

Velocidades de lectura de subtítulos en alemán y español de películas norteamericanas: estudio de caso

JOSÉ LUIS MARTÍ FERRIOL

Universitat Jaume I
martij@trad.uji.es

Recibido: 28 de julio de 2011

Aceptado: 27 de febrero de 2012

RESUMEN

Este artículo presenta resultados de velocidades de lectura de subtítulos en lengua alemana y castellana de tres películas norteamericanas. Dichos resultados se encuentran expresados en los dos parámetros de mayor utilización en este campo: CPS (caracteres por segundo) y WPM (“words per minute”, o palabras por minuto, en inglés), y se han obtenido por medio de una aplicación diseñada a tal efecto por el autor y otros investigadores. El estudio (empírico y cuantitativo) surge con el objetivo de intentar demostrar que el primero de los dos parámetros mencionados (CPS) es más útil para expresar la velocidad de lectura de subtítulos, ya que debe ser independiente de la lengua. Se parte de la hipótesis de que el número de palabras que aparecerán en los subtítulos alemanes será menor, dada la mayor longitud (en términos generales) de las palabras en dicha lengua.

Palabras clave: subtitulación, velocidad de lectura, caracteres por segundo (CPS), palabras por minuto (WPM), películas americanas.

German and Spanish subtitle reading speeds for American films: a case study

ABSTRACT

This article presents results of German and Spanish subtitle reading speeds for three American films. Those results are expressed in the two most common parameters in this particular field: CPS (characters per second) and WPM (Words per minute), and they have been obtained by means of a software application designed to do so by the author and other researchers. This empirical and quantitative study has as an objective to try to prove that the former of the two above mentioned parameters (CPS) is more suited to express subtitle reading speed, since it should be language-independent. An initial hypothesis is formulated: the number of words in German subtitles should be lower, given the fact that German words are (in general terms) longer.

Keywords: subtitling, reading speed, characters per second (CPS), words per minute (WPM), American films.

Sumario: 1. Introducción. 2. Herramientas e hipótesis de trabajo. 3. Descripción del corpus. 4. Metodología. 5. Resultados y discusión. 6. Referencias bibliográficas.

1. Introducción

La subtitulación es posiblemente la modalidad de traducción audiovisual más utilizada y aceptada por el espectador a nivel mundial. En el continente europeo hay sin embargo una serie de países que, por razones históricas, socio-políticas y económicas, cuenta con una tradición eminentemente dobladora. Tanto Alemania como España forman parte de este grupo, aunque la utilización y el consumo de subtítulos va ganando adeptos entre los espectadores de estos países como consecuencia de la irrupción del DVD a finales de la anterior década, la digitalización de la emisión televisiva, la irreversible irrupción de los subtítulos en la publicidad audiovisual y la importancia mayor que cada vez está adquiriendo la accesibilidad de la información, que incluye la utilización de subtítulos para personas sordas o con deficiencias auditivas.

Muchos investigadores han centrado su atención en el doblaje y la subtitulación, ya desde de la segunda mitad de década de los 90 en el siglo anterior. No se pretende ahora realizar un repaso en este sentido, que ya otros autores como Díaz Cintas (2001, 2003) –en el caso de la subtitulación– y Chaume (2004) –para ambas modalidades, aunque con cierto énfasis en el doblaje– lo han llevado a cabo de forma exhaustiva.

Por lo que se refiere al contenido de esta publicación, la presente introducción se debe centrar más bien en el espectador como elemento fundamental de la recepción de los subtítulos, y más concretamente, en su capacidad para leerlos de forma cómoda, expresada en forma de los dos parámetros al uso: caracteres por segundo (CPS) y palabras por minuto (WPM).

Se puede encontrar una revisión más detallada de la bibliografía relacionada con la velocidad de lectura de los subtítulos y su expresión por medio de los parámetros al uso en otro artículo del autor (Martí Ferriol, en prensa), una publicación cuya principal finalidad es presentar la aplicación (o herramienta informática) que se ha empleado en la obtención de los datos empíricos y cuantitativos que son objeto de este artículo. Dicha revisión bibliográfica incluye autores como Karamitroglou (1998), Mayoral (2001), Díaz Cintas (2008), Toda (2009) y Romero Fresco (2009). Todos ellos presentan valores de velocidades de lectura obtenidos para subtítulos que podríamos llamar “convencionales”, en el sentido de que están formados como máximo por dos líneas que pueden llegar a tener unos 35 caracteres. De acuerdo con estudios experimentales, se ha demostrado que un espectador medio necesitaría unos 6 segundos para leer con comodidad un subtítulo de esas características. A partir de esta afirmación, se ha desarrollado lo que se ha dado en denominar la “regla de los 6 segundos”, admitida como un principio estándar, no sólo en el uso, sino

también en la docencia y didáctica de la subtitulación. Romero Fresco (2009) hace referencia a la constatación de estos valores de forma experimental por parte de D'Ydewalle et al. en 1987, por medio de un experimento realizado con la tecnología denominada "eye-tracking", muy utilizada en la actualidad en numerosos experimentos encaminados a investigar la recepción de los textos audiovisuales, así como la relevancia de la información transmitida.

