que la traducción habita y cohabita con otras disciplinas. Los capítulos 4 y 5, que recorren las cuestiones fundamentales en esta materia, nos permiten dibujar el vínculo que une a estas áreas y afirmar, por una parte, que los profesionales de

la traducción tienen que formar parte del proceso de creación de contenidos en la web y se deberían incluir en la cadena de agentes que se ocupa del proceso de creación y evaluación de contenidos audiovisuales en la web. Por otra parte, constatamos que desde la WAI (Iniciativa de Accesibilidad Web) se recomiendan una serie de guías de apoyo, que hemos descrito en este trabajo, para generar subtítulos y audiodescripciones en la web. Esas guías, que de forma prescriptiva indican cómo generar ese contenido alternativo contienen normas que, de nuevo, han vinculado las dos áreas de conocimiento: accesibilidad web y traducción audiovisual.

Por tanto, hemos querido poner de manifiesto que el carácter central del concepto de *norma* también se manifiesta en la accesibilidad audiovisual en la web que, de forma prescriptiva, cuenta con recomendaciones desde distintos ámbitos y geografías que nos han ayudado a mapear la práctica subtituladora en los distintos sistemas.

De hecho, en este sentido, no han sido pocos los esfuerzos realizados para intentar unificar los criterios que dominan la subtitulación; pero la realidad se impone y estamos lejos de estandarizar esta modalidad de traducción audiovisual que se consume de forma global. Por ejemplo, la ESIST⁹⁸ propone un código de buenas prácticas⁹⁹ en subtitulación, redactado por Ivarsson y Carroll (1998) con la intención de que sea de aplicación en el mayor número de países posible; sin embargo, el código es ambiguo en muchos aspectos y eso genera discordancias relevantes. Asimismo, Karamitrouglou (1998) en su artículo “A proposed set of subtitling standards in Europe” ya planteó una serie de pautas con el anhelo de unificar criterios en los países de la Unión Europea para, de esta forma, cohesionar la práctica subtituladora en Europa.

Estas propuestas junto con otras guías, recomendaciones y manuales de subtitulación que han quedado reflejadas en los capítulos 5, 6 y 7, de gran importancia en nuestro estudio, nos han permitido señalar las variables que hemos analizado y describir en qué medida los subtítulos que componen nuestro Archivo, nuestro corpus, coinciden con las indicaciones, pautas o recomendaciones señaladas en estas guías y que son referentes en la práctica subtituladora en Europa.

Con el objetivo de realizar un estudio de estas características hemos creado y analizado el Corpus EMPAC (*EuroparlTV Multimedia Parallel Corpus*) que reúne los subtítulos en inglés y en español de los vídeos emitidos en EuroparlTV entre 2009 y 2017.

⁹⁸ ESIST, European Association for Studies in Screen Translation (www.esist.org)

⁹⁹ Subtitling Code (<http://www.esist.org/ESIST%20Subtitling%20code.htm>)

Queremos poner de manifiesto que esta es la primera vez que se realiza un estudio tan minucioso de la segmentación de los subtítulos mediante el uso de un corpus electrónico de las dimensiones de EMPAC con CQPweb, lo que nos ha permitido extraer los casos de segmentación inadecuada de forma automatizada siguiendo la metodología del grupo de investigación ECPC¹⁰⁰.

Objetivos e hipótesis

Nuestra hipótesis de partida es que el contenido audiovisual (en inglés y español) alojado en el canal de televisión *online* del Parlamento Europeo, EuroParlTV, procura la accesibilidad a través del cumplimiento de los requisitos de la WAI (Iniciativa de accesibilidad web, sus siglas en inglés) y mediante la traducción multilingüe del contenido audiovisual (más en concreto mediante el uso respetuoso de las normas o sugerencias emitidas por organismos o estudiosos en cuanto a los parámetros técnicos de velocidad de lectura, formato y posicionamiento).

De nuestra hipótesis se desprende el siguiente objetivo general:

1. Determinar si los vídeos alojados en la web del Parlamento Europeo son accesibles y valorar si los subtítulos que acompañan a este material audiovisual respetan las normas (técnicas) de esta modalidad de traducción audiovisual.

Para alcanzar el objetivo general de esta tesis, recordamos que hemos propuesto los siguientes objetivos específicos primero en el plano teórico y seguidamente en el plano descriptivo:

Objetivos específicos en el plano teórico

1. Presentar el estado de la cuestión sobre la accesibilidad; en concreto sobre la accesibilidad audiovisual en la web y definir el vínculo que la une a los estudios de traducción.
2. Revisar el marco normativo que regula la subtitulación y las subtitulación para personas sordas (SPS) y mostrar si las recomendaciones que proponen los

¹⁰⁰ El grupo de investigación ECPC (European Comparable and Parallel Corpus of Parliamentary Speeches) es responsable de los siguientes proyectos de investigación con financiación ministerial: HUM2005-03756filo, FFI2008-01610, FFI2012-39389. También es responsable de los siguientes proyectos de financiación por parte de la Generalitat Valenciana o de la Universitat Jaume I: AICO17-082, UJI201-64, UJI2017-25.

organismos de estandarización, guías de buenas prácticas y la academia se ven atendidas en este medio (Internet) y con este tipo de producto audiovisual (vídeos institucionales/traducción institucional).

3. Describir los parámetros técnicos de la subtitulación (velocidad de lectura y formato y posicionamiento, incluida la segmentación) y los valores asignados a cada parámetro y plasmar en qué medida estos valores están presentes en EMPAC

Objetivos específicos en el plano descriptivo

1. Determinar si el contenido audiovisual alojado en EuroparlTV (más en concreto el que hemos seleccionado para conformar el corpus EMPAC) cumple con los requisitos de accesibilidad establecidos por la WAI a partir de las WCAG.
2. Plasmar la evolución de los parámetros (y los valores asignados a cada parámetro) entre 2009 y 2017 en consonancia con la naturaleza diacrónica de nuestro análisis en EMPAC_EN y EMPAC_ES.
3. Describir las coincidencias y divergencias en la presentación de los subtítulos (parámetros técnicos) en EMPAC_EN y EMPAC_ES y su evolución desde el comienzo de las emisiones hasta 2017 (análisis diacrónico).

Recapitulación resultados

Objetivo específico teórico 1

Presentar el estado de la cuestión sobre la accesibilidad; en concreto sobre la accesibilidad audiovisual en la web y definir el vínculo que la une a los estudios de traducción.

Los primeros capítulos de este estudio han servido para vincular, sin fisuras, la accesibilidad y la traducción audiovisual (o textual) y enmarcarlas en los Estudios de Traducción. Este repaso teórico que hemos emprendido nos permite ahora, por una parte, exigir la inclusión de profesionales de la traducción en el proceso de creación de contenidos ya que ha quedado evidenciado que son agentes necesarios en la cadena de la accesibilidad y, por otra parte, nos lleva a cuestionarnos la necesidad de las fronteras (en apariencia) inquebrantables establecidas entre las distintas modalidades/especialidades de traducción reconocidas en el ámbito universitario. Entendemos que, como apuntábamos en el capítulo 1, permeabilizar

las líneas divisorias y flexibilizar el concepto de traducción permitiría establecer puentes entre unas y otras y dotaría a la disciplina del sentido dinámico y cambiante que la define. En lugar de señalar las diferencias consideramos que sería interesante promover estudios que destacaran las similitudes. Son muchos los trabajos que exponen las diferencias de unas modalidades respecto a otras, para así, a nuestro entender, dotarlas de mayor autonomía e independencia, pero ¿cuántos estudios demuestran las conexiones entre la interpretación y la traducción audiovisual? ¿o entre la traducción literaria y la audiodescripción? Quizás, un trabajo conjunto hacia la unidad de la disciplina, sería beneficioso y enriquecedor para todos. Y en este sentido, desde la unidad y la capacidad de expandir las posibilidades de futuros estudios, estamos convenidos de la necesidad de trabajar con otras áreas de conocimiento.

En este sentido, hemos de impulsar el trabajo interdisciplinar para ampliar nuestras fronteras y acercarnos sin miedo a disciplinas más técnicas tales como la creación y desarrollo de contenidos en la web. A nuestro entender, con la intención de procurar calidad en la accesibilidad a los contenidos audiovisuales en web, sería recomendable la implicación de traductores especializados en TAV y accesibilidad. La normativa vigente¹⁰¹, como indicábamos en el capítulo 3, exige que cualquier contenido audiovisual en Internet, para que sea accesible, tiene que incluir SPS y AD.

Objetivo específico teórico 2

Revisar el marco normativo que regula la subtítulos y las subtítulos para personas sordas (SPS) y mostrar si las recomendaciones que proponen los organismos de estandarización, guías de buenas prácticas y la academia se ven atendidas en este medio (Internet) y con este tipo de producto audiovisual (vídeos institucionales/traducción institucional).

Los capítulos 5 y 6 de nuestro trabajo, presentan el marco que define la subtítulos y las subtítulos para personas sordas (SPS) y muestra las recomendaciones y las guías de buenas prácticas que se proponen desde los organismos de estandarización y desde la universidad, en concreto, para Internet. Por otra parte, hemos podido constatar que, efectivamente, las directrices y las normas que se proponen explícitamente para este medio, son coincidentes con las pautas e indicaciones que se hacen en otros canales (TV, DVD, vídeo, etc.). Sin embargo, las indicaciones que desde las WCAG 2.0 se ofrecen en este sentido son muy superficiales. Reconocemos que estas indicaciones (que se recogen en el

¹⁰¹ http://www.w3c.es/Prensa/2012/nota121015_wcag2pas

documento *How to meet WCAG 2.0*) van un paso más allá y constituyen una gran ayuda para los desarrolladores, si bien es cierto que consideramos recomendable elevar el nivel de concreción y responder a preguntas similares a las que proponemos a continuación: ¿cómo se identifica a los hablantes o personajes que aparecen en un vídeo? ¿Y de qué manera se presenta la información contextual y los elementos suprasedgmentales? O bien, ¿cuántos son los caracteres por línea? ¿Cuántos los caracteres por segundo? ¿Cómo segmentar subtítulos? Dejar estas preguntas sin respuesta puede desalentar la tarea de los creadores de contenido pues dificulta la ardua tarea de traducir con las restricciones inherentes a estas modalidades, así como alejarnos de la estandarización necesaria para cumplir con los criterios de calidad.

