



**UNIVERSIDAD
DE GRANADA**

Tesis doctoral:

**EL CONCEPTO DE BÚSQUEDAS SEMÁNTICAS
APLICADO AL TEXTO DE LAS *HISTORIAS* DE
HERÓDOTO**

Doctorando: JAIME RANCHAL BEATO

Director: Dr. PEDRO PABLO FUENTES GONZÁLEZ

Departamento de Filología Griega y Filología Eslava
Programa Oficial de Doctorado en *Lenguas, Textos y Contextos*
(RD 1393/2007)

2017

Editor: Universidad de Granada. Tesis Doctorales

Autor: Jaime Ranchal Beato

ISBN: 978-84-9163-542-0

URI: <http://hdl.handle.net/10481/48333>

El doctorando / *The doctoral candidate* [Jaime Ranchal Beato] y el director de la tesis /
and the thesis supervisor: [Pedro Pablo Fuentes González]

Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

/

Guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisor/s and, as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when their results or publications have been used) have been respected.

Lugar y fecha / Place and date:

Director de la Tesis / *Thesis supervisor*:

Doctorando / *Doctoral candidate*:

Firma / *Signed*

Firma / *Signed*

I. INTRODUCCIÓN

I.1. BÚSQUEDA SEMÁNTICA Y SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN (SRI)

El objetivo de esta tesis es abordar diferentes técnicas de búsqueda semántica y analizar su aplicabilidad a textos escritos en griego antiguo, para lo cual hemos tomado como ejemplo el texto de las *Historias* de Heródoto. Entendemos por *búsqueda semántica* una serie de métodos de extracción o recuperación automática de información. La recuperación de información o *information retrieval* (en adelante *RI*) es un concepto introducido por Mooers 1950 dentro de la ciencia de la Documentación para referirse al proceso de «descubrimiento» de nueva información y separarlo de otros dos procesos: el «almacenamiento» y la «reproducción» de información. Entendemos por *sistema de recuperación de información* (en adelante, *SRI*) una metodología que facilite el proceso de *RI* a partir de una fuente, en nuestro caso, escrita (cf. *infra* II.1. *Búsquedas de información. Búsquedas semánticas*).

La función de un *SRI* es obtener recursos informativos relevantes para una necesidad de información dada, mediante una búsqueda dentro de unos recursos concretos (cf. Moya Anegrón 1996; Mooers 1951). En nuestro caso la necesidad que motiva el uso de un sistema semejante es encontrar la manifestación escrita de un concepto o conceptos, dentro de un corpus de textos. Tomemos un índice como ejemplo, es decir, una lista de términos presentes en un recurso dado, habitualmente un texto, o un corpus de textos. El usuario consulta las correspondencias entre un término o términos elegidos por él como representativos de un concepto o conceptos, que funcionan como criterio de búsqueda, de forma que obtiene la ubicación de uno o varios pasajes mediante los cuales pueda ampliar sus conocimientos sobre el concepto elegido. El origen de los *SRI* está en la creciente dificultad para extraer información útil y manejable de conjuntos de textos cada vez más grandes. Desde la invención de la escritura ha aumentado de modo exponencial la cantidad de información que se puede almacenar, un problema para el que empleamos un término acuñado en los estudios sobre representación visual de la información: la *sobrecarga informativa*¹. Casi tres siglos antes Diderot describe algo semejante: «*Tandis que les siecles*

¹ La expresión *sobrecarga informativa* o *infoxicación* (término propuesto por Cornellá 2004) se refiere a la «incapacidad para comprender y asimilar información provocada por el exceso de esta» (cf. Urbano 2010, 305 y Cornellá 2004, 9). Desde que Toffler 1970, 350-355 denunciara este problema, refiriéndose al mismo como *information overload* (traducido como «sobrecarga de información» en la traducción española de Ferrer Aleu, 247), son muchos los estudios que analizan el problema que supone en diferentes ámbitos y sus posibles soluciones (cf. Shenk 1997, 20-26, y Melgoza *et al.* 2002 para una definición general del problema, así

s'écoulent, la masse des ouvrages s'accroît sans cesse, et l'on prévoit un moment où il seroit presque aussi difficile de s'instruire dans une bibliothèque, que dans l'univers, et presque aussi court de chercher une vérité subsistante dans la nature, qu'égarée dans une multitude immense de volumes» (Diderot & Alembert 1751, s.v. «Encyclopédie», 675). El autor describe un futuro desolador donde el conocimiento universal resulta imposible por la incapacidad de la mente humana de reducir a una escala manejable el peso abrumador de la información. Por otra parte, esta afirmación viene a reforzar, de forma paradójica, la idea sobre la que se justifica la existencia de la *Encyclopédie*, de que la colaboración entre los seres humanos puede superar las limitaciones individuales, y de que la solución al problema arriba mencionado pasa por la selección constante de lo más relevante.

Técnicas que podríamos considerar ejemplos de *RI* existen desde que el ser humano mostró interés por preservar y utilizar el conocimiento adquirido, incluso antes de que lo hiciera en un soporte material. La forma más antigua y conceptualmente más simple de hacer *RI* es buscar manualmente en cada ítem de una colección para determinar si satisface las necesidades de información, un método denominado en el ámbito de la informática (y no sin razón) *fuerza bruta* (cf. Meadow *et al.* 2007, 21). Este método resulta tanto más efectivo cuanto menor es la cantidad de documentos que se ha de revisar, pero resulta inadecuado para grandes colecciones, como ya percibió Diderot. Si una colección es grande y se espera que vaya a ser consultada a menudo, añadir cierto orden a la misma suele facilitar el proceso en tiempo y esfuerzo. Esto significa utilizar algún criterio de ordenación que permita convertir una colección aleatoria en un archivo organizado, o bien crear un índice, un archivo separado que sirva de guía sistemática para los ítems en la colección. Se puede decir que la historia temprana de los sistemas *RI* es la historia de cómo se crearon y consultaron índices.

Antes de la extensión del uso de los ordenadores como herramientas de búsqueda de texto (y de información), consultar uno a uno todos los textos disponibles cada vez que se deseaba encontrar un pasaje en concreto resultaba, además de lento, difícil y la probabilidad de error no era pequeña. En lugar de ello se crearon ayudas «físicas» en forma de marcas, etiquetas o listas representativas de los ítems de toda la colección, y que contenían solo una pequeña parte de la información que contenían, considerada representativa. Podía ser el nombre del autor, de la obra, la primera línea del texto o el lugar de procedencia, pero sin

como Schick *et al.* 1990 para una revisión de su desarrollo en el tiempo); por citar algunos: Hopkins 1995 (biblioteconomía); Edmunds & Morris 2000 (negocios); Wilson 1996 (investigación académica); Yang *et al.* 2003 (representación de información en internet). Para una revisión más extensa de la bibliografía disponible sobre la cuestión remitimos a Eppler & Mengis 2004.

duda era una tarea mucho más fácil e igualmente permitía localizar el texto deseado. La disposición física de los textos, en una biblioteca, por ejemplo, constituye una forma de indexación. Este método tiene sus limitaciones, puesto que sólo puede utilizarse un criterio para ordenar los ítems, mientras que la mayoría de libros y registros de datos cubren múltiples temas y a menudo tienen más de un autor.

Las bibliotecas son el ejemplo más representativo de hasta dónde llegó el grado de elaboración de las técnicas de *RI* en una época «pre-digital», puesto que constituyen un verdadero sistema de ordenación y clasificación de la información destinado a facilitar el acceso a la misma *a posteriori*. No obstante, la diferencia entre la gestión de la información en una biblioteca o en otros sistemas no digitales con respecto a los digitales no es cualitativa, sino de grado, y sus diferencias radican fundamentalmente en las características físicas del soporte en el que se almacena la información, como veremos más adelante.

I.2. ESTRUCTURA DEL SIGNIFICADO: DATOS E INFORMACIÓN

Cuando hablamos de *información* entendemos, por un lado, un conocimiento o conocimientos adquiridos de alguna manera, pero también una «propiedad estadística de los mensajes, sin ninguna referencia al significado o a la semántica»². Consideramos que en el fondo de esta disparidad se encuentra una diferencia de enfoque en cuanto al uso de la idea expresada por el término. No cabe duda de que el término *información* tiene una fuerte carga semántica y que está estrechamente relacionado con una actividad tan relevante para el ser humano como es «conocer». Precisamente, la teoría de la comunicación de Shannon & Weaver 1949 responde a una preocupación por encontrar indicios objetivables de esa *semánticidad* a la que aludíamos dentro del proceso comunicativo. Sin embargo, el «significado» es una cuestión compleja y abierta a debates potencialmente infinitos, debido a la naturaleza contextual del concepto. En consecuencia, recogemos el espíritu práctico de los creadores de la moderna «teoría de la información» al centrar nuestro estudio en las características objetivables del mensaje, pero nos distanciamos de ellos en que no nos negamos a encontrar una explicación semántica a los mismos, de una forma lógica y razonada. Para ello operamos con dos oposiciones básicas:

1. ***Datos e información***: la *información* es un concepto compuesto de unidades más pequeñas, que denominamos *datos*. Un dato debe ser simple, perceptible de forma evidente y preferentemente de naturaleza cuantitativa³.
2. ***Información y significado***: la *información* se opone al *significado*, que sólo existe como un proceso relacional en la mente del ser humano. Mientras que el texto escrito en papel, piedra o almacenado en una sucesión de impulsos eléctricos constituye una mera sucesión de mensajes físicos, únicamente el cerebro humano guarda significados. En consecuencia, el adjetivo «semántico» con el que definimos el tipo de búsquedas de texto estudiadas en esta tesis alude metonímicamente al resultado ofrecido desde el punto de vista del usuario.

² Cf. Shannon & Weaver 1949, 8-9. En la misma línea, para Meadow *et al.* 2007, 1-2, la información es «algo» expresado mediante un conjunto estructurado de símbolos, pero la interpretación o el significado que puedan tener estos depende del receptor.

³ Basamos nuestra definición de *dato* en el concepto de *axioma*, tal y como se utiliza en matemáticas y lógica (cf. Maddy 1988 y Wolff 1963, 47-48), puesto que cumple una función semejante: la de servir de base sólida para la construcción de conocimientos más complejos.

En resumen, lo que todo *SRI* permite detectar es una serie de datos *cuantitativos* en uno o varios documentos, de modo que un intérprete humano pueda «extraer» información, estableciendo *a posteriori* relaciones *cualitativas* entre los mismos. En este sentido, respecto a la intervención de los ordenadores en un *SRI*, no debemos perder de vista lo siguiente: los ordenadores y los programas que se ejecutan mediante ellos facilitan el acceso al material, pero no lo interpretan, del mismo modo que las estanterías y las etiquetas atadas a los papiros de Alejandría no hacían filología por sí mismos, sino que ayudaban al filólogo en su tarea⁴. Los *SRI* han podido desarrollarse enormemente gracias a esta herramienta, impulsados y sustentados por una continua y viva reflexión teórica, a la que aludiremos en más de una ocasión⁵. Sin embargo, la materia prima sobre la que se aplican las soluciones derivadas de este debate incesante y con la que se construyen los instrumentos con los que aplicarlas sigue siendo el texto escrito.

⁴ Como bien dice Dijkstra 1989, 1401: «when all is said and done, the only thing computers can do for us is to manipulate symbols».

⁵ La reflexión en torno a la informática comienza por la propia terminología. De aquí en adelante tomamos la definición del término *informática* que dan Gómez y Méndez & Leal Gil 2001, 102: «el conjunto de conocimientos científicos y técnicos que hacen posible el tratamiento automático y racional de la información por medio de ordenadores o computadores electrónicos».

I.3. MATERIAL DE ESTUDIO: EL TEXTO

Todo *SRI* opera sobre unas «fuentes de información» determinadas. Dentro de la ciencia de la Documentación, el término habitual para referirse a estas fuentes es *documento*. Palomares Perraut 1997, 29-33, analiza la etimología y utilización del mismo, concluyendo a partir de las definiciones de Otlet 1934, 43, Pinto Molina 1993, 64, y Pérez Álvarez-Ossorio 1990, 27 que *documento* tiene tres acepciones fundamentales: es un «objeto material», que «conserva información en un soporte», de modo que pueda ser «transmitida». De acuerdo con Pinto Molina 1993, 66-67, los documentos pueden clasificarse según su *soporte*, su *contenido* o según el *modo de difusión*⁶. Nuestro análisis se centrará en los documentos *textuales*, es decir, en aquellos que utilizan un soporte escrito.

I.3.1. DEFINICIÓN DE TEXTO

El término *texto* alude a una realidad compleja y no existe un consenso absoluto en cuanto a su definición⁷. Puesto que constituye una noción central para nuestra tesis, creemos necesario abordar la cuestión de qué es un *texto*, para poder proceder con firmeza a repasar los métodos mediante los cuales es posible extraer información del mismo. Como ocurre con el término *documento*, la palabra *texto* indica varias cosas diferentes; por un lado es un concepto ligado a la historia de las técnicas de comunicación y expresión humanas, y, por otro lado, se refiere a los objetos creados mediante estas técnicas, es decir, a lo que tradicionalmente llamamos *literatura*⁸. Lo que ofrecemos a continuación es una selección crítica de las perspectivas de estudio del texto más relevantes a nuestro juicio, a partir de las cuales hemos fijado los criterios con los que proponer nuestra propia *definitio*.

Siguiendo una tendencia que surgió durante la Ilustración, diversas actividades hasta entonces eminentemente prácticas han ido reconociendo desde entonces la necesidad de un

⁶ Cf. Palomares Perraut 1997, 34-40 para una descripción más detallada de los distintos tipos de documentos.

⁷ Remitimos a Alcaraz Varó & Martínez Linares 1997, 564-565; Charaudeau & Maingueneau 2002, 549-551, para un panorama de las distintas definiciones de *texto*.

⁸ Cf. Greetham 1992 y 2007, 24-25, para quien el texto es una red, una intersección de teorías y prácticas, tejidas y descosidas en un ciclo continuo de destrucción-creación alrededor de un objeto llamado igualmente *texto*. Sobre *textus* como «unidad coherente del discurso», cf. Quint. Inst. IX 4, 13 y 4, 17 en calidad de precedente.

trasfondo teórico sólido, entre las que se incluyen la filología o la crítica literaria⁹. La lingüística parecía el lugar más apropiado donde buscar fundamentos para una definición «científica» del texto, especialmente dentro de las teorías sobre el funcionamiento de la comunicación humana¹⁰. Con ello el significado de *texto* se distancia del soporte material y pasa a indicar una unidad de comunicación, el vehículo de un mensaje entre emisores y receptores, trasladando el *texto* del entorno de la crítica literaria al contexto de las ciencias de la Comunicación e introduciéndolo dentro de un campo semántico ocupado por otros términos como *oración*, *enunciado* o *discurso*.

Aquellos que ponen el énfasis en el aspecto comunicativo del concepto describen el texto como un fragmento, una sección determinada de la comunicación humana, que de por sí es un proceso infinito¹¹. Así, Van Dijk 1977, 3, relaciona *texto* con una «sequence of sentences» ubicándolo en un nivel superior al de la oración, como hace Slakta 1985. Ambos caracterizan al texto como una *unidad estructural* cuya función es poner límites al *continuum* de la comunicación, de forma similar a la *oración*.

Una relación similar existe igualmente entre *texto* y *enunciado*, que puede dibujarse mediante una clásica oposición saussuriana, donde la palabra *enunciado* actúa como término no caracterizado. En efecto, *enunciado* es un término «primitivo» en tanto en cuanto es un término fundamentalmente práctico, empleado como indicativo general para el material de trabajo de la lingüística, de modo similar al inglés *utterance* (cf. Crookes 1990). Por su propia etimología, comparte con el *texto* su adscripción a las unidades de comunicación, así como la definición habitual de «secuencia», lo que conlleva la necesidad de establecer cuáles son sus límites; para ello se ha recurrido al *significado* como criterio, de modo que el *texto/enunciado* sea una sección del *continuum* lingüístico, dotada de sentido y sintácticamente completa¹². La única diferencia entre *oración* y *enunciado* sería de naturaleza pragmática, en cuanto el primer término no incluye referencia alguna al

⁹ Harrison 2001b, 3, siembra la duda sobre la posibilidad de ejercer la filología sin una base teórica interdisciplinar. En España, obras como el manual de Bernabé Pajares & Hernández Muñoz 2010 tratan de solucionar el problema, tomando como punto de partida la naturaleza fundamentalmente práctica de la crítica textual (cf. 2010, 9-10).

¹⁰ Van Dijk 1977, 3, emplea *texto* para referirse a una unidad abstracta superior a la oración; esta interpretación del texto como «estructura» ha tenido éxito dentro de la informática (cf. DeRose *et al.* 1990).

¹¹ Weinrich 1977, 12-13, describe el *texto* como una simple serie de signos o un sinónimo de *mensaje*.

¹² Remitimos a Gutiérrez Ordóñez 1997, 40-65, y a Osuna García 2006 sobre la difusa línea divisoria entre oración y enunciado.

contexto. Sin este matiz, no sería posible explicar la coexistencia de significados diferentes para una misma estructura oracional (cf. Ducrot & Schaeffer 1995, 250). Para Adam 1992, 15, una posible solución para esta tríada de «secuencias comunicativas» sería limitar la referencia del término *enunciado* al objeto material, observable, en el que se manifiesta el texto, que se convertiría por lo tanto en un objeto *abstracto* (cf. Charaudeau & Maingueneau 2002, 213-215).

Aunque la descripción del texto como unidad significativa dentro del proceso de la comunicación nos permite hablar de textualidad en otros soportes no escritos, como la imagen o el intercambio oral, el material sobre el que trabajan las disciplinas en cuyo seno surgió el debate sobre el texto son objetos escritos. Tanto más cierto es para la filología clásica, dedicada casi en exclusiva al estudio de los testimonios escritos de la Antigüedad¹³. No obstante, es igualmente cierto que la filología se ha beneficiado de los enfoques comunicativos, en particular del concepto de *discurso*.

Fue Guillaume 1973 quien puso de moda el término dentro de la lingüística moderna, con el declive del estructuralismo y el ascenso de las corrientes pragmáticas. Con respecto a la oración, coincide con las definiciones anteriores del texto, en cuanto sería una unidad superior a la misma y formada por un número no definido de oraciones. Por otra parte, continuando con el recurso a las oposiciones, *discurso* se enfrenta a *lengua* porque representa una selección o una restricción de las posibilidades que ofrece la lengua como un sistema general. Para Gardiner 1989, 285, esta selección puede interpretarse como la consecuencia de un contexto concreto, o bien como un patrón que se repite en contextos similares. En el primer sentido y según Guillaume 1973, 71, el discurso hace referencia a manifestaciones de la intención individual del hablante, como puede ser su intención o su actitud con respecto a lo dicho, a lo que se va a decir. En el segundo sentido, el discurso viene a ocupar una posición semejante a la de los campos léxicos, dentro de una esfera superior, ocupada por las oraciones y entidades mayores. Este uso del discurso está especialmente abierto a ambigüedades, puesto que es posible indicar con él no sólo el conjunto de textos (u objetos textuales o enunciados) que reúnen determinadas características, sino el propio sistema de reglas por el que son producidos. Así,

¹³ Cf. Bernabé Pajares & Hernández Muñoz 2010, 9: «el fin último de la filología, que debe ser el conocimiento de una civilización [...] a través de sus textos»; de la misma opinión es Alsina 1963, 172, para quien la filología es «un arte o una ciencia cuya finalidad reside en la edición, explicación e interpretación de los textos».

para Charaudeau & Maingueneau 2002, 180, el discurso *surrealista* describe no solo un conjunto de textos cuyos rasgos permiten atribuirle dicho adjetivo, sino el conjunto limitado de normas que ha producido dicho conjunto.

Discurso, texto y enunciado entrecruzan sus esferas de uso con frecuencia, y los seguidores de la teoría del discurso tampoco han conseguido llegar a una distinción absolutamente clara¹⁴. El recurso habitual consiste en subsumir algunos de los términos como una parte de la definición de otro, pero más interesante nos parece la definición de *discurso* en términos de «texto como unidad comunicativa», enfrentada a *enunciado*, que sería un «texto como unidad lingüística» (cf. Adam 1999, 39, y Guespin 1971, 10).

La extensión del término *discurso* dentro de la lingüística reciente responde a un cambio en la orientación del estudio del lenguaje, especialmente con respecto a su funcionamiento y a su semántica, fruto de la influencia de diversas corrientes pragmáticas. Como hemos visto, la ambigüedad del término no permite el desarrollo de una disciplina específica que lo estudie (por más que se hable de una lingüística del discurso), sino que viene a representar un viraje en la concepción global del lenguaje hacia su consideración como un fenómeno social, en contra de la idea decimonónica del lenguaje como una especie de ser vivo independiente. Este movimiento se asienta sobre una serie de ideas, recogidas por Charaudeau & Maingueneau 2002, 180-184, y que merece la pena mencionar brevemente:

1. El discurso presupone una *organización transoracional*; esto es, la existencia de una cohesión entre las diferentes oraciones que lo componen.
2. La forma del discurso está condicionada por la intención comunicativa del locutor y por el contexto en el que se realiza.
3. El discurso es una forma de acción, y por lo tanto no sólo informa, sino que pretende modificar una situación (cf. Austin 1962, 12-24, y Searle 1969, 16-19).
4. El discurso es *interactivo*, es decir, está inserto en una conversación entre locutor y receptor, tanto si este último participa efectivamente, como en una conversación oral, como si se sobreentiende su participación, en el caso de textos escritos u otras formas de exposición oral donde el auditorio no tiene ocasión de intervenir (conferencias, radio, televisión etc.).

¹⁴ Cf. Benveniste 1966, 266, que iguala los términos *discurso* y *enunciado*.

5. El discurso está *contextualizado*. La relación entre discurso y su contexto es también interactiva, puesto que este último no sólo es necesario para atribuir sentido al discurso, sino que contribuye a definirlo y puede llegar a modificarlo (*cf.* punto 3).
6. El discurso siempre se construye sobre un punto o puntos de referencia, sobre los que las localizaciones cobran sentido. Esta idea puede considerarse una especificación del punto anterior, limitado a la actitud del emisor con respecto a lo dicho, o al contexto real o supuesto.
7. El discurso está regido por reglas de carácter normativo, fijadas por el contexto social en el que se enmarca.
8. El discurso está incluido dentro de un “interdiscurso”. Esta última idea expande el punto 5 de modo semejante a la relación discurso-oración mencionada en el punto 1. Buena parte del contexto de cualquier enunciación está formado a su vez por *discursos*, sin cuyo conocimiento no sería posible la comprensión. Podemos encontrar un paralelo de esta idea en el concepto de *intertextualidad*.

¿Qué nos queda entonces para poder definir el texto de forma específica? Si el enfoque comunicativo-pragmático resulta útil para establecer relaciones con otros términos y conceptos, también es cierto que no hace sino recalcar la dificultad de alcanzar una definición concreta. Hemos visto cómo una solución provisional puede pasar por prestar atención a la forma física del texto, ignorando los puntos conflictivos¹⁵. En efecto, hemos abordado hasta ahora el texto como concepto, pero es evidente que la palabra *texto* hace también referencia en no pocas ocasiones a la realidad física del mensaje. Siguiendo este razonamiento, a continuación exploraremos los orígenes del texto en la Antigüedad y la evolución del texto como concepto y como objeto, así como las técnicas con las que la filología aborda su estudio y que suponen la base de algunos métodos actuales de *RI*.

I.3.2. EL TEXTO ORAL

La definición de *texto* que hemos dado hasta ahora es fruto de la evolución de una serie de conceptos e ideas surgidos en torno a la comunicación escrita en la Antigüedad grecorromana, y desarrollados en el seno de la cultura occidental. El texto es un objeto y un concepto culturalmente contingente, por lo que reconocemos que la descripción del mismo

¹⁵ Incluso para estudiosos como Greetham 1992, 104, partidario de entender el *texto* como un objeto *comunicativo*, la relación del mismo con la escritura es evidente.

sobre la que vamos a sustentar nuestro trabajo no es la única¹⁶. Mediante el reconocimiento explícito de su naturaleza particular, pretendemos purgar nuestro acercamiento a la cuestión textual de todo prejuicio generalista sobre sus funciones, efectos y formas.

El texto, representado habitualmente hoy de forma metonímica por una de sus manifestaciones más comunes, el libro, no siempre tuvo las connotaciones de fijación y prestigio que hoy tiene, ni la escritura fue considerada siempre un medio seguro, fiable y, por lo tanto, veraz, de transmitir la información. La llamada «revolución de la escritura» fue un proceso gradual de cambio de soporte, y no se produjo ni mucho menos sobre el vacío. La cultura griega, como muchas otras, antes de dominar la escritura desarrolló formas de comunicar la información, de crear arte y transmitirlo, y de preservar sus tradiciones e ideas sin necesidad de signo alguno tallado o dibujado. Para una sociedad oral como la griega, *escribir* (o *inscribir*) era una posibilidad más. Como instrumento, la escritura fue mejorada con el tiempo, ampliando la cantidad de usos que se le daban y sobre todo utilizándose de forma más efectiva. Medimos la efectividad de la escritura en función de lo que consideramos son dos de sus funciones principales: el almacenamiento de información y la capacidad de transmitirla¹⁷.

Pero hablar de la oralidad en la antigua Grecia resulta en cierto modo arriesgado y probablemente estemos abocados a cometer, de un modo u otro, algún anacronismo. Venimos de un mundo eminentemente escrito, que considera este método de transmisión de la información uno de los más importantes logros de la humanidad, y sobre el que se sostiene en cierto sentido la práctica totalidad de sus sistemas de almacenamiento, gestión, transmisión y adquisición de conocimiento. ¿Cómo es posible concebir desde esta perspectiva un mundo donde los signos escritos son tenidos por una herramienta secundaria, o, mucho peor, como algo innecesario o incluso pernicioso?

I.3.2.1. La oralidad y la memoria

Oralidad es un término eufemístico. *Oral* significa esencialmente «transmitido de boca en boca», sin que medie ninguna otra forma de comunicación por otro canal que no

¹⁶ Para una revisión del concepto de *texto* fuera de Occidente, remitimos a Eliot & Rose 2007, esp. el capítulo «The book beyond the West» (95-175).

¹⁷ De las cuatro funciones que Wells 1987 asigna a la escritura (*performative*, *functional*, *informational*, *epistemic*) nuestro estudio se centra en las dos centrales: en su utilización como instrumento de comunicación (*functional writing*) y para transmitir información (*informational writing*).

sea el auditivo; así pues, *oralidad* debería referirse al hábito de utilizar solamente la comunicación oral en lugar de la escrita, pero el término fue acuñado como una alternativa para *analfabetización* carente de las connotaciones morales negativas que este otro término trae consigo. Por lo tanto, y a pesar de su pretendida objetividad, la oralidad como concepto es una creación moderna que lleva implícita la existencia de un concepto paralelo y opuesto, expresado por los términos *alfabetización* y *escritura*, y cuyo contenido ya hemos visto que está tremendamente marcado por la cultura moderna occidental¹⁸.

Desde mediados del siglo XX, los testimonios orales han recuperado parte de su valor intrínseco, más allá de los prejuicios heredados de una larga tradición de predominio de la escritura, y es necesario reconocer el papel importante que tuvo en el cambio de actitud la investigación sobre la cuestión en el seno de la filología clásica¹⁹. A pesar del evidente cambio de actitud desde que Milman Parry hiciera tambalear la noción de un *Homerus scribiens*, ha requerido mucho esfuerzo alcanzar un pleno reconocimiento del papel de la oralidad en la antigua Grecia²⁰.

Desde nuestro punto de vista la oralidad es fundamentalmente una forma de comunicación, que genera unos hábitos que afectan tanto a la forma de entender el proceso comunicativo, como a la manera de crear y expresarse. Sólo así podemos entender la oralidad como un fenómeno que trasciende la simple comunicación, y hablar en consecuencia de culturas orales. Entender lo que representa la afirmación de que Grecia era una cultura oral significa conocer las bases sobre las que se construirá el edificio de la escritura: las formas de expresión, de composición literaria, de comunicación oficial y personal, de preservación de la memoria colectiva y de las creaciones dignas de entrar en ella. Todo ello está marcado por los límites que las posibilidades del medio oral imponían, y son el punto de partida del texto escrito.

Evitados los prejuicios, el segundo obstáculo para su estudio es, evidentemente, la ausencia de testimonios directos, y la necesidad de decidir y aislar qué huellas podemos encontrar de lo oral en los textos que han llegado hasta hoy. Por desgracia, no existen unos criterios generales que podamos aplicar sin dudar. En su ausencia, nos vemos obligados a

¹⁸ Tomamos esta ideas sobre las connotaciones de *oralidad* y *alfabetización* de Thomas 1992, 15-20.

¹⁹ Fruto del aumento del interés hacia el papel de la oralidad en la Antigüedad surgen trabajos como el de Thomas 1989, Andersen 1987.

²⁰ Cf. entre otros muchos Thomas 1992 y Harris 1991 para un panorama actual de la oralidad y la escritura en la Antigüedad, así como Halverson 1992; Calinescu 1993.

indagar pistas sobre cómo y dónde obtenían los autores antiguos su información, aplicando el conocimiento que poseemos de otras sociedades mejor conocidas. En este sentido resultan particularmente útiles los trabajos antropológicos que tratan la tradición oral en general, puesto que aportan métodos y principios bien asentados con los que afrontar el problema²¹. En el caso griego debemos partir de un nivel aún más básico, dado lo poco que sabemos con seguridad de la naturaleza de su tradición oral y de sus medios de transmisión.

Además de en la épica homérica, se han encontrado huellas de la tradición oral griega en la documentación de algunas polis, especialmente Atenas, como la conservada en torno a la expulsión de los Pisistrátidas, que, junto a los comentarios de Tucídides sobre la ἀκροή (tomada como sinónimo de «tradición oral»), son un foco de investigación sobre esta cuestión²². Aún así las perspectivas adoptadas resultan anticuadas, quizá por la influencia aún grande de trabajos como el *Atthis* de Jacoby 1949²³, y, si bien el debate se desarrolla en términos de transmisión oral, lo hace a través de presunciones modernas sobre la falsedad y el engaño en la propaganda política antigua, dando por sentado el funcionamiento de la tradición oral.

Escritura y oralidad deben ser examinadas en paralelo. Con frecuencia el excesivo celo metódico las ha relegado a posiciones enfrentadas, distinguiendo un periodo oral de otro posterior, «alfabetizado», mediante franjas artificiales de tiempo²⁴. No disponemos de suficientes datos para sostener de manera sólida que la evolución de una cultura oral a otra escrita fuera un proceso uniforme. En efecto, el grado, la extensión y las implicaciones de la alfabetización cambian a lo largo del tiempo, y de una sociedad a otra; tanto más si cabe en el seno de las multifacéticas comunidades griegas durante su larga historia. Lo que sí podemos afirmar es que una división estricta entre una sociedad *oral* y otra *escrita* resulta inapropiada. Atenas, por ejemplo, guardaba registros escritos y sin embargo el funcionamiento de la democracia y de sus instituciones descansaba en la tradición y comunicación orales (*cf.* Thomas 1989, 196-221). Además de ello, buena parte de la historia escrita

²¹ *Cf.* Vansina 1973 y 1985; ambos trabajos son una buena introducción al estudio antropológico de la oralidad, especialmente el segundo, en el que se presta más atención a los procesos mediante los que se forman las tradiciones orales.

²² *Cf.* Thomas 1989, 3, n. 3 y 1992, 112.

²³ Para un estudio detallado de las tradiciones orales antiguas que conocemos, remitimos al estudio de la oralidad en Atenas realizado por Thomas 1989, cap. 5.

²⁴ *Cf.* Pfeiffer 1968, 16-57, sobre el surgimiento de una *cultura del libro* en la Atenas clásica y su uso como ejemplo de esta pretendida evolución.

griega, especialmente en los primeros siglos de la literatura, derivaba de tradiciones orales anteriores. Por lo tanto, para comprenderlas debidamente debemos prestar atención a todo lo que nos pueda indicar cómo eran las fuentes orales que los testimonios escritos reflejan, y a la interacción entre unas y otros.

En las últimas décadas los filólogos clásicos han prestado cada vez más atención a los elementos de la cultura griega que dependían de la interacción hablada en lugar de la escrita, una tendencia que arranca con el trabajo sobre poesía oral de Parry 1930. A través de su estudio sobre los métodos de composición utilizados por los bardos de Yugoslavia en su tiempo, llegó a la conclusión, por entonces sorprendente, de que la *Ilíada* y la *Odisea* de Homero eran en esencia poesía oral. A partir de esta idea, hoy generalmente aceptada, y aunque la atención siga centrada fundamentalmente en Homero, resulta cada vez más evidente que muchas obras literarias de la Antigüedad griega eran escuchadas en lugar de leídas; presentadas y transmitidas de forma oral incluso aunque existiera una versión escrita²⁵.

Además de manifestarse en la épica homérica, en un sentido más amplio, la tradición oral era lo que proporcionaba a la mayoría de los griegos un conocimiento de su historia. Los textos griegos parecen indicar con frecuencia la existencia de una tradición oral o algún recuerdo transmitido de esta forma, en comentarios cuya presencia no se limita a los historiadores²⁶. En el caso de estos, particularmente Heródoto y Tucídides, en lugar de una sensación encontramos un reconocimiento explícito de la fundamentación de sus obras en testimonios orales²⁷.

²⁵ Cf. Havelock 1986, 47. El debate sobre la oralidad no se puede abordar seriamente sin distinguir entre oralidad de la composición, de la ejecución y de la transmisión. En el caso de Homero Havelock otorga un papel importante a la escritura en el proceso de composición, si bien Signes Codoñer 2004, 123, se manifiesta contrario a esta opinión.

²⁶ Cf. Pi. Sch. in Ol. 8 100a, 101; X. Vect. 4, 14 s.; Ephor. *FGH* 70, F9. También Isoc. 12, 149 ss. y 4, 30; Arist. *Rh.* 1390a, 5-10 y 1395a, 2-5; Plb. 38, 1, 6. Sobre la ἀκοή como medio de recepción de información es interesante Hdt. 2, 156; Th. 1, 20, 1 y 73, 2, 73, 6, 53, 3 y 60, 1), que se complementa con algunas referencias al testimonio visual directo (Hdt. 2, 99 y 147). Para ello remitimos a Hunter 1973, 27, n.25. Los términos expresan una distinción vital entre la comunicación oral predominante y los testimonios oculares (cf. Schepens 1975).

²⁷ La insistencia en las fuentes orales de la historia griega arranca con el monumental trabajo de Jacoby sobre Heródoto (1913, 413), a quien Murray 1987 dedica un tratamiento bastante detallado. En este sentido Momigliano 1972 y 1966b), así como Finley 1975 y 1985 insisten en la falta de atención de los historiadores hacia los textos escritos. En este sentido, otros estudiosos como Lang 1984, West 1985, y Nagy

Pocos trabajos examinan en detalle la tradición oral en el mundo antiguo por sí misma (*cf.* Thomas 1989), y entre ellos pocos van más allá de la mención de la importancia de la tradición oral, y citan quizá algún trabajo antropológico general²⁸. En efecto, el trasfondo teórico de este tipo de investigaciones se asienta sobre los descubrimientos de la antropología, que lleva tiempo discutiendo los problemas de la tradición oral y desarrollando complejos métodos para reunir, evaluar e interpretar los testimonios que ofrece. No obstante, los antropólogos estudian por lo general sociedades que aún existen y para las que es posible coleccionar grandes cantidades de datos *frescos*, un panorama muy diferente al que nos enfrentamos cuando estudiamos el mundo antiguo. Por lo tanto, la contribución de la antropología en este aspecto para la filología o la historia antigua es relativa, en cuanto sus observaciones y recomendaciones son más apropiadas para el trabajo de campo que para el análisis de textos. Además de esto, enfrentados a nuestro conocimiento del mundo griego, las sociedades, intereses, motivos narrativos y, por ende, la propia naturaleza de las tradiciones orales descritas por los antropólogos parecen muy distantes del mundo de Homero, Tucídides o Platón. En consecuencia, la actitud de los estudiosos del mundo antiguo no suele ir más allá de la reafirmación de generalidades sobre la fiabilidad de la transmisión oral en aquellas ocasiones donde no disponemos de fuentes escritas, estas son fragmentarias o están claramente basadas en testimonios orales.

El análisis de la oralidad no se reduce a un estudio del contenido referencial de la información transmitida de forma oral, ni debe plantearse en términos de simple fiabilidad. Así, por ejemplo, se suele pensar que el sistema de transmisión oral de la información se mantenía gracias a una mejor capacidad memorística de los miembros de aquellas sociedades, afirmación completamente falsa²⁹. Si genéticamente no eran diferentes, sus habilidades o capacidades potenciales debían ser las mismas que las de cualquier ser humano actual. Sí parece probable, en cambio, que tuvieran mejores mecanismos mnemo-

1987 tratan el texto de Heródoto como una forma de «narrativa oral». Sobre los «cuentos populares», implícitamente orales, Aly 1969 es todavía útil.

²⁸ En su introducción a las fuentes para la historia más antigua de Esparta, Cartledge 2002 apenas menciona el hecho de que la mayor parte de los testimonios deberían haber sido orales (2002, 43-44; desde una perspectiva herodotea Evans 1982, 146-152, trata de valorar la fiabilidad de las fuentes orales de Heródoto basándose en estudios sobre tradición oral en África; sobre las fuentes orales de la historia de Grecia, *cf.* Hunter 1982, 308-310 Murray 1980, 16-35 y 1987; Davies 1989 trata la oralidad en un tono semejante al hablar de la historicidad de las tradiciones sobre la guerra de Troya.

²⁹ Henige 1982, 67, afirma que la creencia en una memoria fabulosa es «largely untested and untestable». Esta habilidad se suele medir mediante la capacidad de «recuerdo libre» (*free recall*), es decir, de extraer de la memoria ítems aprendidos previamente en un orden aleatorio (*cf.* Waugh 1961).

técnicos a su disposición, aunque estudios en sociedades modernas que recurren a menudo al aprendizaje de memoria han demostrado que su uso no mejora esta capacidad, lo que ha llevado a algunos a cuestionar por completo la existencia de una memoria literal en sociedades orales³⁰.

Podemos servirnos, al menos en parte, de las conclusiones alcanzadas en el campo de la psicología cognitiva en torno al funcionamiento de la memoria, siempre y cuando seamos razonablemente prudentes³¹. El mecanismo por el que se crean y se reproducen los recuerdos afecta a la información que se transmite, es decir, a los datos del mensaje y a su expresión lingüística. El texto escrito actúa como soporte de un conjunto organizado de formas de tratar y comunicar la información que denominamos *literatura*, cuyo paralelo, predecesor y modelo es la memoria, en cuanto funciona como soporte de un sistema de almacenamiento, tratamiento y comunicación de la información basado en la oralidad. Es un instrumento, una herramienta en la misma categoría que la escritura o la movilidad de las manos, y como ambas se adapta a las necesidades del usuario.

1.3.2.1.a. El lenguaje formular

Como ya hemos mencionado, la filología comenzó a interesarse por el efecto de la oralidad y la memoria en los textos antiguos con el trabajo de Parry sobre el «lenguaje formular». Por desgracia, Parry no fue capaz de dar una definición que verdaderamente aislara el concepto de *fórmula* de otros muchos fenómenos ya conocidos, unidos todos por su carácter repetitivo. No obstante, su trabajo sirvió para que se replanteara el funcionamiento del proceso literario en la Antigüedad ¿Qué papel tenía el poeta? ¿Ejercía todo el control sobre la obra o algún otro elemento externo podía influir en ella? Parte de la respuesta ha venido desde la pragmática, una esfera tradicionalmente relacionada con lo oral, sobre la que se ha construido la teoría del texto como un «acto de habla» (*cf.* Searle 1969).

Parry 1930 redescubre en la poesía épica homérica una faceta vital no sólo de la literatura griega, sino de su cultura en general. Sin embargo, la propia visión de Parry sobre

³⁰ Baddeley 1983, 204-207, y Cole & Scribner 1981 rechazan la existencia de un concepto de *literalidad* en la Antigüedad; en contra de la opinión de estos últimos, *cf.* Goody 1987, cap. 8.

³¹ El funcionamiento del cerebro es un tema complejo y en gran medida aún desconocido; *cf. infra* I.3.2.2. «La memoria como soporte de la información».

la cuestión fue bastante limitada y ha generado no pocos errores³². Entre 1928 y su muerte prematura en 1935, Parry escribió una serie brillante de artículos en los que defendía la idea de que la *Ilíada* y la *Odisea* de Homero eran una muestra de poesía oral tradicional; es decir, el producto de una larga tradición, en lugar de la creación de un único genio poético. Aunque ya lo hemos mencionado brevemente, debemos dedicarle unas páginas en mayor detalle a su contribución al estudio de la oralidad, incidiendo en los aspectos que más nos interesarán.

Parry y su discípulo Albert Lord acudieron a algunos bardos del sur de Yugoslavia entre 1930 y 1950, donde pudieron observar cómo un poeta oral componía sobre la marcha y, en particular, cómo empleaba un cúmulo de piezas, temas y fórmulas fijadas por la tradición para ayudarse a componer al mismo tiempo que cantaba. El detallado análisis del texto homérico que hizo Parry a continuación parecía revelar un sistema similar de fórmulas tradicionales. Homero, por lo tanto, era un poeta oral. Su teoría demostró ser revolucionaria, y, gracias al éxito que cosechó, Homero es hoy generalmente reconocido como el representante por antonomasia de la poesía oral griega. El concepto de «poesía oral», también conocido como la *teoría de Parry y Lord*, ha sido aplicado a otras tradiciones poéticas de naturaleza épica o arcaica (antiguo nórdico, anglosajón, épica africana, poesía de Kara-Kirghiz, por ejemplo), pero en el mundo filológico tuvo el efecto de llamar la atención sobre el aspecto oral de la cultura griega, poco tratado hasta entonces y que se ha convertido en el eje de la comprensión de lo griego para algunos, como Havelock 1986, 15, que sitúan el comienzo de la pérdida de importancia del factor oral en una fecha tan tardía como el siglo IV a.C.³³.

El trabajo original de Parry se concentraba en un análisis exhaustivo de una de las características de la épica, el llamado «epíteto homérico». Cada personaje homérico posee un epíteto descriptivo y a la vez ornamental: γλαυκῶπις Ἀθήνη (Il. 1, 206), πολύτροπον (Od. 1, 1), πόδας ὠκὺς Ἀχιλλεύς (Il. 1, 58), etc. Parry defendía que tras el uso y la descripción de estos epítetos se escondía un sistema extremadamente complejo. Los epítetos cambiaban en función no solo de las exigencias de la narrativa, sino a menudo de simples razones métricas. Cada epíteto constituía una unidad que encajaba perfectamente

³² Lo que sigue está basado fundamentalmente en el resumen de la cuestión hecho por Thomas 1992, 31-36.

³³ Thomas 1992, 89-91 se apoya en Finnegan 1977, 73, y Jensen 1980, 22-28, para lanzar varios ataques contra la tesis de Parry y Lord: en su opinión, el debate sobre la oralidad continúa siendo demasiado generalizado, acrítico e incluso tautológico.

en una sección particular del hexámetro; si el mismo personaje era mencionado en otra parte del verso, se le asignaba otro epíteto que encajara mejor en la métrica. Estas fórmulas no se limitaban a palabras aisladas o sintagmas, sino que podían ocupar un verso completo, con variaciones potencialmente infinitas para cada héroe. En estos casos la fórmula solía coincidir también con un bloque coherente semánticamente, aunque, en el caso de descripciones recurrentes, estos bloques podían ocupar varios versos y repetirse tal cual.

Parry defendía que a cada sección del hexámetro le correspondía estrictamente una frase o *fórmula*, y que rara vez se duplicaban. La perfección del sistema que imaginó en términos de economía y complejidad era tal que no parecía posible que fuera obra de una sola persona. Homero sería el eslabón final de una larga serie de rapsodas, que fueron creando y perfeccionando una tradición formular hasta alcanzar el estado que observamos en la épica homérica. Parry convirtió la economía y la extensión en un principio central para la poesía oral, y creía firmemente que podríamos llegar a descubrir que la *Ilíada* y la *Odisea* no eran más que fórmulas artísticamente encadenadas.