Using eye-tracking technology, he tested three different presentation times for subtitles: two lines of 32 characters in 4 seconds (approximately 192 wpm), 6 seconds (130 wpm) and 8 seconds (96 wpm) respectively. The object of this study was to ascertain if the six-second rule (a full two-line subtitle displayed on screen for 6 seconds and shorter subtitles scheduled proportionally), accepted as common practice in most subtitling countries, could be validated by empirical research on reading speed. His results leave little room for doubt, the six-second rule being identified as setting the appropriate reading speed for the participants. This rule has later on been supported by other scholars such as Díaz Cintas (2003), who applies it to longer lines than the ones referred to by D'Ydewalle (72 characters instead of 64), thus setting the recommended speed at 144 wpm (12 cps). (Romero Fresco 2009:114)

Como se puede constatar, la ya mencionada "6 second rule" parece ser la que goza de mayor aceptación entre los espectadores. Si se traduce dicha velocidad de lectura a valores de los parámetros que utilizaremos en nuestro estudio empírico, se obtienen resultados de 12 CPS y 144 WPM

2. Herramientas e hipótesis de trabajo

El proceso de subtitulación consta de dos etapas fundamentales: el denominado "pautado" (o "spotting", en inglés) y la propia creación de los subtítulos. Por "pautado" se entiende la definición de los tiempos de entrada y salida de un subtítulo. Esto significa que cuando el actor del texto audiovisual original comienza a hablar, será necesario marcar el tiempo en el que deberá aparecer el subtítulo en pantalla con la traducción de su diálogo. Asimismo, el subtítulo debe desaparecer, aproximadamente, cuando dicha intervención del diálogo en el texto original ha finalizado. La diferencia entre ambos tiempos, el de salida y el de entrada, define la duración del subtítulo. Se puede afirmar, basándonos en la experiencia profesional y en ensayos previos, que el *pautado*, si es correcto, permite al espectador la lectura del subtítulo sin problemas. Normalmente, dicha duración de los subtítulos suele variar entre 2 y 6 segundos, dependiendo de si el subtítulo tiene una o dos líneas y del número de caracteres (o palabras) que se incluyan en cada una de ellas.

La limitación más frecuente en la fase de creación de los subtítulos, y por tanto la que genera su para algunos especial dificultad, proviene de la combinación de las restricciones espacio-temporales intrínsecas de la modalidad, y radica en el hecho de que por regla general solo se dispone de dos líneas (cada una con 35 caracteres como máximo, normalmente), para traducir un diálogo que se deberá leer con como-

didad en 6 segundos como máximo. Por todo ello, la traducción para subtitulación se caracteriza especialmente por la necesidad de síntesis o condensación de la información del texto audiovisual en lengua origen.

El proceso de subtitulación (pautado y creación de los subtítulos) se puede llevar a cabo de forma manual, o mediante la utilización de programas informáticos. En el mercado hay toda una serie de programas disponibles; algunos son de uso libre, otros se pueden descargar de Internet en una versión de prueba de duración limitada, y otros requieren la adquisición de una licencia, en ocasiones a un precio elevado. Un listado de los principales programas comerciales puede consultarse en Martí Ferriol (en prensa). Además, algunos de ellos proporcionan también valores de los parámetros de velocidad de lectura más al uso (CPS y/o WPM): en ocasiones se trata sólo de uno de ellos, aunque algunos programas ofrecen valores de ambos. Curiosamente, dichos valores no siempre coinciden.

La disparidad de valores obtenidos para los parámetros de velocidad de lectura por medio de los programas convencionales de subtitulación nos animó a desarrollar una nueva herramienta que cumpliera dicha funcionalidad. Como se ha indicado más arriba, los detalles del funcionamiento de la aplicación y las ventajas que presenta están descritos detalladamente en el artículo del mismo autor mencionado anteriormente. Así pues, será este programa de diseño y desarrollo propio el que se va a utilizar para llevar a cabo los cálculos de CPS y WPM de subtítulos alemanes y españoles de las películas que han configurado el corpus objeto de estudio.

En cuanto a la hipótesis de trabajo, esta se ha enunciado en el resumen inicial partiendo de la idea intuitiva de que el empleo del parámetro CPS podría ser más apropiado que el de las WPM, si tenemos la intención de comparar valores de velocidades de lectura de subtítulos en dos lenguas diferentes, ya que la longitud de las palabras puede variar en diferentes lenguas (Mayoral 2001). Desde este punto de vista, parece más apropiado utilizar una unidad más normalizada, como los caracteres (sean letras, signos de puntuación o espacios), sin tener en cuenta su asociación en palabras, que como veremos a continuación para las dos lenguas que nos interesan, puede ser también variable.