Por tanto, si bien es cierto que la WAI ofrece indicaciones a partir de las WCAG 2.0 y más en concreto en el documento *How to meet WCAG 2.0*, en el que se definen y evalúan los criterios de éxito, es decir, cómo crear y evaluar, en nuestro caso, la subtitulación, consideramos que sería muy beneficioso incluir información en el documento que acabamos de mencionar sobre los parámetros técnicos, puesto que como ha quedado reflejado en esta tesis, son esenciales en la subtitulación. Los desarrolladores y creadores de contenido en la web utilizan las indicaciones de *How to meet WCAG 2.0* para generar el contenido y, posteriormente, para evaluar si se ha hecho de forma correcta. Las guías de estilo que recomiendan para la subtitulación desde la WAI, las de DCMP, están alineadas con el resto de pautas, normas y recomendaciones que encontramos en subtitulación; sin embargo, no es necesario aplicarlas para cumplir los requisitos de los criterios de éxito.

Objetivo específico teórico 3

Describir los parámetros técnicos de la subtitulación (velocidad de lectura y formato y posicionamiento, incluida la segmentación) y los valores asignados a cada parámetro y plasmar en qué medida estos valores están presentes en EMPAC

En el capítulo 7 resumimos los parámetros en subtitulación y definimos aquellos que estudiamos en profundidad; es decir, los parámetros técnicos de la subtitulación (velocidad de lectura y formato y posicionamiento, incluida la segmentación). En este capítulo planteamos también las variables que vamos a estudiar, al tiempo que hacemos una revisión teórica de las mismas, y también presentamos los valores asignados para describir en qué medida se hallan presentes en el corpus EMPAC.

Objetivo específico descriptivo 1

Determinar si el contenido audiovisual alojado en EuroparlTV; concretamente, si el que hemos seleccionado para conformar el corpus EMPAC cumple con los requisitos de accesibilidad establecidos por la WAI a partir de las WCAG.

El análisis inicial de nuestro estudio, dividido en tres fases, nos permite concluir que:

Fase 1: Análisis para detectar si los vídeos incorporan subtítulos

En el corpus EMPAC_EN el 95,8% de los vídeos incorporan subtítulos mientras que en EMPAC_ES el 98,2% de los vídeos incorporan subtítulos. Los resultados de esta fase indican que, efectivamente, la UE está haciendo grandes esfuerzos por acercar los contenidos a los ciudadanos.

Fase 2: Análisis para detectar si son subtítulos convencionales o subtítulos para personas sordas

Nuestro análisis demuestra que todos los subtítulos de EMPAC_EN y EMPAC_ES son convencionales y no incluyen la información sonora necesaria que los definiría como subtítulos para personas sordas o personas con discapacidad auditiva.

Fase 3: Análisis para determinar si existen audiodescripciones

Nuestro análisis demuestra que todos los vídeos que componen nuestro archivo carecen de audiodescripción.

Este breve resumen de los resultados obtenidos en nuestro análisis puede parecer muy negativo; sin embargo, queremos poner en valor el esfuerzo que desde la UE se hace para procurar que, lingüísticamente, el contenido sea accesible y, por tanto, lo entendemos como un primer paso hacia la accesibilidad plena. Sin embargo, no podemos dejar de constatar que los contenidos audiovisuales alojados en EuroparlTV no cumplen con los requisitos de accesibilidad establecidos por la WAI a partir de las WCAG.

Objetivos específicos descriptivos 2 y 3

2. Plasmar la evolución de los parámetros (y los valores asignados a cada parámetro) entre 2009 y 2017 en consonancia con la naturaleza diacrónica de nuestro análisis en EMPAC_EN y EMPAC_ES.

3. Describir las coincidencias y divergencias respecto a los parámetros técnicos de la presentación de los subtítulos en EMPAC_EN y EMPAC_ES y su evolución desde el comienzo de las emisiones hasta 2017 mediante un análisis diacrónico.

Los objetivos específicos descriptivos 2 y 3 los integraremos en un único apartado, de esta forma daremos respuesta a la descripción de los valores, su evolución histórica y la comparación de las versiones en inglés y en español de forma conjunta.

Atenderemos primero al parámetro velocidad de lectura y a las tres variables que la conforman: número de caracteres por segundo, pausa entre subtítulos y duración. En primer lugar, resumiremos cuantitativamente los resultados obtenidos de cada variable y seguidamente interpretaremos los datos de cada parámetro.

Parámetro velocidad de lectura

Variable número de caracteres por segundo

De los resultados de nuestro análisis obtenemos los siguientes porcentajes en el corpus:

Variable número de caracteres por segundo: análisis global en EMPAC_EN

- El 56,8% de los subtítulos tienen una velocidad de lectura hasta 15 cps.
- El 36,5% de los subtítulos tienen una velocidad de lectura entre 16 y 20 cps.
- El 6,7% de los subtítulos tienen una velocidad de lectura de 21 cps o más.

Variable número de caracteres por segundo: análisis global en EMPAC_ES

- El 61,7% de los subtítulos tienen una velocidad de lectura hasta 15 cps.
- El 31% de los subtítulos tienen una velocidad de lectura entre 16 y 20 cps.
- El 7,3% de los subtítulos tienen una velocidad de lectura de 21 cps o más.

Variable número de caracteres por segundo: análisis por años en EMPAC_EN

- Si bien en 2009 el porcentaje de subtítulos con 15 cps se situaba en el 71% en 2017 desciende al 41,1% evidenciando la tendencia al aumento de la velocidad de lectura.

- El valor entre 16 y 20 cps suponía en 2009 el 26% y alcanza un 41,7% en 2017.
- Por último, el porcentaje de subtítulo con velocidades superiores a los 21 cps pasa del 3 al 17,2%.

Variable número de caracteres por segundo: análisis por años en EMPAC_EN

- En 2009, el porcentaje de subtítulos con 15 cps se situaba en el 74,8% en 2017 descendiendo al 30,2% y evidencia la tendencia al aumento de la velocidad de lectura.
- El valor entre 16 y 20 cps suponía en 2009 el 22,5% y alcanza un 33,7% en 2017.
- Por último, el porcentaje de subtítulo con velocidades superiores a los 21 cps pasa del 2,7 al 36,1%.

Variable número de caracteres por segundo: análisis por tipo de texto en EMPAC_EN

- No hemos encontrado diferencias significativas entre los tipos de texto de nuestro corpus (*Background, Discovery, History, Interview, News*)
- Podemos quizás señalar que *News* con 51,6% de subtítulos con el valor 15 cps, 40,1% en el valor 20 cps y un 8,3% de subtítulos con 21 cps y *Discovery* con 51,5% de subtítulos con el valor 15 cps, 40,7% en el valor 20 cps y un 7,8% de subtítulos con 21 cps tienen velocidades ligeramente más exigentes que el resto.

Variable número de caracteres por segundo: análisis por tipo de texto en EMPAC_ES

- En la versión en español tampoco hemos encontrado diferencias significativas entre los tipos de texto de nuestro corpus (*Background, Discovery, History, Interview, News*)
- Asimismo, *Discovery* y *News* tienen velocidades ligeramente más exigentes que el resto.

Variable pausa entre subtítulos

De los resultados de nuestro análisis obtenemos los siguientes porcentajes en el corpus

Variable pausa entre subtítulos: análisis global en EMPAC_EN

- El 78,6% de los subtítulos tienen una pausa de $\leq 0,16$.

- El 18,3% de los subtítulos tienen una pausa de $\geq 0,16$.
- El 3,1% de los subtítulos no tienen pausa alguna.

Variable pausa entre subtítulos: análisis global *en EMPAC_ES*

- El 79% de los subtítulos tienen una pausa de $\leq 0,16$.
- El 18,1% de los subtítulos tienen una pausa de $\geq 0,16$.
- El 2,9% de los subtítulos no tienen pausa alguna.

Variable pausa entre subtítulos: análisis por años *en EMPAC_EN*

- En 2009 el 84,1% de los subtítulos tienen una pausa de $\leq 0,16$ y desciende hasta el 13,8% en 2017.
- El valor $\geq 0,16$ suponía en 2009 el 14,2% y se mantienen estable a lo largo de los años (entre el 14% y el 18%) En 2017 el 18,1% de los subtítulos tienen una pausa de $\geq 0,16$.
- Por último, el porcentaje de subtítulo con pausa 0 pasa del 1,7% en 2009 al 68,1% en 2017. Sin embargo, es en 2016 cuando comienza el repunte de subtítulos sin pausa. Los años anteriores los porcentajes oscilaban entre el 1% y el 2%, mientras que en 2016 cambia al 8,8%.

Variable pausa entre subtítulos: análisis por años *en EMPAC_ES*

- En 2009 el 85,7% de los subtítulos tienen una pausa de $\leq 0,16$ y desciende hasta el 16,5% en 2017.
- El valor $\geq 0,16$ suponía en 2009 el 13% y se mantienen estable a lo largo de los años (entre el 13% y el 21%). En 2017 el 19,2% de los subtítulos tienen una pausa de $\geq 0,16$.
- Por último, el porcentaje de subtítulo con pausa 0 pasa del 1,2% en 2009 al 64,3% en 2017. Sin embargo, es en 2016 cuando comienza el repunte de subtítulos sin pausa. Los años anteriores los porcentajes oscilaban entre el 1% y el 2%, mientras que en 2016 cambia al 9,1%.

Variable pausa entre subtítulos: análisis por tipo de texto en *EMPAC_EN*

- El tipo de texto *News* tiene el porcentaje más alto (6% frente al 1 y 2% del resto de tipos de texto) de subtítulos sin pausa.
- En *Discovery* e *History*, encontramos los porcentajes (23% y 21% respectivamente) de pausas por encima de 0,16 milésimas ($\geq 0,16$) más altos.

Variable duración de los subtítulos

Variable duración de los subtítulos: análisis global en *EMPAC_EN*

- El 0,1% de los subtítulos tienen una duración de > 1 sg.
- El 93,3% de los subtítulos tienen una duración de 1 y 6 sg.
- El 6,6% de los subtítulos tienen una duración < 6 sg.

Variable duración de los subtítulos: análisis global en *EMPAC_ES*

- El 0,1% de los subtítulos tienen una duración de > 1 sg.
- El 93,2% de los subtítulos tienen una duración de 1 y 6 sg.
- El 6,7% de los subtítulos tienen una duración < 6 sg.