La idea de este sistema formular se apoyaba en las investigaciones sobre la poesía de los bardos yugoslavos llevadas a cabo por el propio Parry. Aquellos «rapsodas modernos» o *guslari* empleaban una serie de fórmulas transmitidas de boca en boca y que conocían de memoria, para componer canciones, a menudo largas, conservadas en la memoria sin necesidad de la escritura. A través de entrevistas y grabaciones de audio, descubrieron que aquellos poetas, en lugar de repetir las canciones palabra por palabra, las recreaban en cada representación, introduciendo sutiles cambios cuya acumulación podía hacer pensar que se trataba de versiones diferentes³⁴. De esta forma consiguieron ilustrar su teoría con un ejemplo palpable, y, sobre todo, revelaron que algunos aspectos, como el papel de la representación y la audiencia eran imposibles de deducir de un texto escrito y estable. La poesía oral tiene una naturaleza fluida y cambiante, pero la escritura habría asesinado el carácter primigenio de la obra.

La analogía con los bardos yugoslavos ha sido duramente criticada, en primer lugar porque sólo es una comparación, no una prueba sólida, y porque su validez se asienta sobre presupuestos poco claros. Uno de estos es la idea de que por el hecho de compartir un mismo método de producción (el lenguaje formular, en este caso), las creaciones resultantes (la poesía épica griega y yugoslava) deben ser semejantes. Así, la poesía oral, por el simple hecho de no usar la escritura, debería ser igual en todo el mundo y en todas las épocas. En

³⁴ El jazz presenta un procedimiento análogo en el terreno musical (cf. Silk 1987, 26).

cambio, existe una gran diferencia de calidad entre las canciones que Parry y Lord grabaron y analizaron, y la poesía homérica, cuya superioridad fue siempre reconocida por ambos³⁵.

Aunque el ejemplo de la poesía yugoslava fue útil para imaginar cómo la *Ilíada* y la *Odisea* surgieron como la culminación de una larga y refinada tradición, la idea dejaba en entredicho el papel individual del poeta. En efecto, los defensores de la aplicación extrema del sistema formular proponen un modelo de poesía como una actividad mecánica, donde la única función del poeta es unir una serie de segmentos lingüísticos según una serie de normas ya dadas. No obstante, esta imagen se ha revelado errónea. Las llamadas «fórmulas» no forman un sistema claro y cerrado, y la descripción que daba Parry del procedimiento no tiene en cuenta los pequeños cambios que se producen con cada nueva interpretación. La tradición sólo pone a disposición del poeta un conjunto de fórmulas, temas y lenguaje tradicional, pero queda a su elección aceptarlo parcial o totalmente, modificarlo y elegir la forma de combinarlos. Tanto el rapsoda griego como el bardo yugoslavo crean nuevas canciones, y al mismo tiempo recrean la tradición anterior. El poema oral es una criatura viva y en continuo proceso de transformación, fruto del intercambio continuo con el público, cuyo papel posiblemente era mucho más activo de lo imaginado y no debe dejar de tenerse en cuenta. Las inconsistencias que apreciamos en el texto escrito conservado de la épica homérica son probablemente indicadores de la coexistencia de diferentes representaciones o versiones del texto, superpuestas o entrelazadas³⁶.

Sin embargo, la tesis original de Parry hacía demasiado hincapié en la interacción oral, e insistía en que el poeta oral no memoriza sino que compone al mismo tiempo que «actúa» o «interpreta» su canción, adaptándose a lo que percibe de su audiencia y del contexto. Seguramente esto procedía del ejemplo de los *guslari* yugoslavos, entre los que ni él ni su discípulo pudieron detectar el tipo de memoria literal que por entonces solía presuponerse en las culturas orales. Ningún poema era exactamente igual al anterior, y los propios bardos ni siquiera parecían mostrar preocupación o entender el concepto mismo de repetición *palabra por palabra* (cf. Small 1997, 4). Tanto más difícil hubiera sido para los rapsodas griegos memorizar la *Ilíada* y la *Odisea*, cuya longitud y complejidad supera a la canción más larga de la tradición yugoslava.

³⁵ Foley 2002, 3, observa que buena parte del material grabado por Parry y Lord no se ha publicado aún.

³⁶ Cf. Nagler 1974, 95-96 y n. 35, que observa cómo la embajada de Aquiles comienza estando formada por tres hombres (*Il.* 9, 168 ss.), luego solo dos, y finalmente tres de nuevo.

Ahora bien, no es posible dejar a un lado el papel de la memoria. Si dividimos el proceso de composición de un poema en niveles, el primero lo ocupa la memorización de determinadas piezas fijas, los epítetos que iniciaron la investigación de Parry, por ejemplo. El segundo nivel se compone de una sucesión no limitada de «interpretaciones» o «construcciones» del poema. Resulta difícil imaginar que las escenas en las que un personaje rompe la narración con un característico discurso, cuyo lenguaje es marcadamente elaborado y diferente del resto de la narración, fueran producto de un solo momento de inspiración. La fluidez que requiere la interpretación oral es fruto de la inmediatez propia de las circunstancias en las que se produce, pero no excluye en modo alguno un proceso de reflexión y reelaboración de lo dicho antes y después de cada representación³⁷. La longitud del poema tampoco es un obstáculo, si aceptamos la posibilidad de que Homero construyera la estructura de su obra a lo largo de varios años, cuya esencia podría mantenerse sin problema en la memoria del poeta. Sobre ella, como si de una tabla encerada se tratara, el poeta añadiría o quitaría lo necesario, con la libertad que le permitía la carencia de una preocupación propiamente dicha por la *exactitud* (cf. Thomas 1992, 39-40).

Las llamadas *fórmulas* permitieron el descubrimiento de toda una faceta ignorada de la cultura helénica, y, aunque fueran utilizadas por Parry como un argumento en defensa del carácter mecánico del quehacer poético, también son la prueba innegable de que la memoria jugaba un papel crucial en la comunicación. La repetición de breves «segmentos lingüísticos» es, además, un procedimiento mnemotécnico habitual, demasiado simple quizá para ser considerado procedimiento mnemotécnico propiamente dicho, pero que sin duda es la base del funcionamiento de la memoria. Al mismo tiempo, es uno de los rasgos más rápidamente perceptibles del lenguaje hablado, y su presencia en un texto escrito suele considerarse como la prueba más fácil y evidente de que un poema ha sido compuesto de forma oral, aunque no por esta razón sino como sinónimo de la idea expresada por la *fórmula* de Parry. En consecuencia, el estudio de la oralidad presente en los textos griegos se ha centrado preferentemente en la detección y análisis de fórmulas³⁸.

La intuición inicial de Parry no recibió la confirmación esperada, y Homero, en lugar de consistir sólo en fórmulas, muestra alrededor de un 35% de *hapax* locales, una cantidad

³⁷ Jensen 1980, 42, observa que entre los mismos *guslari* no era extraño practicar la noche de antes.

³⁸ Un ejemplo de investigación sobre la composición oral centrado en el elemento formular es el de Hainsworth 1968.

muy elevada de palabras «extrañas» como para considerarse parte de un aparato formular³⁹. Ahora bien, la duda más seria que se puede plantear a toda esta teoría es la ausencia de una definición exacta y útil de *fórmula*, cosa que puede cambiar radicalmente los resultados de un análisis estadístico como el anterior. Por ejemplo, versos que repiten parcialmente una fórmula pero añaden otras palabras son variaciones de la fórmula básica, pero ¿qué ocurre con versos u oraciones que comparten un patrón gramatical pero difieren en el léxico empleado? Algunos prefieren hablar en estos casos de «fórmulas estructurales», es decir, fórmulas sintácticas, pero parece más un intento de preservar la denominación de *fórmula* a toda costa. Si cualquier repetición de un segmento sintáctico o léxico es una fórmula, las fórmulas así descritas estarían presentes no sólo en la épica griega, sino en cualquier género literario⁴⁰, es más, serían un rasgo intrínseco del lenguaje humano. El trabajo de Parry, además, adolecía de no tener en cuenta todas las formas en las que se hacía referencia al héroe homérico, sino sólo los epítetos nominales (cf. Shive 1987, 18). Una vez ampliamos el abanico de ejemplos, observamos cómo en la elección del epíteto correcto influyen muchos aspectos, como la adecuación al contexto, a la métrica, o la intención expresiva del autor.

El lenguaje formular pierde su pretendido carácter mecánico y se convierte en una manifestación más del uso de unas herramientas tradicionales por parte del poeta⁴¹. Es más, pierde también su validez como indicador exclusivo del carácter oral de una pieza literaria⁴². No obstante, eliminado el criterio mecanicista, también es cierto que la repetición de ciertos elementos, sintácticos y léxicos, responde a una decisión consciente del autor, motivada por razones internas (gustos personales, educación) y externas (preferencias del público, contexto de interpretación de la obra, influencia de la tradición). El análisis de la presencia de fórmulas en el texto, o, mejor dicho, de *patrones*, resulta interesante y útil para obtener información de múltiples aspectos. El error de Parry y sobre todo de su discípulo, Lord, que sí tuvo tiempo de estudiar las pruebas con calma, fue no reconocer plenamente que la relación fórmula-oralidad no se sostenía por sí misma. Con todo les debemos el

³⁹ Cf. Pope 1985, así como Richardson 1987 sobre símiles y palabras raras. Griffin 1986 y Minchin 2002 indagan en la alteración del procedimiento formular que producen los discursos.

⁴⁰ Como puso de manifiesto Silk 1987, 21, todos los poetas utilizan patrones métricos para escribir.

⁴¹ Thomas 1992, 44, observa que a pesar de la abundancia de fórmulas en la *Eneida* de Virgilio, su presencia debería atribuirse probablemente al peso de la tradición literaria.

⁴² Deben evitarse extremos como los de Lord 1960, 142-144, que llega a fijar un porcentaje de presencia de fórmulas a partir del cual distinguir entre obras orales o escritas.

hecho de haber despertado el interés sobre la naturaleza de los métodos de composición y transmisión literaria en la antigua Grecia.

I.3.2.2. Funcionamiento de la memoria

I.3.2.2.a. La memoria como soporte de información

El descubrimiento de patrones o fórmulas en los textos conservados de la literatura griega, primero en la épica, pero no limitado a ese género, llamó la atención sobre otras formas de creación literaria ajenas al uso de la escritura. El rechazo de la posibilidad de un Homero «analfabeto» pero capaz de crear una obra maestra de la literatura, llena de matices y con una estructura grande y compleja, se sustenta, como hemos visto, en prejuicios sobre las consecuencias intrínsecas del empleo o la ignorancia de la escritura. Cuando las *fórmulas* épicas de Parry salieron del estrecho marco de la épica homérica, resultó evidente que no eran un fenómeno exclusivo de un género, y que su único rasgo realmente definitorio, la *repetición*, es una constante a lo largo de toda la historia de la literatura, escrita u oral. No obstante, persistió el interés por el papel que jugaban las fórmulas en el entramado textual, y especialmente por explicar el grado de complejidad que podían llegar a alcanzar en obras creadas y difundidas de forma oral. La cuestión de fondo, como ya hemos dejado entrever, es el papel de la memoria en el proceso. Si hemos definido el texto como un medio de almacenamiento y transmisión de la información cuyo soporte es el signo escrito, en una cultura oral la memoria de forma análoga funcionaría como soporte. No obstante, antes de proceder a hablar del papel que esta jugaba en la Antigüedad y de la influencia de los hábitos creados por su utilización en las nuevas formas de creación y comunicación construidas sobre la escritura, creemos necesario exponer de manera sucinta lo que hasta ahora puede decirse con seguridad sobre la memoria como habilidad humana. Para ello nos apoyamos en los descubrimientos de la psicología cognitiva, que se detallan a continuación.

I.3.2.2.b. El estudio de la memoria

La memoria es la capacidad de almacenar información para el futuro. Ha sido objeto de interés desde muy antiguo (*cf. infra I.3.2.3. La memoria en la Antigüedad*), y juega un papel importante en varias disciplinas relacionadas con la adquisición del conocimiento en el ser humano, también llamadas *ciencias cognitivas*: lingüística, psicología, teoría de la

información o inteligencia artificial, entre otras⁴³. El debate en torno a la naturaleza de la memoria y su funcionamiento también ha sido intenso en el seno de la filosofía, pero las conclusiones alcanzadas desde esta perspectiva no nos resultan útiles. El estudio filosófico de la memoria está más relacionado con la definición de unos rasgos característicos del ser humano (cf. Esteban Ortega 1997, 31-69), o con la deontología de la educación (cf. Vergés 2008). Por otra parte, en el siglo XX, con el desarrollo de los ordenadores se establecieron paralelos entre el funcionamiento de aquellos y lo que se sabía, o suponía, del pensamiento humano. La investigación avanzó en ocasiones a la inversa de lo esperado, proponiendo modelos de inteligencia artificial cuyas implicaciones en la psicología humana serían puestas a prueba después⁴⁴. El resultado fue la creación de un marco conceptual de las funciones mentales (memoria, almacenamiento, recuperación y procesamiento de la información, por ejemplo), compartido entre la mente humana y los ordenadores (cf. Anderson 2015, 9-10; 13-14).

Dentro de las muchas aplicaciones de esta rama de conocimiento, la estilística cognitiva analiza el idiolecto de un autor, sus rasgos lingüísticos individuales. Nuevas aportaciones sobre el procesamiento del lenguaje en el cerebro nos ayudan a interpretar los datos procedentes del análisis del texto, especialmente si es por ordenador. Los estudios léxicos cuantitativos revelan la presencia de relaciones estrechas entre términos, llamadas también *redes*, y el modelo cognitivo asegura que el cerebro funciona como un grupo enorme de tales redes⁴⁵.

Las elecciones léxicas y sintácticas constituyen marcas con información sobre la procedencia (temporal-espacial o personal) u otros conceptos o enunciados con los que está relacionado el enunciado en el que se producen. Aunque estos rasgos no sean exclusivos de cada enunciado, considerados en conjunto dibujan un marco suficientemente claro como para apreciar en él patrones comunes o marcas que más adelante pueden emplearse, por

⁴³ Se considera a Neisser 1967, 4, el responsable de la acuñación del término *psicología cognitiva*, en relación a su definición de *cognición*.

⁴⁴ Cf. la *máquina lógico teórica* de Newell & Simon 1956 o el *solucionador general de problemas*, desarrollado también por Newell & Simon 1959.

⁴⁵ Anderson 2015, 110-112, recoge algunos experimentos para demostrar que la memoria está estructurada en forma de *red* o *networks*. El concepto de *red* es utilizado en múltiples disciplinas; el término se ha vuelto casi omnipresente en la investigación léxico-semántica, especialmente en trabajos asistidos por ordenador, probablemente por la extensión igualmente enorme del concepto en la informática (cf. Ananda & Srinivasan 1984, y especialmente Börner *et al.* 2007).

ejemplo, para detectar el estilo de un autor determinado. Las herramientas de análisis estadístico como concordancias, listas de frecuencias de repetición de fenómenos o su localización nos dan la posibilidad de detectar estos conjuntos de rasgos o patrones. No obstante, sin un trasfondo teórico que nos permita entender cómo pueden o suelen producirse, nuestro criterio a la hora de decidir qué corresponde a la elección del autor y qué a interpolaciones o cambios a lo largo de la transmisión del texto, se reduce a la pura subjetividad. Puesto que el funcionamiento de la memoria tiene un carácter universal y atemporal, al menos en lo que atañe al ser humano, creemos que las aportaciones sobre su funcionamiento desde otras disciplinas especializadas contribuirá a enriquecer y asentar firmemente nuestro comentario de los datos ofrecidos por herramientas de análisis estadístico del texto.

1.3.2.2.c. La atención y los tipos de memoria

El objetivo principal de la psicología cognitiva son los procesos mentales que afectan al comportamiento, entre los cuales se incluyen la memoria y la atención, íntimamente relacionados entre sí (*cf.* Astle & Scerif 2011; Posner & Petersen 1990). Esta última consiste en un estado de percepción focalizada en un subconjunto de la información percibida disponible. El papel de la atención es pues separar la información relevante de la que no lo es, actuando como un filtro de material previo al resto de procesos mentales (*cf.* Anderson 2015, 53-78). La atención como proceso se puede dividir en dos sistemas de control, uno *exógeno*, responsable de mecanismos de focalización más o menos inconscientes y otro *endógeno*, encargado de formas de atención consciente (aunque pueden producirse también de forma inconsciente) como es la atención selectiva, dividida, local o global (*cf.* Chica *et al.* 2013). Por otra parte, aunque la atención procesa información procedente de los cinco sentidos, aquella que recibimos de la vista y el oído suele tener un papel predominante.

Los datos que pasan el filtro de la atención pasan a ser almacenados temporalmente en la memoria, que representa una especie de «mesa de trabajo» para el resto de procesos mentales. Lancashire 2004 recoge una división habitual en psicología entre dos tipos básicos de memoria (*cf.* Anderson 2015, 124-136):

- a corto plazo, también conocida como **memoria operativa**;
- a largo plazo o **memoria asociativa**, que a su vez se divide en dos:

- **implícita** o inaccesible, que incluye el recuerdo de un procedimiento y la fijación de elementos primarios o priming.
- **explícita** o accesible, a largo plazo, que incluye la memoria episódica y la memoria semántica.

1.3.2.2.d. La memoria a corto plazo

La memoria operativa ofrece almacenamiento a corto plazo para una cantidad limitada de lenguaje sobre el que se puede trabajar conscientemente. Esta forma de memoria no puede separarse de las actividades de procesamiento. Baddeley fue el primero en proponer hace treinta años un modelo muy influyente de memoria operativa dividido en tres partes: una central, ejecutiva, y dos subsistemas, el área visual y el área fonológica o *bucle articulario* (*articulatory loop*; cf. Baddeley & Hitch 1974). La parte ejecutiva dirige el funcionamiento de las otras dos, pero el trabajo mental sobre la expresión lingüística se realiza en el bucle articulario. Este incluye varias regiones del cerebro, incluidos dos importantes centros del lenguaje: las áreas de Wernicke y de Broca, que se encargan del procesamiento semántico y sintáctico, respectivamente.

La denominación de *bucle* hace referencia al movimiento figurado de la información en esta fase del procesamiento, aunque también puede detectarse en el fondo una comparación con el funcionamiento de las memorias volátiles en los ordenadores, donde la información se mantiene *viva* a base del ir y venir de corriente entre posiciones. El *bucle fonológico o articulario* se encarga de la información sonora. Se divide en dos partes: un *almacén fonológico* que guarda impulsos auditivos durante un breve periodo de tiempo, y un componente de *recreación articularia* (el *bucle articulario* propiamente dicho) que se ocupa de repetir o *revivir* los impulsos almacenados antes de que desaparezcan.

La atención ejerce su fuerza selectiva entre el almacén fonológico y el bucle articulario, puesto que las pruebas indican que no existe un filtro para la información sonora que entra en el almacén. El primero actúa como un «oído mental» que recuerda todos los sonidos percibidos, mientras que el bucle representa una especie de «voz» interna o mental que repite o recrea los sonidos; en esta fase del proceso la atención juega su papel al tiempo que se produce una primera interpretación de los sonidos, distinguiéndose entre palabras y sonidos no inteligibles, por ejemplo. La existencia de este módulo de la memoria se asienta sobre varias pruebas, de las que destacamos tres:

- El efecto de la similitud fonológica. Resulta más difícil recordar por separado palabras que suenan parecido. La similitud semántica no tiene un efecto muy marcado, lo que demuestra la idea según la cual la información verbal se almacena fundamentalmente según su forma sonora (*cf.* Baddeley 1966; Conrad & Hull 1964).
- El efecto de la supresión articulatoria. Si interrumpimos a alguien mientras escucha algo, llamando su atención sobre otra cosa, el recuerdo de lo que estaba oyendo desaparece rápidamente. Esto sostiene la existencia del *bucle articulatorio* como medio de mantenimiento de la memoria (*cf.* Baddeley *et al.* 1975).
- La transferencia de información en códigos diferentes. Está demostrado que los adultos *vocalizan* la información visual (por ejemplo la escrita), de forma que pasa a almacenarse como información sonora (*cf.* Murray 1968). Hace relativamente poco Baddeley añadió un componente más al modelo, apodado el *búfer episódico*. Este se encargaría de enlazar distintos tipos de información (visual, espacial, verbal), añadiéndole una cierta secuencia temporal y convirtiéndolos en unidades integradas o *secuencias*⁴⁶.

Sin tener en cuenta excepciones como la que acabamos de comentar, la limitada capacidad de la memoria operativa restringe el trabajo mental consciente a la hora de crear oraciones de forma continuada. Tanto si lo hacemos de forma oral o escrita, utilizamos uno de tres métodos posibles (*cf.* Lancashire 2004, 404):

- a) composición espontánea sin un control consciente sobre el enunciado;
- b) recitación o recreación de algo conservado en la memoria explícita a largo plazo;
- c) recitación de frases previamente construidas y almacenadas temporalmente en la memoria operativa.

1.3.2.2.e. La memoria a largo plazo

Por otra parte, dentro de la memoria a largo plazo, la llamada *memoria asociativa* continúa siendo en cierto sentido un misterio. Lancashire 2004, 404, la identifica con la «Musa», en cuanto representa el modo como unos recuerdos llaman a otros y forman la

⁴⁶ *Búfer* es un término prestado de la informática y designa un tipo de memoria temporal. Baddeley 2000 cree que el *búfer* episódico está relacionado con la fijación de los recuerdos a largo plazo y con los sistemas encargados de dar sentido a la información.

cadena de ideas que concebimos como la inspiración detrás de la expresión lingüística. Lo que sí sabemos es que los recuerdos almacenados de esta forma se recuperan mediante palabras, sucesos o cosas *clave*, enlazados al recuerdo en sí, y que lo traen a la memoria operativa como la caña de pescar al pez.

La memoria asociativa a largo plazo, por otra parte, no almacena enunciados completos. Estos se crean o recrean cada vez que deseamos expresarnos lingüísticamente mediante el trabajo simultáneo de diferentes partes del cerebro. Según Koenig *et al.* 1992, la mente posee un *thesaurus* de conceptos, léxico y conocimiento enciclopédico, en la forma de «morphologically decomposed representations».

El llamado *priming*, o «fijación de (*sc.* recuerdos) *primarios*», representa una especie de comodín en la formación de nuestro almacén de recuerdos. La experiencia sensorial deja improntas en la mente que actúan como referentes primarios, de los que no somos conscientes y a los que no prestamos atención. También es una forma de crear enlaces entre ideas que normalmente no consideraríamos relacionadas. Puesto que las emociones y la repetición juegan un papel importante en la creación de *primarios*, no existen definiciones de conceptos o ideas compartidos con total exactitud entre personas. Es más, en el ámbito del individuo, el efecto de los primarios se deja notar a la hora de buscar en nuestros propios recuerdos, ya que estos modifican la forma en la que se activan los enlaces de memoria que comentábamos antes. Por ejemplo, si se fija una relación primaria entre *amante* e *infidelidad*, la relación etimológica evidente entre el primer término y las características positivas del verbo amar quedará oscurecida en posteriores apariciones de *amante*.

Otra forma de concebir la memoria a largo plazo la divide en tres sub-tipos que parecen seguir una organización jerárquica, en función del nivel de trabajo consciente que se aplica sobre ellos⁴⁷:

- **Memoria procedimental.** Se utiliza para la realización de determinadas acciones. A menudo se activa de forma subconsciente, y como mucho requiere un mínima cantidad de esfuerzo consciente. Incluye información relativa a la respuesta a estímulos, que se activa a través de la asociación con ciertas tareas o rutinas. Utilizamos conocimiento procedimental cuando parece

⁴⁷ Nuestra clasificación de los tipos de memoria a largo plazo se basa en los trabajos de Squire 2004; McRae & Jones 2013; Tulving 1984; Schacter *et al.* 2011, 240-241, y Balota & Marsh 2004, 364-365.

que se produce una respuesta “automática” a una situación o proceso particular.

- **Memoria semántica.** Es el conocimiento enciclopédico que posee una persona. El acceso a la memoria semántica requiere un grado variable de esfuerzo, dependiendo de varios factores como el tiempo transcurrido desde que se adquirió la información, el número de asociaciones o vínculos con otras piezas de información, la frecuencia de acceso etc.
- **Memoria episódica.** Es la memoria relativa a los eventos personales vividos que pueden expresarse de forma explícita. Precisamente debido a la necesidad de poner en relación información almacenada en otros sectores de la memoria con el procesamiento del lenguaje, la memoria episódica es la que más esfuerzo consciente requiere.

1.3.2.2.f. Aportación del estudio de la memoria al análisis del texto

Desde el punto de vista del análisis del texto, lo que hemos dicho de la psicología cognitiva nos ayuda en dos aspectos: nos indica qué resultados esperar, y nos da pistas sobre cómo interpretarlos. A continuación resumimos las ideas más relevantes⁴⁸:

1. Los enunciados lingüísticos son almacenados, procesados y recuperados como secuencias fonológicas, no visuales.
2. La gramática y el vocabulario son inseparables, puesto que las áreas que se encargan de los aspectos sintáctico y semántico del lenguaje funcionan a la vez en su producción.
3. Los bloques sobre los que se construye todo enunciado son “redes”. Estos no consisten en palabras discretas, sino combinaciones de varios tipos de información (sensorial, emotiva, lingüística). Lo que llamamos “palabras clave” sólo constituye uno de los muchos extremos de la red, y su papel predominante dentro de la misma varía de una persona a otra y cambia a través del tiempo y las experiencias.

⁴⁸ Sobre la aplicación de la psicología cognitiva al estudio del texto, remitimos a Téllez Muñoz 2005, 231-274, Heurley 2010, Vizcaíno Ortega 1997 y Luján Atienza 2006.

4. La memoria operativa no puede trabajar sobre expresiones lingüísticas infinitas. La existencia de unos límites, aunque difusos, nos puede ayudar a la hora de fijar el área de trabajo de un análisis estadístico del texto.
5. La memoria operativa se ralentiza cuando trabaja con estructuras sintácticas complejas de modo que prefiere organizaciones de la oración más simples, es decir, paratácticas.
6. No existe un diccionario o enciclopedia convencional que pueda reflejar y documentar las llamadas «zonas de convergencia» de la mente (*cf.* Damasio 1995, 92). Lo más parecido son los campos semánticos, pero incluso este concepto debe adaptarse a la virtualmente infinita variedad de relaciones posibles entre los elementos debido al mecanismo de formación de primarios. Es el principio de *indeterminación semántica*.
7. La memoria cambia y se deteriora con el tiempo, y por lo tanto también lo hacen ciertos rasgos de los enunciados construidos sobre ella.

Las conclusiones del estudio de la memoria por parte de la psicología muestran un panorama muy interesante a la hora de abordar lo que sabemos sobre la llamada composición oral y ciertos rasgos de la oralidad en la antigua Grecia. Los antiguos parecen haber aprovechado de forma intuitiva el funcionamiento natural de la memoria con fines estéticos y comunicativos, aprovechando la fuerza de la repetición y la conexión entre diferentes tipos de sensaciones (visual, auditiva, espacial).

I.3.2.3. La memoria en la Antigüedad

I.3.2.3.a. La memoria como concepto

Bakker 2008, 65, describe la memoria como un concepto y como una función, compleja y *proteica*; en efecto, es un elemento crucial para la comprensión de muchos aspectos «humanos», tales como el cerebro, la educación, el carácter, la identidad y la sociedad, entre otros. Sin embargo, neurólogos, psicólogos de la rama cognitiva, sociólogos, antropólogos, educadores e historiadores parten de una concepción diferente de la memoria y de qué significa recordar. Además, la concepción de la memoria como una facultad cognitiva está inseparablemente conectada con el medio de comunicación, e incluso con la tecnología a él asociada, y esta se encuentra irrevocablemente abierta al

cambio. Desde la informática, por ejemplo, es probable considerar la memoria en su relación con la capacidad de procesamiento, mientras que los filólogos tenderán a ver la memoria como una cuestión de “recuperación”: uno recuerda cuando sabe qué ha leído y dónde estaba.

La memoria, por lo tanto, es «mediática» e históricamente contingente, puesto que los medios (ordenadores, libros, escritura, habla) y sus conceptualizaciones cambian con el tiempo. En la Antigüedad clásica la memoria es un soporte sobre el que crear y transmitir tanto las creaciones artísticas del ser humano como el conocimiento adquirido. La memoria es igualmente una herramienta moralizadora, educativa, en cuanto sólo transmite lo que la colectividad decide conservar, y que en consecuencia tiene posibilidades de adquirir la categoría de «verdad». De este modo el recuerdo se convierte en un lugar de debate, donde el individuo lucha por evitar que sus ideas se pierdan en el tiempo; para hacerse inmortal, debe conseguir convencer al público, presente y futuro, de que merece la pena recordar sus palabras, de que son verdaderas (*cf. infra I.2.3.3. Adaptación del recuerdo: ἀλήθεια, μνήμη y τὸ πρόπον*).

Este debate se desarrolla mediante mecanismos tradicionales creados a través de la experiencia, y que reflejan de forma práctica los efectos del funcionamiento de la memoria tal y como hemos descrito en la sección anterior. La repetición de elementos, o el uso de otros como signos o llamadas de atención, por ejemplo, son métodos naturales que favorecen la retención de unas ideas sobre otras, y que el poeta, el orador o el historiador utilizan de forma consciente para asegurarse de que el auditorio recuerde lo que ellos quieren decir, es decir, que sus ideas sobrevivan.

Ahora bien, ¿en qué consiste exactamente *recordar* para los griegos? La analogía con el mundo de la informática y, antes de la llegada de esta, la influencia del historicismo predominante durante los siglos anteriores nos han dejado una concepción altamente mecanicista de la memoria; recordar es «recopilar información previamente guardada», de forma análoga a la consulta de un documento escrito en una biblioteca. Ahora bien, la memoria humana resulta inconsistente e inexacta en comparación con la resistencia al cambio y al tiempo que ofrece el mensaje escrito, que es presentado como la salvación del conocimiento y la base del pensamiento científico (*cf. Otlet 1934, memex*). Llamamos la atención sobre esta apreciación de la memoria porque no se aprecia en los testimonios de la Antigüedad. La memoria es uno de los pilares fundamentales de la cultura de los griegos,

aunque estos no recurrieran a los mismos criterios por los que valoramos la importancia de su pretendido análogo, la escritura⁴⁹.

1.3.2.3.b. La μνήμη

Un estudio del término μνήμη y otros relacionados etimológicamente con él revela unas connotaciones semánticas diferentes a las que posee la familia léxica de la *memoria* en el español actual. Bakker 2008 indaga en sus raíces más profundas a partir del testimonio de la épica, aprovechando las teorías holísticas para dibujar un retrato de la μνήμη como una fuerza activa, ligada sutilmente a la realidad física e incluso podríamos decir mística. De forma similar a la μανία divina, la primitiva mentalidad griega concibe la memoria como un acto de re-presentación, en el que una persona encarna a un dios o una fuerza que conoce el pasado y lo recrea en el presente. Lejos quedan de esta concepción las nociones de *información* o de la *mente* separada del *cuerpo*; los recuerdos cobran vida y actúan efectivamente en el momento presente: el público ve el penacho de Héctor, siente la cólera de Aquiles en el estómago, escucha el atronador griterío de troyanos y griegos enzarzados en el combate.

El verbo μμνήσκομαι indica un estado de «posesión o encarnación de μένος». La etimología une el término con μένος «fuerza», μέμονα «estoy lleno de μένος», μῆνις «furia», μάντις «adivino» y μαίνομαι «estar poseído» (raíz *mne-h1)⁵⁰. En los ejemplos de uso del verbo en la *Ilíada* lo que se indica con el término no es un acto de recuperación de información desde la memoria, sino el lugar de procedencia del μένος encarnado en el cuerpo del héroe (*Il.* 8, 252; 14, 441; 15, 380). ¿Pero qué es el μένος exactamente? No es un objeto, ni una cualidad exclusiva del ser humano, puesto que también pueden tenerla ríos, el viento, la luz del sol, el fuego y otros fenómenos naturales⁵¹. Aún así, está conectado con dos órganos o facultades homéricas: el θυμός y la φρόνην o las φρόνες.

⁴⁹ Sobre la memoria en Grecia, remitimos a Notopoulos 1938; Nilsson 1943; Miccoli 1989; Borgeaud 1988; Bouvier 1988; Tardieu 1988 y Basset 1988.

⁵⁰ La forma μνη- no sólo significa «recordar», como indica su presencia en μνάομαι «deseo (tener una mujer)» o μνηστήρ «perseguidor». En la *Ilíada* μνάομαι aparece como equivalente de μμνήσκομαι (*Il.* 11, 71; 16, 697, 771), con una estructura sintáctico-semántica diferente: el objeto directo en acusativo indica el objetivo de la acción, mientras que el genitivo propio de μμνήσκομαι indica el origen de la acción, de acuerdo con la interpretación de Bakker 2008, 68.

⁵¹ Agua: *Il.* 12, 18; viento: *Il.* 5, 524; *Od.* 5, 478; el sol: *Il.* 23, 190; *Od.* 10, 160; 19, 440; *h. Ap.* 371, 374; fuego: *Il.* 6, 182; 23, 238; 24, 792; *Od.* 11, 220.

Atenea insufla μένος en el θυμός de Telémaco, lo que conduce a una actividad acelerada de sus φρένες, además de eliminar el dolor (*Od.* 1, 320-323; *cf. Il.* 15, 59-61)⁵². Aunque usualmente se traduzcan indistintamente por «fuerza», «mente» o «corazón», su significado originario probablemente era mucho más físico y relacionado con acciones fisiológicas básicas, como respirar: el μένος sería el aire que entra y sale de los pulmones, los φρένες⁵³.

El medio por el que se producen estos efectos del recuerdo es la palabra hablada (*cf.* García 2002, 35, n. 13), cuyo poder resulta, literalmente, palpable. Este poder se manifiesta únicamente en el contexto de la interacción oral, y constituye un proceso bidireccional. El poeta mueve al público, busca influir y provocar, extendiendo el μένος traído al presente por medio de su voz hasta su público, cuya reacción provoca una respuesta en el autor. Ya hemos mencionado que el medio oral permite que se produzcan modificaciones sobre la marcha, y uno de sus propósitos naturales es mantener vivo el interés del público hacia el mensaje del poeta. El diálogo pervive en la composición literaria como una de las formas más básicas y naturales de la comunicación oral: para García 2002, 38-39, el texto se convierte así en un diálogo *en diferido*, por cuanto la respuesta del público y la consecuente modificación de la forma del mensaje emitido no se producen con la misma inmediatez que en la interacción oral coloquial⁵⁴.

El pensamiento griego sintió pronto interés por el concepto de *palabra hablada*, por el λόγος, especialmente a partir de la primera sofística, como demuestra el papel central que aquel juega en las primeras teorías pseudo-científicas del conocimiento y del mundo. La *palabra* posee una fuerza, una δύναμις especial, que la tradición había encauzado y protegido con mecanismos de naturaleza ritual. El rango de estos mecanismos oscila entre la protección sagrada del emisor o del mensaje y la fijación de algunos elementos del diálogo, entre la expresión lingüística a través de fórmulas y el contexto de la interacción, mediante la participación obligatoria de personas que cumplieran ciertos requisitos. Ejemplo del primer método sería la prerrogativa divina que poseen por definición los poetas, en cuanto su voz es tanto suya como de la divinidad, gracias a la inspiración que

⁵² En general Homero es la principal fuente de estudio de la relación entre μένος, θυμός y φρένες, donde aparece con frecuencia en expresiones formulares coordinadas; *cf.* Bakker 2008, 71.

⁵³ Sobre la psicología homérica remitimos a Sánchez Lasso de la Vega 1963, 237-253; *cf.* también Plamboeck 1959 y Miller 2009, entre los estudios más recientes.

⁵⁴ Sobre la función social de la memoria remitimos al estudio de Carawan 2008 sobre los μνημόνες; así como a Scodel 2008 sobre la presencia y el uso de la μνήμη en el drama ático a través de la interacción con el público.

reciben. Por otra parte, el funcionamiento de la institución de los *μνημόνες* es un buen ejemplo de la importancia del contexto para el correcto funcionamiento de la palabra, y de cómo se utilizaban las costumbres no escritas para garantizarlo.

1.3.2.3.c. Adaptación del recuerdo

Aunque algunos se mantienen contrarios a la idea de que el auditorio o los lectores consintieran la modificación de discursos que seguramente aún recordaran con anterioridad a su publicación, esta opinión se sostiene sobre una concepción de la *fidelidad* y la *verdad* que no necesariamente coincide con la de los antiguos griegos⁵⁵. Para empezar, la exactitud no era una cualidad prioritaria en un documento histórico ni mucho menos preocupaba desde el punto de vista estético, y la visión antigua de lo que era la *ἀληθεία* era muy distinta de nuestra «verdad científica»⁵⁶. En este sentido, Hubbard 2008, 186, recoge la opinión de Dionisio de Halicarnaso, quien planteaba la posibilidad de que algunos discursos escritos se basaran en modelos teóricos o ficticios, referidos incluso a casos o situaciones hipotéticas pero manteniendo la apariencia de una representación oral real (*cf.* D. H. 49, 11-13).

En este sentido las diferentes versiones de la *Apología* de Sócrates (no sólo la de Platón y Jenofonte, sino también la perdida *Apología* de Lisias (*cf.* D.L. 2, 40; Cic. *or.* 1, 231) y la versión de la acusación de Anito escrita por Polícrates (*cf.* Isoc. 11, 4), nos hacen pensar en la maleabilidad del medio escrito: no era tan importante conservar las palabras exactas tal y como fueron dichas entonces, como escoger o poner en sus labios aquellas que serían apropiadas al *carácter* del orador y al significado implícito en su predicamento. Incluso un historiador tan escrupuloso como Tucídides dice que no puede recordar las palabras precisas de los discursos que escuchó y que tampoco podría hacerlo nadie; en su lugar, afirma que «ὡς δ' ἂν ἐδόκουν ἐμοὶ ἕκαστοι περὶ τῶν αἰεὶ παρόντων τὰ δέοντα μάλιστα εἰπεῖν, ἔχομένῳ ὅτι ἐγγύτατα τῆς ξυμπάσης γνώμης τῶν ἀληθῶς λεχθέντων, οὕτως εἴρηται» (Th. 1, 22)⁵⁷. El hecho de que esperemos que la versión escrita sera un

⁵⁵ En contra de una alteración posterior de los textos de cara a su publicación, *cf.* Dorjahn 1935, 293-295 y Lavency 1964, 190-192.

⁵⁶ Sobre la construcción del concepto antiguo de verdad, *cf.* Detienne 1967, 121-143. En general los griegos se preocupaban más porque las palabras resultaran apropiadas al contexto; en el caso del drama o de un discursos, este contexto sería el ἦθος de quien las pronunciara.

⁵⁷ Remitimos a Dover 1973, 21-27, Cogan 1981, xii-xvii y Hornblower 1987, 45-72, para un análisis detallado del método de adaptación o recreación de los testimonios orales seguido por Tucídides.

reflejo simple de la versión oral refleja lo que Derrida denuncia como «metafísica de la presencia» (cf. Derrida 1981, 63-155). Resulta mucho más productivo e inteligente considerar lo escrito como un medio innovador por derecho propio, con un conjunto de estrategias comunicativas y objetivos genéricos propios⁵⁸.

Las versiones publicadas de los discursos forenses claramente tenían un público diferente al del discurso pronunciado ante el jurado. Un ejemplo de esto es la sección del *Contra Timarco* en la que Esquines defiende una versión noble y platónica de pederastia en contraste con la auto-indulgencia de Timarco (Aeschin. 1, 132-140). Según Hubbard 2008, 193, dicha sección debió de aparecer exclusivamente en la versión escrita, puesto que responde a ataques personales contra el propio Esquines, también por pederastia, en los discursos pronunciados por la defensa. La exposición viene acompañada de abundantes citas literarias (Aeschin. 1, 141-159) y estaría dirigida sin duda a una clase elevada y educada (es decir, «leída») más que a las masas.

1.3.2.3.d. Impacto de la escritura en el concepto antiguo de memoria

Es bastante posible que los fragmentos comentados de la *Poética* y la *Retórica* reflejen los cambios en la forma de componer y acceder a la literatura que trajo consigo el aumento de la importancia de la escritura entre los siglos IV y III a.C. La obra literaria es cada vez más un objeto físico, en cuya adquisición y lectura tienen cada vez menos cabida el diálogo con el público y las circunstancias de representación. Como señala García 2002, 46, fruto de esta transformación es un cambio en la fuente de autoridad del discurso. Entendemos como tal la fuente de significado del discurso, y aquí retomamos aquel uso que ya antes mencionábamos del término *discurso* como un concepto doble: por un lado referido a la forma de componer, a las normas; por otro al objeto creado de este modo. Es decir ¿Quién dictamina que un pasaje requiere tal o cuál interpretación? ¿Quién confirma la veracidad de lo dicho? ¿Quién puede influir en la elección de determinadas figuras? Apartados de su hábitat natural, los elementos poéticos transferidos a la prosa, por ejemplo, se convierten en elementos retóricos: recursos mediante los cuales el autor «impresiona» al auditorio, porque percibe que este aprecia lo *extraño* (τὸ ξενικόν). El trasvase al medio

⁵⁸ Tanto Aristóteles (*Rh.* 3, 12, 1-2) como Plutarco (*Dem.* 9) dan por hecho que los discursos orales y los escritos son estilísticamente diferentes y se dirigen a auditorios diferentes; al respecto, cf. Worthington 1991, 57.

escrito quiebra la naturaleza dialógica que poseía la poesía y el *significado*; este último pasa a estar bajo el control absoluto del poeta⁵⁹.

1.3.2.3.e. Μνήμη y mnemotecnica

La memoria, en un sentido más próximo al nuestro, como actividad consistente en la recuperación precisa y a voluntad de información almacenada en la mente, era una *habilidad* menor, relegada más al terreno de la curiosidad y alejada del curriculum educativo habitual. *Mnemotécnica*, a pesar de su origen griego, es una palabra moderna, utilizada por primera vez en el siglo XIX, y mientras que el término actual indica una habilidad, técnica o método destinado a mejorar la memoria natural, en la Antigüedad clásica cualquier forma de conseguir este propósito era considerada artificial (cf. Small 1997, 72). Aristóteles hace depender la memoria de la sabiduría: ἡ μνήμη δὲ καὶ ἐμπειρία καὶ ἀγχίνοια ἦτοι ἀπὸ τῆς φρονήσεως ἐκάστη αὐτῶν ἐστίν, ἢ παρέρπεται τῇ φρονήσει (VV 1250a 35). El papel que juega la memoria es importantísimo en la Antigüedad, no solo como herramienta de construcción del conocimiento, sino también como sustento de la sociedad (en la medida en que transmite sus valores, costumbres e historias que la fundamentan), y como instrumento de creación⁶⁰.

La primera mención real a algo parecido a nuestra moderna mnemotécnica es la historia sobre la invención de la misma por Simónides de Ceos (c. 556-468 a.C.), y de la que circulan varias versiones (Cic. *or.* 2.351-354 y Quint. 11, 2,14). Resulta bastante similar a la técnica del «palacio mental», consistente en seleccionar un lugar físico, reproducirlo en la mente y colocar en un lugar específico aquello que se desea recordar, todo ello de forma que el orden en el que se «visiten» luego esas posiciones represente el orden lógico de las cosas recordadas⁶¹. Quintiliano extiende la explicación del por qué de este método

⁵⁹ Para García 2002, 43-50, la interpretación de la invocación a las Musas como una validación de la autoridad divina y absoluta del poeta es una prueba de que el uso del texto escrito habría alienado el texto de su contexto social.

⁶⁰ Cf. *Rhet. ad Her.* 3, 28; Cic. *inv.* 2, 160 y *Tusc.* 1, 65; Plin. *Nat.* 7, 88; Quint. 1, 3, 1; Plut. *Mor.* 9e.

⁶¹ Algunas técnicas mnemotécnicas aprovechan procesos mentales habituales relacionados con los sentidos y el movimiento para facilitar la recuperación posterior de los recuerdos (cf. Higbee 2001, 144-157, Worthen & Hunt 2011, 37-44, y Presley, Levin & Delaney 1982). Según Putnam 2015, 131, un *palacio mental* o *memory palace* es «a mental map of a building or walk that you know well, such as your house» en donde se *colocan* los recuerdos; la efectividad de esta técnica se basa en la asociación entre una ubicación física en un espacio conocido y aquello que quiere recordarse.