Para el caso del alemán, se ha estudiado ampliamente la facilidad de composición nominal característica de esta lengua para formar palabras compuestas. Oster (2006) lleva a cabo un repaso histórico exhaustivo de los esfuerzos de investigación en este sentido. La autora agrupa los estudios sobre aspectos específicos de la composición nominal en lengua alemana en cinco grandes grupos: los pragmáticos, textuales, cognitivos, los compuestos en textos especializados y los estudios contrastivos. En el apartado de conclusiones, se pone de manifiesto que existen todavía aspectos interesantes que podrían analizarse por lo que a los compuestos nominales alemanes se refiere, en campos como la terminología o el análisis del discurso especializado. En cualquier caso, y según la autora, la composición se presenta de forma recurrente en lenguajes de especialidad (técnicos o administrativos) o de registro formal, donde se muestra como un recurso especialmente potente para incrementar la redundancia y cohesión del texto. Para el lenguaje objeto de esta investigación, de carácter general y propio de la subtitulación, modalidad que se caracteriza por la síntesis explícita de la información, conse-

cuencia de las restricciones espaciales y temporales, es posible que la utilización de compuestos nominales no sea una de sus características más sobresalientes. Además, los subtítulos son en realidad una representación en modo escrito de los diálogos que los actores interpretan en el texto audiovisual. Por todo ello, es posible que la frecuencia de aparición de compuestos sea inferior, si se compara con los otros tipos de textos mencionados en este párrafo.

La mencionada característica intrínseca que muestra especialmente la lengua alemana (en comparación con la española) para crear compuestos nominales, podría llevar a extraer una conclusión hasta cierto punto lógica: las palabras en alemán, en general, serán más largas que en castellano. La utilización de corpus paralelos en la investigación favorece sin duda este tipo de comparaciones, pero para ilustrar de forma rápida y sin complicaciones esta afirmación, nos basaremos en un experimento mucho más sencillo, el que presenta Quesada-Granja en forma de presentación en la siguiente página web: <http://www.slideshare.net/quesadagranja/distribucion-por-longitud-de-las-palabras-de-diferentes-idiomas-presentation>.

Por medio de la utilización de corpus paralelos de 2500 palabras en diferentes lenguas, obtenidos a partir de textos de carácter general en Wikipedia, y mediante el empleo de un programa de cálculo general ampliamente reconocido a nivel mundial (Matlab), el autor presenta resultados de frecuencias de aparición de palabras con distinto número de letras, y compara las lenguas por grupos o familias. Entre sus conclusiones, el autor constata que los patrones de distribución de frecuencias de palabras por número de letras varían de unas lenguas a otras, y entre sus familias. Más concretamente, los valores medios de longitudes presentados son 4,94 letras/palabra en español, y 6,10 en alemán (dichos resultados son muy similares a los obtenidos en un estudio muy anterior -de 1939-), lo que podría corroborar por otra vía la diferentes capacidades de composición de las dos lenguas mencionadas anteriormente.

Sin embargo, otro argumento que podría favorecer la enunciación de la hipótesis de trabajo en sentido contrario podría deducirse de la idea de que existe una relación inversa entre la longitud de las palabras y su grado de utilización. Para el caso del español, se han encontrado dos ejemplos de documentación que apunta en esta línea, una para el caso de la lengua general, y otro más específico (para un corpus de subtítulos), que es el que en realidad más nos interesa.

En una presentación del Taller de Lingüística del Congreso de la Asociación Europea de Profesores de Español, que tuvo lugar en Alcalá en el año 2003 (<http://www.slideshare.net/RdelMoral/lxico-y-expresiones-del-espaol-actual-y-su-aprendizaje>), se afirma que suele darse una relación inversa entre la longitud de una palabra y su utilización, y que entre las cuarenta palabras más frecuentes del léxico español (que se listan, aunque sin citar la referencia explícita de dónde ni cómo se han obtenido estos datos), no aparece ningún sustantivo. Dado que la diferente capacidad de composición entre el alemán y el castellano se muestra especialmente en dicha categoría gramatical, podría ser posible que esta diferencia (aunque corroborada en el trabajo anterior y expresada por una longitud media de palabra ligeramente superior en textos de tipo general en alemán), no se reflejara en la utilización de la lengua para subtítulos, si la cantidad de sustantivos presentes en ellos tendiera a minimizarse.

En la página web [http://es.wiktionary.org/wiki/Wikcionario:Frecuentes-\(1-1000\)-Subt%C3%ADtulos_de_pel%C3%ADculas](http://es.wiktionary.org/wiki/Wikcionario:Frecuentes-(1-1000)-Subt%C3%ADtulos_de_pel%C3%ADculas) se presentan resultados de frecuencias de aparición de las 1000 palabras con una frecuencia mayor en un corpus específico de 22 millones de palabras obtenidas de subtítulos de películas. Entre las 40 primeras palabras (la primera es “de” con un valor de 41000, y la cuadragésima es “bien” con un valor de 3216) no aparece ningún sustantivo (lo que confirma los valores expuestos en el párrafo anterior para la lengua en general). Curiosamente, el primer sustantivo en aparecer en la lista es “dios”, en la posición 103, con una frecuencia de 1213.