Variable duración de los subtítulos: análisis por años en *EMPAC_EN*

- En 2009 no hay subtítulos con una duración de menos de un segundo y los porcentajes se reparten de la siguiente forma: el 89,6% permanecen en pantalla entre 1 y 6 segundos y 10,3% por encima de los 6 segundos. Hasta el 2016 los porcentajes son prácticamente iguales a los de 2009.
- En 2017, se produce un pequeño, aunque significativo, repunte en los subtítulos que duran menos de 1 segundo (1,2%) y una reducción considerable en los de más de 6 segundos (2,9%).

Variable duración de los subtítulos: análisis por años en *EMPAC_ES*

- En 2009 no hay subtítulos con una duración de menos de un segundo y los porcentajes se reparten de la siguiente forma: el 90,3% permanecen en pantalla

entre 1 y 6 segundos y 9,6% por encima de los 6 segundos. Hasta el 2016 los porcentajes son prácticamente iguales a los de 2009.

- En 2017, se produce un pequeño, aunque significativo, repunte en los subtítulos que duran menos de 1 segundo (1,1%) y una reducción considerable en los de más de 6 segundos (2,8%).

Variable duración de los subtítulos: análisis por tipo de texto *en EMPAC_EN*

- En todos los tipos de texto el porcentaje de subtítulos que tienen una duración de > 1 sg está entre el 0,1% y el 0,2%.
- En todos los tipos de texto el porcentaje de subtítulos que tienen una duración entre 1 y 6 sg está entre el 91% y el 93%, excepto en *Discovery* que alcanza el 97%.
- El 9% de los subtítulos del tipo de texto *Interview* tienen una duración por encima de los 6 segundos, mientras que el resto de tipos está entre el 3% y el 7%.

Variable duración de los subtítulos: análisis por tipo de texto *en EMPAC_ES*

- En todos los tipos de texto el porcentaje de subtítulos que tienen una duración de > 1 sg está entre el 0,1% y el 0,3%.
- En todos los tipos de texto el porcentaje de subtítulos que tienen una duración entre 1 y 6 sg está entre el 91% y el 93%, excepto en *Discovery* que alcanza el 96,8%.
- El 8,9% de los subtítulos del tipo de texto *Interview* tienen una duración por encima de los 6 segundos, mientras que el resto de tipos está entre el 3% y el 7%.

Variable duración de los subtítulos: análisis global de subtítulos cortos, medios, largos *en EMPAC_EN*

Esta variable, que ha sido analizada con el único objetivo de describir qué duración es la más habitual en los subtítulos de nuestro corpus nos muestra que:

- El 1,8% de los subtítulo tiene una duración de entre 0 y 2 segundos (cortos).
- El 54,1% de los subtítulo tiene una duración de entre 2 y 4 segundos (medios).
- El 36,6% de los subtítulo tiene una duración de entre 4 y 6 segundos (largos).
- El 7,4% de los subtítulo tiene una duración por encima de los 6 segundos.

Variable duración de los subtítulos: análisis por años de subtítulos cortos, medios, largos en *EMPAC_EN*

- En 2009 el 1,7% de los subtítulos eran cortos, el 46,5% medios y el 40,3% largos, mientras que un 11,5% superaban los 6 segundos de permanencia en pantalla.
- En 2017 se produce un incremento muy importante en el grupo de subtítulos cortos alcanzando el 12,8%.
- El grupo de subtítulos medios, largos y por encima de los 6 segundos aumenta gradualmente desde 2009 a 2017.

Variable duración de los subtítulos: análisis de subtítulos cortos, medios, largos y cps *EMPAC_EN*

- El 59% de los subtítulos cortos se exponen a velocidades de lectura de hasta 15 cps, el 20,8% entre 16 y 20 cps y el 20,2% a 21 cps o más.
- El 45,5% de los subtítulos medios se exponen a velocidades de lectura de hasta 15 cps, el 43,5% entre 16 y 20 cps y el 10,9% a 21 cps o más.
- El 64,6% de los subtítulos largos se exponen a velocidades de lectura de hasta 15 cps, el 34,2% entre 16 y 20 cps y el 1,2% a 21 cps o más.
- El 99,8% de los subtítulos largos se exponen a velocidades de lectura de hasta 15 cps y tan sólo el 1,2% entre 16 y 20 cps.

Conclusiones Parámetro velocidad de lectura

Tras la recapitulación de los resultados obtenidos en la variable velocidad de lectura en cps, podemos afirmar que casi la mitad del corpus *EMPAC* supera los 15 cps, es decir, están por encima de lo que marca la norma UNE y DCMP. Sin embargo, queremos apuntar que en *EMPAC_EN* la velocidad de lectura es menos exigente que en *EMPAC_ES*, por lo que los subtítulos serían menos accesibles en la versión en español. De hecho, es preocupante que el 36,1% de los subtítulos de *EMPAC_ES* en 2017 tengan un valor de 21 cps o más y tan solo el 30,2% tengan un valor de 15 cps. Con todo, la tendencia es que el número de cps aumenta con los años y podemos afirmar que son menos accesibles.

Los resultados obtenidos en la variable pausa entre subtítulos indican, a partir de los datos totales del corpus, que tan sólo un 1,7% (EN) y un 1,2% (ES) no respetan la norma de dejar un espacio de al menos 4 *frames* (0,16 milésimas de segundo) por lo que podemos afirmar que se cumple la norma de pausa entre subtítulos. Sin embargo, en 2016 se aprecia un cambio en la tendencia y es en 2017 cuando se produce un cambio radical y más del 60% de los subtítulos no tienen pausa, tanto en EMPAC_EN como en EMPAC_ES.

La última variable que analizamos en este parámetro, duración de los subtítulos, indica que tanto en EMPAC_EN como en EMPAC_ES prácticamente la totalidad de los subtítulos tienen una duración entre 1 y 6 segundos, por lo que esta variable sí que está alineada con los valores propuestos en este estudio. Sin embargo, el análisis pormenorizado por años manifiesta que el número de subtítulos con una duración inferior a 1 segundo repunta en 2017 en ambas versiones. Aquellos subtítulos con una duración superior a los 6 segundos es estable a lo largo de los años, excepto en 2017 que se reduce considerablemente.

Por último, en nuestro corpus EMPAC, la duración más habitual de los subtítulos es de entre 2 y 4 segundos (54,1%) seguidos del grupo de 4 a 6 segundos (36,6%). Sin embargo, en 2017, el grupo de 2 a 4 segundos es superior respecto al total y casi el 60% de los subtítulos son medios y el 12,8% cortos, por lo que en este año casi el 73% de los subtítulos tiene una duración máxima de 4 segundos.

La diferencia entre las versiones en inglés y en español es escasísima puesto que se trabaja con una *Master File* y la duración de los subtítulos viene determinada por igual para ambas versiones de subtítulos.

Por último, en los resultados del análisis por tipo de texto de las distintas variables no se aprecian diferencias significativas entre ellos.

Parámetro formato y posicionamiento

Variable número de líneas

Variable número de líneas: análisis global del corpus EMPAC_EN

- El 23,8% de los subtítulos tienen una línea.
- El 76,2% de los subtítulos tienen dos líneas.
- El 0% de los subtítulos tienen una línea.

Variable número de líneas: análisis global del corpus EMPAC_ES

- El 25,1% de los subtítulos tienen una línea.
- El 74,9% de los subtítulos tienen dos líneas.
- El 0% de los subtítulos tienen una línea.

Variable número de líneas: análisis por años del corpus EMPAC_EN

- Entre 2009 y 2016 más del 73% de los subtítulos tienen dos líneas.
- En 2017 hasta el 83,2% de los subtítulos pasan a tener una línea.

Variable número de líneas: análisis por años del corpus EMPAC_EN

- Entre 2009 y 2016 más del 79,3% de los subtítulos tienen dos líneas.
- En 2017 hasta el 79,3% de los subtítulos pasan a tener una línea.

Variable número de líneas: análisis por tipo de texto del corpus EMPAC_EN

- En todos los tipo de texto (programas) se replican los resultados globales con porcentajes por encima del 70% de los subtítulos de dos líneas y el resto de una línea.

Variable número de líneas: análisis por tipo de texto del corpus EMPAC_ES

- En todos los tipo de texto (programas) se replican los resultados globales con porcentajes por encima del 70% de los subtítulos de dos líneas y el resto de una línea.

Variable número de caracteres por línea

Variable número de caracteres por línea: análisis global del corpus EMPAC_EN

- El 56,3% de los subtítulos tienen 32 cpl.
- El 22,5% de los subtítulos tienen 37 cpl.
- El 17,3% de los subtítulos tienen 42 cpl.
- El 3,9% de los subtítulos tienen ≥ 43 cpl.

Variable número de caracteres por línea: análisis global del corpus EMPAC_ES

- El 61,1% de los subtítulos tienen 32 cpl.
- El 22,8% de los subtítulos tienen 37 cpl.
- El 14% de los subtítulos tienen 42 cpl.
- El 2,1% de los subtítulos tienen ≥ 43 cpl.

Variable número de caracteres por línea: análisis por años del corpus EMPAC_EN

- Entre 2009 y 2016 los subtítulos con 32 cpl suponían entre un 59% y un 53% del total. En 2017 el porcentaje desciende al 34,1%.
- Entre 2009 y 2016 los subtítulos con 37 cpl suponían entre un 23% y un 21% del total. En 2017 el porcentaje desciende al 11,9%.
- Entre 2009 y 2016 los subtítulos con 42 cpl suponían entre un 15% y un 20% del total. En 2017 el porcentaje desciende al 9,6%.
- Entre 2009 y 2016 los subtítulos con ≥ 43 suponían entre un 1% y un 8% del total. En 2017 el porcentaje asciende al 44,4%.

Variable número de caracteres por línea: análisis por años del corpus EMPAC_ES

- Entre 2009 y 2016 los subtítulos con 32 cpl suponían entre un 68% y un 50% del total. En 2017 el porcentaje desciende al 34,1%.
- Entre 2009 y 2016 los subtítulos con 37 cpl suponían entre un 21% y un 25% del total. En 2017 el porcentaje desciende al 11,9%.
- Entre 2009 y 2016 los subtítulos con 42 cpl suponían entre un 10% y un 18% del total. En 2017 el porcentaje desciende al 9,6%.
- Entre 2009 y 2016 los subtítulos con ≥ 43 suponían entre un 0,3% y un 5% del total. En 2017 el porcentaje asciende al 49,4%.