(11, 2, 17), y recalca el papel de la experiencia. El fenómeno en el que se basa la efectividad del invento de Simónides es lo que Gallagher 1993, 127-138, llama «aprendizaje dependiente del estado» (*state-dependent learning*). Las técnicas mnemotécnicas clásicas se resumen fundamentalmente en el sistema de lugares comunes (τόποι y *loci*) y su combinación con la imaginería mental.

El círculo socrático parece criticar estos métodos en relación a otro famoso amante de la mnemotecnia, Hippias de Élida. En el *Simposio* de Jenofonte, Sócrates recurre a la ironía para criticar las posibilidades reales de que tales métodos funcionen (X. *Smp.* 4, 62). Con la misma ironía, Platón presenta a Sócrates de nuevo enfrentado a Hippias, indicándole paradójicamente que «ha olvidado» la técnica de la que tanto se vanagloriaba (Pl. *Hp. Mi.* 368d). Para Platón lo importante no es recordar *mucho* sino acordarse de lo *importante*⁶². El filósofo se muestra escéptico con cualquier sistema que no le resulte “natural”, que no provenga de la propia naturaleza predefinida del ser humano. Este carácter natural de la memoria se manifiesta dentro de la teoría platónica como una consecuencia de la inmortalidad del alma: el aprendizaje (μάθησις) es simplemente una recolección (ἀνάμνησις) de lo que ya aprendimos en épocas pasadas (Pl. *Ph.* 72d-e). Para ello resulta útil la dialéctica, pero no otro método, como puede ser la escritura, si se concibe como herramienta memorística⁶³. Aristóteles, en cambio, recupera la idea del uso de lugares como guía para la memoria, y extiende el concepto, creando la teoría de los τόποι, que sería desarrollada después por los romanos y extendida así a lo largo de la civilización occidental⁶⁴.

1.3.2.3.f. Las huellas de la μνήμη en el texto

Ya hemos mencionado brevemente la posibilidad de que la cultura griega, al igual que otras en la Antigüedad, desarrollara técnicas que permitieran sacar el mejor partido a la memoria. Las *fórmulas* descubiertas por Parry sólo eran la punta del iceberg de un complejo entramado de técnicas, afianzado por la costumbre y mantenido por el uso cotidiano. Ahora bien, es la expresión de lo relevante, de lo importante para cada individuo y cada sociedad, lo que dio sentido y forjó el uso de esto que llamamos la *oralidad*, ὡς μήτε τὰ

⁶² Es lo que parece implicar cuando dice: ὥστ' ἐννοῶ ὅτι εἰκότως σοι χαίρουσιν οἱ Λακεδαιμόνιοι ἅτε πολλὰ εἰδότες, καὶ χρώνται ὥσπερ ταῖς πρεσβύτισιν οἱ παῖδες πρὸς τὸ ἠδέως μυθολογεῖσθαι (Pl. *Hp. Ma.* 285e).

⁶³ Para un resumen general de la opinión de Platón sobre la memoria, cf. Coleman 1992, 5-14.

⁶⁴ Cf. Small 1997, 78-84 sobre los τόποι, y 1997, 85-109 sobre la aportación romana a este recurso.

γενόμενα ἐξ ἀνθρώπων τῷ χρόνῳ ἐξίτηλα γένηται (Hdt. 1, 1). En nuestra opinión, la investigación de la oralidad en la Antigüedad no tiene tanto que ver con el uso o no de la palabra hablada, como con el funcionamiento de la memoria y con el papel que esta juega en la interacción del ser humano con la sociedad en la que vive.

Dentro del estudio de la literatura, las implicaciones del uso de la memoria tal y como hemos descrito afectan fundamentalmente a la estructura de la narrativa, y a la forma en la que esta se construye sintácticamente y por medio de elementos léxicos. Los antiguos nos muestran a través de sus textos un conocimiento práctico del funcionamiento de la memoria, que utilizan para tejer un discurso que traslade con firmeza su mensaje a un público que no necesariamente tendría la posibilidad de volver a leer con calma el texto. De esta forma, la composición del texto mantiene viva en cierto sentido parte de aquel contexto conversacional de la literatura que mencionábamos arriba. La acumulación en serie de numerosas escenas pequeñas, por ejemplo, es un procedimiento que se sustenta en el funcionamiento de la memoria episódica, formada a través de las experiencias rutinarias. Se trata de un mecanismo innato y común a oyente y poeta, cómodo de utilizar para este último y que le facilita establecer una conexión con el público. Otros tipos de memoria que subyacen al proceso de creación poética son la memoria implícita y la memoria explícita (*cf. supra* 1.2.2. *Funcionamiento de la memoria*). Dentro de esta última, la memoria episódica juega un papel importante puesto que ayuda a fijar los recuerdos en una secuencia fácil de poner en palabras, donde cada paso activa el recuerdo siguiente. Por otra parte, ciertos patrones constituidos no por palabras, sino por acciones (verbales y no-verbales), pueden ser el reflejo del funcionamiento de la memoria implícita, que almacena series estandarizadas de procedimientos en lugar de información léxica.

En la sección introductoria de su artículo, Bonifazi 2008, 35-40 se apoya en la hipótesis defendida por la psicología cognitiva de que reconstruimos el espacio narrativo para mejorar nuestra capacidad de comprensión. Dentro de la poesía homérica, esto se traduciría, según este estudioso, en una intención de plasmar escenas específicas, rostros y detalles, en lugar de proporcionar una visión panorámica del conjunto. En consecuencia, podemos observar en los textos una «sintaxis del movimiento» que articula estas visualizaciones y hace la narración mucho más vívida, mediante discontinuidades y cambios de foco en lugar de mediante marcos fijos y descripciones estáticas. El poeta tradicional sigue una narrativa paso a paso en lugar de avanzar como si lo hiciera a través de un mapa⁶⁵. Para ello se sirve de elementos léxicos, partículas y adverbios fundamentalmente, que, si bien care-

⁶⁵ *Cf.* Ryan 2003, 233, para una crítica de la interpretación hodológica frente a la «cartográfica».

cen de un contenido semántico claro, pueden indicar no sólo un punto de inflexión o al menos, de cambio, en la narración, sino también el tipo de relación que el narrador percibe entre diferentes partes del discurso. Bonifazi realiza un análisis de αὖ, αὖτε, αὖτις, αὐτάρ, αὐτίκα y αὐτοῦ como marcadores del discurso desde el punto de vista de la teoría pragmática. Otros términos como αὖτις, αὐτίκα y αὐτάρ pueden funcionar tanto en el plano proposicional, indicando contenido semántico, como en el plano pragmático.

Su estudio muestra que el primer grupo de marcadores suelen aparecer en posición inicial, mientras que los adverbios proposicionales aparecen en mitad de la oración. Bonifazi considera que las palabras mencionadas contienen un significado procedimental, es decir, indican discontinuidades visuales entre secciones de la narración épica. La comparación con las señales de carretera es original pero ilustrativa del efecto cognitivo que estos elementos pueden producir, afectando a áreas del cerebro específicamente relacionadas con lo visual. Bonifazi 2008, 48, también toma prestada la terminología del cine, y habla de cambios entre diferentes enfoques y niveles de *zoom*, relacionados con el contenido de las escenas, de forma que sean visualmente impactantes.

Por otra parte, Scodel 2002 extiende las implicaciones posibles de estos marcadores a la esfera de la emotividad y les atribuye la función de introducir elementos importantes en la narración o, como la estudiosa misma denomina, «momentos de impacto». Al igual que Bonifazi recurre a una comparación original y habla de un efecto de «memoria de bombilla» (*flashbulb memory*). En este caso, no obstante, se trata de una expresión sacada de la psicología que alude a ciertos recuerdos que permanecen en la memoria con extraordinaria fijación y viveza, sin necesidad de un esfuerzo consciente de la persona por conservarlos⁶⁶. Pueden producirse por varias razones, pero la presencia de una fuerte respuesta emocional suele ser el detonante y la consecuencia cuando dichos recuerdos son reavivados. Scodel centra su contribución en un tipo de σήματα homéricos asociados a escenas de ἀνάγνωσις o reconocimiento⁶⁷. El criterio de selección de signos seguido por el autor sería la emoción que generan en los personajes, que, a su vez, constituye un fuerte impulso narrativo.

Mediante σήμα indicamos, por un lado, objetos o cosas «palpables»: una lápida, un rayo, una cicatriz, unas marcas, una cama, etc. Estos tienen en principio una fuerza referen-

⁶⁶ El concepto de *flashbulb memory* fue acuñado por Brown & Kulik 1977, para referirse al tipo de recuerdos vívidos que se crean en circunstancias especiales y con un alto impacto emocional (cf. McCloskey, Wible & Cohen 1988).

⁶⁷ El tema ya había sido tratado por Nagy 1983, por ejemplo.

cial en cuanto apuntan hacia algo no contenido en ellos, pero también, paralelamente, son referencias simbólicas, que trascienden su materialidad y comunican directamente un mensaje. Steiner 1994, 11-16, distingue dos clases:

1. signos sólo comprensibles para unos pocos con conocimientos previos;
2. signos basados en convenciones, de forma que cualquiera, en principio, los puede comprender.

Basándonos en la división de Scodel 2002, 101-103, podemos ofrecer la siguiente clasificación de $\sigma\eta\mu\alpha\tau\alpha$:

Según la materialidad del signo y aquello que indican:

- permanentes (una tumba, una cicatriz);
- temporales;
- ocasionales (*obvii*);
- portentos divinos ($\theta\alpha\acute{\upsilon}\mu\alpha\tau\alpha$).

Según los requisitos por parte del receptor:

- inferenciales:
 - necesitan decodificación;
 - sólo necesitan el reconocimiento de su categoría de signo (en este caso actúan como simples enlaces o marcadores mnemotécnicos, carentes de un carácter simbólico);
- semióticos (varían según la intención del emisor y el público que puede interpretarlos; suelen modificar el contexto necesario para su comprensión, y pueden apuntar tanto al pasado como al futuro).

Descubrir rasgos de aquel funcionamiento primigenio de la literatura basado en la memoria y la oralidad es una tarea difícil, que debe abordarse con extrema cautela. El modelo dibujado por Parry, Lord y sus seguidores presenta un sistema excesivamente bien delimitado, que a nuestro parecer pretende camuflar un panorama más difuso y mucho menos homogéneo y claro. Hemos dejado claro que un mundo forjado por la comunicación oral es eminentemente una realidad viva y cambiante, abierta a las modificaciones y centrada en la adecuación a las circunstancias fijadas por el momento de la representación, pero también por las normas de una fuerte tradición previa. Esta tradición, sin embargo,

lejos de ser un conjunto de normas fijas y exactas, constituye un conjunto de patrones y esquemas generales, tanto más abiertos a la interpretación cuanto mayor sea la precisión léxico-sintáctica que pretendamos atribuirles. Literalidad y memoria no van necesariamente unidas de la mano en este contexto.

El texto escrito, o simplemente, el texto, del que hablaremos a continuación, no surge de manera independiente y aislada. La división tan marcada que percibimos hoy entre nuestra civilización, tan dependiente del signo escrito, y una lejana y misteriosa cultura sostenida sobre el intercambio hablado se ha extendido a menudo de modo indebido a épocas pasadas, afectando a nuestro juicio negativamente a la correcta interpretación de la realidad. En consecuencia, la metodología que sigamos debe evitar al máximo caer en este anacronismo y limitarse a trabajar con datos y modelos exegéticos defendibles del modo más objetivo posible. La descripción tan somera que acabamos de realizar de rasgos del uso de la memoria y de la oralidad en el texto pretende ilustrar este punto. La historia de los métodos de composición literaria escrita y oral constituye un *continuum* especialmente difícil de delimitar, y son pocos los rasgos susceptibles de describirse como exclusivamente orales. El peso de la tradición, por ejemplo, ha permitido que técnicas originadas por las necesidades de un contexto oral hayan pervivido siglos o incluso milenios después de perder parte de su utilidad, como pueden ser los recursos retóricos basados en la repetición a cualquier nivel. Lo oral, por otra parte, no ha muerto en todo el tiempo que media entre los griegos y nosotros, ni mucho menos fue aniquilado por la escritura. Como decíamos al hablar del funcionamiento de la memoria, la psicología ha demostrado que dentro de nuestra mente seguimos convirtiendo el signo mudo en sonido, dándole voz a los silenciosos libros y, por qué no, escribiendo como si habláramos ante un auditorio invisible.

I.3.3. EL TEXTO ESCRITO

Desde el siglo VIII a.C se produce en Grecia un desarrollo y extensión gradual de la escritura. El proceso es lento, en la medida en que, como hemos visto, la sociedad griega continúa durante varios siglos funcionando sin necesidad alguna del nuevo sistema. Se ha prestado mucha atención a las etapas más tempranas de uso del alfabeto, aunque no tanto a su aplicación pasado el momento de la primera introducción de los signos escritos. En este sentido, por ejemplo, las exposiciones que se pueden encontrar sobre la cuestión se centran en los ejemplos más inteligibles de esta primera escritura, omitiendo otros mucho menos comprensibles, pero que sin duda deben jugar un papel igualmente importante en un análisis del impacto que debió de producir en Grecia la introducción del alfabeto. Conse-

cuencia de esa omisión es la controversia sobre el agente motivador de su aparición, que habitualmente se asocia a la necesidad de fijar o conservar las creaciones poéticas, pero que investigadores como Thomas 1992, 48-50, han puesto en duda a la luz de otras pruebas (*cf.* Robb 1978 y Powell 1989).

La forma de los signos gráficos así como el funcionamiento básico del llamado alfabeto son una herencia fenicia, adquirida probablemente a través del contacto con aquellos mercaderes procedentes de la costa levantina. Los griegos modificaron el valor fonético que ciertos signos tenían para adaptarlos a sus necesidades, mientras que en algunos casos directamente crearon nuevas formas, aunque conservaron en una forma adaptada la mayor parte de los signos semíticos originales⁶⁸. Precisamente la forma de estos signos, cuya variedad geográfica y diacrónica es de sobra conocida, ha sido utilizada con frecuencia para dilucidar esta cuestión. En todo caso, la introducción del alfabeto no produjo una transformación de la noche a la mañana, y su influencia no debía llegar muy lejos, probablemente por su limitado contexto de empleo⁶⁹.

La escritura ostenta el honor de ser una herramienta «revolucionaria», sobre todo la escritura alfabética, por cuanto representa una mejora de los sistemas de escritura anteriores, así como un instrumento de desarrollo intelectual. El argumento que sostiene esta afirmación es su simpleza aparente con respecto a los silabarios, ideogramas o sistemas mixtos tan extendidos en el momento de su aparición. En efecto, desde este punto de vista, el alfabeto consigue una asociación mucho más precisa entre sonido y forma gráfica, lo que no sólo lo convierte en un sistema fácilmente adaptable a cualquier idioma, sino que facilita su aprendizaje, y con ello la rápida extensión del conocimiento mediante la escritura. El razonamiento de los que defienden esta postura va más allá, y llegan a considerar que el desarrollo del pensamiento abstracto y analítico griego es el resultado de los mecanismos mentales de abstracción que requiere la identificación de letras y sonidos característica del alfabeto⁷⁰.

⁶⁸ Se ha escrito mucho sobre el lugar y el momento exactos en los que el alfabeto griego fue creado; al respecto, *cf.* Signes Codoñer 2010.

⁶⁹ Sobre las funciones de la escritura *cf.* Hoz 2004, así como Signes Codoñer 2009.

⁷⁰ Esta opinión fue defendida originalmente por Goody 1968, 44 y Havelock 1982, 3-39.

I.3.3.1. La escritura en Grecia

Encontramos muestras de escritura desde el II milenio a.C. en la civilización minoica, en primer lugar los jeroglíficos cretenses, seguidos de una escritura denominada «Lineal A», ambas sin describir⁷¹. Poco después se desarrolló otra escritura silábica basada en esta última y apodada en consecuencia «Lineal B» y utilizada fundamentalmente para escribir griego micénico. Se han encontrado documentos tanto en Creta como en la Grecia continental, cubriendo un periodo que va desde *c.* 1400 hasta el 1250 a.C., y un poco más de tiempo en la península (*cf.* Olivier 2012, 20). El contenido de los documentos conservados se reduce prácticamente al área de la administración (cuentas, catálogos, listas), aunque también se encuentran algunos ejemplos de naturaleza religiosa o militar. Todos ellos parecen haber sido registros temporales, conservados únicamente por la circunstancia del incendio accidental del archivo. Por desgracia, no poseemos testimonios de un uso más amplio de la escritura⁷². Se ha planteado la posibilidad de un uso de la escritura sobre soportes menos duraderos, como la madera, o el cuero, e incluso se ha llegado a considerar que la forma un tanto cursiva de los signos podría indicar el empleo del pincel como herramienta de escritura predominante⁷³.

Con el derrumbe de la civilización micénica en el siglo XII a.C., la escritura desaparece. Chipre es una excepción, puesto que hacia mediados del siglo XI a.C. se encuentran allí ejemplos de una escritura silábica, inspirada en la escritura minoico-chipriota presente en la isla desde el siglo XVI a.C. En el resto del territorio griego la escritura pereció, al menos en un soporte duradero; puede que durante algún tiempo se empleara algún tipo de material desechable, pero no tenemos constancia de ello. Los siguientes testimonios escritos que encontramos muestran un tipo de escritura con características completamente diferentes, en contenido, tipología y difusión. Tal y como había ocurrido con el Lineal B y con el silabario chipriota, que eran modificaciones del Lineal A, el alfabeto griego derivaba de

⁷¹ La escritura jeroglífica y la lineal A coexisten durante un tiempo: la primera está atestiguada en el periodo *c.* 2000-1600 a.C., mientras que los ejemplos de lineal A datan del periodo entre *c.* 1850-1450 a.C. (*cf.* Olivier 2012). La mayor parte de los documentos parecen ser transacciones, libros de cuentas y sellos personales o familiares; por otra parte, se han encontrado inscripciones en Lineal A no sólo en Creta, sino también en otros lugares de la península, así como en otras islas. Sobre el descubrimiento de las escrituras lineales A y B remitimos a la monografía de Chadwick 1973, además de Heubeck 1979; Melena 2001, 8-79, y Ruipérez & Melena 1990.

⁷² *Cf.* Palaima 2004 para una revisión más detallada de la documentación conservada en Lineal B.

⁷³ *Cf.* Mylonas Shear 1998; por su parte, Palaima 2004, 170-172, sugiere que se trataría de resúmenes administrativos.

un sistema de escritura no-griego, el fenicio. La nueva herramienta se extendió con rapidez, y hacia finales del siglo VIII a.C. la mayor parte del territorio habitado por los griegos utilizaba alguna variante del alfabeto⁷⁴. Las variantes locales persistieron durante un tiempo relativamente largo, y las tradiciones griegas que describen las expansiones y mejoras del alfabeto a cargo de distintos héroes (como Palamedes, Epicarmo o Simónides, inventor también este último de la mnemotécnica) son claramente un reflejo de esta diversidad.

1.3.3.1.a. Orígenes

Los propios griegos desarrollaron teorías sobre el origen de la escritura alfabética, que podemos resumir en dos grupos según la función asignada a la nueva creación⁷⁵. Un primer grupo está formado por aquellas historias, atestiguadas por primera vez en los primeros escritores milesios (Anaximandro, Dioniso, Hecateo), así como en Hérodoto, que defienden un origen extranjero de la escritura, y que no atribuyen al invento ninguna función específica⁷⁶.

La transmisión de esta escritura a Grecia se produce durante la edad heroica, como un invento egipcio o fenicio. Los mediadores en el proceso habrían sido Dánao o Cadmo, los fenicios en conjunto, o incluso Fénix hijo de Agenor y hermano de Cadmo; el lugar de la *transmisión*, Tebas o la Argólida. No obstante, las figuras de los inventores son bastante oscuras y no están adornadas con suficientes rasgos personales como para considerarlos candidatos auténticos a semejante papel. Sin duda, el uso jonio de los términos φοινικία y φοινικογραφία, con el significado de «letras» y «escribir» respectivamente, atestiguados

⁷⁴ Remitimos a Amadasi 1998 para un resumen excelente de las formas de escritura en el Mediterráneo entre el 1200 y el 800 a.C. En lo que se refiere a Grecia, Jeffery 1990 continúa siendo un recurso fundamental. Cf. también Wachter 1996; Lazzarini 1998; Wirbelauer 2004 y Wilson 2009.

⁷⁵ Sobre los orígenes del alfabeto griego remitimos a Signes Codoñer 2004, 17-65.

⁷⁶ Cf. Anaximand. Hist. *FGH* 9, fr. 3; Dion. Mil. *FGH* 687, fr. 1; Hecat. *FGH* 1, fr. 20, centrado en la figura de Dánao; Herod. 5, 57–61, en Cadmo. Sobre el origen fenicio del alfabeto, cf. S. *Poimenes*, fr. 514 Radt. Sobre Cadmo o los fenicios, cf. D.T. *Schol.* 183, 1 (= Éforo, *FGH* 70, fr. 105a), 184, 20 (= *FGH* 70, fr. 105b), 190, 19–20; Clem. Al. *Strom.* 1, 16, 75, 1 (*FGH* 70, fr. 105c); Dionys. Scyt. *FGH* 32, fr. 8 (= D.S. 3, 67, 1); D.S. 5, 58, 3; Non. *D.* 4, 259–64. El ateniense Anticlides añade al egipcio Menes al elenco de posibles inventores: Anticl. *FGH* 140, fr. 11a y b (= Plin. *Nat.* 7, 193) y D.T. *Schol.* 183, 11.

en dos inscripciones de Teos de principios del siglo V a.C., ayudaron a desarrollar teorías etiológicas que apuntaron en el sentido de un origen oriental⁷⁷.

La búsqueda de una explicación al significado de los términos φοινικία y φοινικογραφέω variaba enormemente, desde la interpretación según la cual indicaba una relación con los fenicios, hasta una referencia al uso de pintura roja o al hábito de escribir en hojas de palma (cf. D.T. *Schol.* 184, 23)⁷⁸. Por otro lado, la existencia de tales términos no explica en sí misma la extensión de las historias sobre un origen oriental, en particular egipcio, de la escritura. La razón por la que figuras extranjeras como Dánao o Cadmo se encuentran asociadas a la escritura seguramente sea su pertenencia a civilizaciones consideradas extremadamente antiguas. En el trasfondo de todo esto podemos intuir el contacto temprano y estrecho entre Jonia y Egipto, desde el asentamiento de Náucratis hasta el movimiento de mercenarios griegos en la zona, o los viajes de Hecateo, y la noción de una sabiduría muy antigua en Egipto, de cuya conexión con Grecia Dánao representaría el enlace mítico.

Como ya dijimos, estas historias comparten la característica de que no asignan un objetivo particular a la escritura: esta simplemente es algo que se inventa, o se transmite. El relato, bastante largo, que ofrece Heródoto sobre Cadmo merece una atención particular, por cuanto establece una relación entre la escritura y el lenguaje que representa, así como por el hecho de hacer referencia a los cambios que los signos experimentan con el tiempo, descritos con el verbo μεταβάλλω y μεταρροθμίζω (cf. Hdt. 5, 57 - 5, 61)⁷⁹. La contribución de los griegos se centraría en un cambio de ὀρθμός impuesto a un material preexistente; no obstante, sigue sin haber una sola mención a una función concreta de la escritura a lo largo del *excursus*. La única excepción es un fragmento de Critias (fr. 2, 6 Diels-Kranz), cuya interpretación es discutida, pero que apunta hacia una función mnemotécnica (cf. Ceccarelli 2013, 65). Mientras que la mayor parte de los testimonios a partir del siglo V

⁷⁷ La epigrafía revela varios términos relativos a la escritura y que parecen denotar una procedencia fenicia del alfabeto: ποινικάζεν y ποινικαστάς en dos inscripciones cretenses de finales del s.VI a.C.; el término Φοινικήια presente en las *Teiorum dirae* (480-450 a.C.), como observaron Meiggs & Lewis 1969, 30, 35-41; por último, Φοινικογραφέω en la línea 1, 20 de una inscripción similar en Teos, también fechada en torno al 480-450 a.C., y publicada por Hermann 1981 (= *SEG* 31, 985).

⁷⁸ Cf. Ahl 1967, sobre las hojas de palma, que sugiere que pueden referirse a unas tablillas micénicas con forma de palma, y no a las palmas fenicias.

⁷⁹ Al respecto cf. Jeffery 1967, 153; Corcella 1986, 54-55. Sobre el significado del verbo μεταρροθμίζω como un cambio de la secuencia gráfica de las letras similar a las modulaciones de la voz, cf. Calame 1993, 788-195 y Magini 1998, 23-24.

a.C. favorecen una interpretación de la escritura como un recurso de apoyo a la memoria, la duda recae en la posibilidad de que Critias, un oligarca, no haya recurrido a un juego de palabras con el término ἀλεξίλογα (γράμματ' ἀλεξίλογα, *que ayudan al pensamiento, auxiliar del pensamiento*), en torno al cual gira el debate y que probablemente se trate de un neologismo propio.

La segunda teoría sobre el origen de la escritura se asienta sobre un grupo de historias que adjudican el invento a un héroe griego o a un dios. Los nombres de estos inventores son tantos que en ocasiones parecen intercambiables, e incluso incluyen figuras históricas: Prometeo, Orfeo, Museo, las Musas, Hermes, Sísifo, Palamedes, así como Pitágoras y Simónides. La lista abarca diferentes épocas y contextos⁸⁰. Los puntos de unión entre estos personajes y la escritura varían de un caso a otro. En ocasiones, el descubrimiento forma parte de un propósito civilizador más general (Prometeo, Palamedes, Hermes); en otras, las cualidades del inventor, como puedan ser la astucia o la inteligencia (Hermes, Sísifo), la memoria (Simónides), o la música (Orfeo, Museo) explican su conexión con el descubrimiento. Dentro de esta teoría, la función de la escritura adquiere una importancia mayor, posiblemente por hallarse inserta en una serie de invenciones civilizadoras, y por lo tanto, necesitada de un papel que desempeñar en este contexto. Así, en esta historias las letras sirven fundamentalmente para preservar la memoria y promover el arte. De este modo afirma el titán Prometeo en el drama que lleva su nombre (A. Pr. 454-461)⁸¹:

ἦν δ' οὐδὲν αὐτοῖς οὔτε χεῖματος τέκμαρ
οὔτ' ἀνθεμώδους ἦρος οὔτε καρπίμου
θέρους βέβαιον, ἀλλ' ἄτερ γνώμης τὸ πᾶν
ἔπρασσον, ἔστε δὴ σφιν ἀντολὰς ἐγὼ
ἄστρων ἔδειξα τὰς τε δυσκροίτους δύσεις.
καὶ μὴν ἀριθμόν, ἕξοχον σοφισμάτων,
ἐξηύρον αὐτοῖς, γραμμάτων τε συνθέσεις,
μνήμην ἀπάντων, μουσομήτορ' ἐργάνην.

⁸⁰ Sobre los primeros inventores, *cf.* Kleingünther 1933, además de Kurke 1999, esp. 248–54, y Gera 2003, 112–181, esp. 113–25. Sobre el intercambio de nombres, *cf.* A. Pr. 460–461: Prometeo es aquí el inventor de la escritura, pero un escolio al verso 458 (= *TrGF* 3, fr. 182a) afirma que el propio Esquilo había asignado este mérito a Palamedes en otra parte. El escolio puede referirse, no obstante, a la teoría sobre el movimiento de las estrellas y no a la escritura. Sobre Orfeo y Palamedes, *cf.* Detienne 1990, 81–93.

⁸¹ La tradición manuscrita del v. 461 está dividida entre μνήμην ἀπάντων y μνήμην θ' ἀπάντων. La segunda lectura sirve de conexión entre el verso 461 y el 460: Prometeo habría enseñado a los hombres la unión de las letras, así como la memoria, madre del arte de las Musas. *Cf.* Ferlauto 1990 para un análisis más detallado.

Prometeo da a los hombres herramientas para orientarse en el espacio y en el tiempo, enseñándoles a discernir el orto y el ocaso de las estrellas, la aritmética, y la escritura (σύνθεσις ὀνομάτων), conocimientos denominados en conjunto σοφίσματα (v. 459). La escritura, entendida como la habilidad para unir signos gráficos, es definida por su capacidad de ayuda a la memoria y como algo que ayuda a producir poesía. Reemplaza de este modo a Mnemósine como la madre de las Musas⁸². El uso del adjetivo ἐργάνη (*industriosa*) para calificar la memoria resultado del uso de la escritura es particularmente interesante, porque puede relacionarse con la aparición de un lenguaje técnico musical que pone especial énfasis en la construcción de la poesía, en lugar de en su representación, como Ford 2002, 131-139, ha demostrado⁸³. Merece la pena destacar que, aunque paradójicamente la historia que atribuye el descubrimiento del alfabeto a Prometeo terminó prevaleciendo en la tradición occidental, ha sido más a menudo considerada una innovación de Esquilo (junto con la lista de regalos hechos a la humanidad por el titán), inspirada en algunos de los descubrimientos generalmente atribuidos a Palamedes⁸⁴.

A partir del siglo V a.C., Orfeo aparece también como posible inventor, así como Lino y Museo⁸⁵. El motivo que impulsa el descubrimiento es, en este caso, probablemente la poesía y la sabiduría. Tal es la explicación que supuestamente estaba consignada en una inscripción sobre la tumba de Orfeo, y que fue recogida después en un discurso atribuido a Alcídante (*Alcid. Od. 24*):

⁸² Cf. Nieddu 2004, 47. Por su parte, Ferlauto 1990 sugiere una interpretación de γράμματα como στοιχεῖα (*elementos significativos* en lugar de simples *signos gráficos*), y traduce la aportación del titán como una «combinación de letras articuladas oralmente»: lo que Prometeo habría dado a la humanidad sería el lenguaje articulado. Semejante interpretación es poco probable, aunque el aspecto fónico del alfabeto griego es lo suficientemente fuerte como para que las primeras descripciones del mismo lo identifiquen con el lenguaje hablado.

⁸³ Para la tragedia, remitimos en particular a Ford 2002, 137-138. Como indica Piccaluga 1988, 41, Prometeo no crea las letras, sino que enseña a la humanidad a ensamblar (= comprender) elementos pre-existentes en la naturaleza.

⁸⁴ Cf. Kleingünther 1933, 83 y n. 38; Sommerstein 2000, 122. Sobre la evolución de la figura de Prometeo, cf. Saïd 1985, 115-130.

⁸⁵ Sobre la figura de Lino como descubridor de la escritura, cf. Boardman 1992 y Bremmer 1999; sobre Museo: Kauffmann-Samaras 1992 y Heinze 1999; y por último, sobre el papel de Orfeo en la invención de las letras remitimos a Linforth 1931 y 1941; Calame 2000 y Graf & Johnston 2007. Desde antiguo Orfeo ha sido también considerado el maestro de Lino (cf. E. *Sch. Rh.* 895 = Asclep. Tragil. *FGH* 12 F 6b; D.S. 3, 67, 2), así como de Museo (cf. D.S. 4, 25, 1).

γράμματα μὲν δὴ πρῶτος Ὀρφεὺς ἐξήνεγκε, παρὰ Μουσῶν μαθὼν, ὡς καὶ ἐπὶ τῷ μνήματι αὐτοῦ δηλοῖ τὰ ἐπιγράμματα· Μουσῶν πρόπολον τῆδ' Ὀρφέα Θοῆκες ἔθηκον, ὃν κτάνεν ὑψιμέδων Ζεὺς πολόεντι βέλει Οἰάγρου φίλον υἱόν, ὃς Ἡρακλῆ' ἐδίδαξεν, εὐρῶν ἀνθρώποις γράμματα καὶ σοφίην.

Por otra parte, al trasladar el origen de las letras fenicias al terreno griego, los autores debían encontrar una explicación plausible y paralela al origen de la denominación de las hojas de palma: ἐν φοινίκων πετάλοις (*Sud.* φ 787 = Adler 4, 769, 10). Para ello se busca un origen etimológico en Fénix, maestro de Aquiles, o también en el otro Fénix, hijo de Europa y Pronapes (o Pronapo), que habría sido asesinado por Radamantis a causa del descubrimiento de las letras (*cf.* Ceccarelli 2013, 68-71). En opinión de Schneider 2004, 135-136 y Piccaluga 1988, este hecho resulta interesante por cuanto muestra un lado negativo en el descubrimiento de esta nueva herramienta, junto con las historias de Palamedes, Prometeo o Lino.

Si analizamos la historia de Pronapides (=> Pronapes) como una actualización ateniense de la tradición cadmea de la importación de la escritura, como hace Ceccarelli 2013, 71, podemos observar en ella un reflejo del papel que jugó Atenas en la extensión de la herramienta. En efecto, desde allí se introdujo una reforma del alfabeto en el 403 a.C., cuya intención era la adopción de la variante jonia (milesia), y que marcaría el comienzo de un único sistema de escritura alfabética griego (*cf.* D'Angour 1999). El cambio debió de provocar un debate importante, parte del cual se ve reflejado en el texto de los dramas contemporáneos, que introducen abundantes menciones a la escritura y sus implicaciones⁸⁶. La relación entre ambas esferas, escritura y drama, será constante desde este momento.

La diferencia entre las dos teorías sobre el origen del alfabeto que hemos comentado parecen subrayar dos concepciones de la escritura, que preceden y vaticinan los puntos en torno a los cuales girará el debate posterior al respecto. En particular, los partidarios de la invención autóctona realzan el papel de los signos escritos como una extensión de la memoria humana, más resistentes al tiempo, y, con menos frecuencia, dejan entrever su utilidad como herramienta creativa⁸⁷.

⁸⁶ D'Angour 1999 y Grassl 1972 sugieren que la existencia de varias versiones del alfabeto es la razón del extenso excursus herodoteo sobre la llegada de la escritura a Grecia con Cadmo. Para un análisis de la presencia de la escritura en el drama ateniense remitimos a Ceccarelli 2013, 183-265.

⁸⁷ Ceccarelli 2013, 357-364, realiza un resumen detallado de la opinión de los antiguos sobre el origen de la escritura.

1.3.3.1.b. Usos

A pesar de la pretendida claridad de los relatos sobre la invención de la escritura, probablemente nunca sepamos el motivo original de su creación. Por otra parte, es muy posible que se dieran varias circunstancias a la vez, más aún si descartamos el origen único a favor de un origen múltiple (*cf.* Thomas 1992, 57). Con todo, las posibilidades que se barajan de forma tradicional son de nuevo dos: la escritura habría surgido para facilitar el uso comercial, resultado quizá de la herencia fenicia, o bien como método de preservación de la poesía. Curiosamente, los poemas homéricos sólo mencionan una vez el uso de la escritura, con una función diferente: la de constituir una carta⁸⁸. En efecto, en el canto sexto de la *Ilíada*, Glauco narra cómo Preto, rey de Tirinto, envió a su huésped Belerofonte a Licia con una misiva que comportaría la muerte de este, como venganza del rey ante la falsa acusación de seducción de su esposa Antea (*Il.* 6, 168-170):

πέμπε δέ μιν Λυκίηνδε, πόρεν δ' ὄ γε σήματα λυγρά,
γράφας ἐν πίνακι πτυκτῷ θυμοφθόρα πολλά
δείξει δ' ἠνώγει ᾧ πενθερῷ, ὄφρ' ἀπόλοιτο.

La referencia ha sido objeto de estudio desde la Antigüedad, y las interpretaciones sobre el significado de aquellos σήματα λυγρά pasan porque se tratara de dibujos o bien de signos pictográficos, ἱερὰ ζώδια, como los empleados por los egipcios⁸⁹. Incluso si ese fuera el caso, los signos deberían haber servido para transmitir un mensaje, una idea muy cercana a la escritura, aunque no fuera alfabética⁹⁰. Resulta justificado extendernos en el comentario de estas dos menciones por cuanto parecen reflejar una línea de pensamiento que está en la base del rechazo posterior a la escritura que apreciamos en autores como Platón. Se trata de una oposición entre dos formas de comunicación: una directa frente a otra indirecta o «mediada» (*cf.* Ceccarelli 2013, 62). Esta última puede producirse mediante la intervención de otras personas que transmitan un mensaje (ἄγγελοι, κήρυκες) o bien a través de la escritura, como parece ser el caso de los signos de Belerofonte. Aunque este segundo tipo de comunicación juega un papel importante en la épica y en el mundo griego

⁸⁸ Remitimos a Heubeck 1979, 127-128 y Steiner 1994, 10-16, sobre la importancia de diferenciar entre mensajes basados en signos lingüísticos y σήματα. Ceccarelli 2013, 60, menciona *Il.* 7, 175-189 como ejemplo de esta cuestión.

⁸⁹ *Cf. Sch. Hom. Il.* 6, 168-169a, que se apoya en el sentido de γράφω como «dibujar». Remitimos a Heubeck 1979, 128-143, para un análisis detallado del escolio; *cf.* también Graziosi & Haubold 2010, 124-125.

en general, podemos apreciar, especialmente en la *Ilíada*, la validación de un modo idealizado de discurso en primera persona, protagonizado por hombres adultos de la élite, que se opone a otras formas de comunicación indirecta, tanto oral como escrita, y que sería considerada inferior y «femenina» (cf. Bassi 1998, 44).

Con respecto a las personas que dominaban la nueva técnica, no parece haber existido en Grecia una *casta* de escribas o personas dedicadas exclusivamente o de forma oficial a la redacción de documentos, como demuestra el gran abanico de usos constatados desde muy pronto, desde inscripciones públicas a los *graffiti*, dedicatorias y *dipinti* (escritura pintada sobre la cerámica antes de su horneado). Dos de los primeros grafitos conservados están en verso, al igual otros más largos del periodo arcaico: la llamada *copa de Néstor*, encontrada en Pithecusa, al sur de Italia, de finales del siglo VIII a.C., y el *vaso de Dipilón*, hallado en Atenas y aproximadamente de la misma época⁹¹. El propósito del texto en ambos ejemplos parece estar a medias entre el gusto estético y la fórmula mágica de bendición, combinando una descripción del objeto con los buenos deseos del escritor, sancionados, en el segundo caso, con una mención a la divinidad.

Andersen 1989 ya expuso en su día que al principio la escritura estuvo fundamentalmente al servicio de la palabra hablada. Esta afirmación respondía a un intento de reforzar la idea de que la escritura bebe de los hábitos orales que la precedieron, en lugar de destruirlos, opinión esta última bastante extendida hasta no hace mucho⁹². La inmortalidad y la permanencia son atributos tradicionalmente relacionados con lo escrito; las lápidas con el nombre de la persona enterrada o las ofrendas dedicadas, tan frecuentes desde el siglo VII a.C., suelen entenderse como una muestra del deseo de perpetuación del individuo mediante la propia permanencia de la escritura. Esta era, o sería, inmutable, ya fuera una maldición o un monumento, y el mensaje así conservado viviría independiente de la presencia física del autor. Un pensamiento semejante subyace al texto de un epitafio del siglo VI a.C.: «A todos los hombres que pregunten, mi respuesta es que Andrón, hijo de Antífanos, me dedicó como una τίθη»⁹³. No obstante, corremos el riesgo de olvidar que estos objetos

⁹⁰ Para Ceccarelli 2013, 59, la recepción de la tablilla habría constituido un *signo* (σήμα) en sí mismo, como parece sugerirse en la expresión σήμα κακόν (cf. *Il.* 6, 176-178).

⁹¹ Cf. Jeffery 1990, 233, n. 1; 76, n. 1, y 68.

⁹² Sobre la relación entre escritura y oralidad cf. Svenbro 1988b, 8-44.

⁹³ Πᾶσιν ἴσ' ἀνθρώποι[ς] ὑποκρίνομαι ὅστις ἐ[ρ]ωτᾶι, ὡς μ' ἀνέθηκ' Ἄνδρων Ἀντιφάν[ου]ς δεκάτην (inscripción ateniense del siglo VI a.C.), IG 1², 410. La interpretación de ἀνέθηκ' como «dedicar como una τίθη» es de Thomas 1992, 62.

fueron pensados para que sus inscripciones se leyeran en voz alta. Es el sonido de la voz humana en el seno de la interacción hablada lo que otorga la permanencia al texto, de otro modo mudo sin nadie que pudiera leerlo. Así lo recalca con ironía Simónides cuando afirma que la poesía (= canción) es más duradera que una inscripción (*cf.* Simon. *FGE* n. 581). La escritura es por lo tanto una ayuda a la memoria, pero inútil de forma aislada.

También se la utiliza como refuerzo de las tradiciones. Ya existían monumentos en honor a los muertos antes de la invención del alfabeto, en los que un objeto, o bien su ubicación, actúan como marcador o elemento significativo de la persona enterrada o del hecho que se quería recordar⁹⁴. La escritura simplemente aumentó la capacidad de especificidad o representación del objeto (*cf.* Pucci 1988). El mensaje de un texto inscrito en piedra u otro material es, por lo tanto, una entidad compleja, donde lo escrito no sostiene el peso total de la significación. Tampoco debemos ignorar el impacto visual que la simple forma de las letras debía tener en una sociedad poco acostumbrada a ellas, forma que seguramente fue utilizada como un elemento más del mensaje, sea como una llamada de atención o como elemento de poder.

La forma lingüística de estos primeros textos parece a menudo querer dar voz a un objeto inanimado. Muchos grafitos e inscripciones están en primera persona, creando la falsa ilusión, cuando son leídos en voz alta, de que es el objeto el que habla (*cf.* *SEG* 14, 604: «Yo soy la copa de Néstor»; *IG* 14, 865: «Soy el lecito de Tataia»). Svenbro 1988a, 470, profundiza en las razones gramaticales de esta elección de la persona verbal, y afirma que debió de ser un fenómeno mucho más común de lo que se supone⁹⁵. En esta misma línea Thomas 1992, 64, observa que la progresiva aparición de la tercera persona a partir del siglo VI a.C., cada vez en mayor número de testimonios, representaría un cambio en la concepción de la escritura, cada vez más alejada del primitivo pensamiento mágico y más cercana a un uso meramente documental.

De este modo la escritura en la Grecia arcaica estuvo en gran medida al servicio de la oralidad, ya fuera proporcionando un método de *copia* para el verso cantado o recitado, ya fuera dando voz a objetos inanimados. En general la escritura preserva y refuerza hábitos preliterarios de la sociedad o bien extiende las posibilidades expresivas de algunos instru-

⁹⁴ *Cf.* Il. 7, 86-91; 23, 331-332; Od. 4, 584; 11, 76.

⁹⁵ No estamos de acuerdo en que la primera persona implica un mayor grado de claridad semántica con respecto a la tercera, y por lo tanto tomamos sus conclusiones sólo parcialmente.

mentos de comunicación no orales preexistentes, como por ejemplo los monumentos en piedra o madera.

Algunos usos de los grafitos muestran una relación con lo oral diferente de la que acabamos de mencionar. Si los usos anteriores suponen una *imitatio* de la escritura con respecto a la lengua hablada, los abecedarios, letras y nombres sueltos quedan fuera de esta explicación, al no ser algo habitual en la conversación. Sin embargo, elementos agramaticales, ajenos al ámbito de la oralidad real, tales como listas o «etiquetas», además de no ser infrecuentes, parecen ser un desarrollo temprano del uso de la escritura (*cf.* Goody 1977, 74-105).