Desgraciadamente, no se ha podido completar este estudio preliminar contrastivo encontrando información sobre frecuencia de aparición de palabras en la lengua alemana, y por tanto, no se puede afirmar que la relación inversa entre la utilización de las palabras y su longitud sea una ley que puede tener una validez global en lenguas diferentes del castellano.

Para el caso de nuestra investigación, y por lo que a la enunciación de la hipótesis se refiere, se debería encontrar un comportamiento en alemán diferente al presentado en castellano (en el sentido de que los sustantivos no son las palabras más utilizadas), para que la hipótesis enunciada se cumpliera y los valores de velocidad de lectura expresados en WPM fueran inferiores.

Para finalizar este apartado, se puede intentar particularizar esta discusión general sobre longitud de palabras y su frecuencia de utilización por medio de un caso concreto, extraído de la vida real y contextualizado en la lengua de los subtítulos. En Alemania, como en otros países, los tranvías y trenes tienen un pulsador de parada de emergencia, normalmente de color rojo y no muy accesible. En dicho país, el texto que suele acompañar al pulsador es el siguiente: “Notbremse. Mißbrauch strafbar”. Como vemos, se trata de un texto de tan sólo 3 palabras. Si deseáramos realizar una “traducción uno por uno” (Martí Ferriol, 2006: 114), podríamos obtener algo parecido a: “Freno de emergencia. Uso incorrecto penalizado”, una propuesta con un total de 6 palabras. Si por el contrario se tuviera la intención de llevar a cabo una “traducción literal” (Martí Ferriol, 2006: 114), más idiomática, se podría proponer: “Freno de emergencia. Su uso incorrecto se penalizará”, de 8 palabras. En realidad, en el original, “strafbar” indica más bien la posibilidad de penalización, y si quisiéramos expresar un mismo sentido exacto, dicha palabra debería traducirse por la tal vez más especializada “punible” (es decir: “Su uso incorrecto es punible”). En cualquier caso, las diferentes propuestas de traducción siempre arrojan como resultado un valor mucho mayor expresado en palabras (6 u 8, frente a 3).

Partiendo del texto y las traducciones propuestas, supongamos ahora que la imagen del pulsador y el texto original aparecieran en pantalla en una película, en un plano que permitiera claramente su lectura por parte del espectador. En primer lugar, se podría iniciar la discusión sobre si dicho texto se debería traducir para subtitulación o no. A partir de las conclusiones de un trabajo anterior (Martí Ferriol, 2010: 289), se puede afirmar que dicha información tiende a traducirse y a explicitarse más en subtitulación que en doblaje, por lo que se opina que sí debería reproducirse en un subtítulo en la lengua meta (en este caso, el castellano), que además debería aparecer en mayúscula según las convenciones de la profesión. Como ya sabemos, por

otro lado, la propia estructura narrativa de la película puede determinar el pautado de los subtítulos. Si, por ejemplo, al plano del freno de emergencia le sucediera rápidamente un diálogo de los personajes que sin duda habría que subtitular, es posible que no se dispusiera de mucho tiempo para el pautado del texto, lo que determinaría la posible solución de traducción a emplear, de las tres propuestas anteriormente. Este tipo de consideraciones, que son consecuencia de la propia idiosincrasia del texto audiovisual, pueden en muchas ocasiones determinar si es posible realizar una traducción “uno por uno” en un subtítulo de una línea y 3 segundos, o una “literal” en dos líneas y 6 segundos.

Con este ejemplo se ha pretendido ilustrar que las consideraciones más de corte lingüístico que se han intentado proponer al inicio del apartado para intentar apoyar con argumentos sólidos la hipótesis enunciada, pueden tener una validez limitada en el caso de textos audiovisuales, como los que nos ocupan y que se describen en el siguiente apartado.

3. Descripción del corpus

La elección del corpus para un trabajo empírico y descriptivo siempre supone la toma de una de serie decisiones difíciles, y si se desea ser conservador y riguroso a la vez, es mejor utilizar textos conocidos y aplicar criterios claros. Para el presente caso, se ha decidido llevar a cabo cálculos de velocidad de lectura con subtítulos en alemán y castellano de tres filmes que el autor conoce bien, bien sea por su docencia en subtitulación, bien por sus estudios de teoría de los medios audiovisuales, o bien por su investigación anterior. Respectivamente, los títulos elegidos, son *Lolita* (Stanley Kubrick, 1962), *Drugstore Cowboy* (Gus van Sant, 1989) y *Lost in Translation* (Sofia Coppola, 2003).