Variable número de caracteres por línea: análisis por tipo de texto del corpus EMPAC_EN

- El todos los tipos de texto los valores más utilizado son 32 cpl y (entre 56% y 60%) 37 cpl (22% y 21%)

- Los valores 42 (entre 56% y 60%) y ≥ 43 (entre 1% y 2%) mantienen los porcentajes más bajos.
- Observamos que, tanto en EMPAC_EN como en EMPAC_ES, desde 2016 hasta 2017 se produce un cambio en la tendencia de líneas: descienden las de 32, 37 y 42 caracteres mientras que ascienden las de 43 cpl o más caracteres.

Variable causas de segmentación inadecuada

Los resultados cuantitativos del análisis de la segmentación se presentan en tantos por mil, dada la escasa incidencia de casos.

Análisis de la segmentación en inglés (EMPAC_EN) entre líneas de subtítulo: segmentación intrasubtítulo:

- El 7,7 ‰ de las causas de segmentación inadecuada son de la categoría conjunción.
- El 14,8 ‰ de las causas de segmentación inadecuada son de la categoría determinante.
- El 37,7 ‰ de las causas de segmentación inadecuada son de la categoría frase.
- El 22, ‰ de las causas de segmentación inadecuada son de la categoría nombre compuesto.
- El 40,2 ‰ de las causas de segmentación inadecuada son de la categoría preposición.
- El 25,5 ‰ de las causas de segmentación inadecuada son de la categoría sintagma.
- El 59,5 ‰ de las causas de segmentación inadecuada son de la categoría verbo compuesto.

Análisis de la segmentación en español (EMPAC_ES) entre líneas de subtítulo: segmentación intrasubtítulo

- El 13,3 ‰ de las causas de segmentación inadecuada son de la categoría conjunción.
- El 8,5 ‰ de las causas de segmentación inadecuada son de la categoría determinante.
- El 38,8 ‰ de las causas de segmentación inadecuada son de la categoría frase.

- El 8,7 ‰ de las causas de segmentación inadecuada son de la categoría nombre compuesto.
- El 11,6 ‰ de las causas de segmentación inadecuada son de la categoría preposición.
- El 28,9 ‰ de las causas de segmentación inadecuada son de la categoría sintagma.
- El 30,7 ‰ de las causas de segmentación inadecuada son de la categoría verbo compuesto.

Análisis de la segmentación en inglés entre subtítulos: segmentación intersubtítulo

- El 0,04 ‰ de las causas de segmentación inadecuada son de la categoría conjunción.
- El 0,1 ‰ de las causas de segmentación inadecuada son de la categoría determinante.
- El 6,2 ‰ de las causas de segmentación inadecuada son de la categoría nombre compuesto.
- El 2 ‰ de las causas de segmentación inadecuada son de la categoría preposición.
- El 1,6 ‰ de las causas de segmentación inadecuada son de la categoría sintagma.
- El 23,9 ‰ de las causas de segmentación inadecuada son de la categoría verbo compuesto.

Análisis de la segmentación en español entre subtítulos: segmentación intersubtítulo

- El 0,8 ‰ de las causas de segmentación inadecuada son de la categoría conjunción.
- El 0,03 ‰ de las causas de segmentación inadecuada son de la categoría determinante.
- El 16,2 ‰ de las causas de segmentación inadecuada son de la categoría nombre compuesto.
- El 0,1 ‰ de las causas de segmentación inadecuada son de la categoría preposición.
- El 1 ‰ de las causas de segmentación inadecuada son de la categoría sintagma.
- El 5,3 ‰ de las causas de segmentación inadecuada son de la categoría verbo compuesto.

El análisis de este parámetro revela que en el caso de la variable número de líneas por subtítulo, tanto en EMPAC_EN como en EMPAC_ES, el porcentaje de subtítulos de dos líneas, en el cómputo total del corpus, es muy superior al de una línea. Es a su vez significativo que tan solo haya un subtítulo de tres líneas en todo el corpus. Sin embargo, en 2016, la tendencia comienza a cambiar y en 2017 se invierte por completo, tanto es así que el 80% de los subtítulos son entonces de una línea. Por último, no detectamos diferencia destacable atendiendo a la variable tipo de texto en la versión en inglés ni en la versión en español.

La variable número de caracteres por línea (cpl) en EMPAC_EN en términos totales muestra que el 78,8% de los subtítulos tienen el valor 32 cpl o 37 cpl, lo que supone un porcentaje muy alto de subtítulos que siguen la norma establecida de no superar los 37 cpl. En el caso de EMPAC_ES también se da esta circunstancia con el 83,9% de los subtítulos entre estos dos valores. Sin embargo, la evolución histórica de los subtítulos desvela que en EMPAC_EN se pasa del 59,5% de las líneas con 32 cpl en 2009 a, en 2017, el 44,4% de las líneas con ≤ 43 cpl. Distinguimos un comportamiento muy parecido en EMPAC_ES; sin embargo, los cambios son sensiblemente más bruscos, y mientras que en 2009 tan sólo el 0,3% de las líneas tenían el valor ≤ 43 cpl, en 2017 asciende al 49,4%. Por su parte, en el análisis según el tipo de texto no apreciamos diferencias significativas entre EMPAC_EN y EMPAC_ES, y los valores 32 cpl y 37 cpl son los de mayor porcentaje. Con todo, apreciamos que en la versión en inglés existe un porcentaje superior de líneas con el valor 42 cpl y 43 cpl que en la versión en español.

Los resultados de la tercera variable, causas de segmentación inadecuada de los subtítulos, en términos generales y cuantitativos, muestran la escasa incidencia de segmentación inadecuada en EMPAC_EN y EMPAC_ES, de hecho, hemos normalizado los datos por mil para que fuera más sencillo leer los resultados.

En cuanto a la segmentación intrasubtítulo en EMPAC_EN observamos que los verbos compuestos (59,5 ‰) y las preposiciones (40,2 ‰) son las categorías donde más casos encontramos seguidas de la categoría frase (37 ‰). En el caso de la segmentación intrasubtítulo en EMPAC_ES, la categoría con más incidencias es también la de verbo compuesto (30,8 ‰) seguido de sintagma (29 ‰) y de nuevo el tercer lugar lo ocupa la categoría frase (11,6 ‰). Aunque los casos de segmentación inadecuada en inglés y en español son muy pocos, sí que apreciamos que la segmentación en español parece estar ligeramente más cuidada que en inglés.

Por su parte, los casos segmentación intersubtítulo en EMPAC_EN y EMPAC_ES, son testimoniales y como explicamos en el análisis de los datos en el capítulo 11, un análisis manual de las pocas ocurrencias que detectamos, mostró que se debían a errores en el POS.

Conclusiones finales

Entre 2009 y 2016 los porcentajes de cps son similares en EMPAC_EN y EMPAC_ES. Se observa una evolución progresiva hacia subtítulos más rápidos en ambas versiones, pero los subtítulos en inglés, hasta 2016, siempre son un poco más rápidos que los españoles. Pero 2016 es el primer año en el que los subtítulos en español son ligeramente más exigentes y en 2017 se produce un cambio sustancial no solo respecto al resto de años sino también entre EMPAC_EN y EMPAC_ES. De hecho, solo un 30,2% de los subtítulos en español respetan la norma de 15 cps y el 36,1% tienen 21 caracteres o más. En EMPAC_EN, por su parte, si bien el porcentaje de subtítulos con 20 cps se ha doblado respecto a 2009, todavía un 41,1% estarían en el valor normativo (15 cps), y el 17,2% con 21 cps o más.

Esta evolución nos lleva a pensar que en 2009 los subtítulos en español se editaban más que los subtítulos en inglés que, por ser la lengua pivote de las *Master File*, tiende desde sus orígenes a estar más cerca de la transcripción *verbatim* que de la subtitulación. De hecho, por lo que hemos podido comprobar, la mayor parte de los subtítulos en inglés son transcripciones *verbatim*, que con el paso del tiempo simplemente han aumentado la cantidad de cps. Sin embargo, en el caso de los subtítulos en español, los subtítulos están editados hasta 2016 y es aquí donde parece que cambia la forma de generar subtítulos y dejan de editarse, por lo que, si a este hecho le sumamos que los tiempos de permanencia en pantalla de los subtítulos se ha reducido, esto explicaría los porcentajes descritos en el párrafo anterior.

De igual forma, la pausa entre subtítulos, también experimenta un cambio radical en 2017, si bien comienza en 2016, al pasar de tener tan solo un 1% aprox. de subtítulos sin pausa a que más del 60% de los subtítulos no dejen tiempo de transición entre uno y otro, lo que dificulta la lectura y aumenta la velocidad de lectura.

Por su parte, los resultados obtenidos en la variable duración, en la que valoramos si los subtítulos se mantienen en pantalla la horquilla de tiempo que dispone la norma, es decir, entre 1 y 6 segundos, todos los años tienen entre tan solo un 0% o un 0,1% de subtítulos que permanecen en pantalla menos de un segundo, sin embargo se aprecia un ligero repunte en

este tipo de subtítulos en 2017. Sin embargo, por los porcentajes obtenidos, en este caso, constatamos que respetan la conocida regla de los seis segundos.

El análisis del parámetro formato y posicionamiento revela que, en el caso de la variable número de líneas, hasta 2016 entre el 22% y el 26% los subtítulos son de una línea, en cambio, en 2017 pasan a ser el 80% de los subtítulos. Por su parte, el número caracteres por línea, la segunda variable de este parámetro, la evolución histórica de los subtítulos desvela que en EMPAC_EN aumenta considerablemente los cpl en 2017 y se aleja del estándar de 37 cpl, pero EMPAC_ES el aumento es mayor, tanto es así que casi el 50% de los subtítulos superan los 43 cpl.

La web de EuroparlTV, como plasmamos en el capítulo 9, ha experimentado desde su creación hasta 2017 dos cambios importantes; el primero afectó en 2010 a la visualización y a la estética de la web; el segundo se produjo en julio de 2016, cuando la empresa ICF Mostra dejó de encargarse de los subtítulos de EuroparlTV y esta tarea pasó a manos de CPSL. Hay una clara coincidencia en el hecho de que los primeros cambios bruscos en los parámetros técnicos que se aprecian en nuestro corpus corresponden a estas fechas, a los últimos meses de 2016.