Las listas de sacerdotes, de leyes sagradas o los calendarios son precisamente el primer ejemplo de este empleo innovador del signo escrito. El objetivo de estas listas no debió de ser simplemente la datación o su uso como registro, puesto que la memoria siguió ocupando un papel predominante. La propia costumbre de poner el nombre de un funcionario al año (el arconte epónimo de Atenas, por ejemplo) probablemente sea un recurso mnemotécnico⁹⁶. La falta imperante de preocupación por la exactitud favorecía la especulación sobre los datos con los que rellenar estas listas más allá del alcance de la memoria de aquellos vivos en el momento de su redacción. La reelaboración del material o incluso la manipulación directa según las propias opiniones o intereses fue la norma durante mucho tiempo. Paradójicamente, la curiosidad por el pasado favorecida por los sofistas, así como el nuevo interés histórico que detectamos desde finales del siglo VI a.C. debieron de actuar en primer lugar como impulsores de esta *complementación* de las listas, o incluso de su creación desde cero, a falta de material previo. Es de notar que las adiciones no necesariamente eran consideradas falsificaciones, por la propia falta de un sistema de referencia *fijo* y por la tolerancia a los cambios propia de una cultura oral (*cf. supra*, I.3.2.1. *La oralidad y la memoria*). Tampoco debemos descartar un elemento de interés personal por dar una apariencia de conocimiento exhaustivo del pasado. Así, Hipias solía recitar la lista de los arcontes «hasta el tiempo de Solón», a petición de cualquiera que simplemente quisiera escucharla (*cf.* Pl. *Hp. Ma.* 285c). En los historiadores este aspecto sea quizá incluso más fuerte, no sólo a medida que la dependencia de fuentes escritas se hace mayor, sino incluso como un eco perseguido de aquella omnisciencia otorgada al poeta por la Musa (*cf.* el *Catálogo de las naves* de Homero).

⁹⁶ Sobre las primeras aplicaciones atestiguadas de listas como sistemas de datación, *cf.* Jeffery 1976, 34-36; 1990, 59.

Cuando nos referimos al impacto de la escritura en la polis griega, las pruebas indican que afectó fundamentalmente a dos áreas: por un lado, a las listas de funcionarios, y, por otro, a la recopilación y exposición pública de leyes. Estudiosos expertos en el tema ya han hecho suficiente hincapié en el cuidado que debe tenerse a la hora de definir hasta qué punto la maquinaria político-legislativa de la polis arcaica utilizó la escritura (*cf.* Thomas 1989 y 1992). El contexto social y político del momento tuvo un papel importante en el modo como la escritura se introdujo en este terreno, hecho que podemos decir supuso el principio de una lenta revolución.

Nuestros testimonios de un uso público del signo escrito pueden datarse desde mediados del siglo VII a.C., aproximadamente un siglo después de los primeros usos privados atestiguados⁹⁷. Los relatos literarios parecen coincidir con los testimonios epigráficos en cuanto a la finalidad de estos primeros ejemplos: conservar y hacer públicas las leyes. De esta época datarían los primeros y legendarios legisladores: Zaleuco, en Locro, Dracón y Solón en Atenas (*cf.* Thomas 1992, 65-66). Así, con la inscripción de Drero en Creta arranca el fenómeno de las inscripciones públicas, central en el mundo grecorromano.

Se consagraban a la piedra no sólo las leyes civiles, sino también el amplio abanico de lo que podemos llamar «normas sagradas», cuya primera muestra procede de Tirinto, a principios del siglo VI a.C.⁹⁸. Los tratados entre polis aparecen más tarde⁹⁹, y en torno a esta misma época encontramos las primeras listas de sacerdotes y funcionarios, aunque no se realizaban ni mucho menos de forma sistemática¹⁰⁰. A pesar de ello, la memoria pervive como el recurso fundamental para el funcionamiento del sistema durante mucho tiempo, incluso en la Atenas clásica. Así piensa Thomas 1989 cuando afirma que muchas de aquellas listas que los historiadores tomaron como base de su cronología debieron haber

⁹⁷ La muestra más antigua de una ley escrita en piedra procede de Drero, en Creta, *c.* 650-600 a.C. (= *LSAG* 59).

⁹⁸ *Cf.* *LSAG* 9a, donde podemos leer una inscripción de Tirinto, de *c.* 600-550 a.C.

⁹⁹ *Cf.* *LSAG* 61, que representa el ejemplo más antiguo conocido hasta ahora, que se puede datar hacia finales del siglo VI a.C.

¹⁰⁰ *Cf.* *LSAG* 60, 195, con dos inscripciones de Esparta y otras dos de Gerontras, que se pueden datar hacia la mitad del siglo VI a.C.

sido *creadas* a finales del s.V a.C. por personas interesadas en la cronología o las antigüedades¹⁰¹.

El estatus real de las leyes inscritas en piedra es una cuestión compleja. En la época clásica el hecho de que las leyes estuvieran escritas se toma como garantía de igualdad y justicia (*cf.* E. *Supp.* 433ss.; Ar. *Pol.* 1270b), de lo que se deduce que la escritura en la Grecia arcaica sirvió para estabilizar las leyes y protegerlas de la interpretación arbitraria de la aristocracia. No obstante, esta afirmación resulta defendible únicamente para la Atenas clásica, debiendo recurrir a otros criterios para explicar el uso real de las leyes en piedra que encontramos fuera de este marco¹⁰². Si nos acercamos a la cuestión desde el conocimiento de la oralidad tal y como hemos expuesto con anterioridad (*cf. supra*, I.3.2.3. *La memoria en la Antigüedad*), vemos que, en primer lugar, existía un corpus bastante amplio de leyes no-escritas, sancionadas por el consentimiento de la comunidad y transmitidas por el peso de la tradición (*cf. Il.* 18, 497-508), y cuya fuerza puede aún notarse hasta el siglo V a.C. Como su propia definición indica, tras la invención del alfabeto, muchas de estas normas permanecieron a salvo en la mente del colectivo griego, y conocemos su existencia por alusiones a las mismas en aquellas leyes que sí fueron puestas por escrito; su papel es el de dar contexto a las sanciones o procedimientos inscritos en la piedra. De hecho, una explicación bastante plausible del motivo por el que empezaron a escribirse en primer lugar es la necesidad de dar fuerza a normas o leyes de escasa fuerza coercitiva, mediante el impacto visual que el monumento (del que la escritura es sólo una parte) proporcionaba. Esto da sentido además a la escasa atención a la exactitud en la redacción de las propias leyes que podemos observar entre una copia y otra (*cf.* Thomas 1992, 68 y Thomas 1994, 38-39).

La escritura no supuso, al menos en este primer momento, un cambio radical en las costumbres o el pensamiento político-cultural griego, y antiguas instituciones no sólo coexisten, sino que acogen en su seno el alfabeto, produciéndose una curiosa evolución simbiótica. Tras la adopción definitiva de la escritura en época helenística, su función fue progresivamente trasladándose a la esfera de los archivos y de la documentación escrita (*cf.* Lambridunakis & Wörrle 1983, 328-344), pero no por ello perdieron del todo los *μνημόνες* su relación con la memoria y la oralidad, aunque fuera a título de curiosidad.

¹⁰¹ Sobre el interés de los antiguos por la ubicación en el tiempo de hechos, objetos o personas, el caso de Hipias es especialmente interesante (*cf.* Thomas 1989, apéndice).

¹⁰² Remitimos a Bowman & Woolf 1994 para un estudio del empleo de la escritura como instrumento de influencia y poder político.

Una de las primeras utilidades de la escritura fue marcar o etiquetar ofrendas a los dioses, así como *fixar* maldiciones a modo de protección de aquellos objetos sobre los que se inscribían las letras. La escritura extiende y acentúa un método primitivo de obtener la atención y la obediencia del público, es decir, la asociación con la esfera de lo divino de una forma coercitiva. Este fenómeno es especialmente destacado en el ámbito del uso oficial de la escritura. La ciudad de Teos publicó una ley cuya expresión lingüística es la misma de una maldición (*cf.* Hermann 1981; Meiggs & Lewis 1969, 30). Cuando los funcionarios la leían en voz alta, el propósito habría sido seguramente el de hacer efectiva la maldición sobre aquellos que pudieran incumplir el texto, a través del proceso, en cierto sentido mágico, de *realización* de lo dicho a través del sonido de la voz humana. La polis arcaica, por lo tanto, expande un uso privado de la escritura (las dedicaciones y las maldiciones), aportándole monumentalidad y solemnidad. Gracias a la introducción de la escritura en este contexto, esta última fue adquiriendo las características primitivamente asociadas con la voz, hasta adquirir poder por sí misma.

I.3.3.2. La conciencia documental y el cambio de actitud hacia la escritura

I.3.3.2.a. Funciones de la escritura

La invención del alfabeto no parece haber sido una revolución desde el comienzo, sino una especie de *curiosidad* cuyos rasgos y posibilidades se fueron descubriendo a medida que la cultura griega ampliaba el abanico de sus aplicaciones. Desde un primer momento compitió con el lenguaje hablado, en la medida en que se trata de una herramienta íntimamente ligada al lenguaje, puesto que lo representa. El enfrentamiento entre *oralidad* y *literalidad* de los primeros estudios sobre el tema responde a una extensión del valor actual de la comunicación escrita y oral, y en particular a un desconocimiento bastante extendido del funcionamiento de una sociedad completamente oral. Gracias a la conjunción del saber proporcionado por la antropología, con una relectura de los testimonios, no sólo escritos sino también arqueológicos y visuales, de la Antigüedad, nuestra visión sobre el tema ha cambiado bastante (*cf.* Gil 1995, 7-8). En efecto, los trabajos de Thomas 1989, 1992 o Bowman & Woolf 1994, representan este viraje que experimentó la percepción filológica sobre la escritura en Grecia, y nos han servido ya de base para los capítulos anteriores y nos siguen sirviendo ahora.

Toda herramienta nueva requiere una necesidad que justifique su creación. Cuando analizamos los dos grupos de historias sobre la invención de la escritura en Grecia, predo-

minan dos ideas: por un lado, la representación gráfica del lenguaje, y, por otro, la conservación de obras de arte cantadas o recitadas y transmitidas de boca en boca (hablar de literatura en una sociedad oral nos parece en efecto un anacronismo). Podemos decir, pues, que la escritura se introduce en los mecanismos de comunicación orales en calidad de refuerzo visual, ejerciendo dos funciones: la primera es extender el poder de la lengua hablada, dando voz a objetos inanimados, a través de inscripciones o grafitos en primera persona; en segundo lugar, la polis aprovecha la naturaleza visual de la escritura para aumentar el impacto de leyes que necesitaban de un apoyo para su cumplimiento, añadiendo una connotación de herramienta de poder a la escritura.

La adopción de la escritura implica al principio una continua participación de lo oral, mediante la lectura en voz alta de lo escrito. La voz humana da *realidad* a los silenciosos signos escritos. Con el tiempo, este flujo de poder desde lo oral a lo escrito se invierte parcialmente, permitiendo a la escritura adoptar parte de las atribuciones propias de la voz hablada y cobrar importancia por sí misma. Creemos que el aumento de usos documentales del texto es representativo de este proceso. Entendemos por *documental* la función de la escritura como elemento transmisor y conservador de la información, distinto del elemento de poder que muy pronto se detecta en ella, especialmente en las inscripciones oficiales. No obstante, es necesario tener cuidado con respecto al concepto de *documento* que subyace al uso de textos escritos. En efecto, la exactitud de lo transmitido no es un motivo de preocupación entre los griegos. Como ya vimos (*cf. supra* I.3.2.3.c.), ni siquiera Tucídides, considerado el primer investigador serio de la historia, transmite palabra por palabra sus testimonios más impactantes, los discursos. Él mismo reconoce la imposibilidad de realizar semejante tarea, al tiempo que prefiere *reconstruir* la idea contenida en el testimonio, dándole una forma más apropiada a su intención (*cf. Th.* 1, 22). En efecto, la repetición exacta de las palabras rara vez es importante en sociedades orales, principalmente porque no hay ninguna forma de comprobarla¹⁰³. La incorporación del texto escrito en el proceso de creación-recepción de la literatura introdujo un punto de referencia más o menos fijo, que fue adquiriendo autoridad hasta forjar la noción de que entre el *original* y la copia debía existir una estrecha relación formal.

Cuando hablamos del aspecto formal del texto, nos referimos fundamentalmente a la repetición de *palabras*. Sin embargo, ¿qué son exactamente las *palabras*? Unas de las razones por las que las sociedades orales no transmiten de forma exacta sus manifesta-

¹⁰³ *Cf.* Small 1997, 5. Dentro del mundo oral, la idea de repetición literal ocurre en unas pocas y limitadas situaciones, fundamentalmente dentro del mundo mágico-religioso (*cf. Plin. Nat.* 28, 11; *Cic. Har.* 23).

ciones es porque no poseen un concepto claro de *palabra*. Havelock 1982, 8 y 1982, 32 n. 5, afirma que los griegos «originally had no word for a word singly identified, but only various terms referring to spoken sound», y duda incluso de que la palabra como elemento básico del lenguaje fuera un descubrimiento anterior a Platón¹⁰⁴. Según algunos psicólogos, la habilidad de distinguir las palabras aisladamente procede de la habituación a su separación en la escritura¹⁰⁵. Hay pruebas de que los niños no conciben el lenguaje como una sucesión de palabras individuales ni por supuesto entienden estas como unidades separadas de significado. Lo que oyen es el significado completo de un enunciado, y esto solo se percibe después de la combinación de los significados de todas las palabras, algo que se produce de forma automática y sin que sean conscientes. Al hablar, no emitimos las palabras una por una; es decir, no hacemos pausas reales entre las mismas. Al escribir, en cambio, dejamos espacios en blanco entre las palabras, de forma que las representamos de forma discreta. Para Olson 1993 esto demuestra que los sistemas de escritura no fueron creados para representar el lenguaje, sino para comunicar información; el resultado de ello es que son los sistemas de escritura los que crean las categorías en función de las cuales somos conscientes del lenguaje. En otras palabras, la escritura proporciona el modelo que conduce a una particular concepción de cómo funciona el lenguaje y no al revés.

El texto del *Fedro* de Platón resulta instructivo a propósito de la conciencia que tenían los griegos de la relación entre la escritura y la memoria, en particular *Phr.* 274e-275d. De este pasaje podemos destacar tres aspectos¹⁰⁶. En primer lugar, establece una conexión entre la memoria y la escritura (μνήμης [...] φάρμακον ἠύρεθη). En segundo,

¹⁰⁴ Compárese con lo que dice en 1986, 113: «There is probably no attestable instance in Greek of the term *logos* as denoting a single «word», though it is often translated as though it did. The first “word for a word” in the early philosophers seems to have been *onoma* a “name”». Para un análisis más detallado del concepto de *palabra* remitimos a Luque Moreno 2006, 13-16 y a la bibliografía allí recogida.

¹⁰⁵ Cf. Adams 1990, 296-299; Gleitman *et al.* 1989, 97-101. En esta misma línea, Olson 1994, 126-127, sostiene que los niños no distinguen entre paráfrasis y repetición literal hasta los 5 ó 6 años. La separación de palabras no se aplica de modo sistemático hasta la aparición de la minúscula en torno al siglo IX d.C., siendo lo habitual la práctica de la *scriptio continua*; al respecto, cf. Gavrilov 1997; Cavallo & Chartier 2004; Ferreiro 2007.

¹⁰⁶ Para un comentario de este y otros pasajes de Platón relativos a la escritura y la oralidad remitimos a Gil 1995, 23-24 (cf. Lledó 1992 y Brisson 1982). Otros pasajes interesantes sobre el tema son Caes. *Gall.* 6, 14; Sen. *Epist.* 88, 33; Vitruv. 7, pr. 9; así como Ov. *Pont.* 4, 8, 48-52. El argumento de este último pasaje puede retrotraerse hasta Alcídama, rétor contemporáneo de Platón; en efecto, la última de las cuatro razones que Alcídama da para escribir su discurso es «καὶ μνημεῖα καταλιπεῖν ἡμῶν αὐτῶν σπουδάζοντες καὶ τῆ φιλοτιμίᾳ χαριζόμενοι λόγους γράφειν ἐπιχειροῦμεν» (cf. Alcíd. fr. 15 [*Soph.*], 32, 5).

establece asimismo la distinción entre una memoria auténtica y el simple recordatorio (ὑπόμνησις), siendo la escritura útil únicamente en este último sentido. El tercer aspecto es una crítica general al uso, o al abuso, de la escritura en cuanto elimina el elemento conversacional presente en la comunicación oral, y que para Platón está intrínsecamente unido a la posibilidad de adquirir conocimiento. Desde el punto de vista de lo que expusimos sobre la oralidad en Grecia, lo que Platón parece indicar es que la escritura pone en duda la necesidad de un contexto para obtener un mensaje significativo, una de las bases del sistema oral establecido, protegida mediante procedimientos que aseguraban su presencia, como ocurría en el caso de los μνημόνες (cf. Carawan 2008).

Con la alfabetización la memoria tuvo que *aprender* a recordar palabras. En efecto, como bien dijo Norman 1993, 78: «Writing something down doesn't really change our memory; rather, it changes the task from one of remembering to one of writing then, later, reading back the information. In general, artifacts don't change our cognitive abilities; they change the tasks we do». La literatura produjo una necesidad creciente de nueva información, y esta necesidad se desarrolló exponencialmente a través del mayor uso de la literatura. En este mismo sentido, Small 1997 sostiene, basándose sobre todo en los argumentos de Platón, que la dificultad para encontrar *información* fue una de las causas de la lenta aceptación del medio escrito. En efecto, la Antigüedad clásica disponía sorprendentemente de pocos recursos para facilitar la recuperación de información en el medio escrito. La pervivencia de modos y formas de comunicación orales actuó como un importante freno al desarrollo de estas técnicas, basadas en una concepción autónoma de la escritura¹⁰⁷. La forma física del libro, incluso después de que se extendiera hasta convertirse en una realidad más o menos común, mantuvo en líneas generales un cierto halo de misterio, desvelable sólo para unos pocos *amantes de las letras*.

1.3.3.2.b. El libro

La escritura ya jugaba un papel bastante importante en las culturas orientales contemporáneas a la griega. Desde muy pronto se encuentran en Mesopotamia indicios de archivos en tablillas de arcilla, así como de bibliotecas con textos literarios¹⁰⁸. El movimiento demográfico y político que las invasiones de persas e hititas produjo en la zona impulsó el desa-

¹⁰⁷ Cf. Pl. *Phdr.* 275d: ταῦτόν δὲ καὶ οἱ λόγοι· δόξαις μὲν ἂν ὡς τι φρονοῦντας αὐτοὺς λέγειν, ἔὰν δέ τι ἔρη τῶν λεγομένων βουλόμενος μαθεῖν, ἔν τι σημαίνει μόνον ταῦτόν ἀεὶ.

¹⁰⁸ Sobre el uso de la escritura en Mesopotamia, cf. Kenyon 1951, 6.

rollo de una temprana labor lexicográfica, que, junto con la costumbre de elaborar listas y archivos heredada de los sumerios, constituyen la base de una civilización asentada firmemente sobre la escritura¹⁰⁹. Por esta razón investigadores como Wendel consideraban que los medios materiales para la escritura utilizados por los griegos fueron una herencia de Oriente y no de Egipto (*cf.* Wendel 1949, 85; Kenyon 1951, 44).

La escritura y en particular esa manifestación de la misma que venimos denominando *libro*, conoció en la Antigüedad numerosos tipos de soporte¹¹⁰. Además de la piedra, difícilmente útil para este propósito, tenemos referencias a textos sobre hojas, corteza de árboles, lino, metales (oro, bronce, hierro y plomo); la madera, en forma de tablillas blanqueadas o cubiertas de cera; fragmentos de cerámica o arcilla (*cf.* Ceccarelli 2013, 23-59 y Roemer 2007, 85). No obstante, en lo que se refiere a Grecia ninguno de estos materiales fue empleado para la producción literaria. Quizá se diera algún caso aislado, como la lámina de plomo que los beocios del monte Helicón enseñaron a Pausanias en la que se había copiado parte de los *Trabajos y Días* de Hesíodo (*cf.* Paus. 9, 31, 4); sin embargo, como observa Roemer 2007, los libros antiguos estaban fundamentalmente escritos en papiro o pergamino, sobre todo el primero. Para Pfeiffer 1968, 24-26, las condiciones materiales necesarias para producir libros debían de cumplirse en el mismo momento de la invención del alfabeto, alrededor del siglo VIII a.C., si bien la memoria mantuvo su hegemonía como soporte durante siglos. Siguiendo esta lógica, el mismo Pfeiffer observa que la historia sobre la edición de Pisítrato de los poemas homéricos o la tradición recogida por Ateneo de que aquel y Polícrates eran ávidos coleccionistas de libros (*cf.* Ath. 1, 4) bien pudieron ser una creación de época alejandrina (*cf.* Pfeiffer 1968, 25).

Roemer 2007, 88, repasa algunas de las primeras referencias literarias y artísticas a los βίβλια, ninguna de ellas anterior al siglo V. a.C.¹¹¹. Las primeras representaciones de rollos escritos aparecen en la cerámica ática de figuras rojas, en escenas de la vida acomodada obra de contemporáneos de los dramaturgos clásicos, entre el 490 y el 425 a.C. En algunas de estas piezas cerámicas es posible leer palabras o letras correspondientes a poemas líricos o épicos sobre los rollos pintados¹¹². Jóvenes y maestros de escuela aparecen

¹⁰⁹ Remitimos a Gangutia Elícegui 1977, 6-7, sobre estos primeros diccionarios orientales; sobre lo que podríamos denominar «notas al pie» en las tablillas sumerias, *cf.* Wendel 1949, 2.

¹¹⁰ Sobre los soportes físicos de la escritura y el proceso de redacción y publicación en la Antigüedad, remitimos a Dorandi 2007, cap. 1 y 5.

¹¹¹ Sobre la extensión del libro en la Atenas del siglo V a.C., *cf.* Turner 1952.

¹¹² *Cf.* Beazley 1948; Turner 1952, 13-16; Immerwahr 1964 y 1973.

leyendo, y hacia finales del siglo V a.C. aparecen nombres famosos como Safo, Lino o Museo sobre las figuras así dispuestas. Dos de los ejemplos más ilustrativos son una copa firmada por Duris, en la que un hombre está leyendo un texto que podría ser la *Ilíada*¹¹³; el segundo es una *hydria* donde se lee «Safo leyendo»¹¹⁴. Para Roemer, en ambos casos el propósito no es la representación de un libro en concreto sino mostrar el propio acto de la lectura. Prueba de ello sería el aparente fallo con respecto al uso real del rollo que detectamos en el primer ejemplo: el maestro sostiene el *volumen* de forma vertical, cuando lo habitual es hacerlo en horizontal. El objetivo de este error aparente sería facilitar la lectura de la primera línea de la obra. Tal y como ocurría con el tratamiento de los testimonios orales, la preocupación de la mente griega es transmitir un mensaje, una idea, más allá de la repetición literal.

Además de los testimonios pictóricos, dentro de la literatura y sobre las mismas fechas encontramos unas pocas alusiones al comercio de libros en la Comedia antigua y en los diálogos de Platón. Eupolis, por ejemplo menciona un lugar «donde se venden libros» (*cf. fr. Kassel-Austin 327*). El pasaje es comentado por Pólux, a propósito de las βιβλιοθήκαι, y aclarando de paso que se refiere sin duda a un lugar especializado en el comercio de libros («οὕτω γὰρ τὸν τόπον ‘τὰ βιβλία’ οἱ Ἀττικοὶ ὠνόμαζον, ὥσπερ καὶ τοὺς ἄλλους τόπους ἀπὸ τῶν ἐν αὐτοῖς πιπρασκομένων»; *cf. Poll. 9, 47*). En efecto, si damos crédito a Jenofonte (*An. 7, 5, 12-14*), entre los siglos V y IV a.C. el comercio de libros es un negocio que llega hasta el mar Negro. El proceso de publicación, por otra parte, no debía de ser muy largo, ya que Sócrates adquiere con relativa rapidez los libros de Anaxágoras, una vez las ideas de aquel llamaron su atención (*cf. Pl. Phd. 97b-98b*). No nos merece la misma credibilidad el apunte que hace sobre su precio, probablemente irónico y relacionado con la decepción que le supuso tras su lectura (*cf. Pl. Ap. 26d*)¹¹⁵.

Quizá fueron estas afirmaciones hechas por Platón las que indujeron a Clemente de Alejandría a mencionar a Anaxágoras como el primero en publicar un libro en la lista de primeros inventores¹¹⁶. La relación entre los sofistas y el uso extensivo de libros, o lo que es

¹¹³ *Antikensammlung* de Berlín, n° 2285, c. 490-480 a.C.

¹¹⁴ Museo Nacional de Atenas, n° 1260, c. 440-430 a.C.

¹¹⁵ *Cf. Lewis 1974, 62 y Turner 1952, 21.*, extrañados cuando compararon el coste real del papiro y la copia con la afirmación de Sócrates. Quizá sea una exageración.

¹¹⁶ *Cf. Clem. Al. Strom. 1, 16, 78: οἱ δὲ Ἀναξαγόραν Ἠγησιβούλου Κλαζομένιον πρὸ τῶν διὰ γραφῆς ἐκδοῦναι βιβλίον ἱστοροῦσιν.* Según Pfeiffer 1968, 28, Clemente bien podría haber confundido el sentido de γραφή.

lo mismo, del texto, es de sobra conocida¹¹⁷. Las obras de los sofistas circulaban en forma de libros, caso de Anaxágoras como ya hemos comentado, o de Pródico (*cf.* Pl. *Smp.* 177b), que escribían lo que podríamos llamar «manuales» para sus alumnos, independientemente de que continuaran practicando la instrucción oral tradicional¹¹⁸.

I.3.3.2. Codificación visual

I.3.3.2.a. Filología y metatexto

Una de las más importantes implicaciones de la aparición del concepto de *libro* entre los siglos IV-III a.C. y el aumento exponencial de su uso es la definitiva aceptación de la escritura como un soporte adecuado para la conservación del acervo cultural griego. Al mismo tiempo, un grupo de poetas y eruditos reunidos en torno a la floreciente capital del Egipto ptolemaico se vuelven conscientes de la ruptura que se había producido con los siglos anteriores¹¹⁹. A través de una intensa conversación literaria que estos poetas y eruditos establecen entre sí no solo toma forma la idea de lo clásico, sino que surge la primera disciplina centrada exclusivamente en el texto: la filología (*cf.* Hummel 2000, 427-430).

Los eruditos alejandrinos escogen el texto como el testimonio máspreciado de una cultura que ya no existe, y que por lo tanto es necesario conservar. Su propósito es triple: preservar el pasado, comprenderlo y explicarlo a las generaciones presentes y futuras. El instrumento para lograrlo es igualmente el texto, continuando con ello la tendencia de los sofistas a utilizar la escritura como herramienta de investigación y no sólo como una alternativa a la oralidad. Para ello desarrollan las posibilidades de aplicación de la escritura. Además de extender el uso de los recursos que habían heredado de la comunicación oral, crean nuevos destinados a la expresión de información basados exclusivamente en las características del soporte escrito. Estas técnicas, creadas en principio para el comentario filológico, dejarán una huella indeleble en la forma física del libro en todas sus manifestaciones, apreciable aún hoy.

¹¹⁷ *Cf.* Pfeiffer 1968, 25-30, uno de los principales defensores de esta tesis.

¹¹⁸ Sobre el carácter didáctico de los textos elaborados por los sofistas, *cf.* Pl. *Phdr.* 266d: τὰ γ' ἐν τοῖς βιβλίοις τοῖς περὶ λόγων τέχνης γεγραμμένοις. Sobre la educación oral de los sofistas, *cf.* D.L. 9, 50 y 54 en relación a Protágoras y Pródico y Pl. *Hp.Ma.* 286 b-c referido a Hipias.

¹¹⁹ *Cf.* Fraser 1972, 447-480; Pfeiffer 1968, 87-105.

Se ha escrito mucho sobre los orígenes de la filología, desde una enorme variedad de puntos de vista. A continuación exponemos aquellos aspectos en los que la actividad filológica ha influido sobre la forma física, los usos y la concepción del texto. La constatación de esta influencia será vital para comprender algo tan aparentemente lejano como pueda ser la aparición del texto digital.

A diferencia del papel auxiliar que había jugado la escritura en los usos oficiales del texto hasta entonces, para la filología helenística la forma escrita constituye el centro del mensaje, alrededor del cual se explica el sentido del resto de características del soporte. El objetivo que persiguen estos eruditos alejandrinos, sin embargo, no es simplemente utilizar el lenguaje para obtener una información o explicar unas formas oscurecidas por el tiempo, sino que a esto se le une la intención de conservar el texto como obra de arte. El significado pasa entonces a estar supeditado a la correcta conservación de ese estado original del texto.

La edición se basa muchas veces en el hallazgo de contradicciones con respecto a lo que se supone debería estar en el texto, es decir, con un modelo ideal. Esta forma de ἔκδοσις, por tanto, refleja una preocupación por la naturaleza del significado que oradores y filósofos de siglos anteriores ya habían manifestado¹²⁰. La edición de texto (διόρθωσις) es un fenómeno que sólo puede producirse cuando el único punto de referencia válido para acceder al contenido de un mensaje es la forma lingüística del texto transmitido en un soporte escrito. Es una forma de rectificación razonada basada en la constatación de que existía una diferencia crucial a varios niveles (lingüístico, estético y semántico) entre los textos de siglos anteriores y la producción literaria contemporánea a la filología alejandrina. Este hecho produce un interés doble hacia los textos de esta época: un deseo de comprender un mensaje oscurecido por el tiempo, y también una preocupación por no destruir la expresión lingüística y escrita exacta en el que ese mensaje está contenido. En lo que se refiere a esto último, la filología rompe con la metodología imperante en la literatura griega hasta entonces, basada en la adaptación y la adecuación (*cf. supra I.2.3.3. Adaptación del recuerdo y I.3.3.2.a. Funciones de la escritura*). El resultado final de esta devoción por la preservación del texto es el comentario¹²¹. La explicación de los antiguos a través de las

¹²⁰ Cf. Lys. 10, 7, «οὐ περὶ τῶν ὀνομάτων διαφέρεσθαι ἀλλὰ τῆς τούτων διανοίας». Para Platón es imposible conocer la διάνοια, el auténtico significado que el poeta tenía en mente a la hora de escribir (*cf. Pl. Pr. 347e; Hp. Mi. 365c-d e Io. 540b-c*).

¹²¹ En griego ὑπόμνημα. Pfeiffer 1968, 29 y 213, observa que el término tiene también el sentido de *notas o apuntes* (*cf. Dorandi 2007, 68-76*). En efecto, la terminología técnica en la Antigüedad carece de la

palabras escritas que se nos han conservado representa el culmen de la actividad filológica. La importancia y la singularidad de la filología helenística o alejandrina en el tiempo descansa, por lo tanto, en el rigor y la preocupación por basar la interpretación en el propio texto¹²².

Un rasgo fundamental de la filología alejandrina es su carácter normativo: los filólogos se erigían en jueces de la importancia y del valor de cada poeta (en cierto modo eran herederos de Aristóteles) y presentaban su propio juicio estético como inapelable¹²³. Basándose, por otra parte, en las necesidades de la enseñanza crearon una serie de cánones literarios, unas listas de poetas que eran considerados como *modelo* dentro de cada género (cf. Torres Guerra 2012). Nacía con ellas el concepto de clásicos, de poeta modélico cuya lectura era aconsejable por una serie de motivos y razones. Por otra parte, el comentario de los autores antiguos se convirtió en un foro de debate intelectual sobre una multitud enorme de temas. La *πολυμαθία*, despreciada en épocas anteriores como una falta de conocimiento real de las cosas, se revaloriza bajo la égida de la actividad exegética de los alejandrinos, convirtiéndose incluso en una virtud deseada (cf. Hummel 2000, 271). Este gusto por saber *un poco de todo* terminaría por desembocar en el *epitomismo* tan extendido en la Antigüedad Tardía y en la Alta Edad Media, como único remedio al problema de una sobrecarga informativa abrumadora¹²⁴.

Invencción del metatexto: los signos críticos

El objetivo de preservación de la forma exacta del texto exigido por la filología impulsa el desarrollo de un código de signos escritos específicos que permitan conectar el texto comentado con su explicación, sin afectar a la visualización del primero, como ocurriría si directamente se eliminaran o alteraran los pasajes. La actividad filológica se mueve, pues, dentro de lo que podemos llamar *metatexto* o *paratexto*, desarrollando formas de mejorar la información transmitida mediante los signos escritos y con ello la comprensión del mensaje o mensajes contenidos en el texto. Es un logro importante porque distingue de

preocupación por la univocidad característica del pensamiento científico actual, por lo que las traducciones de términos técnicos griegos deben ser tomadas siempre con cierta prudencia.

¹²² Cf. Porph. *ad Il.* 56, 3, «Ἀξιῶν δὲ ἐγὼ Ὅμηρον ἐξ Ὀμήρου σαφηνίζειν αὐτὸν ἐξηγούμενον»; la frase se atribuye a Aristarco (cf. Schironi 2004, 122, 160, 500). Sobre este pasaje, remitimos a Montanari & Pagani 2011, 21-22.

¹²³ Cf. Alsina 1963, 174.

¹²⁴ Sobre el comentario en la Antigüedad remitimos a Goulet-Cazé *et al.* 2000.

forma explícita diferentes tipos de información en el texto así editado, a través del propio documento¹²⁵.

La filología practicada en la biblioteca de Alejandría durante los siglos III-II a.C. tomó la forma material de un sistema de anotaciones en los márgenes de las columnas de texto en el rollo de papiro (*cf.* Irigoín 1997, 31-39). Estas anotaciones eran un tipo de marcas que señalaban lugares problemáticos en el texto (una línea, una palabra o pasaje), y sugerían acciones como su corrección, eliminación, o traslación a otro lugar, así como algún otro tipo de clarificación. El sistema pretendía conservar el texto transmitido y constituía un nivel de escritura autónomo, es decir, se trataba de un *paratexto*. Este tipo de actividad crítica no-destructiva conduciría finalmente a una nueva copia del texto, ya que contendría las instrucciones necesarias para realizarla. Así pues, junto con la simple lectura del texto existía otra lectura *técnica* del mismo, el comentario, cuya conexión con el primero se establecía mediante los signos mencionados. Sin embargo, estas marcas críticas eran poco explícitas por sí mismas y el descriframiento de su significado concreto dependía del conocimiento por parte del lector de lo que estaba leyendo, en ausencia de una explicación mejor.

Consideremos ahora otra serie de ayudas visuales o gráficas, habituales en nuestra concepción moderna de libro, y cuya función principal es facilitar el acceso al mensaje que este transmite. Son, por lo tanto, métodos de *optimización* de una de las funciones primarias del texto: almacenar información y facilitar su recuperación posterior. Por lo que respecta a los signos críticos empleados por los alejandrinos y recogidos en la tradición medieval nos limitamos a remitir al esquema de Morocho Gayo 2004, 25-27.

1.3.3.2.b. Composición visual del texto

Las obras antiguas carecían de elementos que articularan *visualmente* el texto¹²⁶. Este se presentaba de un modo similar al de las inscripciones, como un torrente continuo de letras (*scriptio continua*) carente de nuestras modernas separaciones en capítulos o párrafos. La división actual en cantos de *Ilíada* y *Odisea*, por ejemplo, no se produjo hasta el periodo helenístico, coincidiendo con el inicio de la costumbre de escribir la poesía en líneas separadas (*στίχοι*). El uso de espacios en blanco para indicar separaciones internas de la frase,

¹²⁵ Sobre la *ἐκδοσις* en la Antigüedad, *cf.* Dorandi 2014.

¹²⁶ En lo referente a la articulación de habla en el texto durante la Antigüedad, remitimos a Luque Moreno 2001 y 2001.

entre palabras, piezas o entre secciones con una coherencia argumental clara, era extraño y no existía ni mucho menos un estándar que indicara el sentido contextual que dichos espacios pudieran tener en caso de aparecer¹²⁷. Los límites entre distintas partes de una obra estaban sujetos a controversia, y en algunos casos coexistían diferentes divisiones, en un mayor o menor número de partes según los criterios existentes.

Títulos

Los títulos van ligados a la necesidad de identificar un volumen entre una gran cantidad de ellos¹²⁸. Surgen como una necesidad cada vez mayor conforme crece el número de obras publicadas, y especialmente apremiante para aquellos que se dedicaban a coleccionar libros, y posteriormente, como una utilidad organizativa de las bibliotecas. En estas circunstancias las formas de identificar una obra podían ser *a)* un número, *b)* el autor, *c)* el tema, *d)* las palabras que dan inicio a la obra o *e)* una combinación de cualquiera de las anteriores formas. Pero cuando el número excedía cierto límite, una expresión simplificada de su contenido se convirtió en el método más deseado; de ahí la creación de los títulos. Al igual que su división en libros o partes, los nombres *Ilíada* y *Odisea* son creaciones posteriores de los alejandrinos.

Small 1997, 29-31, analiza la memoria en la Antigüedad como una forma de *RI* y aborda la cuestión de los títulos como una respuesta a un problema: la identificación de obras dentro de un panorama literario cada vez mayor. En las páginas citadas recoge algunas de las soluciones adoptadas, y que resumimos a continuación. Al principio los títulos se escribían en etiquetas identificadoras dentro del rollo, a menudo al final del mismo, ya que esta era la parte más protegida; sólo hacia el final de la Antigüedad encontramos títulos al comienzo. Sin embargo, no todas las obras eran identificadas de esta forma; en la Atenas del siglo V a.C., la gran cantidad de obras dramáticas hizo más habitual la presencia de un título para cada pieza, pero continuó siendo una rareza en otros géneros literarios. El comienzo de las obras históricas de Heródoto y de Tucídides muestra hacia dónde apuntaban las preocupaciones sobre lo que debía identificar una obra: en ambos casos se expone

¹²⁷ Sobre la polisemia de estos σημεῖα o «signos críticos», cf. Pfeiffer 1968, 115; lo que sigue se basa fundamentalmente en lo dicho por este autor, salvo indicación de lo contrario.

¹²⁸ Sobre los títulos y los recursos visuales de articulación del texto en las obras antiguas remitimos a Fredouille *et al.* 1997.

quién es el autor, de dónde viene y a continuación, un resumen del contenido¹²⁹. Por otra parte, la redacción del título no siempre era tarea del autor, especialmente en casos donde la brevedad o el contexto concreto de la composición lo volvían innecesario a ojos de su creador (*cf.* Plin. *Nat. pr.*, 24, 1).

Segmentación del texto

Citando a Domingo Baguer 2013, 144, a ojos de un lector moderno, un texto antiguo «más que para ser leído, estaba pensado para ser descifrado»¹³⁰. Las divisiones de capítulos mediante números fueron producto de los modernos editores mucho tiempo después, posiblemente como consecuencia de la pérdida de otros medios de caracterizar las diferentes secciones mediante ilustraciones o cambios de color (como ocurría en los manuscritos medievales), que trajo consigo la imprenta (*cf.* Hadas 1954, 13–14; Reynolds & Wilson 1991, 4). La extensión de la lectura silenciosa probablemente hizo más confusos los sistemas de referencia basados en la lectura en voz alta y la memoria, e impulsó el desarrollo de un nuevo sistema de referencias visuales, apoyado en la posición del texto dentro de la página y en el uso de espacios en blanco y signos especiales¹³¹. Al principio representó un intento de expresar gráficamente diferentes valores contextuales del texto (en este sentido podríamos hablar de un uso *pragmático* de la escritura), tal y como se hacía con los colores en la escritura amanuense, pero con el tiempo fue creando su propia idiosincrasia¹³².

¹²⁹ *Cf.* Hdt. 1, 1, «Ἡροδότου Ἀλικαρνησέος ἱστορίας ἀπόδεξις ἦδε»; Th. 1, 1, «Θουκυδίδης Ἀθηναῖος ξυνέγραψε τὸν πόλεμον». Plinio el Joven (*cf.* Plin. *epist.* 5, 6, 42) muestra la preocupación por la impresión que produce el título de una obra; en este pasaje reconoce el uso extendido de los títulos, y encomienda al autor que lea el título (*titulus*) de su obra a menudo, y que se pregunte qué quiere decir con ello y su relación con el tema (*materia*).

¹³⁰ Como observa Small 1997, 13, en el origen del término moderno *párrafo* está la *παράγραφος* (*γραμμή*), una marca creada no para ser útil al lector, sino para que el escriba pudiera calcular cuánto había escrito y, por lo tanto, su paga.

¹³¹ Sobre los espacios en blanco remitimos a Luque Moreno 2006, 271, así como a Saenger 1997, 125. También es interesante sobre el empleo de otros recursos distintos a los signos especiales para clarificar la lectura Barbier 2005, 17-37, 117-131.

¹³² Por ejemplo, las tipografías tienen una importancia tremenda en toda la maquinaria publicitaria actual. Remitimos a Connors 1993 para un estudio de las modificaciones en el mensaje ocasionadas por la elección de un papel o de una tipografía determinados.

Las tablas de contenido o índices, aunque existían en la Antigüedad, eran algo raro (cf. Small 1997, 15-17). La última sección de la introducción de la *Naturalis Historia* de Plinio constituye una especie de *guía de uso* de un índice, prueba de su infrecuencia¹³³. Llamar a un índice como el que figura en Plinio *tablas de contenido* puede ser en cierta medida un anacronismo, puesto que carece de la característica sintaxis nominal que los suele caracterizar; sin embargo, coincide con los actuales índices en que su único propósito es ofrecer una lista de contenidos. Mucho más no podían ofrecer los índices antiguos, en ausencia de un método para remitir al lector a una sección concreta de la obra. Aulo Gelio mejorará la idea de Plinio, ofreciendo algo similar a resúmenes del contenido de cada capítulo (cf. Gell. pr. 4-10).

Signos de puntuación y acentuación

La invención de los signos de puntuación y acentuación se atribuye tradicionalmente a Aristófanes de Bizancio, pero no parece que debamos darle plena credibilidad a esta noticia (cf. Luque Moreno 2006, 255-256). Aunque ya se percibiera su utilidad desde antiguo, no se vuelve habitual su uso en libros hasta el comienzo del siglo X (cf. Small 1997, 19). Los primeros filólogos alejandrinos emplean algunas indicaciones de este estilo en sus ediciones, por su utilidad para facilitar la lectura de la *scriptio continua* y dar una solución al significado de determinados lexemas dudosos. Una función semejante tuvieron los espíritus, aunque su uso no se generalizó hasta la época bizantina. Esta es la situación que vemos en los papiros, donde los acentos y los espíritus, o bien no aparecen, o bien lo hacen de forma diferente a la norma actual (cf. Morocho Gayo 2004, 20)¹³⁴.

¹³³ Cf. Plin. *Nat.* pr. 33 y *Nat* 1, 33a, donde se puede leer el índice en cuestión.

¹³⁴ Morocho Gayo 2004, 27, observa que en ocasiones la acentuación se coloca sobre la sílaba que no necesita la marca, como ocurre en el papiro Hawara de la *Ilíada*, donde se acentúa θεοείδης en lugar de θεοειδής. También llama la atención, en época posterior a la de Aristófanes, ver el acento puesto sobre la primera vocal del diptongo, y no en la segunda, según la norma actual. Otro ejemplo de marca ortográfica de aparición irregular es la iota adscrita, que servía para distinguir la forma de ciertas palabras o casos.

II. METODOLOGÍA

En la primera parte de esta tesis hemos fijado el propósito de nuestra investigación. Tras definir las unidades con las que vamos a operar, *datos e información*, así como el objeto de estudio donde se manifiestan, el *texto*¹, procede en esta segunda parte abordar la metodología que ha guiado nuestro análisis, junto con las herramientas concretas utilizadas.

Si el campo semántico de la información es amplio (*cf. supra I.2. Estructura del significado: datos e información*) tanto o más es el abanico de soluciones que se han propuesto para el problema de su efectiva localización o identificación en textos. Nuestro estudio de la evolución del objeto textual y en especial de su tratamiento a lo largo de los siglos de tradición filológica nos ha llevado a la convicción de que, lejos de la primera impresión que puedan arrojar las innovaciones tecnológicas actuales, la gestión de la información es un problema muy antiguo. Del mismo modo, las respuestas a este problema se remontan muy atrás en el tiempo.

Desde el reconocimiento de nuestras limitaciones, lo que ofrecemos a continuación es una clasificación genérica de los distintos métodos de *RI* a los que hemos tenido acceso. Por *genérica* queremos decir que se halla vinculada al uso de un tipo de herramientas u otras. Seguimos con ello nuestra tendencia a incluir la tecnología como una extensión de procedimientos e ideas ya existentes.

En un primer apartado, titulado *Búsquedas de información*, establecemos finalmente la conexión entre la *información*, objetivo de los *SRI*, y el adjetivo *semántico* con el que describimos el tipo de búsqueda que está en el origen de esta investigación. Desarrollamos igualmente los criterios sobre los que descansa nuestra propuesta de clasificación de los diferentes métodos tratados. Establecemos una línea divisoria general entre métodos de búsqueda «directa» y «mediada», según el modo en el que se enfrenta el usuario con el texto, prestando una mayor atención al segundo grupo, dada la naturaleza de la propuesta de aplicación práctica que hacemos, mediante el uso del ordenador.