El criterio principal a la hora de su selección podría simplificarse de una manera muy sencilla atendiendo a la cantidad de diálogos, con una tendencia claramente descendente entre los tres. Desde un punto más académico, se definió dicha característica en un trabajo anterior (Martí Ferriol 2010: 193) como la “preponderancia del lenguaje filmico”, que en este caso, mostraría una tendencia claramente ascendente en la serie de tres.

Dado que el segundo criterio es más difícil de cuantificar, y con la finalidad de poder establecer unos órdenes de magnitud para su comparación, se presentan a continuación la cantidad de subtítulos de las tres películas, tanto en su versión subtitulada en alemán como en castellano. De las tres películas, la primera tiene una duración superior, pero a la vista de los números a continuación, resultan evidentes las diferencias. La película de Kubrick (con una duración de unos 150 minutos), incluye 2200 subtítulos en alemán, y 2259 en castellano. En la de van Sant (unos 100 minutos), aparecen 1059 subtítulos en alemán y 1155 subtítulos en castellano. Por fin, para la de Coppola, de unos 95 minutos de duración, se han pautado y creado 723 subtítulos en alemán y 815 en castellano.

Las diferencias entre la cantidad de subtítulos en las dos lenguas se puede explicar en función de la diferente cantidad de subtítulos de una o dos líneas, que puede

variar (el pautado, como la traducción, no es una ciencia exacta, y puede haber más de una solución para un mismo problema, véase las Tablas 1 y 2 del siguiente apartado), aunque curiosamente, en los tres casos presentados, los subtítulos son más numerosos en nuestra lengua.

También es necesario explicar la procedencia de dichos subtítulos. Dos son los procedimientos habituales para obtener los subtítulos de una película (véase Martí Ferriol 2009). El primero consiste en adquirir el DVD, digitalizar sus contenidos y extraer los subtítulos con programas al uso. El segundo, más sencillo, consiste en descargarlos de los portales existentes a tal efecto en Internet. Tres de dichas páginas, de donde se han descargado los subtítulos objeto de esta investigación, y que por otro lado son de las que cuentan con mayor cantidad de subtítulos, son:

- www.allsubs.org
- www.Mysubtitles.com
- www.findsubtitles.com

Otra posible discusión, en buena lógica, consiste en plantearse si los subtítulos obtenidos directamente de un DVD son diferentes de los descargados de Internet. Se ha comprobado estadísticamente (utilizando tests de hipótesis) para una única película con subtítulos en 5 lenguas diferentes (en un trabajo de investigación pendiente de publicación), que la calidad de los subtítulos sí varía, en cuanto al seguimiento de las convenciones profesionales al uso se refiere. Sin embargo, en términos de velocidad de lectura, es decir, en cuanto a los valores de CPS y WPM obtenidos en subtítulos de ambas procedencias, se ha comprobado que no hay diferencias estadísticamente significativas.

4. Metodología

A continuación se describe de forma muy sucinta la forma de proceder que se ha seguido para obtener los resultados que se presentan en el apartado siguiente. Una vez localizados y descargados los ficheros con los subtítulos en alemán y en español de las tres películas elegidas, se han seleccionado unos 300 aproximadamente en cada lengua y de cada película (son 6 combinaciones posibles, tres películas en dos lenguas). En consecuencia, el número total de subtítulos analizados se aproxima a los 2000.

Dichos contenidos se han copiado a ficheros de Word que contenían la herramienta de cálculo de velocidades de lectura, una macro en Visual Basic para Word. Tras el necesario ajuste de los formatos que la nueva aplicación requiere, se ha ejecutado la macro, con la consiguiente creación de una tabla con los resultados.

A continuación se muestran ejemplos de dichas tablas con cinco subtítulos en alemán y seis en castellano de *Lost in Translation*, que corresponden al mismo fragmento de diálogos de esta película. Además de presentarse con la intención de mostrar la información generada con la herramienta de cálculo, estas tablas podrían utilizarse como punto de partida para un análisis de tipo cualitativo muy simplificado,

en que se podría ver con algunos ejemplos determinados que la cantidad de palabras en los subtítulos alemanes tiende a ser inferior.

Subtitle Number	Subtitle Text and Time	Duration	Characters	Words	CPS	WPM
60	00:13:23:58 00:13:27:71 Ich weiß nicht, / ich war heute in einem Tempel und...	4,73	51	10	11	127
61	00:13:28:55 00:13:32:21 die Mönche da haben / die ganze Zeit gesungen.	3,66	43	8	12	131
62	00:13:32:30 00:13:36:43 Und ich habe nichts empfunden. / Verstehst du? Und...	4,73	50	8	11	101
63	00:13:36:89 00:13:40:97 Ach, was soll ich sagen, / ich hab's sogar mit Ikebana versucht.	4,68	61	11	13	141
64	00:13:41:06 00:13:45:14 Und John nimmt diese Haarprodukte... / Mein eigener Mann ist mir so fremd.	4,68	71	12	15	154