Creemos que los avances tecnológicos han permitido que, desde finales de 2016, los subtítulos de EuroparlTV se generen de forma semiautomática. Así pues, el proceso sería primero una transcripción en inglés del vídeo; posteriormente, ese texto se traduce al resto de lenguas, pero sin que los traductores tengan en cuenta los parámetros técnicos de la subtitulación. Después, el texto se vincula al vídeo y, a partir del pautado en inglés (que se genera de forma automática y sin tener en cuenta los estándares comúnmente aceptados en subtitulación), aparecen los subtítulos en pantalla. Por lo que hemos podido comprobar, los subtítulos abarcan todo el ancho de la imagen y cambian de línea solo cuando la línea superior se completa.

Esto explicaría el cambio sustancial que hemos evidenciado en los parámetros técnicos que hemos analizado y muestra un panorama actual en el que los subtítulos de la web de EuroparlTV son mucho menos respetuosos con las normas que definen esta modalidad de traducción.

Por tanto, podemos afirmar que los vídeos alojados en la web del Parlamento Europeo no son accesibles para personas con discapacidad sensorial, si bien sí que son accesibles en términos lingüísticos. No obstante, actualmente los parámetros técnicos no están en absoluto alineados con las normas en subtitulación.

Por otra parte, en términos globales, algunos de los parámetros estudiados sí que respetan las normas de la subtítulos, sin embargo, la evolución histórica ha demostrado que no solo las velocidades han aumentado, sino que, desgraciadamente, a día de hoy, ninguno de los parámetros están alineados con la norma.

Futuras líneas de investigación

Consideramos que la tesis doctoral que aquí presentamos puede ser la base para futuras líneas de investigación. Concretamente, nos gustaría poder desarrollar algunos de los siguientes aspectos que, por motivos de espacio, hemos presentado de forma sintética o parcial en este trabajo:

- Estudio en mayor profundidad de la segmentación de los subtítulos.
- Estudios de parámetros de la subtítulos distintos a los estudiados en esta tesis.
- Estudio puramente descriptivo del procesamiento semántico de los componentes de la frase que permita desvelar cómo agrupan los traductores de forma cognitiva los bloques significativos de la frase.
- Estudio comparativo de las *keywords* de EMPAC con las de su corpus hermano ECPC.

La investigación que hemos llevado a cabo con nuestro corpus ha sido, sin duda, intenso y extenso, y nos permite tener un horizonte de expectativas realmente amplio respecto a las posibilidades de explotación del mismo, no únicamente con una finalidad descriptiva sino también con una perspectiva aplicada a la didáctica de la traducción audiovisual y la accesibilidad. Este es, sin duda, uno de los retos que nos planteamos en el futuro más próximo: continuar nuestra investigación para mejorar, en la medida de nuestras posibilidades, el conocimiento de la TAV y la formación de futuros traductores que puedan ofrecer a nuestra sociedad un trabajo de calidad.

BIBLIOGRAFÍA

- ACCEPLAN: Plan de accesibilidad 2003-2010. Libro Blanco. [En línea]. Centro Documental del CEAPAT. Recuperado de <http://www.ceapat.org/mostrardocumento.do?idDoc=200706280027&idSeccion=49>
- AENOR (2004). *Aplicaciones informáticas para personas con discapacidad. Requisitos de accesibilidad para contenidos en la web*. Norma UNE 13980. Madrid: Aenor.
- AENOR (2005). *Audiodescripción para personas con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguías*. Norma UNE: 153020. Madrid: Aenor.
- AENOR (2012). *Subtitulado para personas sordas y personas con discapacidad auditiva*. Norma UNE 153010. Madrid: Aenor.
- Agost de Canós, R. (1999). *Traducción y doblaje: palabras, voces e imágenes*. Barcelona: Ariel.
- Alonso López, F. (Coord.). (2003). *Libro Blanco de la accesibilidad. Por un nuevo paradigma, el Diseño para Todos, hacia la plena igualdad de oportunidades*.
- Arnáiz Uzquiza, V. (2012). Los parámetros que identifican el subtitulado para sordos. Análisis y clasificación. *Monti*, 4, pp. 103-132.
- Arnáiz Uzquiza, V. (2015). Eye tracking Tests in Spain. En Romero-Fresco, P. (Ed.) *The reception of subtitles for the deaf and hard of hearing in Europe*. Berlin: Peter Lang.
- Baker, M. (1992). *In Other Words: A Course Book on Translation*. London: Routledge.

- Baker, M. (1993). *Corpus Linguistics and Translation Studies: Implications and Applications*. En Baker, M. *et al.* (1993). *Text and Technology: In Honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins.
- Baker, M. (1995). Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target. International Journal of Translation Studies*, 7(2). Ámsterdam: John Benjamins Publishing Company, pp 223-243.
- Ballester Casado, A. R. (2001). *Traducción y nacionalismo. La recepción del cine americano en España a través del doblaje (1928-1948)*. Granada: Comares.
- Ballester, A. (2007). La Audiodescripción: apuntes sobre el estado de la cuestión y las perspectivas de investigación. *Tradterm: 13*, pp. 151-169. Recuperado de <http://www.revistas.usp.br/tradterm/article/view/47471>
- Bartoll, E. (2008). *Paramètres per a una taxonomia de la subtitulació*. (Tesis doctoral). Universitat Pompeu Fabra. Barcelona. Recuperado de <http://hdl.handle.net/10803/7572>
- BBC (2009): *Online Subtitling Editorial Guidelines V1.1*. Williams, F.
- Bird, S., Klein, E. & Loper, E. (2009). *Natural Language Processing with Python*. Sebastopol: O'Reilly Media Inc.
- Bowker, L & Pearson, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.
- Brondeel, H. 1994. Teaching Subtitling Routines. *Meta: Translators' Journal*, 39(1), pp. 26-33.
- Calzada Pérez, M. (2007). *El espejo traductológico. Teorías y didácticas para la formación del traductor*. Madrid: Editorial Octaedro.
- Camacho, K. (2005). La brecha digital. En Ambrosi, A., Peugeot, V. y Pimienta, D. (Coords.). *Palabras en juego. Enfoques multiculturales sobre las sociedades de la Información*. Caen: C&F Éditions.
- Cambra, C., Silvestre, N. & Leal, A. (2013). La interpretación de los adolescentes sordos y oyentes de un documental audiovisual: importancia del subtitulado. *Revista de Logopedia, Foniatría y Audiología*, 33(3), 99-108.

- Castells, M. (1998) *La era de la información, Economía, Sociedad y Cultura; La sociedad Red*, Madrid: Alianza Editorial.
- Castells, M. (2000) Tecnologías de la información y desarrollo global. *Estudios de política exterior*, 14, pp.151-168.
- Castro Roig, X. (2004). Solo ante el subtítulo: experiencias de un subtitulador. *La Linterna del Traductor*, 9.
- Cerezo, B. (2012). *La didáctica de la traducción audiovisual en España: Un estudio de caso empírico-descriptivo*. (Tesis doctoral). Castelló de la Plana: Universitat Jaume I. Recuperado de <http://www.tdx.cat/handle/10803/83363>.
- CERMI (2010). *Los jóvenes con discapacidad en España. Informe de situación 2010*. Madrid: Cinca. Recuperado de http://www.intersocial.es/images/stories/pdf/jovenes_con_discapacidad.pdf
- Chaume Varela, F. (1997). La traducción audiovisual: estado de la cuestión. En Vega, M. A. y Martín-Gaitero, M. (Eds.). *La palabra vertida. Investigaciones en torno a la traducción*. Madrid: Editorial Complutense/Ediciones del Orto.
- Chaume Varela, F. (1999). La traducción audiovisual: investigación y docencia. *Perspectives: Studies in Translatology*, 7(2), pp. 209-219.
- Chaume Varela, F. (2003). *Doblatge i subtitulació per a la TV*. Barcelona: Eumo.
- Chaume Varela, F. (2004). *Cine y traducción*. Madrid: Cátedra.
- Chaume Varela, F. (2013). The turn of audiovisual translation: New audiences and new technologies. *Translation Spaces* (2), pp. 107-125.
- Chesterman, A. (1993). From 'is' to 'Ought': Laws, Norms and Strategies in Translation studies. En *Target*, 5, pp. 1-20.
- Chesterman, A. (1997). *Memes of Translation. The Spread of Ideas in Translation Theory*. Amsterdam/Philadelphia: John Benjamins.
- Chica Núñez, A.J. & Rodríguez, A. (2014) Plataforma on-line de evaluación de productos audiovisuales. PRA2. En Amadis 2014. Recuperado de <http://www.cesya.es/sites/default/files/documentos/amadis2014.pdf>

- Clark, J. (2004). *Best practices in online captioning*. Recuperado de <http://joelclark.org/access/captioning/bpoc/>.
- Constitución Española. (1978, 31 de octubre). Boletín Oficial del Estado, n.º 311, de 29 de diciembre de 1978.
- Corpas Pastor, G. (2012). Corpus, tecnología y traducción. En Casas, M. (Dir.) y García Antuña, M. (Ed.). *XII Jornadas de Lingüística*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- Crovi Druetta, D. (2002). Sociedad de la información y el conocimiento. Entre el optimismo y la desesperanza. En *Revista Mexicana de Ciencias Políticas y Sociales* XLV (mayo-agosto). Recuperado de <http://www.redalyc.org/articulo.oa?id=42118502>
- Cumbre Mundial sobre la Sociedad de la Información (CMSI) (2004a), Declaración de Principios: Construir la Sociedad de la Información: un desafío global para el nuevo milenio, Ginebra, mimeo, documento WSIS-03/GENEVA/4- S, 10 pp.
- DCMP (2009). *Captioning Key. Guidelines and preferred techniques*. Recuperado de http://www.dcmp.org/captioningkey/about_c.html.
- Delabastita, D. (1989). Translation and mass-communication: film and T.V. translation as evidence of cultural dynamics. En *Babel*, 35(4), pp. 193-218.
- Delabastita, D. (1990) Translation and the Mass Media. En Bassnett, S. y Lefevere, A. *Translation, History and Culture*. Londres: Pinter, pp 97-109.
- Díaz Cintas, J. (2001). *La traducción audiovisual. El subtítulo*. Salamanca: Almar.
- Díaz Cintas, J. (2003). *Teoría y práctica de la subtitulación (inglés-español)*. Barcelona: Ariel.
- Díaz Cintas, J. (2007a). Traducción audiovisual y accesibilidad. En Jiménez Hurtado, C. (Ed.). *Traducción y accesibilidad: Subtitulación para sordos y audiodescripción para ciegos: nuevas modalidades de traducción audiovisual*. Frankfurt: Peter Lang.
- Díaz Cintas, J. (2007b). Por una preparación de calidad en accesibilidad audiovisual. En *Trans*, II, pp. 45-49.
- Díaz Cintas, J. (2009). *New Trends in Audiovisual Translation*. Búfalo: Multilingual Matters.