Algunos manuales y monografías consultadas sobre *SRI* se centran demasiado en el uso concreto y específico de cierto tipo de herramientas, especialmente informáticas, y no

¹ El excursus sobre la naturaleza del texto (*cf. supra I.3.1. Definición de texto*) cobra aquí pleno sentido, puesto que la clasificación de los métodos de búsqueda que se expone a continuación se asienta sobre las ideas en él expuestas. Debe, pues, entenderse como un requisito indispensable para la argumentación que sostiene nuestro particular empleo de métodos desarrollados en diversas disciplinas.

en los principios generales que motivaron la creación de dichas herramientas. Nuestro análisis sigue una línea más teórica, aunque sin perder de vista el contexto real y material de su aplicación práctica. Seguimos en esto los pasos marcados por los documentalistas movidos por el deseo de fortalecer la integración teórica de las herramientas informáticas en el ejercicio filológico. El manual editado por López Yepes 1996 ha sido uno de nuestros principales puntos de referencia al respecto.

A modo de enlace entre los presupuestos teóricos y el modelo de aplicación práctica que proponemos, hemos dedicado un espacio para exponer ciertos detalles técnicos sobre el funcionamiento del ordenador. El rango de sus aplicaciones a la investigación ha crecido sin descanso en las últimas décadas. Por desgracia, la complejidad cada vez mayor de los mecanismos que hacen posible estos logros dificulta una percepción clara de la esencia básica de este instrumento, que, en lo fundamental, no ha cambiado desde su invención a mediados del siglo XX. Conocer con precisión los límites de sus posibilidades nos ha ayudado a establecer con firmeza los límites de nuestro uso del mismo.

II.1. BÚSQUEDAS DE INFORMACIÓN Y BÚSQUEDAS SEMÁNTICAS

El concepto de «búsqueda semántica» con el que titulamos esta tesis está basado en el concepto de «recuperación de información», habitual en las ciencias de la documentación y en informática. Entendemos por «búsqueda semántica»:

1. por un lado, el proceso por el que se adquiere información (*cf. supra, I.2. Estructura del significado: datos e información*) a partir de una serie de datos;
2. por otro lado, la expresión *búsqueda semántica* puede referirse también a una metodología concreta que permita extraer, recuperar o *crear* información a partir de dichos datos.

Nuestra descripción de *búsqueda semántica* se basa en el presupuesto de que la expresión es un mero trasunto de *SRI* (*cf. supra, I.1. Búsqueda semántica y Sistemas de Recuperación de Información [SRI]*). La caracterización de un tipo de búsquedas mediante el adjetivo «semántico» es un influjo de la expresión *web semántica*, que Berners-Lee, Hendler & Lassila 2001 definen como «a new form of web content that is meaningful to computers». Más que un concepto, se trata de toda una línea de investigación dentro de las tecnologías de la información, cuyo propósito es desarrollar el proceso de *RI* por ordenador para que ofrezca resultados más relevantes para el usuario humano. A día de hoy las soluciones a este planteamiento son fundamentalmente dos: desarrollar modelos de procesamiento de datos por ordenador que se asemejen a la forma en la que el ser humano procesa la información, y, de manera paralela, convertir la *información* obtenida por el ser humano en *datos* procesables por un ordenador².

A lo largo de este capítulo trataremos las diferentes formas posibles que puede adoptar dicha metodología. La clasificación de los métodos de búsqueda que aquí presentamos está basada en nuestra experiencia con varios tipos de *SRI* y en las conclusiones obtenidas tras el análisis de la naturaleza tanto del objeto textual como de los medios tradicionales de utilización del mismo como vehículo de información.

² Un ejemplo de adaptación del funcionamiento del ordenador al pensamiento humano son las *ontologías*, que analizamos más adelante (*cf. infra II.3.2. Ontologías*). La conversión de información textual o de otro tipo a un formato manejable por ordenador se basa en el trabajo con *metadatos* (*cf. Gartner 2016, 1-13*), cuyo ejemplo más importante son los llamados *lenguajes descriptivos* (*cf. infra II.2.4. Lenguajes descriptivos*).

Para fijar unos criterios que guiarán nuestra clasificación hemos partido de la acepción (*a*), estableciendo generalizaciones a partir de los ejemplos tratados. Hemos tenido en cuenta también diferentes puntos de partida teóricos y el tipo de resultados que ofrecen³. Estos pueden adoptar una forma compleja o bien presentarse como una visualización o representación alternativa de datos básicos y cuantitativos.

No todos los métodos descritos a continuación representan lo que hemos denominado «búsqueda semántica». Medimos la validez de un método de *RI* como sistema de búsqueda semántica por su capacidad para generar directamente en el usuario la información deseada, o al menos ayudar a ello de manera más directa y sencilla que el texto o textos originales sobre los que se aplique la búsqueda.

II.1.1. CRITERIOS DE CLASIFICACIÓN

Hemos abordado la tarea de la clasificación a partir de dos criterios básicos. Por un lado, hemos tenido en cuenta quién es el sujeto que realiza el procesamiento de los datos, y, por otro lado, el modo en el que se aborda el continente de dichos datos, el texto. A la hora de profundizar en esta división hemos utilizado como criterios el método concreto de manipulación de los datos, el tipo de resultado obtenido y el objetivo perseguido. En primer lugar hemos establecido una división global atendiendo al modo en el que se opera con el texto. Si la búsqueda es una operación que se aplica directamente sobre el texto, hablamos de «búsquedas directas». Si, por el contrario, el texto es abordado a través de alguna forma de modificación del mismo, ya sea producto del esfuerzo del propio ejecutor de la búsqueda o aprovechando el trabajo realizado por otros, hablamos de «búsquedas mediadas». Aunque el enfoque cuantitativo nos lleva a prestar más atención a técnicas que requieren el uso del ordenador, hemos intentado no caer en un excesivo apego a lo tecnológico, ofreciendo un punto de vista global e independiente. Evitamos así aquella tendencia denunciada por Rodríguez Adrados 1977, 213, a ignorar la investigación tradicional por parte de los precursores de investigación con ordenadores en filología.

³ Con respecto a la teoría, la ciencia de la Documentación es naturalmente la que ha mostrado más interés por el tema, desde su fundación: *cf.* Otlet 1934, 372-387. Ya en esta obra fundacional se aprecia un deseo por proporcionar un fundamento teórico a una actividad eminentemente práctica; en este sentido, el manual editado por López Yepes 1996 mantiene un buen equilibrio entre los enfoques teórico y pragmático, aunque el contraste entre el título elegido por Moya Anegrón 1996 para su contribución, «Sistemas de *RI* documental» y el hecho de que solo mencione las bases de datos informatizadas como ejemplo indica que aún queda trabajo por hacer.

Nos hemos centrado en los principios rectores de cada técnica, independientemente de la herramienta concreta utilizada. Esto nos ha permitido afianzarnos en nuestra idea de que las herramientas digitales no representan tanto una novedad como una extensión y expansión de métodos que ya existían. Tomamos aquí algunas de las ideas apuntadas en el capítulo sobre el objeto textual (*cf. supra, I.3. Material de estudio: el texto*) y las reinterpretemos como métodos de *RI* desarrollados teniendo en cuenta las características del soporte material usado en cada momento.

Dicho esto, nuestra clasificación de los métodos de búsqueda no es excluyente; es decir, la aplicación de uno deja abierta la posibilidad para el uso de otros. En líneas generales puede apreciarse un cierto desarrollo evolutivo, en el que métodos más complejos se construyen sobre técnicas más simples. Así, por ejemplo, las búsquedas mediadas requieren en determinadas partes del proceso el uso de técnicas de búsqueda directa (lectura comprensiva o búsqueda de coincidencias). En particular, las tres técnicas recogidas bajo la rúbrica de «búsqueda mediada» son esencialmente complementarias: un resumen es al mismo tiempo una forma de reducción y adaptación de información; las ontologías, en las que haremos especial hincapié como método elegido para su aplicación al texto de Heródoto, representan un método de búsqueda semántica que combina la reducción, la recreación y el establecimiento de relaciones.

La división que ofrecemos aquí, por lo tanto, aunque tiene en cuenta el uso de herramientas para el procesamiento de los datos, no implica un tipo concreto de instrumento, sino que es un modelo de su funcionamiento. Mientras cumpla las condiciones con respecto al modo de operar y a su capacidad de ofrecer el resultado descrito, resulta indiferente el empleo de una herramienta digital o analógica.

II.1.2. BÚSQUEDA *DIRECTA*

El método más sencillo y natural de afrontar el problema de la búsqueda de información es el tratamiento de los datos que componen el texto sin más intermedios que el texto y el propio intelecto, de ahí la calificación de este tipo de métodos de *RI* como «directos». Tratamos este tipo de técnicas en primer lugar, puesto que representan el nivel más básico de búsqueda semántica.

II.1.2.1. La lectura

Leer constituye, en esencia, un acto de búsqueda y recuperación de información (cf. Pinto Molina 1996, 318-319). Leer implica un proceso de decodificación de un sistema de signos para obtener unos datos, de cuya puesta en relación surge la información⁴.

Chartier 2004, 139-140, ofrece una serie de ejemplos literarios sobre la percepción del aumento de obras publicadas que trajo la imprenta y nos hace pensar que el objeto textual no es depositario *per se* del conocimiento, sino que es necesaria la intervención correcta del lector. Sin embargo, ninguno de estos ejemplos entra en la otra cuestión que nos preocupa, la de si el conocimiento existe en el texto, y por lo tanto la lectura consiste en un acto de *descubrimiento*, de *decodificación*, o, por el contrario, el conocimiento se produce en el lector tras la lectura atenta.

Existen más de diez verbos para indicar la noción de nuestro verbo «leer» en griego antiguo, atestiguados desde el siglo V a.C. (cf. Svenbro 2004, 71). Son el instrumento principal para conocer la percepción íntima de lo que significaba enfrentarse a un texto escrito en la Antigüedad. Su abundancia y frecuentes cambios muestran una realidad lectora estrechamente unida a la oralidad vigente⁵.

Dentro de este vocabulario, la combinación frecuente del prefijo *ανα-* hace pensar en la lectura como un acto dificultoso de desciframiento, como sugiere Chantraine 1950, 115. Es el caso del verbo *ἀνανέμειν*, que además parece tener una fuerte relación con la *vocalización* de los signos gráficos, cuando para Simónides indica «citar» un dicho (cf. Simon., fr. 37, 11-12), o en Heródoto se emplea para «recitar» (Hdt. 1, 173). Pero si hay un candidato para el verbo «leer» en griego, de modo general, es *ἀναγιγνώσκω*. Está relacionado con un acto de *desciframiento* y más aún, de *reconocimiento* del texto. Leer produce un efecto de aparición de una información en la mente similar al que, en el drama, producía la realización de las circunstancias (los *datos* diríamos nosotros) que se han ido desvelando a lo largo de la obra; cuando el personaje combina o procesa finalmente todos esos datos, alcanza un estado de revelación o *ἀνάγνωσις* (cf. Svenbro 2004, 77-78).

⁴ Van Dijk 1978, 178, ilustra la complejidad del aparentemente simple acto de la lectura, analizándola como un proceso de análisis progresivo que puede dividirse en cuatro pasos (cf. Kintsch & Van Dijk 1978).

⁵ Cf. *supra*, I.2.3.4 *Impacto de la escritura en el concepto antiguo de memoria* y I.2.3.6. *Las huellas de la μνήμη en el texto*.

La lectura silenciosa es una *rara avis* en la Antigüedad. Uno de los primeros en romper una lanza a favor de la existencia de este modo de lectura entre los griegos fue Knox 1958, quien desveló las primeras evidencias textuales de que algunas personas eran capaces de leer en silencio en la antigua Grecia (1958, 432-435). Unos pocos verbos parecen indicar una relación no-oral con el texto: ἀνηλίσσειν (X. *Mem.* 1, 6, 14), διηξιέναι (Aesop. 276) o incluso ἐντυγχάνειν y συγγίγνεσθαι⁶.

Si consideramos la lectura como una técnica de *RI*, podemos establecer algunas distinciones de acuerdo con el modo en el que se aborda el texto. Como hemos desarrollado en el capítulo sobre el texto como fuente de información (*cf. supra, I.3. Material de estudio: el texto*), el texto utiliza diferentes códigos para representar y almacenar datos; en particular, el texto escrito ha desarrollado un modo específico de codificar datos mediante elementos visuales. Sin embargo, este código puede utilizarse en mayor o menor medida, dependiendo de si la lectura se realiza en voz alta o no. Existe una conexión estrecha en la forma de leer y la forma del objeto textual, sin importar cómo hayamos definido exactamente su significado. Así, en Grecia, la *scriptio continua* favoreció una lectura en voz alta, reduciendo el papel del texto al de mera guía sobre la que aplicar toda una serie de recursos prosódicos no recogidos en el mismo (*cf. Cavallo 2004, 121*). La aparición del códice, por otra parte, marca el inicio de una nueva forma de lectura donde la distribución de los elementos en el espacio de la página, que hemos llamado «código visual», cobra cada vez más importancia (*cf. Chartier 2004, 141*).

Leer se convierte en una actividad cada vez más relacionada con la vista, y naturalmente el acto de comprensión implicado en la lectura se asocia con más fuerza al establecimiento de relaciones entre elementos gráficos (*cf. Svenbro 2004, 91; Chartier 2004, 141*). Esta tendencia tiene como punto final la noción de *hipertexto*, que lleva el empleo del medio visual hasta el extremo y donde la exposición lineal es sustituida por una de tipo tabular⁷.

Aquellos elementos visuales que expresen datos útiles para la comprensión del texto una vez leído (signos de acentuación y de puntuación) son más propensos a aparecer en textos destinados a la lectura en voz alta que aquellos cuyo contenido no puede expresarse a

⁶ Es Chantraine 1950, 122-126 y 118, quien los incluye en esta área semántica.

⁷ *Cf. Vandendorpe 1999, 49-69. Chartier 2004, 144*, habla incluso de un cambio en la forma de argumentación provocado por la fragmentación en apariencia inherente a este formato de texto. Nuestra opinión al respecto la expresamos más adelante en el capítulo sobre el texto digital (*cf. infra II.2.2. El texto digital*).

través de la voz (separaciones de párrafos, cambios de formato, delimitación de un espacio para las notas). Por otra parte, un modo de lectura carente de representación auditiva se apoyará con más fuerza en el código visual para desentrañar los datos contenidos en el texto.

El desarrollo de formas de codificación de datos en el texto, desde la carencia de signos auxiliares de una lápida griega en *scriptio continua* hasta el complejo sistema de relaciones visual-espaciales utilizadas en las páginas de los libros actuales puede, pues, ser interpretado no sólo como el detonante de cambios en los métodos de *RI* aplicados al texto, sino más bien como una consecuencia de los mismos.

Dickey 2007, 123, analiza las formas de codificar diferente tipos de datos en los escolios griegos y muestra cómo la filología griega antigua ya tanteaba el uso de elementos gráficos para dividir áreas de interés diferentes en el texto, como por ejemplo la separación entre palabras, sintagmas o caracteres. No obstante, la mayoría de las indicaciones que en un texto moderno se realizan mediante signos especiales o cambios de formato, como cursiva o negrita, en griego antiguo se codificaban mediante una sintaxis específica o por la presencia de determinadas palabras⁸. El orden es otro instrumento importante de codificación de información, aún sin estar reforzado por el formato de lista al que estamos acostumbrados en la actualidad (*cf.* Dickey 2007, 113-114).

Si consideramos el mismo proceso desde el presente hacia el pasado, tener en cuenta el modo en el que se accedía a la información en el momento de composición de la obra puede resultar muy útil a la hora de establecer los elementos y rasgos más relevantes para diseñar un plan de búsqueda semántica. Un texto compuesto para ser leído en voz alta, como tradicionalmente se considera la obra de Heródoto, naturalmente mostrará su estructura a través de rasgos léxicos y sintácticos relacionados con la expresión oral⁹.

II.1.2.2. Búsqueda de coincidencias

Cuando la fuente de información es un texto impreso, la lectura es el único método directo de *RI* disponible. La aparición del texto digital abre la puerta a un segundo método,

⁸ Dickey 2007, 111 y 113, aduce como ejemplos de esta práctica el empleo del dativo o de preposición para señalar citas.

⁹ Un *SRI* diseñado para un texto compuesto de forma oral debe tener en cuenta los elementos léxicos y sintácticos propios de este tipo de textos (*cf.* Dickey 2007, 107-123).

en la línea de la lectura parcial: la búsqueda de coincidencias. Es una técnica que forma parte del proceso de lectura comprensiva descrito arriba. Requiere que el usuario establezca previamente una secuencia de caracteres con un determinado valor significativo asociado a ella; a continuación se procede a identificar esa misma secuencia en el texto original. Esta técnica es la base del funcionamiento de toda obra lexicográfica (léxicos, concordancias y diccionarios) y de cualquier tipo de índice en general¹⁰. Dejamos el análisis de estas herramientas para más adelante (*cf. infra II.1.3.1. Reducción*), y nos centramos aquí en su empleo directo sobre el texto original.

La búsqueda manual de coincidencias es un proceso largo y laborioso, pero sigue siendo el único recurso en ausencia de otro tipo de material auxiliar. Por ejemplo, cuando no existe ningún tipo de obra lexicográfica que recoja el uso de un término o términos en una determinada obra, debemos acudir a una edición del texto, y si esta carece de un índice de referencia, revisar página a página hasta encontrar la coincidencia deseada. Buena parte del proceso de creación de una obra lexicográfica requiere el ejercicio puro y duro de la lectura intensiva de los textos¹¹.

Naturalmente las herramientas informáticas han supuesto un alivio importante para la realización de esta tarea. Desde que Rodríguez Adrados 1977, 212) y, en su misma línea, Rodríguez Somolinos & Álvarez 1991 y Ochoa 1992 reconocieran las posibilidades que ofrecía el ordenador como ayuda, su empleo ha superado algunas de las limitaciones que en aquel momento se consideraban insalvables. Tal es el caso de la automatización de ciertas tareas como la indización o la clasificación de textos, consideradas una actividad netamente atribuida al investigador o filólogo, pero que la estadística y la tecnología han permitido asignar, al menos parcialmente, también al ordenador¹².

¹⁰ Luque Moreno 1987, es un ejemplo de los resultados obtenidos mediante este método.

¹¹ La descripción de la actividad lexicográfica tal como la exponen Lara 1997, 43, y López Facal 1977, 156-157, es ilustrativa al respecto.

¹² La automatización ha sido especialmente útil en áreas como la lexicografía. Una muestra de ello es el programa *Lexis* creado por Goulet 1990 con el objetivo de agilizar la indización de textos y la posterior elaboración de léxicos. Sobre su funcionamiento y ventajas remitimos a la reseña de Fuentes González 1993. La existencia de un número creciente de textos antiguos en un formato digitalizado ha reducido considerablemente el número de casos en los que no resulta posible beneficiarse de las capacidades del ordenador (*cf. Hockey 1994; Renear 2004 y Hunt, Lundberg & Zuckerman 2010*). Una función especialmente útil para la búsqueda de coincidencias es la que proporcionan las llamadas «expresiones regulares» (*cf. infra II.2.3. Expresiones regulares*).

II.1.3. BÚSQUEDA *MEDIADA*

Una búsqueda mediada es, en oposición a las búsquedas directas, un método de localización y recuperación de información que implica el uso o la creación de un elemento intermedio. Este elemento es producto de una serie de procesos, cuya finalidad es obtener un objeto, textual o no, independiente del texto original, sobre el que resulte más fácil obtener información a través de una consulta directa.

La clasificación de los métodos de creación de este elemento intermedio se basa sobre todo en nuestro análisis particular de diferentes formas de representación de los datos contenidos en el texto. Tomamos el concepto de «mediación» de la teoría de la Documentación (cf. López Yepes 1996, 38). Sin embargo, como ya dejamos claro en el capítulo dedicado al objeto textual (cf. *supra* I.3.3.2.a. *Funciones de la escritura*), no compartimos la definición de «documento» como un ente distinto del texto. Esta distinción no hace sino complicar innecesariamente la tarea del análisis del significado, producto de una determinada combinación de los datos presentes en el texto¹³.

Nuestro análisis revela tres procesos diferentes involucrados en una búsqueda mediada, según el efecto que tienen sobre los datos del texto o el uso que hacen de los mismos. En primer lugar, ciertas técnicas limitan el volumen de datos sujetos a revisión detallada por parte del usuario, por lo que a este lo hemos denominado *método de reducción*. Otras técnicas extraen datos del texto y los representan de formas diferentes, dando lugar a nuevos objetos significativos. Hemos llamado a este método *reconstrucción*. En el último método, la *relación* es común a todas las técnicas de *RI* tratadas, y consiste en el establecimiento de conexiones entre los datos, o en el refuerzo de otras ya marcadas en el texto original.

II.1.3.1. Reducción

Este método se basa en el presupuesto de que el texto es un conjunto de datos de diferente clase y en diferente cantidad. De todo este volumen de datos, cierta cantidad no haría sino generar *ruido informativo* y por lo tanto no es relevante para generar información. En consecuencia, una selección previa de los datos en función de sus características, o

¹³ Las bases de datos en las que se han convertido las bibliotecas o repositorios digitales fundamentan su eficacia en que almacenan *metadatos*, es decir, *datos sobre datos* (cf. Montejo-Ráez *et al.* 2005, 119).

de las características de la información deseada, representa una forma eficaz de disminuir el esfuerzo de procesamiento de una consulta directa del texto¹⁴.

Hay dos formas de reducir el volumen, dependiendo del efecto que causen sobre el texto fuente. A imagen de una operación quirúrgica, hemos dividido las aplicaciones del método de reducción en invasivas y no invasivas¹⁵. Las primeras causan un cambio en el propio texto original, o bien crean un nuevo texto que refleja dicho cambio. En estos casos la reducción de datos es literal y física, puesto que se produce una pérdida del texto que los contiene. Una forma de reducción no invasiva, por el contrario, consiste en limitar los datos considerados en la búsqueda sin modificar el texto original. Ejemplos de formas no invasivas son la segmentación del texto y la agrupación de textos (*cf.* Pinto Molina 1996, 320-321). En el primer caso nos referimos a algo que ya comentamos en la introducción, a propósito de la composición visual del texto (*cf. supra* I.3.3.2. *Codificación visual*). La introducción de espacios entre palabras y de separaciones entre párrafos u otras «áreas de interés» dentro del texto es, desde el punto de vista que acabamos de exponer, una forma de *RI* basada en la *reducción*¹⁶.

II.1.3.1.a. Reducción no invasiva

Segmentación del texto

Habitualmente no se trabaja sobre un texto como si fuera un bloque uniforme. El análisis del texto es una actividad gradual que se aplica sobre unidades discretas dentro de él, tales como palabras, sintagmas, oraciones o unidades discursivas¹⁷. Sin embargo, desde

¹⁴ *Cf.* Van Dijk 1980, 213: «los hablantes no pueden, y no tienen por qué almacenar toda la información proposicional de un discurso dado en el proceso verbal. Por tanto, esta información será, al menos en parte, reducida a las macroestructuras» (la traducción es nuestra).

¹⁵ En realidad, para documentalistas como Pinto Molina 1996, 320, se trata de un proceso siempre negativo: «consiste en la eliminación de unidades de significación (frases y palabras) que son consideradas irrelevantes para el análisis».

¹⁶ *Cf.* Dickey 2007, 123, sobre el uso de los espacios en las obras lexicográficas antiguas para marcar el tema de la discusión o para indicar citas o algún otro segmento que necesitara ser resaltado. No olvidemos tampoco las implicaciones de este uso del elemento visual para la evolución de los modos de lectura, como señala Svenbro 2004, 92.

¹⁷ Ya hemos mencionado los recursos para la segmentación del texto en el medio impreso, así como el propósito y las consecuencias de su utilización, en varios apartados de esta tesis (*cf. supra* I.3.3.2. *Codifica-*

la perspectiva del procesamiento informático, un texto digital es una simple secuencia de caracteres¹⁸. Para que sea posible un procesamiento efectivo del texto, en primer lugar este debe ser dividido en unidades lingüísticas del tipo que acabamos de describir.

Este proceso recibe el nombre de *tokenización*, derivado del término *token* con el que se denomina en inglés a cada una de las agrupaciones de caracteres resultantes¹⁹. Se trata de una tarea relativamente fácil en el caso de lenguas que presentan una segmentación previa, mediante el uso de espacios en blanco o de signos de puntuación, como es el caso del castellano, o en el caso de las ediciones actuales de textos antiguos, por ejemplo. En el texto digital también son utilizados los espacios vacíos o los signos de puntuación como delimitadores por defecto para las unidades lingüísticas²⁰.

Naturalmente surgen problemas cuando los segmentos objeto de la segmentación no se encuentran bien delimitados mediante este tipo de caracteres. Es el caso de un texto con una puntuación poco clara o inconsistente, de palabras unidas por un guión, de la presencia de términos clíticos o de términos abreviados mediante un apóstrofo. Si se trata de un texto en griego antiguo podemos añadir la *coronis* a esta breve lista de inoportunos enemigos de la *tokenización*. Habitualmente este tipo de dificultades se extienden también a la segmentación de oraciones en textos modernos, ya que un mismo signo, el punto, coincide como marca de abreviatura y de final de frase. Por suerte, en este caso el texto griego resulta más fácil de segmentar, puesto que no se da tal coincidencia de uso²¹.

ción visual y II.1.2.1. La lectura). Para nuestra exposición de la aplicación de estos medios en el medio digital seguimos fundamentalmente lo expuesto por Mikheev 2003.

¹⁸ Todos los caracteres no son *informativamente* iguales; como dice Mikheev 2003, 201, es posible establecer clases entre los mismos, según el tipo de datos que representan, según sean signos gráficos, como letras, números o signos de puntuación (llamados «caracteres de contenido»), o bien marcas que indiquen a un determinado software la operación que debe realizar (caracteres de control y de formato).

¹⁹ La traducción castellana del término *token* sería *ítem*, como aparece recogido en el DRAE (s.v. «ítem», 4). Por desgracia no contamos con el verbo derivado correspondiente *itemizar**, ni con un sustantivo *itemización**. Contraviniendo pues el buen gusto, y en aras de la coherencia, nos permitimos el uso de este neologismo.

²⁰ Los procedimientos informáticos de *tokenización* se basan por regla general en el supuesto de que todas las palabras están separadas por alguno de los delimitadores descritos (cf. Mikheev 2003, 202).

²¹ Sobre la segmentación de oraciones y palabras en griego antiguo, remitimos a Luque Moreno 2006, 231-246 para los textos en verso y 2006, 255-262 para un panorama general sobre la cuestión.

La *tokenización* que basa la separación de palabras o de oraciones en marcas gráficas es un tipo de segmentación del texto de «bajo nivel» (cf. Mikheev 2003, 202). A pesar de su carácter aparentemente simple y mecánico, es una fase fundamental para el éxito de cualquier procesamiento complejo que deseemos aplicar con posterioridad y requiere a menudo de una revisión delicada e incluso fatigosa (cf. *infra* II.2.3. *Expresiones regulares*). Tal ha sido nuestro caso en el texto de Heródoto, donde el apóstrofo es usado con frecuencia, y por lo tanto es necesario revisar uno a uno los *tokens* obtenidos, completando la forma correspondiente de modo manual (cf. *infra* III.4.3.1. *Preparación del texto*).

Por otra parte, una segmentación del texto de «alto nivel» pretende localizar y aislar grupos lingüísticos, ya sea dentro de una oración, como sintagmas verbales o nominales, o bien en un plano *supra-oracional*. En este último caso hablamos de crear grupos de oraciones o párrafos, utilizando por ejemplo un criterio semántico, es decir, su relación entre sí como partes que expresan una misma idea. Estas agrupaciones de textos, referidas en la bibliografía al respecto como «teselas textuales» o *text tiles* es el objetivo de la técnica del *topic modeling* que hemos aplicado sobre el texto de las *Historias*²².

El trabajo de Fischer-Starcke 2010 sobre la obra de Jane Austen requirió de una segmentación del texto de alto nivel. La solución que propuso (2010, 144-146) fue observar el uso del léxico, basándose en la idea de que existe una correspondencia entre la homogeneidad léxica y la coherencia temática de dos o más fragmentos analizados. El resultado fue un estudio muy detallado de las relaciones entre las diferentes obras de un corpus (cf. Fischer-Starcke 2010, 176-183), diferente en algunos aspectos al estudio tradicional basado en el análisis directo de la obra, es decir, mediante lo que hemos denominado una «búsqueda directa».

La segmentación del texto puede responder a diferentes propósitos y el alcance de su aplicabilidad real depende mucho del estado de los datos de partida. Una razón para dividir el texto puede ser analizar en detalle la evolución de los temas tratados a través de su expresión léxico-lingüística, o centrarse en los puntos de transición de un tema a otro. No hay duda de que un conocimiento más profundo de la estructura del texto facilita la interpretación del mismo. ¿Es posible extraer alguna conclusión sobre la intención del autor o su pensamiento a través de la manifestación lingüística concreta de estas divisiones? Este es

²² Orasan 2000 ofrece una solución diferente para textos en inglés mediante un algoritmo basado en el llamado *aprendizaje-máquina*, combinado con un sistema de reglas, y que ofrece un grado de eficiencia prometedor. Se aprecia una tendencia a combinar diferentes métodos (cf. Sang *et al.* 2001), y la revisión humana en determinadas fases suele ser crucial.

un ejemplo del tipo de preguntas que hemos intentando responder al abordar el texto de *Heródoto*.

Volviendo a los detalles técnicos y prácticos del uso de esta técnica, la mayoría de los programas informáticos diseñados para segmentar un texto utilizan espacios en blanco para delimitarlos; es decir, insertan espacios en blanco. Esto puede resultar problemático si se quiere mantener la forma del texto original. Otro problema añadido es el caso de aquellos segmentos que contienen espacios en blanco en su interior. Una solución en estos casos es sustituir por guiones bajos este tipo de espacios en el interior de grupos de elementos que funcionan como una unidad. Naturalmente este tipo de remedios deben ser considerados como parches o remiendos en ausencia de una solución que elimine el problema de raíz. En este sentido la alternativa más «limpia» y eficiente es la utilización de un lenguaje descriptivo (*cf. infra II.2.4. Lenguajes descriptivos*) para anotar los límites de las unidades en cuestión. Así lo propone Mikheev 2003, 204, que destaca la posibilidad de incluir más información asociada a las marcas de separación, mediante los instrumentos dispuestos por este tipo de lenguajes.

Un ejemplo de las herramientas a las que se refiere Mikheev son las etiquetas utilizadas en el lenguaje XML. A continuación mostramos un ejemplo en el que las palabras están delimitadas por la secuencia de caracteres <w> </w>; como veremos (*cf. infra II.2.4. Extensible Markup Language*), el lenguaje XML recurre a abreviaturas y convenciones gráficas para describir el texto. En este caso *w* es la abreviatura de *word*, e indica que la secuencia de caracteres es una *palabra*; *c* se utiliza para especificar la *clase de palabra*, que a su vez se explicita mediante una abreviatura (*adv*, *adverbio*; *part*, *partícula*; *n*, *sustantivo*; *v*, *verbo*):

```
<w c=adv>οὐτιω</w> <w c=part>μὲν</w> <w c=n>Πέρσαι</w> <w  
c=v>λέγουσι</w>
```

El proceso de segmentación de un texto puede automatizarse a través de programas que definen las características de los límites entre segmentos y aplican sistemáticamente una serie de etiquetas como en el ejemplo anterior. Para ello se utilizan expresiones regulares, un tipo de lenguaje que describimos más adelante con el que es posible seleccionar caracteres o secuencias de caracteres mediante ciertos rasgos comunes. El éxito o fracaso de este tipo de sistemas radica en la capacidad de su diseñador para encontrar una defi-

nición lógica, bien de los segmentos, bien de la separación entre los mismos, que describa con exactitud todas sus apariciones²³.

En el caso de la segmentación de frases, seguimos la misma línea de similitud léxica empleada por Fischer-Starcke 2010, 145-146. En su caso escogió como base para la delimitación una serie de elementos léxicos que pueden funcionar como enlaces entre las partes, algo que denomina «cohesión léxica». La idea es que los patrones formados por la aparición conjunta de elementos lingüísticos (no necesariamente léxicos) forman una especie de hilo que da coherencia y une no sólo las palabras en sí, sino también aquellos lugares del texto donde aparecen.

Por desgracia, el problema general del enfoque *distribucional* es que un análisis simple de la frecuencia de aparición conjunta de dos términos léxicos cualesquiera arroja una cantidad de resultados sencillamente enorme. Planteada de este modo la segmentación del texto contraviene el espíritu de reducción del esfuerzo de análisis con el que hemos calificado esta técnica. Es necesaria, pues, alguna forma de restricción adicional.

Según Mikheev 2003, 214-125, la mayoría de los sistemas de segmentación automática utilizan la frecuencia de las palabras tal cual; dicho de otro modo, consideran únicamente la forma *lematizada*. Ante esto, todas las soluciones que se plantean consisten en alguna forma de agrupación previa de las palabras cuya frecuencia conjunta se quiere analizar; es decir, implican una reducción previa a la reducción. La división de las palabras por campos semánticos, aunque lógica a primera vista, requiere un grado de intervención del ser humano demasiado elevado para ser realmente práctica. El marcado sintáctico, por otra parte, se plantea como una solución muy prometedora. El número de categorías sintácticas no es muy elevado, y pueden establecerse ciertas reglas formales que permitan a un programa informático hacer un trabajo aceptable.

En el caso particular del griego, las esperanzas en el marcado sintáctico depositadas por Mikheev 2003, 215, se han plasmado en una plataforma *online* dentro de la iniciativa *Perseids*, llamada *Arethusa*, donde cualquiera con un mínimo conocimiento de griego y la voluntad necesaria puede contribuir a crear un corpus de textos con anotación sintáctica, morfológica e incluso una traducción. Aunque la iniciativa es admirable, no hace sino reconocer la limitada confianza que podemos depositar aún en la deseada «automatización». El

²³ Mikheev 2003, 206, menciona como posibles soluciones la incorporación de sistemas automáticos de decisión que permitan modificar la expresión regular empleada, o bien elegir entre varias expresiones predefinidas, según sea más conveniente.

camino hacia el éxito en este sentido pasa necesariamente por la colaboración de todos, pero, en nuestra opinión, requiere una dedicación y un esfuerzo superiores al beneficio que podemos obtener²⁴.

Agrupamiento, clasificación y catalogación

Higbie 2010 considera que la división en libros de obras clásicas como la de Homero o Heródoto, herencia de la filología alejandrina, es una forma de facilitar la gestión de información y en definitiva el trabajo con estos textos. Describe el problema que causaba la falta de un estándar de referencia para citar las obras y, al mismo tiempo, observa que los autores antiguos reconocen la existencia de unidades menores dentro de una obra, de naturaleza temática (*cf.* Higbie 2010, 3-5; 6).

Señala igualmente que muchas obras fueron objeto de varias divisiones (*cf.* 2010, 14-16), aunque en algunos casos tenemos constancia de que el autor había establecido una clasificación propia desde el momento de su composición, como un recurso útil para la redacción y exposición de un material extenso (caso de historiadores como Éforo o Polibio). En líneas generales, un tipo de división formal, como es la utilizada hoy en día, responde a una forma de recepción y creación del texto escrita y, por lo tanto, centrada en una expresión lingüística cuantificable. La memoria y la oralidad predominantes en la Antigüedad favorecen una división del texto en unidades temáticas relativamente abiertas a la *variatio* lingüística, desde el punto de vista no sólo del autor que organiza la exposición de sus propias ideas, sino de aquellos interesados por su obra tiempo después, y que utilizan el recuerdo de las ideas expresadas como índice y referencia.

Otra forma de reducción de información es la agrupación de diferentes textos representada por cualquier tipo de antología o selección de fragmentos. Palmer 2004 realiza un estudio sobre el género de las recopilaciones de textos siguiendo una orientación semántica. Considera que las colecciones temáticas son una tendencia fuerte dentro de las Humanidades Digitales, aunque siguen la línea trazada de la actividad documental tradicio-

²⁴ Köntges 2016, creador del programa *ToPan* (*cf.* Köntges 2016b) promueve el uso de archivos XML con etiquetado sintáctico para un uso más productivo del *software*. No indica, sin embargo, la extrema escasez de este material. A fecha de 2017, en *The Ancient Greek and Latin Dependency Treebank* (<https://perseusdl.github.io/treebank_data/>) la selección de textos antiguos con un etiquetado sintáctico completo es bastante reducida: algunas tragedias de Esquilo y Sófocles, *Odisea* e *Ilíada*, la obra de Hesíodo y algunos discursos de Lisias; de Heródoto lamentablemente solo está disponible el libro primero de las *Historias*, lo que nos ha impedido aprovechar este material en nuestro análisis.

nalmente realizada por las bibliotecas²⁵. En este sentido, Palmer 2004, 363, aprecia un deseo por superar su tradicional carácter auxiliar y crear obras que contribuyan de un modo más activo a la investigación.

II.1.3.1.b. Reducción invasiva

Con respecto a las técnicas invasivas, consideramos ejemplos habituales todos aquellos usos del texto cuyo propósito es eliminar partes del mismo. Consideramos que existen tres tipos de representaciones posibles que cumplen esta condición: las listas, en cualquiera de sus denominaciones (los diversos tipos de índices); la expresión simplificada de ciertas fórmulas o frases habituales (abreviaturas y notación bibliográfica)²⁶; y, por último, lo que denominamos *segmentación invasiva*, en referencia a las citas literales o las concordancias que ofrecen sólo un fragmento literal del texto original²⁷.

Abreviaturas

El corte de la forma lingüística de una palabra probablemente es el ejemplo más inmediato de la economía del esfuerzo en el acto comunicativo. Requiere de la colaboración del receptor, que debe ser capaz de adelantar la forma completa del elemento abreviado para poder decodificar correctamente el mensaje. Por esa razón suelen ser frecuentes en contextos que limitan el vocabulario o el número posible de interpretaciones, como es el caso de los lenguajes técnicos y profesionales. La filología no escapa a esta

²⁵ La selección de extractos o *excerpta* formaba parte del proceso de composición de ciertas obras en la Antigüedad. Sobre esta técnica en general remitimos al artículo de Van den Hoek 1996, así como a Dorandi 2016 y Reydam-Schils 2011. Como observa Dorandi 2016, 46, el tema empezó a llamar la atención de filólogos a finales del siglo XIX, con estudios como el de Münzer 1897, 1-133, sobre Plinio el Viejo, o Skydsgaard 1968, 101-116, sobre Varrón; más recientemente podemos citar el trabajo de Mejer 1978 16-19, dedicado a Diógenes Laercio, o el de Astarita 1993, 26-31, que analiza el proceso de selección de extractos en Aulo Gelio.

²⁶ Un ejemplo muy cotidiano son las abreviaturas utilizadas en las ediciones críticas (*cf.* Maurer 2012; Wellington 2003; Bernabé Pajares & Hernández Muñoz 2010, 155-165). La notación bibliográfica está relacionada con los llamados «lenguajes documentales» utilizados en biblioteconomía (*cf.* Gil Urdiciáin 1996a, 1996b).

²⁷ Las citas constituyen un acto de selección de texto, que después es insertado en un contexto diferente, y como tal podrían considerarse una forma invasiva de reducción del mismo tipo que las comentadas en este apartado. Sobre la cuestión remitimos a Bernabé Pajares & Hernández Muñoz 2010, 166 y Finnegan 2011, 141-147.

norma, y hace un uso activo de las abreviaturas. Incluso una breve recopilación de aquellas empleadas habitualmente en un aparato crítico como la que ofrecen Bernabé Pajares & Hernández Muñoz 2010, 155-165, en calidad de anexo a su manual sobre edición crítica nos da una idea de uso extendidísimo de este recurso²⁸.

Las abreviaturas son un recurso en constante crecimiento. En el caso de las publicaciones científicas, como lamenta Wellington 2003, ix-x, la dificultad para garantizar la comprensión de las mismas es cada vez mayor, especialmente en situaciones de «homonimia» referencial, cuando los acrónimos de diferentes revistas coinciden. No podemos dejar de mencionar aquí la lista de abreviaturas de *L'Année Philologique*, gratamente actualizada cada año gracias al esfuerzo de sus redactores.

Palabras clave

La definición de *palabra clave* depende, evidentemente, de las cualidades que nos permitan definir algo como *clave*. Sin embargo, se trata de un adjetivo tan evidente, al menos en la superficie, que apenas se suele dar una explicación más seria del mismo, a pesar de las evidentes implicaciones que tiene para el establecimiento de unos criterios de selección²⁹.

Scott & Tribble 2006, 56-59, estudian las diferentes perspectivas al respecto, analizando diversos trabajos centrados en la detección y uso de palabras clave. La conclusión que extraemos de toda su exposición es que «palabra clave» es un concepto que trata de aislar la esencia de una idea compleja como es la de «relevancia», en una expresión léxica simple. Sin embargo, tal y como se deduce de la definición de unidades de trabajo que propusimos en la introducción (*cf. supra* I.2. *Estructura del significado: datos e información*), considerar algo como relevante o no es un acto perteneciente a la fase de interpretación de los datos. La relevancia es, pues, «información». La técnica de las palabras clave podría entenderse como recurrir a información para detectar información, una contradicción que debemos evitar.

²⁸ El uso crítico de abreviaturas es un tema complejo y con una larga tradición; sobre la cuestión remitimos al manual clásico de Gardthausen 1911, 1-22, así como a Allen 1889 y Chassant 1884.

²⁹ La definición de *palabra clave* que da Delcourt 1989, 285, es un ejemplo del tipo de definición al que nos referimos: «a key word with regard to a sub-corpus is a word of which the frequency in that sub-corpus is significantly higher than expected under the hypothesis that its use and the variable “part of the corpus” are mutually independent».

Para evitarla se hace necesario fijar una serie de rasgos cuantificables que permitan calificar y detectar posibles elementos clave. En primer lugar es necesario tener un rango de comparación, por lo que los datos deben distribuirse en conjuntos semejantes; por ejemplo, configurando una lista de todas las palabras presentes en dos o más textos (*cf.* Fischer-Starcke 2010, 65-66).

II.1.3.2. Reconstrucción

La delimitación de un área de trabajo o la simple reducción a unos elementos no es siempre suficientemente clarificadora. El objeto resultante de esta operación carece además de la fluidez y continuidad del texto original.

Un segundo método de *RI* que corrige estos defectos es la reconstrucción de la información del texto. Este método produce un objeto a partir de los datos obtenidos del texto original. La ventaja de esta aproximación es que se obtiene una reducción del *ruido informativo*, además de que el usuario posee una libertad mayor para utilizar los recursos de codificación del texto para expresar de un modo más explícito su interpretación de los datos, es decir, la información que ha percibido. El objeto resultante no tiene por qué ser un nuevo texto, sino una representación gráfica del mismo.

El producto de este método son objetos que no sólo contienen parte de los datos originalmente presentes en el texto, sino que el emisor o creador de este objeto nuevo añade otros nuevos. Habitualmente el propósito de esta modificación de la forma del texto es facilitar la percepción de aquellos datos considerados relevantes o interesantes (*cf.* Martínez Comeche 1996, 39, 43).

Mientras que el método de reducción se basa en liberar la carga de atención necesaria para comprender los datos del texto, a menudo las ideas complejas no son fácilmente divisibles en segmentos. Las ideas son producto de la combinación de varios datos, y sólo pueden expresarse al completo mediante una expresión que recoja los datos necesarios y facilite su puesta en relación. En este sentido, el método de reconstrucción ofrece como ventaja un objeto que contiene todos los datos relevantes y los relaciona explícitamente.

Sin embargo, el método de reconstrucción establece relaciones entre los datos de manera diferente al texto fuente. Estas relaciones pueden servir para reforzar la adquisición de las ideas originales o para crear información nueva (*cf.* Garrido Arilla 1996, 232).