Tabla 1: Ejemplos de tabla generada con valores de velocidad de lectura de subtítulos en alemán

Subtitle Number	Subtitle Text and Time	Duration	Characters	Words	CPS	WPM
92	00:13:23:90 00:13:26:39 No sé -- / Hoy fui a un santuario.	2,49	30	7	12	169
93	00:13:26:53 00:13:27:97 Y...	2,04	4	1	2	29
94	00:13:28:84 00:13:31:63 había unos monjes / que estaban cantando.	2,79	38	6	14	129
95	00:13:31:77 00:13:35:71 Y no sentí nada, ¿entiendes?	3,94	28	5	7	76
96	00:13:35:84 00:13:38:64 Y... no sé...	2,8	13	3	5	64
97	00:13:39:51 00:13:43:04 Incluso probé el Ikebana. / Y John usa productos para el cabello.	3,53	62	11	18	187

Tabla 2: Ejemplos de tabla generada con valores de velocidad de lectura de subtítulos en castellano

Resulta patente la diferente numeración de los subtítulos. La cantidad total para este film es superior en el caso de los subtítulos en español (como se puso de manifiesto en el apartado anterior), y este hecho ya se hace patente en este momento determinado de la película, es decir, transcurridos unos 13 minutos. También se pone de manifiesto el diferente pautado, al observar cómo el mismo fragmento del diálogo puede aparecer en subtítulos diferentes, agrupado con otros fragmentos (de hecho, el subtítulo número 64 de la Tabla 1 incluye una última frase adicional, que aparece en el subtítulo 98 de la versión española y no se ha mostrado). Sin embargo, en subtítulos con contenido comparable (como el 60 y el 92, o el 61 y el 94), se observa que los valores de velocidad de lectura expresados en WPM son similares o inferiores en el caso de la versión subtitulada en alemán. Este breve ejemplo de cómo se podría llevar a cabo un análisis de tipo cualitativo refuerza la necesidad de desarrollar una metodología de análisis cuantitativa, si se desea compensar estas posibles diferencias en pautado y número de subtítulos, llevando a cabo un cálculo con un corpus que contenga un número de subtítulos mayor (como los casi 2000 de este estudio).

Para poder llevar a cabo el análisis cuantitativo propuesto, los resultados de velocidad de lectura (expresados en CPS y WPM) se han transferido a un programa de tratamiento estadístico de datos (Minitab for Windows), donde se han realizado las pruebas que han generado los resultados siguientes.

5. Resultados y discusión

Las siguientes tablas muestran los resultados obtenidos por la macro para los subtítulos de las 3 películas del corpus en las dos lenguas, expresados tanto en CPS (Tabla 3) como en WPM (Tabla 4).

LENGUA	FUENTE	# DE SUBTÍTULOS	MEDIA	DESV. ESTÁNDAR	MEDIANA	Q1 - Q3
DE	CPS_LOLITA	321	13,2	4,2	14	10 - 16
DE	CPS_DRUGSTORE	320	11,8	3,8	13	10 - 14
DE	CPS_LiT	257	10,8	3,7	12	9 - 14
ES	CPS_LOLITA	327	12,9	4	13	10 - 16
ES	CPS_DRUGSTORE	370	12,3	5,7	12	8 - 16
ES	CPS_LiT	376	11,6	5	12	8 - 15
Total / Media		1971	12,1	4,4	12,7	
Max.			13,2	5,7	14	16
Min.			10,8	3,7	12	8
% Variación			20	45	16	

Tabla 3: Resultados de velocidad de lectura expresados en CPS

LENGUA	FUENTE	# DE SUBTÍTULOS	MEDIA	DESV. ESTÁNDAR	MEDIANA	Q1 - Q3
DE	WPM_LOLITA	321	136	46	136	100 - 167
DE	WPM_DRUGSTORE	320	131	46	132	105 - 160
DE	WPM_LiT	257	117	41	122	90 - 143
ES	WPM_LOLITA	327	135	50	135	97 - 170
ES	WPM_DRUGSTORE	370	141	66	138	92 - 180
ES	WPM_LiT	376	132	62	129	88 - 178
Total / Media		1971	132	52	132	
Max.			141	66	138	180
Min.			117	41	122	88
% Variación			18	48	12	

Tabla 4: Resultados de velocidad de lectura expresados en WPM

En primer lugar, conviene explicar los cálculos estadísticos utilizados para obtener los valores presentados en las tablas, para, a continuación, analizar las implicaciones de los mismos en la hipótesis de partida enunciada.

Como puede observarse, el número total de subtítulos analizados es 1971; para cada película en cada lengua se han realizado cálculos de CPS y WPM. También se puede apreciar en las tablas, como se indicó anteriormente, que la cantidad de subtítulos para una película puede variar de una lengua a otra. En cada uno de los tres filmes se fijó un momento determinado en la película, y se estudiaron los subtítulos desde el inicio hasta dicho tiempo. La disparidad más evidente se muestra en el caso de *Lost in Translation*, y se debe a la gran diferencia en la frecuencia de aparición de subtítulos de una única línea en la versión subtitulada al castellano.