- Díaz Cintas, J. (2012). Los subtítulos y la subtitulación en clase de lengua extranjera. *Abehache*. 3. Recuperado de http://discovery.ucl.ac.uk/1451727/1/Diaz-Cintas_abehache_12.pdf
- Díaz Cintas, J. & Orero, P. (2003). Postgraduate Courses in Audiovisual Translation. En *The Translator*, 9(2), pp. 371-388.
- Díaz Cintas, J. & Remael, A. (2007). *Audiovisual Translation: Subtitling*. Manchester/Kinderhook: St. Jerome Publishing.
- Díaz Cintas, J., Orero, P. & Remael, A. (Eds.) (2007). *Media for All: Subtitling for the Deaf, Audio Description and Sign Language*. Amsterdam/New York: Rodopi.
- D'Ydewalle, G., Van Rensbergen, J. & Pollet, J. (1987). Reading a Message When the Same Message is Available Auditorily in Another Language: The Case of Subtitling. En O'Reagan, J. K. & Lévy-Schoen, A. (Eds.). (1987). *Eye Movements: From Physiology to Cognition*. Amsterdam/New York: Elsevier Science.
- Eco, U. (2003). *Mouse or Rat? Translation as Negotiation*. London: Weidenfeld & Nicolson.
- EDAD (2008). *Encuesta sobre discapacidades, autonomía personal y situaciones de dependencia 2008*. Madrid: Instituto Nacional de Estadística. Recuperado de <http://www.ine.es/jaxi/menu.do?L=0&type=pcaxis&path=/t15/p418&file=inebase>
- Eguíluz, F., Merino, R. Olsen, V. Pajares, E. & Santamaría, J. M. (1994). *Trasvases culturales: Cine, literatura, traducción*. Universidad del País Vasco. Vitoria.
- Esté, H. J. (2011) *Reading second language subtitles: A case study of South African viewers reading in their native language and L2-English*. (Tesis doctoral). North-West University, Vanderbijlpark, South Africa.
- Even-Zohar, I. (1978). The Position of Translated Literature Within the Literary Polisystem. En Holmes, J. S., Lambert, J. & Van den Broeck, R. *Literature and Translation: New Perspectives in Literary Studies, with a basic Bibliography of Books on Translation Studies*. Leuven: Katholieke Universiteit Leuven.
- Even-Zohar, I. & Toury, G. (Eds.). (1981). Theory of Translation and Intercultural Relations. Número especial de *Poetics Today*, 2(4). Stanford: Duke University Press.

- Even-Zohar, I. (1990). Polysystem Studies. En *Poetics Today* 11(1). Recuperado de https://www.tau.ac.il/~itamarez/works/books/Even-Zohar_1990--Polysystem%20studies
- Evert, S. & Hardie, A. (2011). Twenty-first Century Corpus Workbench: Updating a Query Architecture for the New Millennium. En *Proceedings of the Corpus Linguistics 2011 Conference*. Birmingham: University of Birmingham.
- Fodor, I. (1976). *Film Dubbing: Phonetic, Semiotic, Esthetic and Psychological Aspects*. Hamburg: Helmut Buske.
- Fuentes-Luque A. (2015) Institutional Audiovisual Translation: A (Shop) Window on the World. En Baños Piñero, R. y Díaz Cintas J. (Eds.). *Audiovisual Translation in a Global Context. Palgrave Studies in Translating and Interpreting*. London: Palgrave Macmillan.
- Gambier, Y. (1996) La traduction audiovisuelle, un genre nouveau?, En Gambier, Y. (ed.) *Les transferts linguistiques dans les médias audiovisuels*. París: Presses Universitaires du Septentrion, pp. 7-12.
- Gambier, Y. (1997) [1994]. *Language Transfer and Audiovisual Communication. A Bibliography*. 2.^a ed. Turku: Centro de Traducción e Interpretación, Universidad de Turku.
- Gambier, Y. (1998). *Translating for the Media*. Turku: Centre for Translation and Interpreting.
- Gambier, Y. (2003). Working with Relay: An Old Story and a New Challenge. En González, L. P. (Ed.). *Speaking in Tongues: Language across Contexts and Users*. Valencia: University of Valencia Press.
- Gambier, Y. (2003b). Screen Transadaptation. Perception and Reception. En *The Translator*, 9(2), pp. 171-189.
- Gambier, Y. (2009). Challenges in Research on Audiovisual Translation. En Pym, A. & Perekrstenko, A. (Eds.). *Translation Research Projects 2*. Tarragona: International Studies Group.
- García Izquierdo, I. (2005). Corpus electrónico, género textual y traducción: metodología, concepto y ámbito de la Enciclopedia electrónica para traductores. En GENTT. *Meta*, 50(4). Recuperado de <https://doi.org/10.7202/019918ar>

- García Yebra, V. (1982). *Teoría y práctica de la traducción*. Madrid: Gredos.
- Gerber-Morón, O. & Szarkowska, A. (2018). Line breaks in subtitling: an eye tracking study on viewer preferences. En *Journal of Eye Movement Research*, 11(3):2.
- Gómez Hernández, J. A., Calderón Rehecho, A. y Magna Wals, J. A. (Coord.). (2008). *Brecha digital y nuevas alfabetizaciones. El papel de la biblioteca*. Madrid: Universidad Complutense. Biblioteca Complutense, documentos de trabajo 2008/1.
- González-Iglesias, D. (2012). *Desarrollo de una herramienta de análisis de los parámetros técnicos de los subtítulos y estudio diacrónico de series estadounidenses de televisión en DVD*. (Tesis doctoral). Universidad de Salamanca. Salamanca.
- Goris, O. (1993). The Question of French Dubbing: Towards a Frame for Systematic Investigation. En *Target*, 55(2), pp. 169-190.
- Gottlieb, H. (1997). *Subtitles, Translation & Idioms*. Copenhagen: University of Copenhagen.
- Hardie, A. (2012). CQPweb: Combining Power, Flexibility and Usability in a Corpus Analysis Tool. En *International Journal of Corpus Linguistics*, 17(3), pp. 380-409.
- Hefer-Jordaan, E. (2013). Reading Second Language Subtitles: A Case Study of Afrikaans Viewers Reading in Afrikaans and English. En *Perspectives*, 21, pp. 22-41.
- Hermans, T. (Ed.). (1983). *The Manipulation of Literature: Studies in Literary Translation*. London/Sydney: Croom Helm.
- Hermans, T. (Ed.). (1985). *Second Hand. Papers on the Theory and Historical Study of Literary Translation*. Belgium: Alw-Cahier.
- Hermans, T. (1991). Translational Norms and Correct Translations. En K.M. van Leuven-Zwart & T. Naaijken (eds.). *Translation Studies: The State of the Art : Proceedings of the First James S. Holmes Symposium on Translation Studies*, pp. 155-170.
- Hermans, T. (1999). *Translation in System: Descriptive and System-Oriented Approaches Explained*. Shanghai: Shanghai Foreign Language Education Press.
- Hernández Bartolomé, A. I. & Mendiluce Cabrera, G. (2005). New Trends in Audiovisual Translation: The Latest Challenging Modes. En *Miscelánea: a journal of english and american Studies*, 31, pp. 89-104

- Hernández, M. & Montes, E. (2002). Accesibilidad de la cultura visual: límites y perspectivas. *Integración. Revista sobre ceguera y deficiencia visual*, 40, pp. 21-28.
- Holmes, J. (1988). The Name and Nature of Translation Studies. En *Translated! Papers on Literary Translation and Translation Studies*. Amsterdam: Rodopi.
- Hurtado, A. (1994). Modalidades y tipos de traducción. *Vasos Comunicantes*, 4, pp. 19-27
- Hurtado, A. (1996). La traductología: lingüística y traductología. En *TRANS: Revista de traductología*, 1, pp. 151-162.
- IMSERSO (2011). *Envejecimiento activo. Libro Blanco*. Madrid: Ministerio de Sanidad y Política Social, Secretaría General de Política Social y Consumo Instituto de Mayores y Servicios Sociales. Recuperado de www.imserso.es/InterPresent1/groups/imserso/documents/binario/8088_8089libroblancoenv.pdf
- INTECO (2009). *Guía de recomendaciones de accesibilidad y calidad web*. Madrid: Inteco.
- Ivarsson J.(2004). A short technical history of subtitling in Europe. En *Transedit*. Recuperado de <http://www.transedit.se/history.htm>
- Ivarsson, J. & Carroll, M. (1998). *Subtitling*. Simrishamn: Transedit. [1992. *Subtitling for the Media: A Handbook of an Art*, Estocolmo: Transedit].
- Jakobson, R. (1959). On Linguistic Aspects of Translation. En Brower, R. A. (Ed.). *On Translation*. Cambridge: Harvard University Press, pp. 232-239.
- Jiménez Hurtado, C. (Ed.). (2007). *Traducción y accesibilidad*. Frankfurt: Peter Lang.
- Karamitroglou, F. (1998). *A Proposed Set of Subtitling Norms in Europe*. Recuperado de <https://translationjournal.net/journal/04stndrd.htm>
- Karamitroglou, F. (2000). *Towards a Methodology for the Investigation of Norms in Audiovisual Translation. The Choice between Subtitling and Revoicing in Greece*. Amsterdam: Rodopi.
- Kiraly, D. (1995). *Pathways to Translation: Pedagogy and Process*. Kent: Kent State University Press.
- Kooltra, C. M., Peeters, A. L. & Spinhof F, H. (2002). The Pros and Cons of Dubbing and Subtitling. *European Journal of Communication*, 17, pp. 325-354.