II.1.3.2.a. Reconstrucción textual

La reconstrucción textual es aquella cuyo producto final es de naturaleza textual y, por lo tanto, utiliza los códigos propios del texto para su construcción (*Cf. supra I.3.3.2. Codificación visual*). Según la intención de la reconstrucción, establecemos dos tipos: los resúmenes (y métodos semejantes), que se centran en el texto original, y la reconstrucción *exegética*, que añade nuevos datos y relaciones con la intención de «explicar» o expandir la información del texto original.

De menor a mayor nivel de alteración de la forma original, los principales ejemplos de reconstrucción textual son:

1. El resumen.
2. Diferentes formas resultado de la exégesis del texto, como pueden ser desde la edición crítica o el comentario.
3. Las obras lexicográficas (diccionarios) y enciclopédicas representan el grado más extremo de reconstrucción, así como un producto mixto que combina rasgos resultantes de un método de reducción (el formato de lista, los índices), un uso importante del código visual (distribución de áreas de interés dentro del epígrafe de cada lema), así como representaciones visuales: diagramas, gráficos y mapas.

Resumen

La técnica del resumen se centra en preservar las ideas originales del texto y para conseguirlo limita las modificaciones de la forma que efectúa sobre la forma del texto original. Estos cambios puede ser de tipo reduccionista, como por ejemplo eliminar ciertas figuras de estilo que dificulten la comprensión; o bien producen una *variatio* con respecto al original a favor de una *lectio facillior*, alterando elementos léxicos o construcciones sintácticas. En la Antigüedad, la *epitomización* es una actividad necesaria, en cierta medida, por la dificultad que entrañaba la copia manual de libros, especialmente si se trataba de obras de longitud considerable³⁰.

³⁰ La extensa obra de Tito Livio es uno de los casos donde la transmisión indirecta a través de *epitomes* representa nuestra única forma de acceso a varias partes de su obra (*cf.* Oakley 2016). Podemos citar también el caso de las *diatribas* del moralista de inspiración cínica Teles (s. III a.C.), de las que sólo conservamos

El objeto resultante es similar a los objetos obtenidos mediante de un proceso de reducción, como las antologías y recopilaciones. No obstante, el resumen y sus variantes constituyen, estrictamente hablando, un texto *diferente* al original. El criterio seguido por aquel que realiza el resumen debe ser observado y analizado con la máxima atención a la hora de considerar qué es precisamente lo importante. De lo contrario se favorece la extensión de errores previos (*cf.* Bernabé Pajares & Hernández Muñoz 2010, 146. Es una actividad eminentemente personal, difícil de someter a un conjunto de normas (*cf.* Pinto Molina 1996, 322).

Edición crítica y comentario

Una edición es la acción de preparar un texto de acuerdo con los principios de la *ecdótica* y la filología, así como el objeto textual resultante de dicha acción (*cf.* Bernabé Pajares & Hernández Muñoz 2010, 169). Editar un texto es una actividad compleja, y puede parecer a veces que consiste en una «modificación de una modificación», como observan Bernabé Pajares & Hernández Muñoz 2010, 31-32.

En efecto, el propósito de una edición crítica es restituir la forma anterior a que se aplicara sobre el texto una reconstrucción. Aunque se disponga de argumentos para justificar las decisiones tomadas, en primer lugar, la edición no deja de consistir en aplicar ciertos cambios a la forma de un texto determinado; en segundo lugar, nunca se puede tener la total seguridad de que los cambios realizados correspondan a la forma original³¹.

Obras lexicográficas

Por último, las obras lexicográficas son el último ejemplo y quizá el más radical de reconstrucción textual. Los diccionarios, en efecto, son considerados de forma natural sistemas de recuperación de información, especialmente desde que la informática se ha convertido en una aliada omnipresente en el todas las fases de creación de una obra semejante³².

extractos en Estobeo sacados ya de una versión epitomizada por un tal Teodoro: *cf.* Fuentes González 1989, 184; 1998, 3-9; 2016.

³¹ La existencia de dicho texto original en la Antigüedad no deja de ser problemática. Sobre la cuestión de las variantes de autor, remitimos, por ejemplo, a Dorandi 2007, 83-85.

³² Remitimos al respecto a Woolridge 2004, 77; Rodríguez Adrados 1977, 2003 y Rodríguez Adrados & Rodríguez Somolinos 1994.

La estructura de un lema de diccionario representa una labor de siglos por aprovechar al máximo los recursos visuales y lingüísticos establecidos por el marco de la escritura y el libro. Combina casi todos los ejemplos de técnicas de *RI* mencionados (abreviaturas, citas, resumen, catalogación y agrupación de textos) y aprovecha los tres métodos básicos de manipulación de los datos del texto que hemos reseñado (reducción, reconstrucción y relación) para ofrecer un compendio de conocimiento en torno a cada palabra en el menor espacio posible.

II.1.3.2.b. Reconstrucción visual

Por reconstrucción visual entendemos aquella cuyo producto final es un objeto construido fundamentalmente sobre un código visual, como un mapa o un diagrama; pueden contener texto, pero la fuerza expresiva de este tipo de recursos radica en otro tipo de signos y códigos. En efecto, el conocimiento puede ser representado como imágenes mentales en nuestros pensamientos, como palabras habladas o escritas, en forma gráfica o en imágenes, como cadenas de caracteres o como colecciones de señales eléctricas o magnéticas dentro de un ordenador. La efectividad de la representación visual se puede defender desde los presupuestos de la lingüística cognitiva; en particular, desde la descripción que hace este modelo de la forma en la que se construye el lenguaje y su relación con la comprensión de conceptos.

De la extensa bibliografía sobre el tema, la obra que más ha influido en nuestra investigación ha sido el manual de Talmy 2000, que recoge, resume y analiza los puntos más relevantes de la investigación anterior. En particular nos ha sido útil el concepto de *esquemmatización* referido a la utilización del espacio visual como una herramienta de adquisición de conceptos y desarrollo de su expresión lingüística (*cf.* Talmy 2000, 176-219).

La esquematización es «a process that involves the systematic selection of certain aspects of a referent scene to represent the whole, while disregarding the remaining aspects» (*cf.* Talmy 2000, 177). La distribución de ideas en un espacio virtual es un proceso habitual y clave en el aprendizaje. Como el mismo Talmy expone (2000, 180-181), los elementos clave de este código visual-espacial utilizados para este tipo de reconstrucciones al que nos referimos son fundamentalmente tres: el tamaño, la posición (relativa), y los signos de unión o «camino».

Siguiendo con la exposición de Talmy 2000, 97-254, las reconstrucciones o representaciones visuales del texto tienen el mismo propósito que una reconstrucción textual.

Sirven para destacar cierto tipo de datos que producen determinadas ideas que se consideran importantes, centrales, o útiles. Para ello emplean un código diferente al textual para expresar la mayor o menor relevancia de los datos (mediante el tamaño), o bien para explicitar que existe una relación entre los mismos (mediante líneas o «caminos», o mediante su posición relativa en un espacio euclideo³³). En este sentido, un esquema visual nunca puede ser global, esto es, no se puede otorgar a todos los elementos la misma relevancia, puesto que esta noción sólo se percibe a través de una comparación (positiva o negativa) con otros elementos. De este modo se interpreta la relación entre las respectivas posiciones de los datos, dentro de una representación visual.

Fabrikant & Buttenfield 2001 estudian los beneficios cognitivos de la representación visual de información geográfica, uno de los objetivos más importantes de la investigación en información geográfica, comúnmente referida como *GIS (Geographic Information Science)*. Los resultados de este tipo de trabajos han mostrado ser útiles en multitud de áreas y no necesariamente relacionadas con la geografía, como la toma de decisiones, la planificación, o en general cualquier actividad humana que implique cierto uso del espacio³⁴.

No hay duda de que vivimos una cierta revolución «visual» o «explosión gráfica» como la denomina Tufte 1997. La exposición de ideas, argumentos o conocimiento en general por medio de imágenes, gráficos o mapas está a la orden del día y es raro el proyecto que pretenda llamar la atención y no emplee estos recursos. Para llegar a esta situación la definitiva implantación de un sistema de interacción visual con el ordenador, la *GUI (Graphical User Interface)*, ha debido de ejercer una considerable influencia, habida cuenta de la omnipresencia del ordenador como herramienta de plasmación y transmisión de los conocimientos. Más allá del factor tecnológico, ha sido el reconocimiento del papel importantísimo que cumple lo visual en el pensamiento humano, desde la comprensión

³³ Un *espacio euclideo* es un tipo de espacio geométrico donde se satisfacen los axiomas de Euclides: 1) se puede trazar una recta entre dos puntos cualquiera dentro de dicho espacio; 2) cualquier segmento puede prolongarse indefinidamente en cualquier sentido; 3) es posible trazar una circunferencia de cualquier radio con centro en cualquier punto; 4) todos los ángulos rectos son congruentes; 5) por un punto exterior a una recta, se puede trazar una única línea paralela a la recta dada.

³⁴ Fabrikant & Buttenfield 2001, 264, recogen bastante bibliografía al respecto.

hasta su expresión lingüística³⁵, lo que impulsa el desarrollo de estrategias de representación, gestión y recuperación de información basadas en relaciones espacial-visuales.

La llamada «metáfora espacial» (cf. Talmy 2000, 99-103; Langacker 2008, 27-54) resulta especialmente útil para visualizar relaciones complejas entre datos, por ejemplo, para la representación de conceptos o ideas que abarcan múltiples áreas de conocimiento. Su éxito se basa en el aprovechamiento de los mecanismos naturales de la cognición humana, basados en el desplazamiento por el espacio y en la asignación de un significado procedimental al movimiento y a ciertas propiedades de los objetos representados en dicho espacio (cf. Skupin 1998, 1).

Fabrikant & Buttenfield 2001, 266, denomina a este proceso de creación representaciones visuales «espacialización» y consiste, según la definición de Skupin & Buttenfield 1997, 117, en la «proyección de información con una alta dimensionalidad en un espacio de menor dimensionalidad». Es decir, el objeto visual resultante implica una simplificación o reducción de la realidad representada. Los criterios por los que se produce dicha simplificación pueden ser semánticos (en función de los atributos o características de los objetos representados) o geométricos (en función de las relaciones jerárquicas o lógicas entre los objetos), o incluso una combinación de ambos.

La creación de representaciones visuales por ordenador consiste en una variante del proceso denominado *ordenación*. Relacionado con el alemán *Ordnung* («disposición», «orden»), puede definirse de manera simple como el proceso por el que determina la ubicación de un objeto en un espacio de n dimensiones, en virtud de la cantidad o del tipo de sus características (cf. Fabrikant & Buttenfield 2001, 267). Para que la *ordenación* resulte efectiva como método de expresión del conocimiento, es importante reducir al máximo el número de dimensiones utilizadas, de modo que resulte fácil de comprender (cf. Kent & Coker 1994, 162). Evidentemente, el proceso de *ordenación* no solo supone una reducción del número de datos sino una modificación radical de la estructura, de la forma original en la que estos datos se encontraban; en nuestro caso, del texto. Esta es la razón por la que hemos incluido esta técnica dentro del método de reducción.

Tanto el nivel de reducción de los datos como la forma final que adopte su representación reflejan los criterios y la percepción del investigador o crítico que los

³⁵ Cf. Hampe 2005, Kukkonen 2013, Langacker 1987, Leguizamo León & García Carrasco 2011, Neisser 1967, Olson 1994 y Talmy 2000 por hacer solo una selección de trabajos que abordan el papel de lo visual en psicología, la lingüística o el análisis literario.

realiza. Es por esto por lo que todos estos recursos de reconstrucción visual suelen emplearse como herramienta de generación de hipótesis o de consulta exploratoria de los datos (cf. Fabrikant & Buttenfield 2001, 267; Moretti 2005, 56).

Las representaciones visuales más frecuentes en la investigación literaria son aquellas que dan nombre al célebre libro de Moretti 2005: mapas, grafos y diagramas en árbol. Los grafos son también denominados habitualmente «redes», y resultan especialmente útiles para reflejar las conexiones entre diferentes conceptos, y una herencia de la investigación en inteligencia artificial (cf. García Martul & Franco-Álvarez 2009, 135). Los diagramas en árbol, por otra parte, son más adecuados para representar conceptos que mantienen una relación jerárquica entre sí, como por ejemplo la que mantienen entre sí los diferentes manuscritos utilizados para una edición crítica³⁶.

A continuación dedicamos un espacio un poco más extenso a los mapas, o mejor dicho, a un conjunto de representaciones visuales de datos unidas por la común denominación de *mapa*. En efecto, esta es una de las formas habituales de representación de los datos obtenidos mediante algoritmos de análisis semántico o *topic modeling* como los que hemos utilizado para el análisis del texto de las *Historias*. Creemos, pues, interesante detenernos a describir dos ejemplos de uso del mapa como metáfora informativa: el primero desde la perspectiva del uso del mapa como herramienta de comprensión; el segundo, como constructo informático destinado a facilitar el procesamiento de la información por ordenador.

Mapas conceptuales

Los mapas conceptuales tienen su origen en las teorías sobre la psicología del aprendizaje de David Ausubel que fueron enunciadas en los años 1960 y desarrolladas por Joseph D. Novak en 1984 para la didáctica de disciplinas científicas (cf. Novak & Gowin 1984, 7-12). Su objetivo es representar relaciones entre conceptos en forma de proposiciones. Los conceptos están incluidos en cajas o círculos, mientras que las relaciones entre ellos se explicitan mediante líneas que unen sus cajas respectivas. Las

³⁶ El diagrama en árbol es la forma estándar de representación del *stemma codicum* de cualquier edición crítica. Remitimos a Bernabé Pajares & Hernández Muñoz 2010, 68-69) para un par de ejemplos. Resulta aquí especialmente interesante mencionar de modo más específico, en el marco del empleo de los procedimientos de edición crítica asistidos por ordenador, los inspirados en la biología. Se trata de los métodos filogenéticos, utilizados para la clasificación de los seres vivos según su parentesco y su evolución. Uno de estos métodos es la llamada cladística o clasificación por ramas (del griego κλάδος = rama). No en vano

líneas, a su vez, tienen palabras asociadas que describen cuál es la naturaleza de la relación que liga los conceptos (*cf.* Castillo Blasco 2006, 32-33). Otra manera de decirlo es que son una especie de «gráficas cognitivas» o «redes» de conceptos (*cf.* García Martul & Franco-Álvarez 2009, 135). Estas redes se componen de nudos, que representan conceptos y están marcados por puntos o vértices, así como de enlaces, denominados «arcos» (*arcs*) y «márgenes» o «extremos» (*edges*) y que representan las relaciones entre los conceptos³⁷.

Los mapas conceptuales ayudan a exponer un panorama general sobre una materia y permiten desarrollar mejores mecanismos de representación y recuperación de los datos, puesto que destacan los datos y las relaciones que se establecen entre los mismos, de un modo más visible que una exposición escrita tradicional. Además, se puede navegar entre los datos de manera relativamente sencilla, ya que es posible asociar o ligar los conceptos con cada uno de los nodos del mapa e ir agregando recursos a cada uno de los temas o subtemas sin que ello interfiera con la navegación en sí.

Topic maps

Según Castillo Blasco 2006, 33-34, los *topic maps* representan una herramienta de acceso y recuperación de información registrada en la web. Es una iniciativa relacionada con la web semántica³⁸ y los sistemas de representación y gestión de metadatos. Su punto de partida es describir de forma precisa los recursos de información distribuidos a través de internet. Según Moreiro González *et al.* 2003 dicha herramienta es un estándar que se está implantando en la web y que posibilita la navegación conceptual³⁹.

ese principio filogenético es el que está en el origen de la representación en forma de *stemma*. En el campo de la crítica textual, con todo, estos métodos siguen teniendo todavía una presencia muy marginal.

³⁷ Las redes semánticas o asociativas son otro ejemplo de este tipo de simbolismo, cuyo origen remonta a la notación gráfica de la lógica simbólica propuesto por Peirce 1988 y los grafos teóricos de Sowa 2000, 476-489, que por cierto aplicó con interesantes resultados a la poesía oral griega (*cf.* Sowa & Sowa 1974). Los mapas conceptuales han sido bien acogidos como elemento didáctico: *cf.* la aplicación *CMap* disponible en Internet (<<http://cmap.ihmc.us>>).

³⁸ Sobre el concepto de «red semántica», remitimos a Shadbolt *et al.* 2006. Particularmente interesante para nuestro campo de estudio es el artículo de Berners-Lee & Hendler 2001 sobre las implicaciones que este nuevo modelo de gestión de información en internet tiene sobre la difusión científica.

³⁹ Concretamente, los *topic maps* aparecen registrados bajo diferentes epígrafes de la normativa ISO (ISO/IEC 13250:2003; 13250-2:2006; 13250-3:2013; 13250-4:2009; ISO/IEC 13250-5:2015 y 13250-6:2010), accesibles en <<http://www.iso.org>>. El «marco de descripción de recursos» o *RDF* (*Resource Des-*

Este tipo de representación visual de conceptos apareció a finales de los años 1990 con independencia de la web. Se trataba de un modelo de gestión de información destinado a facilitar la creación y el intercambio de índices que se incluyen al final de los libros y manuales, algo cada vez más necesario en el ámbito de la documentación técnica de *software* (cf. Eito Brun 2003, 144). Estos índices contienen una lista de términos ordenados alfabéticamente y relaciones cruzadas entre ellos, y tienen como objetivo localizar información puntual entre las páginas de los libros impresos. Posteriormente el modelo evolucionó hacia otras estructuras, como la de los tesauros, hasta llegar a ser una herramienta considerada en la web como forma de organización, representación y gestión del conocimiento (cf. Moreira González *et al.* 2003)⁴⁰.

Ahora bien, como observan García Martul & Franco-Álvarez 2009, 139, en su sentido más técnico, los *topics maps* consisten en una forma de estructurar hiperenlaces por ordenador de modo que sea posible generar y representar visualmente relaciones entre los mismos, semejantes a las que un humano genera al observar un mapa auténtico. Si los mapas conceptuales son un modelo teórico de organización y visualización de conceptos, los *topic maps* son un constructo o modelo informático creado a partir de la misma idea, que incluye detalles técnicos para su aplicación concreta (cf. *Id.* 2009, 142-148).

II.1.3.2.c. Relación

En el método de *reconstrucción* hemos insistido en el establecimiento de relaciones entre datos como un rasgo fundamental. En efecto, uno de los puntos fuertes de las representaciones visuales es su fuerza expresiva para declarar explícitamente que existe una conexión entre los datos. Esto nos lleva al tercer y último método general de *RI* que podemos aplicar al texto, y sin el cual no podríamos obtener ningún tipo de información: el establecimiento de relaciones.

Se puede decir que la relación es un *archimétodo* o un *metamétodo*, ya que actúa como principio rector de los dos anteriores, que en el fondo se basan precisamente en la existencia de una relación más fuerte con determinados elementos y no con otros (la *reduc-*

cription Framework) es un estándar con funcionalidades semejantes desarrollado por el *World Wide Web Consortium* (cf. Alam *et al.* 2015, 3).

⁴⁰ En este sentido, Garshol 2004 muestra cómo la flexibilidad de los *topic maps* les permite representar el resto de técnicas de organización y descripción de la información en un estudio sobre los *metadatos*, los tesauros y las taxonomías.

ción), o en la creación de nuevas relaciones más evidentes (*reconstrucción*). Entendemos que algo se refiere al campo semántico de la «relación» cuando se refiere a conceptos que implican el establecimiento de uniones entre dos o más elementos, como pueden ser los conceptos de «estructura», «inferencia», «deducción», como vemos en Pinto Molina 1996, 302-310, a lo largo de su descripción introductoria sobre el análisis documental, o en la monografía de Villén Rueda 2006, 35. La mención del establecimiento de «relaciones» es explícita al hablar de la creación de índices o «indización» (*cf.* Pinto Molina 1996, 322).

Aunque el concepto así definido es prácticamente omnipresente en casi cualquier teoría sobre el lenguaje, nuestro uso del mismo como criterio de clasificación de técnicas de *RI* se basa en dos principios. En primer lugar, establecer relaciones posee una utilidad práctica como vehículo de comunicación; es decir, como una forma de trasladar la atención del lector o receptor de unos datos a otros. En segundo lugar, la unión de dos o más elementos es la base del empleo simbólico del lenguaje y de su capacidad para agrupar y dar coherencia a datos aislados. La «metáfora espacial» es un ejemplo de esta interpretación del término «relación»; sin ir más lejos es el mecanismo que guía y posibilita, según la lingüística cognitiva, la creación del lenguaje (*cf.* Talmy 2000, 19-24; Langacker 2008, 3-5), así como su aprendizaje (*cf.* Talmy 2000, 119-122).

No nos detendremos más en estos principios teóricos, y pasamos a describir algunos ejemplos de técnicas de *RI* basadas en la *relación*. A la hora de clasificar un ejemplo como *relacional*, hemos tenido en cuenta la existencia de un deseo de resaltar una conexión, tanto en su propósito como en su diseño (el segundo principio), así como producir objetos que efectivamente muevan al usuario, ya sea de una idea a otra, ya sea de un lugar físico del texto a otro (el primer principio). En consecuencia, hemos dividido las técnicas descritas en dos subgrupos, denominados *enlaces* y *conjuntos*.

Enlaces

En el grupo de los enlaces incluimos aquellos objetos textuales o visuales cuyo objetivo es trasladar la atención del usuario de un lugar a otro o de una idea a otra. En este sentido los ejemplos más claros de enlace son algunos signos auxiliares utilizados durante la edición de un texto, las referencias bibliográficas y, dentro del texto digital, los *hiper-enlaces*.

Índices y catálogos

Un índice es «una serie de puntos de acceso organizados que guía desde la información conocida por el usuario a una información adicional previamente», según la definición de Villén Rueda 2006, 35. Un índice pretende servir de «mapa» del contenido informativo de un texto, y para ello debe ser resultado de un análisis exhaustivo del mismo, como señala Mulvany 2009, 19. Dependiendo del interés que tenga su creador por un tipo de datos u otros se obtienen diferentes tipos de índices: de materia, bibliográfico, *index nominum* o *index rerum* etc.

En la ciencia de la documentación, el proceso de creación de índices es denominado *indización*⁴¹. Esta consta de dos fases, según Lancaster 1998: *a)* análisis conceptual, y *b)* traducción. La primera consiste en un estudio del contenido del texto, bien mediante una lectura comprensiva o mediante cualquier otro método o conjunto de métodos como los que hemos descrito hasta ahora, a la que sigue una selección de los conceptos o ideas más importantes. Luego, se escogen una serie de términos o expresiones con los que describir dichos conceptos, y que actuarán como enlaces, nexos o puntos de acceso en el índice final.

El aparato crítico

El aparato crítico es una parte fundamental de la estructura física de una edición crítica, aislada visualmente por un área de interés propia al final de cada página⁴². Es un complemento de la edición, que trata de recoger con el máximo de rigor posible y en el mínimo espacio todas aquellas lecturas presentes en alguno o varios manuscritos utilizados pero desestimadas por el autor. Puede ser positivo, si recoge la lectura elegida además de las otras variantes, o negativo si sólo incluye estas últimas (*cf.* Bernabé Pajares & Hernández Muñoz 2010, 100).

Lo incluimos en esta sección de técnicas de enlace porque remite a otras copias del texto, a otras versiones o a la opinión de otros editores o del mismo editor. El movimiento que indican los enlaces de un aparato crítico es, pues, unas veces físico, cuando implica

⁴¹ Como bien aprecia Villén Rueda 2006, 38, existe una importante confusión terminológica dentro de la ciencia de la Documentación: *indización*, *catalogación* y *análisis documental* hacen referencia prácticamente al mismo procedimiento, esto es, a la interpretación de los datos (información) del texto y su plasmación en un lenguaje especial denominado «lenguaje documental», a veces estrictamente normalizado y tipificado (*cf.* Pinto Molina 1996, 301; Gil Urdicián 1996a, 1996b; Garrido Arilla 1996, 242).

⁴² Para una descripción histórica de los orígenes del aparato crítico como elemento formal de una edición, remitimos a Pagani 2011, 25; 42 y 52, así como a Montana 2011, 111-116).

desplazamiento de un manuscrito a otro u a otra edición, y otras intelectual, cuando proporciona una forma alternativa del texto.

Para la expresión de los datos, el aparato crítico ha terminado desarrollando un código propio, formado por abreviaturas (*cf.* Bernabé Pajares & Hernández Muñoz 2010, 155-165) y signos especiales (*cf.* Morocho Gayo 2004, 20). Un impulso no menos importante para esta evolución hacia la mayor expresión de contenido en el menor espacio posible han sido las limitaciones de espacio físico impuestas por el formato del libro impreso.

Conjuntos: campos semánticos

A diferencia de los enlaces, el establecimiento de relaciones en este tipo de técnicas tiene como propósito crear, demostrar o resaltar la cohesión entre varios elementos; es decir, crear grupos o conjuntos. Para la definición de *campo semántico* nuestra principal fuente ha sido el artículo de Martínez 2003, en el que recoge las distintas expresiones de conceptos lingüísticos que incluyen el término *campo* y que a su vez glosan la evolución de un concepto con ya casi ochenta años de edad, desde que fuera acuñado por Trier 1931.

Carmona Vázquez 2000 detalla la manera habitual de trabajar con campos semánticos en Filología Griega: según el método *distribucional*, seguido por la escuela de Rodríguez Adrados, observa que «cada elemento (unidad de significado) de la cadena hablada entra en dos tipos de relaciones con los demás: una paradigmática y otra sintagmática [...] Del componente paradigmático del significado surgen los *semas*, que diferencian una respecto a otras las palabras de un campo semántico; del sintagmático surgen las *clases de palabras*, que también diferencian entre sí las distintas unidades de contenido» (*cf. Id.* 2000, 47). La primera división sigue un criterio lógico mientras que la segunda, uno funcional.

El motivo ulterior que impulsa esta organización de los significados de una palabra es su exposición clara bajo el lema de un diccionario, siguiendo unos criterios claros. Rodríguez Adrados 2000, 290, rechaza específicamente el criterio cronológico y el criterio de la «relevancia», se entiende que en los casos en los que esta se base en la sola percepción del lexicógrafo. «Distribución» puede considerarse un sinónimo de «contexto»⁴³. Uno y otro término quieren decir que para determinar la información que proporciona una palabra es necesario ubicarla dentro de un conjunto donde se oponga a otras en función de los datos que poseamos sobre su utilización (datos de tipo diatópico, diastrático, diafásico y,

⁴³ El contexto según el cual se analiza la distribución de una palabra puede ser sintáctico, como piensa Rodríguez Adrados 2000, 293-294, o de otro tipo (*cf.* Carmona Vázquez 2000, 49).

por supuesto, sintáctico). Por esta razón los campos semánticos representan inevitablemente *conjuntos* de datos.

Este enfoque demuestra la preocupación que existe por el establecimiento de unos criterios objetivos para la constatación y el empleo efectivo de relaciones entre palabras. Esta preocupación no es exclusiva de la lexicografía, y su posible solución constituye el trasfondo teórico de casi todos los métodos de *RI* y búsqueda semántica, especialmente si utilizan el ordenador⁴⁴.

Tesauros

Los tesauros son una herramienta para la organización del vocabulario de una determinada ciencia o ciencias. Según Gil Urdiciáin 1996b, 354, el tesauro representa el prototipo de un tipo de lenguajes documentales basado en la combinación de diferentes expresiones conceptuales llamadas «descriptores», para las que han sido definidas previamente sus relaciones semánticas posibles. Es una forma de *reconstrucción* de información, basada en una fuerte *reducción* controlada del vocabulario utilizado, y con un fuerte énfasis en el establecimiento y definición de *relaciones*.

La definición que da la norma internacional *ISO 2788:1986* es más técnica, pero ilustra los puntos más importantes para nuestra exposición. Según dicha definición, de acuerdo con la función que cumple, un tesauro es un instrumento de control terminológico que traduce a un lenguaje sistémico o documental el lenguaje natural empleado en los documentos y por los usuarios. Según su estructura, se trata de un vocabulario controlado y dinámico de términos relacionados semántica y jerárquicamente, que se aplica a un campo específico del conocimiento.

Con respecto a los campos semánticos, la propuesta del tesauro es una simplificación controlada de las formas de indicar 1) los conceptos y 2) los vínculos que los unen⁴⁵. En este último sentido se nota una preocupación por extender la simple noción de «relación»,

⁴⁴ Remitimos, por ejemplo, al artículo de Hliaoutakis *et al.* 2006, 2, enfocado hacia la web semántica y cuya exposición del problema coincide bastante con nuestro punto de partida para la definición de búsqueda semántica. La tesis de Martínez Alonso 2013 interesa también a nuestro propósito.

⁴⁵ Como observa Martínez Alonso 2013, 163-164, la polisemia es sin duda uno de los principales escollos que resta por superar de cara a un posible análisis del contenido de documentos objetivo y (semi)automático.

de forma que resulte descriptiva por sí misma. Para ello el tesoro contempla únicamente tres posibles tipos de relación:

- a) Relación de equivalencia: se establece entre los términos preferentes y no preferentes cuando se considera que uno o más términos se refieren al mismo concepto; es decir, se trata de una especie de sinonimia.
- b) Relación jerárquica: se expresa en forma de grados de subordinación o *superordinación*⁴⁶ de unos conceptos con respecto a otros.
- c) Relación asociativa: se establece entre términos que, no siendo equivalentes, pero tampoco relacionables jerárquicamente, sí que poseen un grado de conexión suficiente como para que sea necesario explicitarla.

Evidentemente, las relaciones asociativas, como señalan Saadani & Bertrand-Gastaldy 2000 son las más difíciles de definir y constituyen la categoría relacional que más críticas ha recibido, ya que, a pesar de las directrices que regulan su interpretación, su atribución suele fundarse en un análisis intuitivo que expresa el punto de vista de quien la establece.

Es necesario distinguir entre un tesoro como constructo y la representación del mismo, que puede ser textual o visual. García Martul & Franco-Álvarez 2009, 151, hacen una interesante comparación entre tesauros y *topic maps*, demostrando su evidente similitud y también ciertas ventajas expresivas de los últimos. A su vez, unos y otros guardan relación con otro método de *RI* característico de la web semántica, que comentamos con más extensión más adelante: las ontologías (*cf. infra II.3.2. Ontologías*).

⁴⁶ La superordinación es la reiteración de una unidad léxica por otra de significado más general.

II.2. PROCESAMIENTO DE INFORMACIÓN ASISTIDO POR ORDENADOR

Por *procesamiento* entendemos el hecho de producir un cambio o transformación en un objeto o serie de objetos, o la obtención de un objeto nuevo a partir de los cambios realizados. El conjunto de las acciones realizadas constituye un *proceso*. En un sentido más específico, utilizamos «procesar» como sinónimo de «manejar» un texto por ordenador.

Desde la definición de las unidades de trabajo que expusimos al comienzo de nuestra tesis, el objeto del procesamiento al que nos referimos a lo largo de esta han sido los datos. Estas unidades elementales son transformadas para generar información a través de diferentes tipos de representaciones. Por último, estas representaciones pueden ser tratadas, a su vez, como suma de datos sobre los que trabajar, o bien como un conjunto que funciona como un «dato»; es decir, como unidad elemental de un razonamiento. Este será el objetivo prioritario de nuestra exposición a continuación.

Dependiendo del tipo de actividades o acciones (operaciones) que implique el procesamiento, estas son susceptibles de ser realizadas por una máquina, como los cálculos matemáticos, o bien su ejecución es una capacidad exclusiva del ser humano, como por ejemplo la interpretación cualitativa y en general la creación de ideas «emergentes» (*cf.* Pinto Molina 1996, 309-310). De acuerdo con lo que acabamos de exponer, el sujeto del procesamiento de los datos puede ser un ser humano, pero también un ordenador, en determinados casos.

En este capítulo se reafirma nuestra concepción del ordenador como una herramienta auxiliar, analizando sus características y en consecuencia sus posibilidades reales. También contiene una descripción más en profundidad de los aspectos técnicos que han condicionado nuestro uso del ordenador.

II.2.1. EL ORDENADOR

Un ordenador es una máquina que realiza cálculos, de ahí que se llame *computer* en inglés. La historia del ordenador como máquina de cálculo automático se remonta al siglo XVII con el *motor diferencial* y el *motor lógico* diseñados por Charles Babbage. Aunque el primero llegó a convertirse en un prototipo funcional, el segundo, descrito en nuestro siglo como «el primer ordenador», o incluso como «un ordenador de otro mundo», permaneció como un proyecto (*cf.* Campbell-Kelly 2009, 65; Campbell-Kelly *et al.* 2004, 41-61).

Sin duda, la Segunda Guerra Mundial fue el periodo que marcó el definitivo inicio de la era de la computación, bajo el pretexto de intentar desarrollar una máquina que descifrara los códigos alemanes, y como resultado del desarrollo exponencial de las comunicaciones a distancia y de nuevos métodos de almacenamiento de información mecánicos (tarjetas punteadas), y electromagnéticos (basados en «pulsos» o en la preservación de un estado de carga eléctrica determinado en cintas magnéticas).

La máquina que hoy conocemos como *ordenador* es el resultado del desarrollo de máquinas capaces de realizar de forma autónoma cálculos cada vez más complejos, y de las investigaciones llevadas a cabo durante la primera mitad del siglo XX en la teoría de la comunicación. Durante este tiempo se desarrollaron nuevas formas de almacenar datos que sirvieran de alimento al ordenador, así como herramientas y aparatos que permitieran introducir los datos en el ordenador y reproducir el resultado de los cálculos de una forma comprensible por el ser humano⁴⁷. Un ordenador es, pues, un concepto, un modelo que agrupa una serie de instrumentos diseñados para operar al unísono.

II.2.1.1. Estructura física: *hardware*

La contribución del físico y matemático húngaro-estadounidense Von Neumann (1903-1957) fue fundamental para diseñar la arquitectura básica de este sistema, y que pervive apenas sin cambios en los ordenadores actuales⁴⁸. La idea original de Von Neumann, interesado entre otras cosas en la creación de autómatas (un campo de investigación que ligeramente tocamos al hablar de las expresiones regulares), imitaba la estructura de un ser vivo (*cf.* Campbell-Kelly *et al.* 2004, 76). El ordenador se compone, pues, de un sistema de memoria, una unidad de procesamiento, un centro de control, y de «órganos» con los que comunicarse con el usuario humano. Estos últimos son denominados habitualmente «dispositivos de entrada» (*input*) y «dispositivos de salida» (*output*), en relación a su capacidad para permitir la comunicación de datos al ordenador y la representación de los resultados de su procesamiento.

⁴⁷ Para un análisis más detallado remitimos, además de a la monografía de Campbell-Kelly *et al.* 2004, al conciso resumen hecho por Sagredo Fernández & Espinosa 1996, 159-165.

⁴⁸ En realidad, la llamada *máquina de Turing*, en honor a su creador a finales de los años 1930, el matemático británico Alan Turing, ya era capaz de realizar operaciones sobre datos, es decir, de «procesar» con algunos componentes menos (*cf.* Campbell-Kelly *et al.* 2004, 60).

Estos cinco elementos constituyen el «cuerpo físico» del ordenador, el *hardware*. Conocer esta arquitectura básica nos permitirá comprender cómo un ordenador obtiene los datos, cómo los procesa, y finalmente cómo los representa. Sólo de esta forma podemos plantear un uso lógico, razonable y productivo de esta herramienta.

II.2.1.2. Funcionamiento: *software*

Un ordenador no tiene voluntad y, sin embargo, *piensa*, en el sentido de que realiza operaciones sobre datos, no diferentes de algunas que también realiza el ser humano⁴⁹. El ordenador trabaja con datos numéricos. Lo que vemos a través de la pantalla es una representación visual diseñada para asemejarse a la de un texto, basada en las órdenes que el ordenador ejecuta de acuerdo a unos códigos creados por el ser humano. Tanto los datos como las órdenes le son proporcionados por un usuario humano, bien directamente, bien a través de un lenguaje que «traduce» lo que el usuario humano quiere decir a un «lenguaje» comprensible por la máquina.

El único lenguaje, por llamarlo así, que «entienden» los ordenadores es el código binario, es decir, alternancias de 0 y 1 que representan cargas positivas y negativas asociadas a cambios físicos (*cf.* Laue 2004, 147), como pueden ser cambios de posición en un sistemas mecánico, o bien a cambios de carga eléctrica, como en el caso de las memorias electromagnéticas actuales⁵⁰.

Estos cambios son la base del funcionamiento de la memoria del ordenador. La memoria es un elemento fundamental donde están guardados todos los códigos necesarios para la realización de todas sus funciones. A través de estos códigos el ordenador interpreta diferentes clases de órdenes, para conseguir una representación visual en la pantalla, para crear procedimientos de interacción con el usuario. El llamado *Sistema Operativo* o *SO*, es un marco general creado mediante este procedimiento y que sirve de canal de comunicación o *interfaz* con el ordenador (*cf.* Laue 2004, 148). Aquellos conjuntos de órdenes concretos destinados a un fin específico son lo que se denomina *software*.

⁴⁹ Tal debió de ser el pensamiento de fondo que impulsó a Turing a crear una máquina de resolución general de problemas, la ya referida *máquina de Turing* (*cf.* Campbell-Kelly *et al.* 2004, 59-60). Compárese con la opinión de Rodríguez Adrados 1977, 212, de corte radicalmente diferente pero igualmente cierta: «un ordenador no piensa, desbroza datos».

⁵⁰ Sobre el origen del código binario, es decir, de la representación alfanumérica mediante alternancias de 0 y 1, remitimos a Heath 1972 y Aschoff 1983.

Utilizar un ordenador, pues, consiste en emitir una serie de órdenes para que el ordenador realice cálculos matemáticos sobre unos datos, tales como suma, sustracción, etcétera. Datos y órdenes se introducen mediante los dispositivos de entrada, que habitualmente consisten en la pareja de teclado y ratón o mediante una unidad de memoria externa, donde ya están almacenados⁵¹.

Nuestro interés en el uso del ordenador radica en su capacidad para trabajar con textos que, como ya hemos indicado, son conjuntos complejos de datos. Las operaciones matemáticas que podamos realizar sobre ese conjunto de datos son igualmente complejas y responden a un modelo previo de:

- a) la forma de los datos de entrada;
- b) las operaciones que deben realizarse sobre ellos;
- c) el tipo de datos de salida que ofrecerán.

Estas operaciones, por lo tanto, deben seguir un orden para ser efectivas. En lógica y matemáticas, estos conjuntos de operaciones abstractas consecutivas con un propósito determinado son llamadas «algoritmos». Siguiendo a Chabert 1999, 1, se pueden definir de forma simple como instrucciones paso a paso que, realizadas de forma sistemática y mecánica, producen un resultado determinado con anterioridad. El principal uso, así como la idea que les dio origen es la resolución de problemas, como precisa el mismo Chabert (1999, 479). En este sentido, los algoritmos son la forma más efectiva de procesamiento de los datos mediante ordenador⁵².

Los llamados «lenguajes de programación» no son sino una forma de escribir algoritmos de modo que un ordenador pueda ejecutarlos. Las dificultades surgen pues en relación al tipo de «lenguaje» que pueda utilizarse para esta tarea, y en particular en torno a la sintaxis de dicho lenguaje. En efecto, la sintaxis representa el modo de unión de dos elementos, así como la función que desempeña cada uno de ellos. Un «lenguaje de programación» no es más que un modelo tipificado de sintaxis, en la que el significado y las

⁵¹ Sagredo Fernández & Espinosa 1996, 165-180, ofrecen en su artículo un panorama bastante completo de los diferentes tipos de sistemas de almacenamiento digital, desde los antiguos disquetes a los discos ópticos. Las unidades *USB (Universal Series Bus)* aparecieron después de la redacción de este artículo, pero en lo más esencial su funcionamiento no es muy diferente.

⁵² Remitimos a Striegnitz *et al.* 2003 para una descripción detallada de los principales tipos de algoritmos utilizados en lingüística computacional.

reglas de combinación de cada elemento están perfectamente definidas. El ordenador «lee» este conjunto de órdenes, verifica que el texto que las contiene (también llamado «guión» o *script*) se ajusta a las normas de la sintaxis elegida, y procede a ejecutar uno a uno los pasos descritos consultando para ello un *compilador* o manual con el que «traduce» las proposiciones escritas en el *script* a operaciones de cálculo concretas (cf. Chabert 1999, 479).

Al hilo de esta cuestión exponemos a continuación lo más relevante para el estudio de textos griegos en ordenador, desde el punto de vista de la representación visual del texto griego, así como desde la perspectiva de los métodos de almacenamiento de datos y de los métodos disponibles para su procesamiento.

Al contrario que en la sección anterior, nuestra intención aquí no es ofrecer un panorama general sino centrarnos en aquellos instrumentos efectivamente empleados en nuestra investigación, aprovechando la clasificación de los métodos de búsqueda semántica realizada previamente para analizar sus ventajas e inconvenientes.

II.2.2. EL TEXTO DIGITAL

Según Gómez y Méndez & Leal Gil 2001, 101, el término *digital* «se dice del computador u ordenador en que todas las representaciones de magnitudes se traducen en números, con los cuales opera para realizar los cálculos con datos susceptibles de adoptar estados discretos a través de sistemas, dispositivos o procedimientos». A partir de este adjetivo se crea el verbo transitivo *digitalizar* que indica el acto de codificar numéricamente la información convirtiendo una información analógica (texto, imagen o sonido) en un código digital susceptible de ser almacenado y manipulado mediante un ordenador⁵³. Dentro de la clasificación de tipos de documentos ofrecida por Palomares Perraut 1997, 36, se distinguen documentos *textuales* de otros *legibles por máquina*, en los que la información está registrada de forma digital⁵⁴. La misma autora observa que las *Tecnologías de la Infor-*

⁵³ Mencionamos sólo una pequeña selección de los trabajos que presentan los usos del término en el sentido descrito que nos hemos encontrado durante nuestra investigación: Canet Vallés 2014, 16-17; Spence 2014, 44; Rodríguez Somolinos 2008, 416; Gangutia Elícegui 2007, 759 y López Jimeno 2000, 262.

⁵⁴ La expresión *machine-readable-text* o *texto legible por máquina*, también referida habitualmente por sus siglas en inglés, *MRT*, se refiere a un texto cuyos datos están codificados de forma que un ordenador pueda ejecutar operaciones sobre ellos. La conversión de textos a este nuevo soporte ha sido una de las áreas de trabajo fundamentales para humanistas digitales y todos aquellos interesados en aprovechar las capacidades del ordenador para el estudio del texto. Sobre esta cuestión en general remitimos a Rommel 2004, 92,

mación y la Comunicación (TIC) han cambiado las formas de producción, almacenamiento, recuperación y difusión de la información. Fruto de estos cambios algunos, como Reinhardt 1994, 91-104, hablan del *documento virtual*. Es un tipo de documento digital que nace a partir de la interconexión de ordenadores mediante redes y de la utilización de aplicaciones informáticas para enlazarlos en lo que se conoce como un *sistema hipertexto* o *hiper-medio* (cf. Reinhardt 1997, 40-42)

Este proceso, sin embargo, dista mucho de ser todo lo sencillo que el término «digitalización» puede indicar. La omnipresencia de diferentes dispositivos de entrada de datos: mecánicos (teclado, ratón) o táctiles, a partir del último decenio, y especialmente la respuesta inmediata que producen en la pantalla puede hacernos olvidar que en cada pulsación se producen una serie de pasos, y que la inmediatez es sólo superficial. Laue 2004, 146-147, aprovecha precisamente el ejemplo de la pulsación de una tecla y la aparición de un signo en la pantalla para ilustrar el funcionamiento del proceso en el *hardware*. Lo más interesante de dicho proceso es el mecanismo de «traducción» que se produce entre la pulsación de la tecla y la representación visual de un signo determinado en la pantalla. El teclado consiste de un circuito de corriente que se interrumpe al presionar la tecla, enviando a la *BIOS (Basic Input / Output System)* del ordenador una señal, que es interpretada de acuerdo a con un «manual de equivalencias» como la secuencia de bits correspondiente a determinado signo (cf. Laue 2004, 147).

II.2.2.1. Código Beta y Unicode

El último paso antes de la aparición final del signo deseado en la pantalla es un proceso de «interpretación» de datos, es decir, una tarea correspondiente ahora al software y no al hardware. El problema, en el caso del texto griego, radica en que los responsables de redactar los primeros «manuales de traducción», como hemos denominado a las tablas de codificación de caracteres, no eran precisamente filólogos griegos ni tenían un especial interés por representar letras griegas.