Los resultados obtenidos de los dos parámetros de velocidad de lectura se analizaron por medio de un programa de cálculos estadísticos, donde se realizaron cálculos descriptivos generales. En concreto, para cada serie y cada uno de los dos parámetros se obtuvieron valores de medias, desviaciones estándar, medianas y valores de Q1 y Q3 (primer y tercer cuartil). Cada una de las dos tablas muestra valores de dichos estadísticos en cada serie. Después se calcularon los valores máximos y mínimos de ellos, así como sus porcentajes de variación, con la idea de elegir como parámetro más apropiado para representar la velocidad de lectura aquel para el que dicha variación fuera menor.

A continuación se exponen las conclusiones que se pueden extraer de dichos valores numéricos:

- Los valores medios en CPS (12,1) y WPM (132) son muy próximos (especialmente el primero) a los citados en la bibliografía para la mencionada “ley de los 6 segundos”. En consecuencia, se corrobora la vigencia de esta ley en la práctica subtituladora del mundo real.
- Los valores medios de las medias y las medianas (tanto para CPS como para WPM) son muy próximos entre sí, lo que en términos estadísticos nos permite deducir un comportamiento “normal” de los datos, que implica aleatoriedad y bajo grado de manipulación.

- En cuanto a los porcentajes de variación de ambos parámetros, se observan valores del mismo orden de magnitud (20% para CPS y 18% para WPM). Dicho resultado refuta la afirmación intuitiva inicial, en el sentido de que expresar la velocidad de lectura en CPS puede tener ventajas por su menor variación entre lenguas. Estos valores nos indican que no es el caso, al menos con este corpus, esta cantidad de subtítulos y estas dos lenguas.

Para poder comprobar el cumplimiento de la hipótesis inicial, formulada afirmando que el número de palabras en los subtítulos en alemán sería inferior que en castellano, se ha llevado a cabo un test de hipótesis (“T-test”) con las dos series de valores obtenidos para cada una de las tres películas (en las dos lenguas), mostrados en la Tabla 2 (136, 131 y 117 –alemán– frente a 135, 141 y 132 –español–). A continuación se muestran los resultados de la prueba llevada a cabo con el programa estadístico, así como un gráfico ilustrativo.

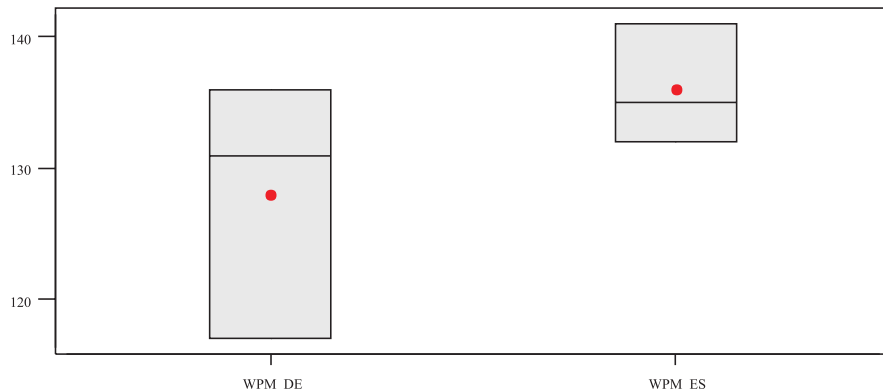
Two Sample T-Test and Confidence Interval

Two sample T for WPM_DE vs WPM_ES

	N	Mean	StDev	SE	Mean
WPM_DE	3	128,00	9,85	9,85	5,7
WPM_ES	3	136,00	4,58	4,58	2,6

95% CI for mu WPM_DE - mu WPM_ES: (-35,0; 19,0)

T-Test mu WPM_DE = mu WPM_ES (vs not =): T = -1,28 P = **0,33** DF = 2



Aunque la media de velocidad de lectura expresada en WPM es inferior en alemán que en castellano (128 frente a 136), los resultados del test de hipótesis muestran un valor de $p=0,33$, lo que nos indica que la hipótesis nula no se cumple, es decir que ambas series de datos no pertenecen a poblaciones diferentes. De acuerdo con los criterios estadísticos al uso, se necesitaría un valor de p inferior a 0,05 para

poder afirmar con un 95% de certeza que las medias de las dos series son estadísticamente diferentes. A esta misma conclusión se puede llegar por medio del análisis visual del gráfico de cajas (“Boxplot” en inglés) resultado de la prueba, donde se observa la superposición del lado superior de la caja de los valores para los subtítulos alemanes con el lado inferior de la caja de los subtítulos españoles. Cuando el resultado de una prueba de este tipo arroja valores de p inferiores a 0,05 (95% de confianza), dicha superposición entre las cajas no se observa.