- Lambert, J. (1990). Le sous-titrage et la question des traductions. Rapportsur une enquete. En Arntz R. y Thomé Gisela (coord.) *Übersetzungswissenschaft. Ergebnisse und Perspektiven*. Tübinga: Günter Narr, pp. 228-238.
- Lecuona, L. (1994). Entre el doblaje y la subtitulación: la interpretación simultánea en el cine. En F. Eguíluz. *Transvases culturales: literatura, cine, traducción* 1. Vitoria: Publicaciones de la Universidad del País Vasco, pp. 279–286.
- Ley 34/2002, de 11 de julio de 2002, de Servicios de la sociedad de la información y de comercio electrónico (LSSICE). Boletín Oficial del Estado, n.º 166, de 12 de julio de 2002.
- Ley 51/2003, de 2 de diciembre de 2003, Sobre igualdad de oportunidades, no discriminación y accesibilidad universal de las personas con discapacidad. Boletín Oficial del Estado, n.º 289, de 3 de diciembre de 2003.
- Ley 49/2007, de 26 de diciembre de 2007, por la que se establece el régimen de infracciones y sanciones en materia de igualdad de oportunidades, no discriminación y accesibilidad universal de las personas con discapacidad. Boletín Oficial del Estado, n.º 310, de 27 de diciembre de 2007.
- Ley 56/2007, de 28 de diciembre de 2007, de Medidas de impulso de la sociedad de la información. Boletín Oficial del Estado, n.º 312, de 29 de diciembre de 2007.
- Lomheim, S. (1999). The Writing on the Screen. Subtitling: A Case Study from Norwegian Broadcasting (NRK), Oslo. En Anderman, G. y Rogers, M. (eds.). *Word, text, translation. Liber Amicorum for Peter Newmark*. Clevedon: Multilingual Matters, pp. 190-207.
- Lorenzo, L. & Pereira, A. (2001). *Traducción subordinada inglés-español/galego II: el subtitulado*. Vigo: Servicio de Publicaciones de la Universidad de Vigo.
- Lorenzo, L. & Pereira, A. (2012). Subtitulado para sordos: un reto y una necesidad: pasado, presente y futuro del SPS en España. En Martínez Sierra, J.J. (coord.): *Reflexiones sobre la traducción audiovisual. Tres espectros, tres momentos*. Servicio de publicaciones Universitat de Valencia.
- Luyken, G. M. et al. (1991). *Overcoming Language Barriers in Television: Dubbing and subtitling for the European Audience*. Dusseldorf: The European Institute for the Media.

- Mace, R. L., Hardie, G. J. & Place, J. P. (1990). Accessible Environments: Toward Universal Design. En Preiser, W., Visher, J. & White, E. (Eds.). *Design Interventions: Toward a more Human Architecture*. New York: Van Nostrand Reinhold.
- Madariaga Ortuzar, A. (2004). El Diseño para Todos: Accesibilidad a las TIC. En Jacob, I. (Coord.). *Capacidades humanizadoras de las TIC. Libro de Comunicaciones de las V Jornadas de Informática y Sociedad*. Bilbao: Facultad ESIDE, Universidad de Deusto.
- Mank, D. (2000). Inclusión, cambio organizacional, planificación centrada en la persona y apoyos naturales. En *Integra*, 3(8), pp. 1-4.
- Martí Ferriol, L. (2006). *Estudio empírico y descriptivo del método de traducción para doblaje y subtitulación*. (Tesis doctoral). Castellón: Universitat Jaume I.
- Martí Ferriol, J. L. (2017). Nueva aproximación al cálculo de velocidades de lectura de subtítulos. En *TRANS. Revista de traductología*, 16, pp. 39-48. Recuperado de <http://www.revistas.uma.es/index.php/trans/article/view/3210>.
- Martínez Sierra, J. J. (2004). *Estudio descriptivo y discursivo de la traducción del humor en textos audio-visuales. El caso de Los Simpson*. (Tesis doctoral). Universitat Jaume I. Castellón. Recuperado de <http://www.tdx.cbuc.es/>.
- Martínez Sierra, J. J. (2011). De normas, tendencias y otras regularidades en traducción audiovisual. En *Estudios de Traducción*, 1, pp. 151-170.
- Mason, I. (1989). Speaker Meaning and Reader Meaning: Preserving Coherence in Screen Translating. En Kölmel, R. y Payne, J. *Babel The Cultural and Linguistic Barriers between Nations*. Aberdeen: Aberdeen University Press, pp. 13-24.
- Mateo, M. (1997). Performing Musical Texts in a Target Language: The Case of Spain. Comunicación presentada en *EST Congress*. Universidad de Granada. Granada. Septiembre 23-26.
- Mayoral Asensio, R. (1997). Sincronización y traducción subordinada: de la traducción audiovisual a la localización de software y su integración en la traducción de productos multimedia. En Mayoral, R. y Tejada, A. (coord.). *Actas del Primer Simposium de Localización Multimedia*. Granada. Universidad de Granada.
- Mayoral Asensio, R. (1999). *La traducción de la variación lingüística*. Soria: Diputación Provincial de Soria/Marcial Pons. 1997, tesis doctoral de la Universidad de Granada.

- Mayoral Asensio, R. (2001a). Campos de estudio y trabajo en traducción audiovisual. En Duro Moreno, M. (Coord.). *La traducción para el doblaje y la subtitulación*. Madrid: Cátedra.
- Mayoral Asensio, R. (2001b). El espectador y la traducción audiovisual. En Chaume Varela, F. y Agost de Canós, R. *La traducción en los medios audiovisuales*. Castellón: Universitat Jaume I.
- Mayoral Asensio, R. (2001c). *Aspectos epistemológicos de la traducción*. Castellón: Universitat Jaume I.
- Mayoral Asensio, R. (en prensa). Por una renovación en la formación de traductores e intérpretes: revisión de algunos de los conceptos sobre los que se basa el actual sistema, su estructura y contenido. En *Sendebarr*, 12.
- McEnery, T. & Wilson, A. (1996). *Corpus linguistics*. Edimburgh: Edimburgh University Press.
- Merino, R. & Rabadán, R. (2004). Introducción. En Toury, G. (2004). *Los estudios descriptivos de traducción y más allá. Metodología de investigación para estudios de traducción*. Madrid: Cátedra. pp. 17-33.
- Meyer, C. F. (2002). *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.
- Ministerio de Trabajo y Asuntos Sociales (2003). *II Plan de acción para las personas con discapacidad 2003-2007*.
- Ministerio de Sanidad y Política Social, Dirección General de Coordinación de Políticas Sectoriales sobre la Discapacidad. (2009). *III Plan de acción para las personas con discapacidad 2009-2012*.
- Molés-Cases, T. (2016). Compilación y análisis de un corpus paralelo para la investigación en traducción. Proyecto con Déjà Vu, TreeTagger y Corpus Workbench. *Revista de Lingüística Teórica y Aplicada*, 54(1), pp. 149–174.
- Moreno, L. et al. (2008). *Accesibilidad a los contenidos audiovisuales en la web: Una panorámica sobre legislación, tecnologías y estándares (WCAG 1.0 y WCAG 2.0)*. Madrid: Editorial del Real Patronato sobre Discapacidad.

- Muñoz Cabrera, R. (2017). *Glosario militar de la Segunda Guerra Mundial. Trabajo de extracción terminológica y de lingüística de corpus mediante el análisis de textos de época*. (Tesis doctoral inédita). Sevilla: Universidad de Sevilla. Recuperado de <https://idus.us.es/xmlui/handle/11441/63984>
- Navarrete Moreno, F. J. (1997). Sistema AUDESC: el arte de hablar en imágenes. En *Integración*, 23, pp. 70-75.
- Neves, J. (2005). *Audiovisual Translation: Subtitling for the Deaf and Hard-of-Hearing*. (Tesis doctoral) Roehampton University. Roehampton. Recuperado de <http://roehampton.openrepository.com/roehampton/bitstream/10142/12580/1/neves%20audiovisual.pdf>
- Neves, J. (2007). A world of change in a changing world, En Díaz-Cintas, J., Orero, P. y Remael, A. (ed.) *Media for All. Subtitling for the Deaf, Audio Description, and Sign Language*. Amsterdam/Nova York: Rodopi, pp. 91-100.
- Neves, J. (2009). Interlingual subtitling for the Deaf and Hard-of-Hearing. En Díaz-Cintas, J. y Anderman, G. *Audiovisual Translation: Language Transfer on Screen*. Basingstoke: Palgrave MacMillan, pp. 151-169
- Nida, E. (1959). Principles of Translation as exemplified by Bible Translating. En *On Translation*. Ed. Reuben Arthur Bower. Harvard: Harvard University Press, pp. 11-13.
- Nord, Ch. (1991). Scopus, Loyalty and Translational Conventions. *Target*, 3(1), pp. 91-109.
- Nord, Ch. (1998). La unidad de traducción desde un enfoque funcional, en *Quaderns. Revista de traducción*, 1, pp. 65-77
- OMS, Organización Mundial de la Salud y Grupo del Banco Mundial (2011). *Informe Mundial sobre la Discapacidad*. Recuperado de https://www.who.int/disabilities/world_report/2011/report.pdf
- OMS (2011). *World report on disability. WHO & The World Bank*. Recuperado de http://www.who.int/disabilities/world_report/2011/accessible_es.pdf
- ONU (1948). *Declaración Universal de Derechos Humanos*. Asamblea General de Naciones Unidas. 10 de diciembre 1948.