En efecto, la representación del texto griego fue uno de los principales escollos para la aplicación del ordenador al campo de la Filología Griega. Tuvo una primera solución mediante el uso de caracteres latinos, cuyo ejemplo más famoso es el código *beta*, utilizado para la codificación de los textos incluidos en el *Thesaurus Linguae Graecae (TLG)*

y Hockey 1994. Schmidt 2010 presenta un caso práctico, analizando el proceso de conversión al formato digital del Nuevo Testamento.

(*cf.* Heslin 2001). Por desgracia, el código *beta*, aunque legible, no es cómodo y tampoco compatible con muchos programas comerciales de análisis de texto, lo que ocasiona la incómoda necesidad de recodificar el texto para poder usarlo con otros programas⁵⁵.

La cuestión se solucionó definitivamente con la implantación del estándar *Unicode*⁵⁶, cuya adopción ha traído numerosas ventajas porque facilita el almacenamiento y procesamiento correctos de textos en griego antiguo. Es el estándar de codificación de caracteres que siguen todos los proyectos digitales en Filología Clásica y Humanidades Digitales actuales, aunque todavía circulan versiones en código *beta*, como la distribuida dentro del paquete de textos en *XML* de *Perseus*⁵⁷. En nuestra investigación utilizamos para el estudio por ordenador una versión codificada en *Unicode*.

II.2.3. EXPRESIONES REGULARES

La locución «expresión regular» se emplea en el contexto informático como una forma de notación diseñada para representar patrones susceptibles de ser localizados por el ordenador. Una versión más formal de esta definición describe las «expresiones regulares» como una secuencia de símbolos referidos a constantes y operadores que indica un conjunto de secuencias de caracteres, llamadas habitualmente «cadenas»⁵⁸, y paralelamente como las operaciones que desean realizarse sobre ese conjunto (*cf.* Hopcroft *et al.* 2001², 4 y 83-85; Sipser 2006, 63).

La bibliografía sobre las «expresiones regulares» abunda en manuales de uso generales, pero hemos encontrado pocas referencias a aplicaciones concretas con textos. Las referencias al uso de las «expresiones regulares» en este sentido se enmarcan dentro de proyectos más complejos, donde la aplicación de esta técnica es un paso necesario para el

⁵⁵ Sobre los inconvenientes que ocasionaba el uso de código beta para manejo de textos griegos por ordenador, *cf.* Crane 2004, 51; Goulet 1990, 95 y Fuentes González 1993, 90, al hilo de la notable aportación de Goulet a la resolución de estos problemas.

⁵⁶ *Unicode* es una forma de codificación de caracteres definida por el estándar internacional *ISO/IEC* 10646-1993 (*cf.* TEI-P5, xxvi). A diferencia de los sistemas anteriores, el sistema *Unicode* asigna a cada carácter un código y organiza los caracteres pertenecientes a un mismo sistema de escritura en bloques; de esta manera separa la información relativa al carácter de su representación gráfica, que corre a cargo de otros programas.

⁵⁷ Dicho paquete esta disponible en <<https://github.com/PerseusDL/canonical>>.

⁵⁸ *Cf.* Astrid Ciro 2014, s.v. «cadena de caracteres», 467.

correcto funcionamiento de un algoritmo (*cf.* Brunova 2013; Chacón Beltrán 2008). Especialmente difícil ha sido encontrar referencias sobre su aplicación a textos escritos en griego o latín. Una excepción es Heslin 2000, desarrollador de *Diogenes*, un software para la gestión del corpus de textos griegos *TLG*⁵⁹. Nuestra exposición de los detalles sobre el funcionamiento y utilización de las «expresiones regulares» se basa fundamentalmente en las monografías de Fitzgerald 2012 y Watt 2005.

Por otra parte, aunque la generalización del estándar *Unicode* haya resuelto muchos problemas de interoperabilidad, se trata de un estándar abierto, y la compatibilidad no siempre es completa. Dentro de *Unicode* los caracteres se agrupan en conjuntos o bloques⁶⁰ pero la distinción entre los signos pertenecientes a uno u otro conjunto de caracteres no siempre es apreciable a simple vista, y puede ocasionar problemas a la hora de aplicar una «expresión regular», como advierte Heninger 2004. Así, por ejemplo, el acento agudo y el τόνος del griego moderno son visualmente idénticos pero se codifican de forma diferente (U00B4 y U0384, respectivamente). Un sistema de lematizado automático que tenga en cuenta los acentos arrojaría un error al identificar palabras que un humano no tendría ningún problema en asociar con su lema correspondiente. Para evitar esta clase de errores, algunos desarrolladores de software prefieren ignorar completamente los datos relativos a los acentos en pro de la compatibilidad; es el caso del motor de búsqueda basado en *Perl* utilizado por el *software Diogenes*.

II.2.3.1. Historia

Las «expresiones regulares» nacieron como un uso particular de los lenguajes formales creados a mediados del siglo XX durante el desarrollo de una lógica de funcionamiento para sistemas autónomos o autómatas. En concreto, la notación inventada por Kleene 1956, 8-13, para representar su concepto de «eventos regulares» se considera habitualmente su antecesor directo. Como observan Striegnitz *et al.* 2003, 33-39, desde el campo de la teoría de autómatas, este tipo de notación se extendió al terreno de los lenguajes formales, donde

⁵⁹ El *software Diogenes* fue creado por Heslin, de la Universidad de Durham, en colaboración con el proyecto *Perseus*. La información más reciente sobre su funcionamiento puede encontrarse en su propia página <https://community.dur.ac.uk/p.j.heslin/Software/Diogenes>. En particular, *Diogenes* fue concebido como una versión más moderna de otras piezas de *software* anteriores creadas para la consulta y gestión de la versión en CD-ROM del *TLG* (el *TLG-E* fue la última versión publicada en este formato), como el famoso y aparatoso sistema *Ibycus* o el *software Musaios*.

⁶⁰ La lista completa puede consultarse en <<http://www.unicode.org/charts/>>.

se ha transformado según las necesidades específicas de su uso, dando lugar a diversos estándares.

Thompson 1968 adaptó el sistema de notación original de Kleene para utilizarlo en el editor de texto *QED* como método para localizar patrones en archivos de texto. A raíz de esto comienza a popularizarse su uso, también para el análisis del léxico en diseño de compiladores (cf. Johnson *et al.* 1968; Hopcroft *et al.* 2001², 84). Actualmente, las «expresiones regulares» son un modelo de lenguaje formal regulado por el estándar *POSIX* fijado por la *IEEE (Institute of Electrical and Electronics Engineers)*, que establece dos «módulos» o codificaciones: el estándar básico (*BRE, Basic Regular Expressions*) y el estándar extendido (*ERE, Extended Regular Expressions*).

Sin embargo, además de estos dos modelos, existe otro modelo de «expresiones regulares», desarrollado en el entorno del lenguaje de programación *Perl*, y que ha terminado por convertirse en un estándar *de facto*. Uno de los programas más habituales para el manejo de la base de datos del *TLG*, el *software Diogenes*, está escrito en *Perl* (cf. <<https://www.perl.org>>), y utiliza «expresiones regulares» para ejecutar búsquedas siguiendo dicho estándar.

II.2.3.2. Funcionamiento

El funcionamiento de las «expresiones regulares» se basa en el uso de una serie de símbolos con valores diferentes a los habituales, llamados *metacaracteres*. Se trata de caracteres o secuencias de caracteres que el ordenador interpreta como símbolo de un conjunto determinado de datos o como órdenes, y no como un simple símbolo gráfico (que sería su valor literal). Si se utilizan para expresar diferentes tipos de operaciones lógicas o matemáticas, son denominados «operadores».

Ejemplo de un operador es una barra vertical para separar alternativas; así «gris|gres», indica que una coincidencia con cualquiera de los dos términos es válida. El mismo efecto puede conseguirse mediante paréntesis combinado con la barra vertical, de forma que sólo quedan delimitados los caracteres diferentes («gr(ile)s»).

Otro grupo de *metacaracteres* sirven para indicar el número de veces que aparece un elemento, por lo que son denominados «cuantificadores». Los más frecuentes son el signo de interrogación (?), el asterisco (*) y el signo de suma (+). El valor concreto puede variar ligeramente dependiendo del estándar utilizado, pero, en general, el signo de interrogación y el asterisco indican un número de repeticiones igual o superior a cero, mientras que el

signo de suma representa un número igual o superior a uno. Para indicar un número determinado, o bien un límite superior o inferior de repeticiones, se utilizan llaves: {*n*}, donde *n* indica el número de veces que debe repetirse el elemento.

Estos *metacaracteres* sólo son una serie de piezas clave, pero no existe una manera determinada de combinarlas, sino que su uso depende del modo en que el usuario haya planificado su búsqueda. Por otra parte, las expresiones regulares se construyen utilizando signos escritos que pueden tener un valor literal, además de funcionar como un *metacarácter*. Puede ser necesario encontrar una secuencia de caracteres o cadena que comience por un paréntesis. Si construimos una expresión regular que incluya dicho paréntesis, el ordenador entraría en conflicto, puesto que el paréntesis tiene un valor como *metacarácter* para agrupar conjuntos de caracteres, y, en consecuencia, no lo incluiría en nuestra cadena de búsqueda. Para solucionar este problema, se han fijado unas secuencias «de escape», que indican al ordenador que el signo escrito a continuación debe ser interpretado con su valor literal y no como un *metacarácter*.

Mediante estos signos se construye una expresión equivalente a la cadena o serie de cadenas de caracteres que deseamos localizar en un texto. Al igual que en el lenguaje natural, tanto las «expresiones regulares» como otros tipos de lenguajes artificiales como los que describimos más adelante poseen una sintaxis, es decir, un conjunto de normas sobre la forma correcta en la que pueden combinarse los distintos componentes del código⁶¹.

La estructura básica de cualquier «expresión regular» consiste en uno o varios «átomos», unidos para formar conjuntos mediante *metacaracteres* (como el paréntesis), cuantificadores y operadores. Los «átomos» equivalen a caracteres literales: por ejemplo, escribir «*l*» permitirá encontrar única y exclusivamente el carácter «*l*» en el texto. Sin embargo, la utilidad de las «expresiones regulares» radica en su capacidad para expresar de una forma muy concisa todo un conjunto de valores posibles. Son especialmente útiles, por lo tanto, para encontrar términos de los que conocemos algunas características pero no su forma literal precisa.

Nuestro uso de las «expresiones regulares» ha estado limitado a la preparación del texto para su análisis posterior. Sin embargo, dado que el objetivo general de este capítulo

⁶¹ La validez de una estructura sintáctica, en estos casos, se mide en una escala bipolar: o es válida o no lo es. Una expresión regular es válida si cumple exactamente con las normas de combinación establecidas y, por lo tanto, produce el efecto deseado; en caso contrario, la expresión es incorrecta y no produce ninguna acción.

era ofrecer un panorama de los sistemas de búsqueda de información, hemos creído necesario hacer mención de un instrumento de trabajo básico para cualquier investigador que se acerca por primera vez al análisis del texto por ordenador⁶². Si tuviéramos que describirlo según la clasificación que hemos propuesto (*cf. supra*, II.1.1. *Criterios de clasificación*), sin duda el método que más lo define es el de *reducción*, puesto que delimita un campo de trabajo (la cadena o cadena de caracteres deseada).

II.2.4. LENGUAJES DESCRIPTIVOS

La codificación del texto en el ordenador, como vimos en el apartado sobre la representación visual del texto y en el apartado sobre el funcionamiento del ordenador (*cf. supra*, I.3.3.2. *Codificación visual* y II.2.1.2. *Funcionamiento: software*), revela que inconscientemente tendemos a identificar la representación visual del texto digital con lo que se ve en la pantalla y con el propio texto digital. Sin embargo, una observación más atenta de los datos que permiten esa representación abre un campo de posibilidades para utilizar el texto digital como un medio de almacenamiento de datos no codificables con anterioridad. Con esto queremos decir que el texto digital permite codificar de forma explícita cierto tipo de datos sin sobrecargar necesariamente la representación gráfica del texto. Esto es debido a la forma en la que el ordenador distingue entre datos con órdenes para representar el texto de una manera y datos sobre lo que es el texto. Este proceso de «enriquecimiento» de los datos en la versión digital de un texto se denomina *marcado*.

Siguiendo a Schmidt 2010, 338, interpretamos el término *marcado* (*markup*) desde dos puntos de vista. Por un lado, es un neologismo derivado de las marcas que dejaban los editores de imprenta para indicar las operaciones necesarias para conseguir un determinado aspecto final del texto impreso (*cf. Raymond et al. 1992, 1; Goldfarb 1997*. Por otro, Coombs *et al.* 1987, Sperberg-McQueen 1991 y Raymond *et al.* 1992 consideran que los espacios entre palabras o los signos de puntuación ya constituyen un tipo de «marcado» *sensu lato*, en el sentido de que proporcionan una información extra al texto. En este sentido, el «marcado» se identifica, al menos en parte, con el código visual que describimos como la principal innovación de la implantación de la escritura.

Como observa Renear 2004, 219, el *marcado* es «información formalmente distinta de la secuencia de caracteres resultado de la transcripción digital de un texto, que sirve para

⁶² El sistema de detección de variantes morfológicas en latín descrito por Schinke *et al.* 1998 responde perfectamente a la forma de trabajo habitual con «expresiones regulares».

identificar rasgos lógicos o físicos (marcado descriptivo), o bien para controlar su procesamiento posterior (marcado procedimental)»; este es el sentido más específico del término *marcado*, y así lo utilizamos a partir de ahora.

La definición que aparece en el manual de la última versión de la guía de la *Text Encoding Initiative*, de la que hablaremos a continuación, es mucho más específica y desarrolla su rasgo más importante: el «marcado» es «any means of making explicit an interpretation of a text. Of course, all printed texts are implicitly encoded (or marked up) in this sense [...] Encoding a text for computer processing is, in principle, like transcribing a manuscript from *scriptio continua*; it is a process of making explicit what is conjectural or implicit, a process of directing the user as to how the content of the text should be (or has been) interpreted» (TEI-P5, xxv).

II.2.4.1. Historia

Filólogos y humanistas en general fueron los primeros en desarrollar sistemas de «marcado». Los primeros trabajos convertidos al formato digital fueron codificados en lo que conoce como «texto plano»⁶³. Hacia mediados de los años 1970, ya existían varios sistemas de codificación desarrollados para proyectos específicos (cf. Ott 1979), con el objetivo de trasladar al medio digital las obras de grandes autores. Estos consistían en códigos, separados del resto del texto mediante símbolos especiales como «&», «!», «%» u otros, mediante los que se indicaba las entradas del índice, el número de línea, eliminación de palabras etc.⁶⁴.

Sin embargo, cada proyecto utilizaba un sistema propio de «marcado», lo que dificultaba sobremanera el intercambio de resultados o el aprovechamiento del material creado por otros investigadores. El proyecto *Gutenberg* (cf. Lebert 2008) o el *Oxford Text Archive* nacieron con la intención de remediar esta situación caótica. En el caso de la Filología Griega, hasta la aparición y extensión del estándar *Unicode*, el código *beta* creado por el *TLG* era el estándar *de facto*.

⁶³ Citamos en particular el *Index Thomisticus* del jesuita Roberto Busa (cf. Busa 1975; Renear 2004, 232). Para otros trabajos de codificación como «texto plano» remitimos a Froger 1968 y Gilbert 1973.

⁶⁴ Otro ejemplo es el lenguaje *COCOA* creado para el *Oxford Concordance Program* (cf. Russell 1967). Para una descripción somera de su funcionamiento remitimos a Schmidt 2010, 338.

La creación del lenguaje *GML* por Charles Goldfarb en 1971, uno de los primeros lenguajes de «marcado» generalistas, es decir, destinados a cualquier tipo de aplicación, así como de sus sucesores, el *SGML* y finalmente el *XML*, marcaron el progresivo final de esta situación de coexistencia inoperante de diferentes sistemas de codificación, como señala Schmidt 2010, 340.

II.2.4.2. Ventajas

En primer lugar, el propósito del lenguaje *GML* fue crear un estándar que garantizara el fácil intercambio de datos y la interoperabilidad de los archivos así generados entre diferentes entornos informáticos (*cf.* Renear 2004, 227; TEI-P5, xxvi). Las dos ventajas principales con respecto a otros tipos de codificación son el soporte para la *RI*, así como para la aplicación de procedimientos analíticos por ordenador, como explican DeRose *et al.* 1990, 11-14.

La primera ventaja deriva de la estructura que adopta un documento marcado mediante un lenguaje descriptivo. Este ordena jerárquicamente el contenido del texto (estrictamente hablando lo que estructura es la *forma* o los *datos* del texto) mediante unos códigos fáciles de identificar por el ordenador y cuyo contenido puede ser personalizado para responder teóricamente a cualquier necesidad de descripción que requiera el texto. Estos códigos, así como la estructura del texto pueden ser aprovechados por un algoritmo de análisis para extraer y analizar el texto en función de los datos que hayan sido codificados en el mismo, lo que explica la segunda ventaja (*cf.* Renear 2004, 224).

Fruto de este concepto de «marcado» nacen los lenguajes descriptivos, que son formas de codificar dentro del texto digital datos que describen el contenido del texto. El ejemplo más importante es el lenguaje *XML*, heredero o descendiente del lenguaje *SGML*.

II.2.4.3. Extensible Markup Language

Según el *World Wide Web Consortium (W3C)*, las siglas *XML* representan *eXtensible Markup Language*, y este se define como un subconjunto derivado de un lenguaje descriptivo anterior, el *Standard Generalized Markup Language (SGML)*. En principio la función del *XML* era facilitar la representación y procesamiento en la web de documentos

en *SGML*, de forma parecida a como funciona el lenguaje estándar de codificación de contenido en internet, el *HyperText Markup Language (HTML)*⁶⁵.

Los rasgos más relevantes del lenguaje *XML*, según la *Text Encoding Initiative*, son el énfasis que pone en el marcado «descriptivo», en lugar de en el «procedimental» (como hace *HTML*), su independencia de cualquier *hardware* o *software* específico, y, por último, la distinción que hace entre la *corrección sintáctica* y la *validez*. Este último punto se refiere a la circunstancia de un documento que cumpla con las normas generales de la sintaxis establecida para cualquier documento *XML*, pero que no cumpla con los requisitos concretos fijados por una definición secundaria de esta sintaxis general, llamada *definición de tipo de documento (DTD)*, que veremos más adelante (*cf. infra, II.2.4.6. El esquema XML*).

En lugar de utilizarse directamente para codificar nuevos datos en el texto, el uso de *SGML* y *XML* suele estar regido por definiciones secundarias de su sintaxis general. Dicho de otro modo, ambos son metalenguajes diseñados para definir lenguajes de marcado descriptivos, en palabras de Renear 2004, 226. Esto tiene mucho que ver con su capacidad para representar no sólo información relativa a un texto concreto, sino un modelo o estructura de texto.

A partir del trabajo de DeRose *et al.* 1990, se extiende la idea de que el marcado descriptivo esconde una concepción o modelo del texto «esencialmente correcta», incluso exageradamente correcta para algunos (*cf. Coombs et al.* 1987). Efectivamente, no hay duda de que cualquier texto contiene o representa los datos siguiendo un determinado orden, que podemos interpretar como una estructura (*cf. Renear* 2004, 224). Esta estructura se compone de objetos «lógicos» u «objetos de contenido», que obtienen su identidad de la función que desempeñan en el conjunto. A su vez estos objetos de contenido se organizan de una manera jerárquica, sin que ninguno ocupe a la vez la misma posición o función que otro. Este solapamiento se denomina «anidamiento» en la jerga propia del uso de estos lenguajes de marcado y es un requisito indispensable para el correcto funcionamiento. Por último, los objetos de contenido siguen un orden lineal (*cf. DeRose et al.* 1990, 3-4; TEI-P5, xxvii, xxxiii-xxxiv).

⁶⁵ Schmidt 2010, 339, resume muy bien su historia. Probablemente el *XML* fue inventado por Tim Berners Lee (*cf. Berners-Lee & Fischetti* 1999, 29), aunque se basa fundamentalmente en una versión concreta de *SGML* utilizada en el *Conseil Européen pour la Recherche Nucléaire (CERN)*. Su estructura básica

Este modelo de texto compuesto de objetos funcionales, estructurados, unívocos y expuestos en una secuencia lineal recibe el nombre *Ordered Hierarchy of Content Objects* (el modelo *OHCO* de aquí en adelante), y representa al mismo tiempo el principal argumento esgrimido en defensa de las virtudes del uso de *XML*, y su fallo más importante.

Empezamos por las críticas que se han planteado a este modelo, de modo que podamos construir a partir de ahí un uso sólido de lenguaje *XML*. Tanto Schmidt 2010 como Buzzetti 2002, 72, denuncian que no todos los textos son susceptibles de ser representados como una estructura jerárquica, y que, de hecho, un mismo texto puede interpretarse como un conjunto de estructuras que se solapan entre sí, lo que supone un problema para cualquier sistema de codificación basado en *SGML* (cf. Barnard *et al.* 1988, 266-267).

La propuesta de Buzzetti 2002, 76, consiste en distinguir entre un marcado «fuerte» y uno «débil», donde cada uno es adecuado para representar diferentes niveles del modelo de texto. Desde su punto de vista, la expresión lingüístico-textual y el contenido semántico poseen características diferentes. De los dos tipos, sólo la expresión lingüístico-textual sigue un orden lineal y jerárquico, mientras que el contenido se organiza en una «matriz multidimensional».

Traduciendo esta expresión según los presupuestos teóricos que hemos ido estableciendo, la estructura de los datos de un texto es fundamentalmente lineal y puede representarse mediante un etiquetado del tipo utilizado por un documento *XML*, jerárquico, inserto directamente en el propio texto. La información, por el contrario, consiste en el establecimiento de relaciones entre los datos más allá de los límites impuestos por esa linealidad, y del propio texto. Se produce fuera del texto y es particular de cada intérprete. En este sentido, lo que resulta imposible es representar todas las posibles conexiones; sin embargo, en principio no tiene por qué resultar problemático codificar un solo modelo de conexiones establecido por un intérprete concreto en un momento concreto. Invocando este argumento, sostenemos el uso del lenguaje *XML* como forma de plasmación de nuestra interpretación del contenido semántico del texto.

remonta a su vez al predecesor de estos tres lenguajes, el *Generalized Markup Language (GML)*, desarrollado a principios de los años 1970 por Charles Goldfarb, en IBM (cf. Raymond *et al.* 1992, 16).

II.2.4.4. Funcionamiento

Antes de describir cómo se codifica un documento utilizando el lenguaje *XML*, conviene aclarar la terminología que vamos a utilizar. Salvo mención explícita, todo lo que sigue sobre el funcionamiento, estructura y aplicación de *XML* ha sido adaptado del manual de Bradley 2002.

En primer lugar, cualquier software que necesite procesar *XML* necesita poder distinguir entre los datos indicados mediante el marcado y el contenido real u original del texto. Esta distinción se realiza enmarcando los signos o códigos *XML* entre corchetes angulares, como sigue: `<w>ήμέρα</w>`. El texto entre los corchetes, incluidos los corchetes se denomina etiqueta. Hay dos tipos de etiquetas, de apertura (`<w>`) y de cierre (`</w>`), que sirven para indicar al *software* que debe aplicar a la cadena de caracteres entre una y otra la información o la orden contenida en dichas etiquetas.

Como todo lenguaje, *XML* posee una limitaciones a la hora de su aplicación. La primera de todas es que un documento *XML* es procesado por el ordenador de forma secuencial. El procedimiento mediante el cual el ordenador interpreta esta serie de datos se denomina *análisis* o *parsing*.

En segundo lugar, algunos signos gráficos, como el corchete angular, tienen un sentido especial en el lenguaje *XML*. Si el texto original contiene algún signo gráfico empleado por la sintaxis *XML*, este debe ser marcado de manera especial para que el *software* lo reconozca simplemente como el signo gráfico, y no por su función. El valor de un signo gráfico como tal, sin ninguna función añadida, se denomina «valor literal», y se representa mediante una «secuencia de escape». Esta consiste en una secuencia de caracteres que, o bien se antepone al carácter en cuestión, o bien lo sustituyen por completo, y que el ordenador interpreta como el valor literal del signo cuando el texto es analizado. Por ejemplo, para representar un corchete angular abierto hacia la derecha `<`, en *XML* debe escribirse `<`.

Con respecto a la posible necesidad de codificar estructuras solapadas, actualmente no hay una solución completamente satisfactoria. Schmidt 2010, 343-345, recoge algunas propuestas, pero la salida más habitual, simple y efectiva a este problema sigue siendo plantear un modelo de texto que no requiera ningún solapamiento.

II.2.4.5. El documento *XML*

Un documento *XML* se compone de elementos, atributos y comentarios, organizados formando una estructura.

Elementos: *XML* permite que los documentos se descompongan en unidades más pequeñas y significativas llamadas «elementos». Los elementos son unidades individuales que pueden ser reconocidas dentro del *XML* y procesadas. Cualquier documento *XML* debe contener al menos un elemento, y el primero identifica siempre y engloba todo el documento. De hecho, muchas veces los documentos *XML* consisten en una serie de elementos, mezclados en ocasiones con texto. Un elemento puede descomponerse en dos partes: el contenedor (incluye las etiquetas de apertura y cierre), y el contenido del elemento (lo que hay entre las etiquetas).

Atributos: un elemento puede expresar información adicional sobre su contenido más allá del nombre, como puede ser un tipo especial de párrafo, o una clase particular de sustantivo. Este tipo de metadatos son denominados *atributos*. Un atributo posee un nombre y un valor, enmarcado entre comillas dobles.

Comentarios: es posible añadir comentarios a un documento *XML*, enmarcándolos con las etiquetas `<!-- texto -->`. La secuencia así delimitada no será procesada de ninguna manera. Constituye una buena forma de añadir explicaciones o notas de ayuda en casos donde sea necesario, sin romper la estructura del texto ni afectar a su contenido.

Dicho esto, *XML* no proporciona ninguna forma de expresar el significado de un determinado tipo de elemento salvo la clase de relación que puede mantener con otros elementos. Esta despreocupación por el significado de los elementos textuales es precisamente lo que permite que el lenguaje *XML* funcione independientemente del *software* utilizado (cf. TEI-P5, xxvii-xxix).

La solución que hereda el lenguaje *XML* del *SGML* para expresar este tipo de contenido son las *declaraciones*. Una «declaración» es otro tipo de marcado *XML* utilizado para identificar características especiales o particulares del documento o de alguno de sus elementos. Siguiendo la tónica general del *XML*, las «declaraciones» son una forma de explicitar el tipo de contenido del documento. Existen diferentes tipos de «declaraciones», encargadas de fijar información sobre el documento en general (*DTD*), sobre los elementos y atributos que van a emplearse (de marcado) y, por último, para definir las *entidades*.

Cada documento *XML* posee una estructura lógica y física. La terminología resulta un tanto confusa, pero la referimos aquí por ser la comúnmente aceptada. Por un lado, la estructura lógica de un documento está formada por la sucesión lineal y jerárquica de elementos y atributos delimitados por etiquetas de apertura y de cierre. Por otro, un texto codificado en *XML* puede describirse como un conjunto de definiciones de elementos formales y conceptuales, denominadas «entidades».

Aunque puedan resultar similares, «entidades» y «elementos» difieren en algunos puntos. Para empezar, una «entidad» no se define exclusivamente por una delimitación formal, como sucede con las etiquetas a la hora de delimitar un elemento. Las «entidades» tienen una cierta carga semántica real que va más allá del aspecto formal. El propio documento *XML* en su conjunto es una entidad. Ahora bien, quizá resulte más fácil definir las aludiendo a sus aplicaciones habituales, que son:

1. definir secuencias de escape para ciertos caracteres de marcado;
2. proporcionar un mecanismo para representar caracteres no disponibles directamente en el teclado o en el conjunto de caracteres elegido;
3. separar un documento extenso en fragmentos más pequeños y manejables;
4. crear componentes reutilizables en otros documentos;
5. incluir referencias a otro tipo de datos externos, como imágenes;
6. ayudar a la construcción de *DTD*.

II.2.4.6. El esquema *XML*

Hemos mencionado poco antes la noción de tipo de documento (*cf. supra*, II.2.4.3. *Extensible Markup Language*). Al igual que ocurre en otras disciplinas o en cualquier otro objeto procesable por ordenador, se considera que los documentos *XML* pueden ser de distinto tipo (*cf. TEI-P5*, xxvi). En relación con un texto literario, por ejemplo, un documento *XML* se define fundamentalmente desde el punto de vista de la estructura de sus elementos constituyentes. El aspecto descriptivo de este tipo de lenguajes es simplemente más efectivo si se asocia con accidentes formales del texto. No es de extrañar que los ejemplos habituales en los manuales introductorios consistan en textos poéticos o dramáticos, cuya estructura ya está claramente dividida en líneas, estrofas, cantos etc. (*cf. TEI-P5*, xxix-xxxiii).

Si cierto número de documentos comparten una misma estructura formal, resulta natural que compartan los mismos tipos de *elementos*. El conjunto de elementos, atributos y

declaraciones que representan las características comunes de una clase de textos es lo que llamamos *tipo de documento*. Se comporta como una especie de «archivo de configuración», que el ordenador o *software* concreto pueden utilizar como plantilla para detectar alguna anomalía en la codificación de un texto que se supone está codificado siguiendo sus normas. Este modelo de texto, codificado siguiendo la sintaxis tipificada por el lenguaje *XML*, se denomina habitualmente *esquema XML*.

El uso del término *esquema* viene de las tecnologías de bases de datos, donde hace referencia a las tablas y campos que conforman una base de datos, así como a las limitaciones de los valores posibles que pueden insertarse en cada campo, junto con las relaciones entre tablas. Su significado coloquial como la representación de algo por medio de un diagrama o boceto es igualmente válido.

Resumiendo, además de *elementos*, *atributos*, *comentarios* y *declaraciones*, delimitados por etiquetas (corchetes angulares, normalmente), mediante *XML* es posible codificar, escribir o crear un modelo de la estructura formal, y hasta cierto punto semántica⁶⁶, de un conjunto de textos. Este modelo recibe el nombre de *esquema*. El diseño de *esquemas* es uno de los usos más productivos e interesantes del lenguaje *XML*, porque facilita el procesamiento automático posterior del un texto y porque constituye una expresión detallada y ordenada de la interpretación del mismo. Tres aspectos han de tenerse en cuenta a la hora de diseñar un esquema: el nivel de especificación (si se pueden especificar por adelantado todos los valores de elementos y atributos); la posibilidad de restringir la ubicación de los elementos (si determinados rasgos aparecen sólo en algunas partes o siguen alguna norma para ello); por último, la sintaxis de estas reglas que acabamos de mencionar (de qué modo se pueden combinar unos atributos con los elementos).

Para terminar, *DTD* y *esquema* son dos tipos de representación de un modelo de documento, pero no son exactamente iguales. El primero es una herencia del progenitor de *XML*, *SGML*, y aunque no por ello sea completamente desechable, los *esquemas* han terminado por imponerse como opción preferida en casi todos los proyectos de investigación que utilizan este lenguaje. Siguiendo el buen criterio de la mayoría, en nuestro caso nos hemos decantado por la segunda opción, de acuerdo con los estándares que se han definido en los últimos años, como veremos a continuación.

⁶⁶ En efecto, *elementos* y *atributos* pueden describir accidentes de naturaleza semántica y no sólo formal.

II.2.4.7. La Text Encoding Initiative

Como ya hemos visto antes, *XML* ofrece una libertad teórica enorme a la hora de elegir qué tipo de información representan las etiquetas. Esto, sin embargo, puede dificultar el trabajo de diferentes investigadores sobre el mismo documento en caso de que no compartan el mismo uso de las mismas. Por ello a finales de la década de los 1980, vista la necesidad de un marco común para facilitar la colaboración, se fundó la *Text Encoding Initiative (TEI)* (cf. Schmidt 2010, 340-341; Hockey 2004, 12-15).

Concretamente, la *TEI* es una organización internacional creada en 1987 con el objetivo de desarrollar unas líneas guía que sirvieran de estándar de codificación de textos en *XML* para su uso en Humanidades y Ciencias Sociales. Estas guías se encuentran disponibles en internet bajo el título de *TEI Guidelines*, y se actualizan periódicamente. A lo largo de este trabajo nos referimos a ellas como *la normativa TEI*, o simplemente mediante las siglas *TEI-P5*⁶⁷. Asimismo en 2001 se creó el *Consortio TEI (TEI-C)*, que en la práctica es una evolución institucional del comité fundador original de la iniciativa.

La normativa *TEI* no ofrece *esquemas* concretos sino un conjunto de módulos destinados a convertirse en los «ladrillos» de futuros *esquemas*. La libertad descriptiva intrínseca al lenguaje *XML* limita las posibilidades de creación de un estándar, de modo que la solución ideada por la *TEI* fue recoger y tipificar las piezas básicas con las que se construye cada uno de los *esquemas* posibles. Este conjunto de definiciones no está cerrado completamente, y la *TEI* acepta propuestas que eventualmente producen una actualización de las guías (cf. Schmidt 2010, 341). En este sentido, la normativa *TEI* proporciona herramientas para codificar teóricamente cualquier tipo de textos y prepararlos para su análisis mediante un abanico bastante grande de aplicaciones informáticas. Su naturaleza flexible, unida a los infinitos puntos de vista que puedan establecerse de modo general sobre un texto, motivan que en la práctica casi todos los *esquemas* que siguen la normativa *TEI* contengan alguna modificación de la misma⁶⁸.

⁶⁷ A menudo se emplea la abreviatura *TEI* para referirse a la propia normativa, así como al conjunto de esquemas que describe. El código «P5» presente en sus siglas es una referencia a la versión actual, disponible para su consulta *online* en <<http://www.tei-c.org/Guidelines/P5/>>.

⁶⁸ Existe una lista de proyectos de investigación y digitalización de textos que utilizan las guías de la *TEI* disponible en <<http://www.tei-c.org/Activities/Projects/>>.

II.2.5. LENGUAJES PROCEDIMENTALES

Para cerrar el capítulo sobre el uso del ordenador, dedicamos un espacio al método mediante el cual se dan órdenes al ordenador para que realice los cálculos de los que es capaz sobre los datos introducidos. Hemos denominado a esta forma de comunicación con la máquina, «lenguaje procedimental», a imagen del «lenguaje descriptivo» ejemplificado por *XML* y semejantes. Sin embargo, en el entorno informático, la denominación de este tipo de códigos es «lenguajes de programación».

II.2.5.1. Historia y definición

Los primeros lenguajes de programación se crearon con el propósito de controlar las acciones realizadas por un ordenador, como señala Hudak 1989, 359, de quien tomamos lo que sigue. Los primeros lenguajes trataban de abordar la cuestión analizando la estructura física del ordenador. Conforme el aparato mecánico de este se hacía más complejo, la investigación en lenguajes de programación se fue desplazando desde la simple sustitución del lenguaje binario hacia formas que reflejaran diferentes modelos de trabajo. Es decir, los lenguajes de programación son la *abstracción* de una idea sobre lo que puede hacer el ordenador y sobre la forma en la que un ser humano puede expresar esa acción o conjunto de acciones.

Una «abstracción» es una visualización o representación de una entidad que incluye únicamente sus rasgos más importantes, como la define Sebesta 2012, 474-475. Dicho de otra forma, una abstracción es una forma de clasificar datos o cosas según sus características comunes. En programación, el proceso de abstracción es una poderosa herramienta contra la increíble complejidad de las acciones de cálculo que desean expresarse.

Este concepto de «abstracción» es coetáneo a los primeros diseños de lenguajes de programación. Un elemento básico del funcionamiento de cualquier programa informático, los subprogramas (acciones o procedimientos que utilizan sólo parte de las funciones o métodos establecidos por el funcionamiento general del programa) son, en esencia, abstracciones, puesto que proporcionan una manera de especificar un determinado proceso sin necesidad de dar detalles sobre cómo este debe producirse.

La abstracción no se limita a los procesos, sino que los datos utilizados también pueden abstraerse. El lenguaje *COBOL*, allá por los años 1970, ya incluía una estructura de datos, es decir, una forma de clasificar los mismos y referirse a ellos figuradamente

(cf. Sebesta 2012, 72-73, de quien tomamos lo que sigue). El término técnico para referirse a esto es *tipo de dato abstracto*, y se refiere a una estructura de datos, almacenada como un archivo, por ejemplo, pero que incluye un subprograma para manipular los datos que contiene. Sintácticamente (es decir, respecto al modo en el que se aplica al escribir la orden), las «abstracciones» se comportan como unidades separadas en las que se puede trabajar sin necesidad de otras partes del programa. Es más, se pueden utilizar para *invocar* (referirse) a un elemento concreto de toda esa clase definida por el *tipo de dato abstracto* y trabajar con él sin necesidad de manifestar específicamente todo el conjunto de datos que forman parte de la clase, pero manteniendo la noción de que pertenece a ella. Por ejemplo, si definimos una clase con todos los *verbos* de un texto, *verbo* sería el *tipo de dato*. Durante el procesamiento, podemos «llamar» a todos los datos que forman parte de ese tipo, o solo unos cuantos, sin que el ordenador «olvide» que se trata de un verbo. Es un procedimiento muy cómodo cuando se trata de un documento extenso, y más aún cuando se requiere efectuar diferentes acciones. Sin este tipo de abstracciones sería necesario escribir una larga e incómoda serie de órdenes. Este elemento concreto al que nos referimos representa una variable o instancia con respecto a su clase, y se denomina *objeto*.

De esta idea surge un modelo de programación *orientado a objetos*⁶⁹. Un lenguaje de programación orientado a objetos deber estar preparado para ofrecer tres funciones básicas: permitir la definición de tipos de datos abstractos; permitir la *herencia*; y una forma de dar las órdenes llamada *enlace dinámico* (*dynamic binding*). Hemos definido ya el primer requisito, veamos los otros dos:

Herencia: la presión sobre los desarrolladores de aplicaciones informáticas para aumentar su productividad les llevó a pensar, desde mediados de los años 1980, que la forma más rápida de conseguirlo era poder aprovechar los logros que otros ya habían conseguido, en lugar de empezar desde cero cada vez. Naturalmente la posibilidad de reutilizar los tipos de datos fue una de las primeras opciones en este sentido. El problema es que en general, pocas veces era posible emplear el mismo tipo de datos tal cual, sino que algún tipo de pequeña modificación resultaba casi siempre necesario. Realizar estas modificaciones requiere de una comprensión profunda de los criterios y métodos empleados para la definición original del tipo de dato, además de la resolución de otros problemas técnicos. Un segundo problema es que los tipos de datos tienen todos el mismo nivel jerárquico, y ello dificulta su utilización para resolver problemas que requieran la definición de subcategorías de datos. Retomando el ejemplo anterior de los verbos, esto significa que no

⁶⁹ Remitimos a Sebesta 2012, 72-73; 525, para una descripción de los primeros lenguajes de este tipo.

podríamos operar solo con los verbos en aoristo o imperfecto. El mecanismo de la herencia permite que un tipo de datos reutilice, ya sea los datos exactos, ya sea algunas funciones o características de otro tipo de datos. Además de facilitar el desarrollo de sistemas más complejos, establece nuevos tipos de relaciones entre las categorías de objetos, y así se habla de elementos *padres* y *ancestros* (aquellos de los que se hereda un rasgo) y elementos *hijos* (que efectivamente heredan un rasgo de otro).

Enlace dinámico (*cf.* Sebesta 2012, 526-528): es a la vez resultado de la implementación del método de herencia, y también una forma de mejorar su aplicabilidad. El enlace dinámico separa los cambios producidos en una subclase de datos de los que se aplican sobre la clase *padre*. Esto facilita la expansión o evolución de un sistema, y la preservación de su funcionalidad. Para proseguir con el ejemplo de los verbos, consideremos que tenemos una subclase aoristo y otra imperfecto que contienen verbos en indicativo, y queremos fundirlas en una subclase conjunta denominada tiempos de pasado. Gracias al marcado dinámico, podríamos hacerlo sin necesidad de volver a relacionar la nueva subclase con la clase *padre verbo*; es decir, sería posible seguir aplicando las mismas acciones definidas para la clase *padre* en la nueva subclase *hija*.

En esta investigación hemos empleado dos tipos de lenguajes procedimentales orientados a objetos muy semejantes: *Python* y *R*. A continuación exponemos sus características fundamentales.

II.2.5.2. *Python*

Python es un lenguaje relativamente reciente (*cf.* Sebesta 2012, 99-100; Lutz & Ascher 2004). Diseñado originalmente por Guido van Rossum en el *Stichting Mathematisch Centrum* de los Países Bajos a principios de los noventa, ahora mismo es propiedad de la *Python Software Foundation*⁷⁰. *Python* es un lenguaje empleado en multitud de ámbitos como la administración de sistemas, la programación de gráficos generados por ordenador etc. Es un lenguaje de código abierto, es decir, gratuito y libre de ser modificado por cualquiera, además de funcionar en cualquier sistema operativo.

La sintaxis de *Python* no está basada en ningún lenguaje previo, e incluye tres clases de estructuras de datos: listas; listas inmutables, llamadas *tuplas*⁷¹; y archivos denominados

⁷⁰ El lenguaje *Python* está accesible en <<https://www.python.org/psf/>>.

⁷¹ En matemáticas *tupla* es una lista ordenada de elementos y el término procede de una generalización de la secuencia *dupla*, *tripla*, *cuádrupla*, *quíntupla*, ... *n-tupla*.

diccionarios. *Python* es un lenguaje orientado a objetos, pero incluye las capacidades de búsqueda de patrones o expresiones regulares que posee el lenguaje *Perl*. Por esta razón se ha convertido en uno de los lenguajes de programación más extendidos en lingüística computacional. En nuestra investigación utilizamos algunos *scripts* escritos en *Python*, que describiremos en el capítulo final.

II.2.5.3. *R*

Bajo la simple y críptica denominación *R* se esconde un sistema de análisis computacional muy potente, y bastante similar en su funcionamiento básico a *Python*⁷². Como aquel, permite realizar operaciones sobre archivos a través de funciones, así como ejecutar conjuntos de órdenes desde un *script*.

El núcleo de *R* es un lenguaje de programación centrado en el uso de conjuntos de funciones o *módulos*. Estas funciones son órdenes que aprovechan capacidades propias de *R*, o bien invocan otros programas o aplicaciones para extender las capacidades originales de este lenguaje. *R* está especializado en procedimientos estadísticos. Junto a este módulo de funciones básicas, incluye otras destinadas a generar un entorno gráfico de trabajo⁷³ o a representar los resultados de forma gráfica. Además de estos módulos es posible añadir nuevas capacidades mediante la instalación de «paquetes» adicionales, creados por investigadores independientes del equipo creador de *R*⁷⁴.

Los puntos más importantes a favor de *R* son la relativa facilidad de su uso, su interoperabilidad con diferentes formatos de datos, incluidos documentos *XML*, y su capacidad para «grabar» o almacenar el proceso de análisis, de un texto en nuestro caso, para su futura reproducción y comprobación por parte de otros investigadores. Para su aplicación al texto de Heródoto hemos seguido las indicaciones técnicas de los manuales de Jockers 2014 y las recogidas en la documentación de la aplicación (*Meletē*)*ToPān* diseñada por Köntges 2016.

⁷² Nuestra exposición sobre el funcionamiento de *R* procede de su propia página web <<https://cran.r-project.org/>>, en ausencia de un manual específico a nuestra disposición.

⁷³ Se trata de *RStudio*, disponible en <<https://www.rstudio.com/>>.

⁷⁴ La mayoría de estos paquetes aparecen recogidos en <<https://cran.rstudio.com/web/packages/>>.

II.3. MÉTODOS DE BÚSQUEDA SEMÁNTICA POR ORDENADOR

A continuación procedemos a describir dos ejemplos de aplicación de los métodos de búsqueda semántica para obtener información de textos. En primer lugar abordamos el *Análisis Semántico Latente (ASL)*, un método de análisis del contenido de textos basado en el estudio y comparación de frecuencias. En concreto, hemos aplicado una técnica basada en este método, denominada «modelado de temas» o *topic modeling*.