Debe ponerse de manifiesto que los resultados presentados podrían ser distintos si al corpus propuesto se añadieran más películas con sus respectivos subtítulos. Este ejercicio se propone ahora como actividad para futuras investigaciones.

Para terminar la exposición de este trabajo de investigación empírico y descriptivo, y dado el corpus y los datos manejados, podemos afirmar que la hipótesis de partida se ha cumplido de forma parcial; es decir, la cantidad de palabras en subtítulos en lengua alemana es inferior que en castellano, aunque dicha afirmación no se puede realizar de una forma estadísticamente rigurosa (o con un grado de confianza estadísticamente significativo).

A la hora de buscar una explicación a este fenómeno, se podría aventurar como causa la información presentada en el segundo apartado, en el sentido de que parece que los valores de un corpus extenso de subtítulos en castellano muestran frecuencias de aparición bajas de los sustantivos, la categoría gramatical de la composición por excelencia en la lengua alemana. En general, esto significaría que los sustantivos tienden a aparecer poco en los subtítulos (hecho constatado para la lengua castellana e implícitamente deducido de los resultados de este estudio de caso para la lengua alemana), aunque esta nueva hipótesis ahora mismo formulada debería a su vez ser comprobada en futuras investigaciones.

Referencias bibliográficas

- Chaume, F., *Cine y Traducción*. Madrid: Cátedra 2004.
- Díaz Cintas, J., *La traducción audiovisual: el subtítulo*. Barcelona: Ariel 2001.
- Díaz Cintas, J., *Teoría y práctica de la subtitulación inglés – español*. Barcelona: Ariel 2003.
- Díaz Cintas, J., «Teaching and learning to subtitle in an academic environment», en Díaz Cintas, J. (ed.), *The didactics of Audiovisual Translation*. Amsterdam/Filadelfia: John Benjamins 2008, 89-105.
- D’Ydewalle, G., J. V. Rensbergen y J. Pollet: «Reading a message when the same message is available auditorily in another language: The case of subtitling», en J. K. O’Reagan y A. Lévy-Schoen (eds.), *Eye Movements: From Physiology to Cognition*, Amsterdam / New York: Elsevier Science Publishers 1987, 313-321.
- Karamitroglou, F., «A Proposed Set of Subtitling Standards in Europe». *Translation Journal* 2 (1998) (<http://accurapid.com/journal/04stndrd.html>).
- Mayoral, R., «El espectador y la traducción audiovisual», en: Chaume, F. y Agost, R. (eds.), *La traducción en los medios audiovisuales*. Castellón: Universitat Jaume I, Servei de Publicacions 2001, 33-48.
- Martí Ferriol, J.L., *Estudio empírico y descriptivo del método de traducción para doblaje y subtitulación*. Tesis doctoral defendida en el Departament de Traducció i Comunicació.

- ció de la Universitat Jaume I 2006 (<http://www.tesisenxarxa.net/TDX-1122106-122044/>).
- Martí Ferriol, J.L., «Herramientas informáticas disponibles para la automatización de la traducción audiovisual (“revoicing”)», *Meta* 54 (3) (2009), 622-630.
- Martí Ferriol, J. L., *Cine independiente y traducción*, Valencia: Tirant Lo Blanch 2010.
- Martí Ferriol, J.L., «Nueva aproximación al cálculo de velocidades de lectura de subtítulos», *TRANS* (en prensa).
- Oster, U. «Tradición y nuevas perspectivas en el estudio del compuesto nominal alemán», *Epos*, revista de Filología XXII, Madrid: UNED (2006), 221-234.
- Quesada-Granja, C., «Distribución por longitud de las palabras de diferentes idiomas» (2009) (<http://www.slideshare.net/quesadagranja/distribucion-por-longitud-de-las-palabras-de-diferentes-idomas-presentation>).
- Romero Fresco, P., «More haste less speed: Edited versus verbatim respoken Subtitles», *VIAL* vol. 6 (2009) (<http://webs.uvigo.es/vialjournal/pdf/Vial-2009-Article6.pdf>).
- Toda, F. & González Iglesias, «Spoken language and ICC: Managing Cultural Diversity in Dubbing and Subtitling in Spain», *Boras NIC 2009* (<http://nic.hb.se/index.php?id=12>).
- Wikcionario, corpus de subtítulos de películas en español ([http://es.wiktionary.org/wiki/Wikcionario:Frecuentes-\(1-1000\)-Subt%C3%ADtulos_de_pel%C3%ADculas](http://es.wiktionary.org/wiki/Wikcionario:Frecuentes-(1-1000)-Subt%C3%ADtulos_de_pel%C3%ADculas)).

Agradecimientos

A la Dra. Ulrike Oster por su orientación y comentarios durante el proceso de elaboración del presente artículo.