- ONU (1993). Normas uniformes sobre la igualdad de oportunidades para las personas con discapacidad. Asamblea General de Naciones Unidas. 20 de diciembre de 1993.
- ONU (2006). Convención sobre los Derechos de las Personas con Discapacidad y Protocolo Facultativo. Resolución 56/168 de la Asamblea General, de 19 de diciembre de 2001. Recuperado de <http://www.un.org/esa/socdev/enable/documents/tccconvs.pdf>
- Orero, P. (Ed.). (2004). *Topics in Audiovisual Translation*. Amsterdam/Philadelphia: John Benjamins.
- Orero, P. (2005). La inclusión de la accesibilidad en comunicación audiovisual dentro de los estudios de traducción audiovisual. En *Quaderns. Revista de traducció*, 12, pp. 173-185.
- Orero, P. (2008) Three Different Receptions of the same Film. The Pear Stories Project Applied to Audio Description. *European Journal of English Studies*, 12(2), pp.179-193.
- Pedersen, J. (2011). *Subtitling norms for television*. Amsterdam: Benjamin translation library.
- Perego, E. (2008). Subtitles and line breaks: Towards improved readability. En Chiaro, D. Heiss, C. Bucaria C. (eds): *Between text and image. Updating research in screen translation*. Amsterdam: John Benjamins, pp. 211-223.
- Perego E., Laskowska, M., Matamala A., Remael A., Robert I. S., Szarkowska A. et al. (2016). Is Subtitling Equally Effective Everywhere? A First Cross-National Study on the Reception of Interlingually Subtitled Messages. En *Across Lang Cult*, 17(2), pp. 205-229.
- Pereira, A. (2005). El subtitulado para sordos: estados de la cuestión en España. *Quaderns. Revista de traducció*, 12, pp. 161-172. Recuperado de <http://ddd.uab.es/pub/quaderns/11385790n12p161.pdf>
- Principios de diseño universal*. Version 2.0-4/1/97 Copyright 1997. NC State University, The Center for Universal Design. Recuperado de http://www.ncsu.edu/ncsu/design/cud/about_ud/udprinciplestext.htm
- Rabadán, R. (1991). *Equivalencia y Traducción*. León: Universidad de León.

- Rabadán, R. & Fernández, P. (2002). *La traducción inglés-español: fundamentos, herramientas, aplicaciones*. León/Valladolid: Universidad, Secretariado de Publicaciones y Medios Audiovisuales.
- Rabadán, R. y Merino, R. (2004). Introducción. En Gideon Toury. *Los estudios descriptivos de traducción, y más allá. Metodología de la investigación en estudios de traducción*. Madrid: Cátedra, pp. 17-33.
- Real Decreto 1494/2007 de 12 de noviembre, por el que se aprueba el Reglamento sobre las condiciones básicas para el acceso de las personas con discapacidad a las tecnologías, productos y servicios relacionados con la sociedad de la información y medios de comunicación social. Boletín Oficial del Estado, n.º 279, de 21 de noviembre de 2007.
- Reiss, K. (1977/89). *Text types, translation types and translation assessment, translated*. En A. Chesterman.
- Reiss, K. (1981/2004). *Type, kind and individuality of text: decision making in translation*. En L. Venuti.
- Remael, A. & Neves, J. (Eds.). (2007). A Tool for Social Integration? Audiovisual Translation from Different Angles. En *Linguistica Antverpiensia, New Series 6*, pp. 11-22.
- Rodríguez, G. (2007). La audiodescripción: parámetros de cohesión. En Jiménez C. *Traducción y accesibilidad : subtitulación para sordos y audiodescripción para ciegos: nuevas modalidades de traducción Audiovisual*. Frankfurt: Peter Lang, pp. 93-110.
- Rodríguez Victoriano, J.M. (2003). *La democratización del conocimiento en el contexto de la globalización neoliberal*. Universidad de Valencia. Valencia.
- Romero-Fresco, P. (2015). Final Thoughts: Viewing Speed in Subtitling. En Romero-Fresco, P. (Ed.). *The Reception of Subtitles for the Deaf and Hard of Hearing in Europe*. New York: Peter Lang.
- Roussou, M. (2003). *La teoría de las normas en el subtulado griego. Análisis de películas de Almodóvar*. (Trabajo de investigación). Universidad de Granada.
- Royal National Institute of Blind People. (2010). *A Comparative Study of Audio Description Guidelines Prevalent in Different Countries*. London: RNIB. Recuperado de https://www.rnib.org.uk/professionals/Documents/International_AD_Standards_comparative%20study_2010.doc

- Sánchez Nieto, M. T. et al. (Eds.). (2015). *Metodología y aplicaciones en la investigación en traducción e interpretación con corpus*. Valladolid: Ediciones Universidad de Valladolid.
- Secretaría de Estado de Servicios Sociales, Familia y Discapacidad (2004). *I Plan de accesibilidad 2004-2012. Por un nuevo paradigma, el diseño para todos, hacia la plena igualdad de oportunidades*.
- Serrano Mascaraque, E. (2008). *Accesibilidad web para discapacitados visuales: Acercamiento y propuestas*. (Tesis doctoral). Alcalá de Henares: Universidad de Alcalá.
- Serrat Roozen, I. & Martínez Martínez, J. M. (2012). ECPC: El discurso parlamentario europeo desde la perspectiva de los estudios traductológicos de corpus. En *Linguamática*, 4(2), pp. 65-73.
- Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. M. (2005). Corpus and Text: Basic Principles. En Wynne, M. (Ed.). *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books. Recuperado en <http://ahds.ac.uk/linguistic-corpora>.
- Silverwood A. (1992). *Audiodescription*. DAM (Disability Arts Magazine), 2-4, 10-14, Recuperado de <https://www.leeds.ac.uk/disability-studies/archiveuk/disability%20arts%20mag/winter%201992.pdf>
- Simpson, R. & Swales, J. (2001). Introduction to North American perspective on corpus linguistics at the millennium. En Simpson R. y Swales, J. (Eds.) *Corpus linguistics in North America. Selections from the 1999 Symposium*. Ann Arbor: The University of Michigan Press, pp. 1-14.
- Snell-Hornby, M. (1988). *Translation Studies. An Interdiscipline*. Amsterdam: John Benjamins.
- Sokoli, S. (2000). *Research Issues in Audiovisual Translation: Aspects of Subtitling in Greece*. (Trabajo de investigación). Barcelona: Universitat Autònoma de Barcelona.
- Sponholz, C. (2003). *Teaching Audiovisual Translation. Theoretical Aspects, Market Requirements, University Training and Curriculum Development*. (Trabajo de fin de carrera). Johannes Gutenberg-Universität Mainz: Mainz. Recuperado de: http://isg.urv.es/library/papers/thesis_Christine_Spon-holz.doc
- Steiner, G. (1975/1980). *Después de Babel*. México: Fondo de Cultura Económica. Traducción de Adolfo Castañón.

- Szarkowska *et al.* (2013): Harnessing the Potential of Eye-Tracking for Media Accessibility. En Grucza, S. *et al.* (eds.): *Translation studies and eye-tracking analysis*. Frankfurt am Main: Peter Lang.
- Szarkowska, A. (2016). *Report on the results of an online survey on subtitle presentation times and line breaks in interlingual subtitling. Part 1: Subtitlers*. SURE Project Dataset. RepOD. London. Recuperado de <http://dx.doi.org/10.18150/repod.4469278>
- Szarkowska A, & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: Evidence from eye movements. En *Public Library of Science ONE*, 13(6). Recuperado en <https://doi.org/10.1371/journal.pone.0199331>.
- Tamayo, A. (2015a). *Estudio descriptivo y experimental de la traducción en TV para niños sordos. Una propuesta alternativa*. (Tesis doctoral). Universitat Jaume I. Castellón. Recuperado de <https://www.tdx.cat/handle/10803/353962>
- Tamayo, A. (2015b). Estudio descriptivo de la subtitulación para niños sordos y con discapacidad auditiva en las cadenas infantiles y juveniles en España. En *Quaderns, Revista de Traducció*, 22, pp. 363-383.
- Taylor, Ch. (2008) What is corpus linguistics? What the data says. *ICAME Journal*, 32, pp. 179-200. Recuperado de <http://sro.sussex.ac.uk/53389/>
- Titford, C. 1982. Subtitling-constrained Translation. *Lebende Sprachen*, 27(3): 113-116.
- Toury, G. (1980). *In Search of a Theory of Translation*. Tel-Aviv: The Porter Institute for Poetics and Semiotics.
- Toury, G. (1985). A Rationale for Descriptive Translation Studies. *Dispositio (The Art and Science of Translation)*, (Vol.7). Michigan: Center for Latin American and Caribbean Studies, pp. 23-39.
- Toury, G. (1991). What are Descriptive Studies in Translation Likely to Yield Apart from Isolated Description? En K. V. Leuven-Zwart, & T. Naaijken (Eds.), *Translation Studies: The State of the Art*. Amsterdam: Rodopi.
- Toury, G. (1995). *Descriptive Translation Studies and Beyond*. Amsterdam/Philadelphia: John Benjamins.
- Toury, G. (1999). A Handful of Paragraphs on «Translation» and «Norms». En Schäffner, C. (Ed.). *Translation and Norms*. Clevedon: Multilingual Matters.

- Toury, G. (2004), *Los estudios descriptivos de traducción y más allá. Metodología de investigación para estudios de traducción*. Madrid: Cátedra.
- Valentí López, P.(2002). La sociedad de la información en América Latina y el Caribe TICs y un nuevo marco institucional. *Revista Iberoamericana de Ciencia, Tecnología, Sociedad e Innovación de la OEI*, 2. Recuperado de: <https://www.campus-oei.org/revistactsi/numero2/valenti.htm>
- Vinay, J. P. y Darbelnet, J. (1976) *Stylistique comparée du français et de l'anglais*. Paris: Didier.
- Voces Merayo, R. (2008). El contenido audiovisual: otro reto para la accesibilidad web. En *BiD: textos universitaris de biblioteconomia i documentació*, 21. Recuperado de <http://bid.ub.edu/21/voces2.htm>.
- Voces Merayo, R. (2010). *Comunicación audiovisual sin barreras: Televisión pública. World Wide Web y Accesibilidad*. (Tesis doctoral). Universitat Autònoma de Barcelona. Recuperado de: <https://www.tdx.cat/handle/10803/4851>
- W3C. (2008). *Web Content Accessibility Guidelines 2.0*. Recuperado de <http://www.w3.org/TR/WCAG20/>

ANEXOS

EMPAC_EN.

https://drive.google.com/file/d/1IqGGwG2HQr6Rbr6Rn2ZDQ9TBTdPlw33_/

EMPAC_ES.

https://drive.google.com/file/d/1aMnV8PQK_vjwYcUOE22wy-rcwSY_1LaL/

ALL_VIDEOS.

<https://drive.google.com/file/d/1J2Q-cJTUxeOc2pOc8rIOme-NyzZvgqlt/>