En segundo lugar abordamos brevemente las *ontologías*, un modelo de categorización y clasificación de información, cuyo origen remonta a la tradición filosófica pero que ha terminado por convertirse en una técnica muy utilizada para la gestión de contenido en diversos ámbitos, como pueden ser bibliotecas o internet. Creemos que los *esquemas XML* que hemos descrito antes constituyen una forma de aplicación válida y concreta del concepto de *ontología*, y explicamos el por qué.

Los dos ejemplos son en cierto sentido antitéticos, puesto que parten de puntos de vista diferentes con respecto a la forma de utilización de los datos del texto. Por una parte, el *ASL* concibe la expresión lingüística como un conjunto de datos cuya utilización real (y en consecuencia su capacidad comunicativa y significativa) está regida por la frecuencia en buena medida. Las *ontologías*, sin embargo, son una descripción lógica de los contenidos semánticos, más o menos independiente de la forma lingüística concreta que estos adopten.

En el *ASL* (y por lo tanto en el *topic modeling*) el texto es el origen de unos resultados obtenidos *a posteriori*. Es lo que Sperberg-McQueen 2004, 167, denomina un «sistema basado en datos» (*data-driven*). Por el contrario, las *ontologías* tratan el texto como la manifestación de un modelo preexistente, y, por lo tanto, son un modelo *a priori*.

Una y otra son formas de análisis del texto complementarias. Una *ontología* (o un *esquema XML*) puede beneficiarse de la solidez argumental que ofrecen los datos cuantitativos obtenidos mediante un análisis del tipo realizado mediante técnicas inspiradas en la teoría del *ASL*. La metodología que proponemos para el estudio del texto de Heródoto es un reflejo de esta idea.

II.3.1. ANÁLISIS SEMÁNTICO LATENTE

El *Análisis Semántico Latente (ASL)* o *Indexado Semántico Latente (ISL)* es un modelo teórico del modo en que se relacionan los datos del texto para producir lo que

denominamos «significado», desarrollado por Deerwester *et al.* 1990 y Hoffman 1999. El *ASL* otorga una importancia fundamental a los datos estadísticos, en particular a la frecuencia de aparición conjunta de términos: en líneas generales, establece una relación de proporcionalidad directa entre la frecuencia con la que dos palabras están cerca y la relación semántica entre las mismas. El *topic modeling* o «modelado de temas» es un conjunto de algoritmos informáticos que extraen y representan datos del texto siguiendo las premisas del *ASL*, ofreciendo algo similar a selecciones temáticas de fragmentos de texto.

II.3.1.1. Topic modeling

El «modelado de temas» es una forma de recuperar información de textos, es decir, un modelo de *RI*, basado no en la selección, reducción o relación de «palabras clave» sino mediante la noción semántica de «tema» (*cf.* Blei 2012, 77).

Ninguna máquina *comprende* realmente un texto como para ofrecer un resumen o asociar diferentes textos según otro criterio que no sea la presencia explícita y formal de un determinado elemento (palabra, atributo, característica, información). Los modelos probabilísticos son un conjunto de algoritmos desarrollados dentro del área de la inteligencia artificial y el aprendizaje (*machine-learning*), cuyo propósito es descubrir y anotar grandes archivos de documentos con información temática. Son métodos estadísticos que analizan las palabras de los documentos originales para establecer conexiones, sin necesidad de anotación previa o etiquetado.

Estas técnicas son útiles en la medida en que estemos de acuerdo en que representan fielmente un determinado modelo de texto. Por otro lado, lo que ofrecen nunca es un resultado *final* sino una ayuda u orientación para una determinada propuesta de clasificación de textos, quizá más objetivable que otras argumentaciones debido a su naturaleza estadística.

El modelo de tópicos más simple es el denominado *Latent Dirichlet Allocation (LDA)* definido por Blei & Jordan 2003. Parte de una idea o «modelo conceptual» generativista del texto. Este modelo es el resultado de la elección de un determinado número de palabras que se corresponden a una serie de *temas*, palabras que después son combinadas siguiendo un patrón desconocido para el algoritmo⁷⁵. Este método (y otros relacionados con él y descritos someramente a continuación) sólo detectan patrones o, mejor dicho, *distribuciones* de palabras (palabras que aparecen cerca unas de otras con una probabilidad mayor que otras). Es

⁷⁵ Es la llamada hipótesis del «saco de palabras», recogida como anexo por Jockers 2014, 163-165.

el usuario humano el que realmente ve en esas distribuciones la manifestación de un tema y los etiqueta en consecuencia.

El *LDA* fue desarrollado para arreglar un problema con un modelo probabilístico previo, el *análisis probabilístico semántico latente (pLSI)* (cf. Hoffman 1999), que, a su vez, era una adaptación estadística del trabajo seminal sobre *análisis semántico latente* (cf. Deerwester *et al.* 1990). Esta metodología se sostiene sobre la descomposición en valores singulares de una matriz de términos de un documento (es decir, factorización de una matriz). Desde este punto de vista, *LDA* es un tipo de análisis de componentes en datos discretos (cf. Buntine 2002; Buntine & Jakulin 2006).

Un *tema* o *tópico* es la distribución de un conjunto de palabras determinado. Su detección se basa en los siguientes presupuestos, como expone Blei 2012, 78-82:

1. Un documento *contiene* varios temas en una proporción variable.
2. Los temas de un texto están especificados antes de que se haya generado el propio documento y tienen un número determinado.
3. Un documento se descompone en un aspecto externo y una *estructura escondida* o *interna* que consiste en:
 - a. temas;
 - b. distribución de temas por documento;
 - c. asignación de temas por palabra y por documento.
4. Concepción *generativista* del texto.
5. El orden de las palabras no importa.
6. El orden de los documentos no importa.

La fórmula final a la que lleva el modelo del método de *LDA* implica un número de posibilidades demasiado grande para ser calculado con la tecnología actual (no imposible por completo, pero sí para el usuario común, incluso para un investigador con recursos). Por esta razón su aplicación se basa en aproximaciones. Los algoritmos se dividen en dos categorías: los basados en muestras y los *variacionales*. Asunción *et al.* 2009 hacen un buen análisis de los pros y contras de ambos acercamientos.

II.3.2. ONTOLOGÍAS

Las *ontologías* son un tipo de representación semántica del texto, en cuanto son un modelo de la estructura del contenido. Se trata de un concepto trasladado desde la filosofía a la informática, y que se ha creado su propia parcela de investigación y aplicación en este

campo, especialmente por sus múltiples aplicaciones como *SRI*, o como instrumento para crear *SRI*.

La comparación explícita entre las ontologías y el modelo de texto subyacente a la estructura de un archivo *XML* defendida por la *Text Encoding Initiative* tal como vemos en Sperberg-McQueen 2004, 162, fue una inspiración fundamental para nosotros a la hora de desarrollar un método de *RI* aplicable al texto de las *Historias*.

En las últimas décadas, el término *ontología* se ha convertido en sinónimo de una importante área de investigación en Informática (cf. Sánchez Ruenes 2007, 1-2). Como es sabido, la *ontología* es una parte de la filosofía dedicada al estudio de la naturaleza y organización de la realidad (cf. Sandborg-Petersen & Øhrstrøm 2010, 70-71). Ahora bien, de forma técnica podemos definirla como una «especificación formal y explícita de un concepto compartido», en palabras de Studer *et al.* 1998, 184; es decir, como una representación de alguna clase de fenómeno o conjunto de datos compartido por un grupo de individuos, mediante un lenguaje que sea comprensible por un ordenador (formal).

La definición de Neches *et al.* 1991, 40, se asemeja a la noción de *esquema XML*: «an ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary». Una *ontología* es una definición de los diferentes datos que conforman un determinado objeto y de la relación que guardan entre sí. Comparte pues con un *esquema* la división del contenido en clases (tipos de elementos), la expresión de relaciones, y la representación de datos individuales dentro del conjunto (*instancias* en una ontología; *entidades* en un esquema XML).

Ontologías y *esquemas* realizan lo que Sperberg-McQueen 2004, 161, denomina un acto de *clasificación*; esto es, asignan datos a diferentes clases, creando con ello conjuntos de datos. Una *clase* es, pues, un conjunto o agrupación de datos, elementos u objetos que comparten una propiedad. Clasificar tiene dos propósitos fundamentales:

- Al crear clases o categorías, sirve para *definir* el contenido del texto, ya que los rasgos utilizados para crear la clasificación son resultado del análisis del mismo, es decir, de su *interpretación*.
- Al separar los datos por grupos, permite observar con atención qué rasgos son más relevantes para el objetivo general de la interpretación.

Por estas razones encontramos que un esquema de clasificación, en nuestro caso un *esquema XML*, pero también una *ontología* o similar, constituye un instrumento de *RI*. En efecto, mediante su descomposición en datos discretos (método de reducción) y la reorganización de los mismos en función de sus propiedades compartidas (*reconstrucción y relación*), dicho esquema cumple con todas las características de un sistema de búsqueda de información.

El resultado de este proceso es un modelo de texto, que, a su vez, puede ser representado, bien textualmente, bien visualmente, con un mapa como los generados por diferentes herramientas de *topic modeling* (cf. Köntges 2016), donde los rasgos identificativos utilizados se representan como ejes de un espacio geométrico⁷⁶.

Para terminar listamos algunas reglas que hemos tenido en cuenta para la expresión de nuestra interpretación del texto de Heródoto a través de los medios que hemos venido reseñando:

1. En general se debe evitar la clasificación cruzada.
2. Toda clasificación debe tener un propósito claro y definido.
3. Toda característica empleada como base para un eje en el plano lógico debe:
 - a. distinguir un objeto de otro, es decir, producir al menos dos subclases;
 - b. ser relevante para el propósito de la clasificación;
 - c. estar definida y ser comprobable; una clasificación debe tener en cuenta las condiciones en las que se va a producir;
 - d. ser permanente (o tender a ello); evitar la necesidad de redefinición o reclasificación;
 - e. prever todos los posibles valores, incluyendo casos en los que no es posible comprobarlo (desconocido o no especificado).
4. Debe explicitarse el orden de ciertos valores.
5. Deben especificarse características comunes.
6. Debe permitirse un nivel de profundidad variable.
7. Deben permitirse la extensión semántica o la adición de nuevos conceptos.

⁷⁶ Aunque las descripciones teóricas de este proceso hablan siempre de «espacio de n -dimensiones», por la limitación física que impone el soporte físico de la representación, estas se reducen a dos o, como mucho, a tres.

II.4. CONCLUSIÓN. METODOLOGÍA ELEGIDA

Recapitulando, nuestra intención ha sido ofrecer un panorama de los fundamentos teóricos presentes en cualquier proceso de *RI* (*cf. supra*, II.1), seguida de los condicionantes impuestos por la tecnología de que disponemos para aplicarlos (*cf. supra*, II.2). En la última sección (*cf. supra*, II.3), hemos presentado las técnicas que vamos a aplicar efectivamente en el texto, analizando su utilidad de acuerdo con los principios expuestos anteriormente: la técnica del *topic modeling*, basada en el llamado *análisis semántico latente*, así como la codificación mediante *XML* de diversos datos relativos a la estructura de los elementos que componen el texto.

III. ANÁLISIS DEL TEXTO

Nuestra descripción de los métodos de *RI* es aplicable a cualquier tipo de texto. Sin embargo, el objetivo de nuestra investigación ha sido estudiar el modo en el que un *SRI* puede mejorar el estudio de los textos de la literatura grecorromana, de modo que hemos escogido como ejemplo práctico sobre el que poner a prueba los métodos reseñados una obra en concreto, las *Historias* de Heródoto¹. De los numerosos aspectos susceptibles de estudio en la obra de Heródoto, nos hemos centrado en el «contenido» de la obra, es decir, en la «información» o «informaciones» que pueden generarse a partir de los «datos» presentes en ella, siguiendo la definición de estos conceptos que establecimos en la introducción (*cf. supra* I.2. *Estructura del significado: datos e información*).

III.1. LAS *HISTORIAS* DE HERÓDOTO

Las *Historias* de Heródoto son una obra ideal para estudiar la aplicación de los métodos de *RI* descritos anteriormente, por la especial relación entre forma y contenido que se da en ella. De hecho, la utilización de un hilo argumental para entretener una serie de noticias de muy diversa índole es una de las razones por las que se la considera fundadora del género historiográfico. Por otro lado, además de un conjunto de informaciones históricas o simplemente curiosas sobre hechos, personas y lugares, Heródoto entretiene su opinión personal sobre las causas profundas de los conflictos, sobre el comportamiento humano, sobre el cambio constante, aprovechando hábilmente los recursos literarios a su alcance. Desde el punto de vista de nuestra investigación, esto nos ha llevado a abordar la tarea impuesta desde una perspectiva triple: *a)* detectar las unidades temáticas o informaciones expresadas en la obra; *b)* identificar el mayor número posible de datos asociados a cada una de estas unidades; *c)* escoger de entre estos datos aquellos que resulten más adecuados para su uso como marcadores del significado mediante el etiquetado descriptivo que ofrece el lenguaje *XML*.

Si bien no hemos encontrado investigaciones anteriores a la nuestra que aborden el estudio de Heródoto desde el punto de vista de los *SRI*, no deja de ser inmensa la biblio-

¹ Aunque Schrader 1988 traduce el título de la obra en singular, *la Historia*, probablemente a raíz de la frase inicial, «ἱστορίας ἀπόδεξις τάδε» (Hdt. 1, 1), hemos decidido mantener la forma plural con la que habitualmente es denominada.

grafía en torno a uno de los escritores más famosos e importantes de la Antigüedad. Cuestiones como la temática de la obra, su estructura formal, o los recursos literarios empleados por el historiador nos han proporcionado una base muy sólida de la que partir. En particular nos hemos apoyado mucho en la investigación sobre la segmentación del texto. Como ya mencionamos, la segmentación es una técnica de *RI*, inseparable del concepto moderno de escritura como instrumento de comunicación y consistente en la separación visual de secciones de texto, de acuerdo con unos criterios². Su aplicación al texto de las *Historias* ha sido desde antiguo un tanto problemática, ya que interrumpe en varios lugares la cohesión del argumento, lo que ha dado lugar a otras propuestas de segmentación, más «lógicas». Este es un campo de trabajo perfecto para poner a prueba la capacidad del lenguaje *XML* para representar diferentes tipos de estructuras textuales.

III.1.1. EL AUTOR

Heródoto no da mucha información sobre sí mismo en las *Historias* (*cf. pr.* 1), por lo que la principal fuente de datos sobre su biografía sigue siendo la *Suda*³. Parece ser que nació en Halicarnaso⁴, y, por el nombre de su padre, Lixes, y de su paisano Paniasis, es posible que fuera de origen cario⁵. Habría apoyado una tentativa fallida de expulsar al tirano Lígdamis, lo que le habría valido un exilio temporal a Samos; más tarde volvería a su

² *Cf. supra II.1.3.1. Reducción.*

³ *Cf. Sud.* η 536, vol. 2, p. 588, 15 – 589, 2 Adler; π 248, vol. 4, p. 24, 20 – 25, 3 Adler. En lo referente a la persona de Heródoto, junto con el artículo que le dedica Schrader 1988, 505-508, seguimos la introducción que escribió Rodríguez Adrados a la traducción de Schrader (*cf. Schrader* 1992, 7-67) en Gredos, además de la introducción de Legrand 1932 a su edición del texto, por el detallado análisis de los datos a nuestro alcance que contiene. Remitimos asimismo a dos obras recientes de conjunto: *cf. Bakker, De Jong & Van Wees* (eds.) (2002); y *Dewald & Marincola* (eds.) (2006).

⁴ Las ruinas de Halicarnaso se encuentran hoy bajo la ciudad de Bodrum, en el suroeste de Asia Menor. Además de la *Suda* y del adjetivo Ἀλικαρνησέος presente en la mayoría de manuscritos (Hdt. pr.), la mayoría de autores antiguos apoyan el origen halicarnasense de Heródoto (*cf. infra* n.6 para una lista de excepciones); una inscripción en la base de un busto que adornaba la biblioteca de Pérgamo, construida por Eumenes II a mediados del s. II a.C. menciona a un Heródoto «de Halicarnaso» junto a Paniasis (*cf. IG* 12, 1, 145); por lo demás, el propio Heródoto conoce bien la historia local del lugar (*cf.* 1, 144), y siente una admiración un poco tendenciosa hacia Artemisa, princesa de Halicarnaso y una mujer excepcional que llegó a participar en los consejos del Gran Rey (*cf.* 7, 99; 8, 68-69; 87-88; 93; 101-103).

⁵ Es posible que Paniasis fuera incluso familiar de Heródoto (*cf. Sud.* η 536, que lo describe como ἑξαδελφός del historiador. Por otra parte, Legrand 1932, 8-9, cree ver en el tono positivo con el que describe a los carios (*cf.* 1, 171, 174; 5, 118-121; 8, 19, 22 y 5, 66) un reconocimiento silencioso de sus orígenes.

patria coincidiendo con el derrocamiento definitivo de Lígdamis, c. 454 a.C. No sabemos muy bien por qué, pero emigraría después a la colonia de Turio. Allí debió de ejercer algún tipo de actividad destacada, motivo por el cual parte de la tradición literaria posterior le asignó el gentilicio Θούριος⁶.

Pasó buena parte de su vida viajando y entre sus destinos estuvo Atenas, donde es posible que formase parte del círculo de Pericles⁷. Miembro de este grupo fue Sófocles, quien dedicó una oda a Heródoto (cf. Anth. Lyr. 1, 3, 79) y que además hace referencia varias veces a las *Historias*⁸. El retrato que Heródoto hizo de Adrasto (cf. 1, 34 ss.) y Jerjes (libros 7 y 8) ha dejado una influencia importante en la tragedia (cf. Meister *et al.* 2006). Su andadura le llevó a visitar los lugares de Grecia donde se produjeron las batallas más importantes de las guerras médicas, como las Termópilas, Platea o la llanura de Maratón. Sin embargo, es más conocido por las «excursiones» que realizó fuera de territorio heleno, a Cirene, a la Magna Grecia, a la región del Mar Negro, a Olbia, (cf. 4, 17), o subiendo el curso del río Hípanis hasta la región controlada por los escitas (cf. 4, 81). Recorrió Oriente Próximo, donde visitó Tiro, la capital fenicia, (cf. 2, 44), observó de primera mano el Éufrates (cf. 1, 185) y Babilonia, al menos en parte (cf. 1, 178 ss.), aunque no llegó a Persia. Pero el más famoso de sus viajes, y el que dejó en él una impronta mayor fue la visita a Egipto, país al que dedica casi buena parte del libro segundo.

Meister *et al.* 2006 y Legrand 1932, 11-12, se basan en la edad aproximada de Heródoto cuando estalló la guerra del Peloponeso para datar su nacimiento y su muerte, ya que algunos de los primeros sucesos del enfrentamiento entre atenienses y lacedemonios encontraron un eco en las *Historias* (cf. 7, 91; 7, 137, 233; 9, 73)⁹. Por otra parte, si se considera

⁶ La presencia de Heródoto en Turio está confirmada por Strab. 14, 2; 16; Plut. *De ex.*, 13; sobre la actividad que la Suda le atribuye durante dicha estancia (cf. Sud. π 248), Schrader 1988, 506, da por hecho que Heródoto obtuvo la ciudadanía, aunque para Legrand 1932, 12-15, eran necesarios unos requisitos que no podemos estar seguros de que el historiador cumpliera.

⁷ Como indica Schrader 1992, 22, no se puede afirmar con total seguridad la pertenencia de Heródoto al círculo de intelectuales que se formó en torno al líder ateniense. Sobre los viajes de Heródoto, cf. Jacoby 1913, col. 247 ss.; Hauvette 1894, 16 ss.

⁸ Cf. Soph. *Ant.* 903 = Hdt. 3, 119.

⁹ El asesinato de los embajadores espartanos enviados a Tracia a manos de los atenienses (cf. 7, 137) es una de las pocas menciones claras que hace Heródoto a la guerra «κατὰ τὸν Πελοποννησίων καὶ Ἀθηναίων», es decir, a la guerra del Peloponeso. El autor explica las causas de este conflicto bajo su habitual paradigma de «falta-reparación»: la ira de Taltibio se habría extendido de «forma sobrenatural» (θειότατον) sobre los descendientes de estos mensajeros, exigiendo una reparación proporcional. Para Wilson 2015b, 645-

que algunos pasajes de los *Acarnienses* de Aristófanes son una parodia de la obra herodotea (cf. *Ach.* 523 ss., y *Hdt.* 1, 4), la fecha de publicación se retrotraería, y con ello la de la muerte del historiador¹⁰.

III.1.2. HERÓDOTO EL HISTORIADOR

La variada naturaleza de las noticias que transmite Heródoto, unida al efecto negativo que tuvo la comparación con Tucídides en cuanto historiador influyeron negativamente en la opinión que se tenía de Heródoto en la Antigüedad. Se le acusó de plagio de Hecateo¹¹, su predecesor, y sobre todo, de «mentiroso»; en efecto, Plutarco, en su opúsculo *de Herodoti malignitate*, lo acusa de no haber dado una versión fiel a la realidad de ciertos hechos, o de haber inventado directamente su relato (cf. *Plut. Herod. mal.* 11; 33; 39; 46). Esta idea es la responsable de los términos despectivos con los que se ha definido la persona de Heródoto (μυθολόγος, λοποιός, ληρώδης, ψεύστης) o su actividad literaria (ψεύδестαι, μυθογραφείν, παραδοξολογείν, σχεδιάζειν)¹².

Más allá de la opinión de unos y otros, el mismo Heródoto explica con sus propias palabras los principios metodológicos que sigue en un par de pasajes: μέγχι μὲν τούτου ὄψις τε ἐμὴ καὶ γνώμη καὶ ἱστορίη ταῦτα λέγουσα ἐστὶ [...] προσέεται δὲ αὐτοῖσιν τι καὶ τῆς ἐμῆς ὄψιος (cf. 2, 99). Heródoto se siente obligado a transmitir la información que

646, lo que sigue tras la mención de los hijos de los espartanos muertos hasta el final de la sección es un pasaje insertado después y falto de una revisión que lo integrara correctamente en su contexto.

¹⁰ Meister 1990, 210, n. 14, recoge una lista de investigadores favorables a una fecha de publicación tardía. Sobre las posibles referencias literarias a Heródoto en Aristófanes, Legrand 1932, 22-24, duda de la autenticidad de algunas y cree que en general sólo demuestran que parte de las *Historias* se habían publicado ya en época del cómico.

¹¹ Un ejemplo de esta acusación es el *Περὶ τῆς Ἡροδότου κλοπῆς* del gramático Polión, donde este afirma que Heródoto tomó de la *Periégesis* de Hecateo la descripción del fénix, del hipopótamo y del cocodrilo.

¹² Legrand 1932, 147-149, juzga a Heródoto un pensador poco profundo y considera que el verbo σχεδιάζω indica en este caso una exposición superficial de los hechos, como hacía ya Diodoro (cf. *Diod.*, 1, 38): περιφανῶς ὁ συγγραφεὺς σχεδιάζων εὐρίσκειται. Sin embargo, resulta muy difícil concebir la posibilidad de que Heródoto se hubiera «inventado» la enorme cantidad de referencias que se encuentran en las *Historias*. Esta ha sido, sin embargo, la opinión de muchos investigadores, incluso en tiempos recientes, como West 1985, Hartog 1988, Hall 1991 y Armayor 1985, apoyados en el trabajo de Fehling 1971; en contra de esta opinión, remitimos a la crítica detallada que hace Pritchett 1993 de algunos de los pasajes utilizados por los partidarios de la «deshonestidad» herodotea.

recibe¹³, y el resultado de la aplicación de estos principios es el hecho de que con frecuencia se expongan tradiciones divergentes o contradictorias, una junto a la otra. Lejos de restarle valor a su trabajo, esta forma de exposición representa quizá el único logro «erudito» de Heródoto, su capacidad para dar una forma definida y una función narrativa a todo un *maremagnum* de datos orales de diversa clase y procedencia.

Con respecto a las fuentes, la proporción de datos que Heródoto recoge procedentes de testimonios orales es bastante superior a los derivados de documentos escritos. Ejemplos de este último grupo son algunas inscripciones, como la «columna de la serpiente de Delos»¹⁴, las recopilaciones de oráculos y algún pasaje de Hecateo¹⁵. No hay constancia de que utilizara listas de vencedores o de cargos públicos como referencia¹⁶, y, por otra parte, su desconocimiento del idioma le habría impedido utilizar los registros persas y egipcios. Con todo Heródoto no siempre es claro estableciendo el origen de la información que transmite.

Además de estas fuentes, no debemos olvidar que, como observa Schepens 1975, Heródoto es un historiador «de campo», y que en muchas ocasiones su relato se apoya en observaciones personales (τῆς ἐμῆς ὄψιος) y en su propia experiencia. Esto es más frecuente en las partes de naturaleza etnográfica o geográfica, mientras que las partes «históricas» se sostienen sobre la tradición oral (cf. Meister 1990, 211, n. 34), a la que había tenido acceso, según Lendle 1992, 44, a partir de eruditos locales (2, 28, 1; 125, 6; 4, 76, 6; 8, 65, 6) o de ciertos colectivos, como sacerdotes, los habitantes de un lugar (los egipcios, los escitas, los cartagineses, etc.), o de una ciudad (los atenienses, los corintios o los Cireneos, por ejemplo).

¹³ Cf. 7, 152: ἐγὼ δὲ ὀφείλω λέγειν τὰ λεγόμενα, πείθεσθαι γε μὲν οὐ παντάπασιν ὀφείλω, καί μοι τοῦτο τὸ ἔπος ἐχέτω ἐς πάντα λόγον. Para Schrader 1992, λόγος indica aquí la «totalidad» de la obra; Powell 1938, s. v. λόγος, 10, cita otros cinco pasajes donde se emplea el término de modo semejante: 1, 5, 3; 1, 95, 1; 2, 123, 1; 4, 30, 1; 7, 171, 1.

¹⁴ Cf. Hdt. 8, 82. El texto de la inscripción en la que se basa este pasaje ha sido publicado por Meiggs & Lewis 1969, 27.

¹⁵ Por ejemplo, el pasaje del libro 2, 70-73, parece proceder de Hecateo (= FGH 264, fr. 324 b).

¹⁶ Estas listas serían la fuente de los nombres de los trescientos espartanos fallecidos en las Termópilas o de los capitanes jonios que participaron en la batalla de Salamina (cf. 7, 224; 8, 85).

III.1.3. IMPORTANCIA DE SU OBRA

La influencia de Heródoto en la literatura grecorromana posterior ha sido enorme, como muestra la tesis de Riemann 1967, o Meister 1990, 40 ss. Por un lado su obra contiene, en una forma embrionaria, los elementos retóricos, dramáticos y pragmáticos que configurarían el género historiográfico durante el periodo helenístico, y dio un impulso al desarrollo de obras especializadas de historia, como la de Helánico y Antíoco¹⁷. Aunque no gozó de un gran prestigio como historiador en la Antigüedad¹⁸, su obra debía de ser bastante conocida, a juzgar por la noticia que tenemos sobre un epítome en dos libros de las *Historias* escrito por Teopompo de Quíos (cf. *FGH* 115, fr. 1-4). Su estilo vívido y sencillo (cf. *D. H. Pomp.* 3, 11 y *Th.* 23) le valió el aprecio de muchos como un buen narrador.

En época bizantina Heródoto mantuvo su renombre y no le faltaron imitadores, como señala Rodríguez Adrados en su introducción a la traducción de Schrader 1992, 66. Al menos existieron dos recensiones de las *Historias* durante la Edad Media, de las que derivan las dos familias de manuscritos que hoy tenemos (cf. *infra* III.2. *El texto de las Historias*). En el Renacimiento el hecho más destacado es la traducción latina de Lorenzo Valla (realizada entre 1432 y 1456, y publicada en 1474) así como la publicación de la *editio* aldina (Venecia, 1502). Sin embargo, el conocimiento de la obra de Heródoto después de la Antigüedad casi siempre fue parcial y procedente de fuentes indirectas, como revela la producción literaria en lengua vernácula inspirada en alguno de los temas y personajes de las *Historias*¹⁹.

Por desgracia, hasta bien entrado el siglo XX aún se le consideraba un «inventor de historias», y solo la investigación de las últimas décadas ha conseguido sacar a Heródoto de

¹⁷ El gramático Helánico es responsable de algunos escolios a Homero, cuya calidad hace perder el interés por conocer lo que pudo decir del texto de Heródoto; los fragmentos conservados han sido publicados por Montanari (ed.) 1988, 45-73. Para una revisión más detallada del interés por Heródoto en la Antigüedad remitimos a Wilson 2015b, xxi-xxii.

¹⁸ Sobre la opinión que disfrutó Heródoto en la Antigüedad añadimos la referencia tácita de *Th.* 1, 22 al testimonio ya referido de Plutarco. En este sentido Aristóteles refuta algunas de las afirmaciones de Tucídides (cf. *GA* 3, 22, 522a), pero también se sirve de Heródoto cuando le conviene (cf. *HA* 6, 31, 579b; *Rh.* 3, 9, 19; *Ath.* 14).

¹⁹ Las traducciones son a veces la única vía de comunicación entre los autores antiguos y los escritores de épocas posteriores. En el caso de la literatura española, Heródoto parece haber sido conocido sobre todo a partir de la traducción de Justino. Sobre el influjo de Heródoto en nuestra literatura, cf. Berenguer Amenós 1960, Fernández Galiano 1951, Lida 1949 y Reichenberger 1965. Sobre la recepción de Heródoto debe consultarse hoy Priestley & Zali (eds.) (2016).

la sombra que Tucídides arrojaba sobre él. Cada vez destaca más la amplitud de su concepto de *ιστορίη* y su cuidada atención a la dimensión antropológica del devenir histórico. Lejos de ser un mero compendio de noticias curiosas mal hilvanadas, las *Historias* son una obra bien pensada, que responde adecuadamente a los objetivos para los que se planteó, con las herramientas existentes en el momento de su creación. El avance en otras áreas de la ciencia (psicología, antropología, sociología) ha sido fundamental a la hora de comprender correctamente cuáles eran estas herramientas, como ya hemos visto (*cf. supra I.3.2. El texto oral*).

No se sabe mucho sobre los modelos literarios de las *Historias*. Es posible rastrear en ellas características presentes ya en la logografía jonia anterior, tales como periplos, el interés por la genealogía o por el origen y fundación de ciudades. Además de los logógrafos, Heródoto bebe de la sabiduría popular²⁰, que combina con un buen conocimiento de la tradición literaria griega, a juzgar por las referencias literarias presentes en su obra²¹. La influencia de otros escritores se deja ver también en el lenguaje de Heródoto, donde encontramos formas y giros épicos, fórmulas de corte sentencioso al estilo de elegíacos como Teognis (*cf. Thgn. 43-52*), reminiscencias de los poetas líricos, y especialmente de Homero²².

El principal mérito de Heródoto como escritor es haber dotado de un marco cronológico definido (las guerras médicas) y de un hilo argumental (las causas y la evolución del enfrentamiento entre griegos y bárbaros) a un conjunto de informaciones de muy diversa clase. Como afirma Rodríguez Adrados, se trata del primer escritor en prosa relevante de la literatura occidental (*cf. Schrader 1992, 23*). En efecto, a pesar de las expectativas que genera en el lector un título como las «Historias», la obra de Heródoto, más que una

²⁰ Sobre la influencia de los poetas gnómicos y de los dichos populares en las *Historias*, remitimos a Legrand 1932, 149; además de la bibliografía que allí añade, debemos mencionar el trabajo de Aly 1969 sobre la relación de las *Historias* con el género del cuento popular.

²¹ Heródoto menciona explícitamente obras épicas como la *Ilíada* (2, 116; 117), la *Odisea* (2, 116; 4, 29), y los *Epígonos*, una epopeya del ciclo tebano presuntamente escrita por Homero (4, 32); conoce la obra de Hesíodo (2, 53; 4, 32) y la *Arimaspea* de Aristeas de Proconeso (4, 14); hace referencia a una tragedia y a algunos poemas de Esquilo (2, 156); de los poetas líricos, además de Simónides de Ceos (5, 102; 7, 228) y Píndaro (3, 38), cita a Arquíloco (1, 12), Solón (1, 29-32; 34; 86; 2, 177; 5, 113), Alceo (5, 95) y Safo (2, 135).

²² *Cf.* la selección de pasajes en los que Heródoto copia a Homero hecha por Legrand 1932, 151: 3, 14 < *Il.* 22, 60; 24, 487; 3, 182 < *Il.* 1, 217; 7, 159 < *Il.* 7, 125; 7, 11 < *Il.* 10, 173; 5, 106 < *Il.* 1, 552 y *Od.* 1, 64; 8, 121 < *Od.* 4, 350 e *Il.* 10, 413; 5, 97 < *Il.* 5, 62 y 63.

exposición científica, es un ejemplo de narrativa. Las semejanzas con la novela son varias, desde la riqueza de anécdotas y relatos cuya función primera parece ser la de entretener al público²³, hasta la alternancia entre la narración en tercera persona con diálogos y discursos en estilo directo (cf. Lang 1984,). En efecto, anécdotas, cuentos y demás aspectos «novelcos» están estrechamente ligados al esqueleto argumental de las *Historias*²⁴.

²³ Cf. Schrader 1992, 27-28. Parece ser que al menos algunos fragmentos de la obra de Heródoto fueron objeto de lecturas públicas, si hacemos caso a Eusebio (cf. Eus. *Chronica Arm.* 83). Una de ellas, realizada en Atenas, le habría proporcionado incluso un premio económico (cf. Diyll., *FGH* 73, fr. 3). Rodríguez Adrados, en Schrader 1992, 22, da por bueno este dato, citando a Zenón de Elea como ejemplo de otro autor que dio a conocer su obra mediante una exposición oral. Encontramos algunos ejemplos de pasajes destinados a captar la atención del público en 2, 121, donde se describe la ingeniosa idea del faraón Rampsinito para atrapar a un ladrón utilizando a su propia hija; la historia del anillo de Polícrates (3, 40-45), el carácter de Hipoclides (6, 126 ss.) o el almuerzo caníbal de Harpago (1, 117 ss.).

²⁴ Cf. Immerwahr 1966, 308 ss. Aly 1969 representa una postura un tanto extrema, al comparar las *Historias* con la *Vida de Esopo* o *Las mil y una noches*, como si la obra de Heródoto no fuera más que una suma de «novelas» enlazadas mediante un marco narrativo común.

III.2. EL TEXTO DE LAS *HISTORIAS*

Nuestro resumen de la transmisión del texto se basa fundamentalmente en Wilson 2015b, v-xiii y Wilson 2015, xi-xxvi. El número de manuscritos que contienen parte o la totalidad del texto de las *Historias* ronda los sesenta. La relación entre ellos no está completamente clara, pero por razones prácticas se puede reconstruir un arquetipo con dos ramas descendientes, denominadas habitualmente «familia romana» y «familia florentina», por la procedencia de los manuscritos que les dan origen²⁵.

El códice *Laurentianus Mediceus* LXX, 3, designado mediante la sigla A, suele ser considerado el más importante y el mejor conservado, especialmente en cuanto al texto del libro primero, y es el que da nombre a la familia florentina. En los manuscritos de la familia romana, por otra parte, este libro ha sido abreviado y su principal representante, el códice *Vaticanus gr.* 2369 (sigla D), ha perdido algunos folios. La rama de la familia romana se subdivide en dos vertientes, una representada por D, ya mencionado, y otra representada por el ancestro común de otras copias, cuyo texto, referido mediante la sigla r en las ediciones consultadas, puede reconstruirse a partir de los testimonios proporcionados por los códices *Vaticanus gr.* 123 (sigla R), *Vaticanus Urbinas gr.* 88 (sigla U), *Vindobonensis hist. gr.* (sigla V) y *Vaticanus gr.* 122 (X).

La colación realizada por Wilson ha revelado que el texto de U y X está menos abreviado que el de sus parientes, de modo que sus lecturas han sido tenidas más en cuenta que en anteriores ediciones. De modo semejante, el *Laurentianus Conventi Soppressi* 207 (C), el *Parisinus gr.* 1633 (P), el *Laurentianus plut.* LXX, 6 (T) y el códice *Cantabrigiensis, collegii Emmanuelensis* 30 (S), obra del prolífico copista Andrónico Calisto, han permitido iluminar algunos pasajes oscuros.

Con respecto al testimonio de los papiros, hasta 1965 sólo se conocían veintidós que contuvieran algún fragmento de Heródoto, recogidos y publicados por Pack 1965, 45, n° 462-483, entre los que se incluye el único fragmento conocido de las *Λέξεις* que Aristarco escribió sobre las *Historias* (cf. *P. Amherst* II, 12). En 1981 se publicó el volumen 48 de los papiros de Oxirrinco (cf. *Chambers et al.* 1981, 22-73), que contenía quince papiros más y

²⁵ Wilson observa que Maas estaba firmemente seguro de que este era el *stemma* correcto (cf. Maas 1957, 30), ya que tuvo acceso a las anotaciones que hizo el propio Maas sobre el *stemma* en un ejemplar de la edición de Hude en su poder (cf. Wilson 2011).

restan por publicar otros cuarenta y cinco en el próximo volumen de la misma serie²⁶. Su aportación al establecimiento del texto se limita en la mayoría de los casos a confirmar alguna conjetura (cf. Wilson 2015b, xii, xiii).

III.2.1. Ediciones críticas

Después de la *editio princeps* de Aldo Manucio, Estéfano publicó su propia edición en París (1570), a la que siguieron muchas más; sin embargo, no podemos hablar de ediciones críticas hasta la publicada por Gronovio en 1715. A partir de esta, el trabajo de la filología alemana entre los siglos XVIII y XIX sentó las bases de las ediciones actuales, con ediciones como la de Schweighäuser, Gaisford, Bakker, Dindorf y, sobre todo, la de Stein 1893-1908.

La relación de las ediciones modernas resulta más comprensible si se la divide en función de su procedencia²⁷. En el ámbito francés hay que destacar la edición bilingüe de Legrand 1932-1954, publicada en el marco de la *Collection des Universités de France* y precedida de una extensa introducción especialmente interesante para la biografía de Heródoto. La filología anglosajona ha mostrado un interés creciente por Heródoto en el último siglo, que se ha plasmado en nada menos que tres ediciones, además de en una abundante producción científica. La primera de ellas, en el marco de la colección *Oxford Classical Texts*, obra de Hude y un tanto desactualizada ya a pesar de su reedición posterior en 1927, ha sido revisada recientemente por Wilson. Para ello ha realizado una nueva colación de los manuscritos, además de tener en cuenta los últimos descubrimientos papiáceos. El resultado es una edición crítica en dos volúmenes (cf. Wilson 2015b y 2015c), con un comentario publicado aparte (cf. Wilson 2015). La edición bilingüe de Godley 1922, publicada en el marco de la *Loeb Classical Library*, no es una edición crítica, pero es obligatorio mencionarla por cuanto constituye la base de la única versión digitalizada libre de derechos de autor disponible actualmente (cf. *infra*, III.3.1.2. *Perseus* y *el primer Heródoto en XML*). Terminamos este breve recorrido por las ediciones herodoteas volviendo al entorno alemán, donde Rosén 1987 y 2008, publicó una edición de las *Historias* que destaca por el tratamiento de la conocida y problemática *variatio* ortográfica de

²⁶ Sobre la contribución de los papiros al establecimiento del texto herodoteo remitimos a Papp 1948, así como a la revisión más reciente y actualizada realizada por West 2011.

²⁷ El listado que ofrecemos se basa en Wilson 2015, xi-xii, así como en Schrader 1988, 530.

Heródoto, manteniendo la exhaustividad y el rigor formal de la filología clásica alemana²⁸. En España la obra del historiador está siendo editada en el marco de la colección *Alma Mater (Colección de Autores Griegos y Latinos)*, hasta el libro tercero²⁹.

²⁸ Según Wilson 2015b, xii, la edición de Rosén representa «a step backwards». Para una lista de sus desaciertos remitimos a Renehan 1991.

²⁹ Cf. Berenguer Amenós 1960b y 1971 (libros primero y segundo), y Rodríguez Adrados 2011 (libro tercero).

III.3. ETIQUETADO *XML*

Pasamos ahora a describir nuestra aplicación del lenguaje descriptivo *XML* para codificar información presente en las *Historias*, como una técnica de *RI* que facilite la recuperación y procesamiento posterior de los datos. Comenzamos por abordar el estado actual de la digitalización de textos clásicos, observando su evolución y las soluciones propuestas. Nos centramos en la posición que ocupa Heródoto en este proceso y concluimos analizando en detalle las opciones que ofrece *XML* para codificar características del texto, a través de la normativa publicada por la *Text Encoding Initiative*.

III.3.1. DIGITALIZACIÓN DE LOS TEXTOS CLÁSICOS

Para poder aplicar los métodos de *RI* que hemos elegido es necesario una versión digitalizada del texto. Actualmente existen dos versiones digitales de las *Historias*, una realizada por el equipo del *TLG*, en código beta, y otra disponible en *Perseus*, codificada en *XML*.

III.3.1.1. Heródoto en el *TLG*

El *TLG* es un proyecto de investigación de la universidad de California, y fue fundado en 1972 con el deseo de producir el primer gran corpus digital de textos literarios de la Antigüedad clásica escritos en griego³⁰. Se trata de la mayor biblioteca digital a nuestro alcance, donde es posible consultar la mayoría de textos conservados desde el siglo VIII a.C. hasta la caída de Constantinopla en 1453. Su longevidad le ha permitido experimentar de primera mano la evolución acelerada de la tecnología de los últimos cincuenta años: las primeras versiones del *TLG* se distribuían en cintas magnéticas, que más tarde pasaron a ser CD-ROM, hasta llegar finalmente a la versión actual del *TLG*, accesible a través de Internet.

A pesar del término *thesaurus*, desde el comienzo el *TLG* fue concebido como una *base de datos textual*, centrada en la preservación y difusión de textos literarios en griego antiguo. Aunque el marco temporal que se fijó en un primer momento, desde Homero hasta el siglo II d.C., fue modificado relativamente poco después, el resto de normas sobre su

³⁰ Lo que sigue procede de la información disponible en la propia página web del *TLG* (<<https://www.tlg.uci.edu>>).

funcionamiento no han cambiando: no incluye testimonios epigráficos ni papirológicos, así como tampoco el aparato crítico del texto. Una de las complicaciones técnicas más importantes que tuvo que afrontar el *TLG* fue la codificación de los caracteres griegos. Para ello, Packard desarrolló un sistema denominado *código beta*, que utiliza caracteres latinos y símbolos no alfabéticos para representar el alfabeto griego y otros caracteres presentes en los textos. Es una solución pragmática a un problema que la extensión del estándar *Unicode* ha eliminado casi por completo, aunque no hay indicios de que el *TLG* tenga pensado abandonar el *código beta* en un futuro próximo (cf. *supra* II.2.2.1. *Código Beta* y *Unicode*).

El texto de las *Historias* forma parte del núcleo primigenio de autores y obras codificadas por el equipo de Irvine, y como tal está incluido en la lista de autores desde la primera versión del CD-ROM del *TLG* (letra D), accesible en <<https://www.tlg.uci.edu/>>. La versión del *TLG* es una digitalización de la edición de Legrand. Por desgracia, el texto digital que ofrece el *TLG* presenta algunas desventajas importantes para su utilización en nuestro trabajo, empezando por el hecho de que el uso del *código beta* limita mucho el rango de programas informáticos que pueden utilizarse para trabajar con él. Esto, unido al hecho de que la Universidad de California ostenta todos los derechos sobre esta versión digital, prácticamente impide cualquier tipo de modificación o procesamiento del texto fuera de las herramientas disponibles en el *TLG online*³¹.

III.3.1.2. *Perseus* y el primer Heródoto en XML

La biblioteca digital *Perseus* (*Perseus Digital Library*) fue creada en 1985 con la intención de ser una fuente de recursos para el estudio de las Humanidades bajo el título de *proyecto Perseus*, nombre por el que se la sigue conociendo hoy en día (cf. Smith *et al.* 2000)³². En un primer momento recogía únicamente material griego (textos e imágenes de testimonios arqueológicos y artísticos), pero pronto amplió sus horizontes para incluir material del mundo romano, historia de la ciencia e inglés actual. Además de recopilar material, *Perseus* mostró desde sus inicios un interés creciente por establecer conexiones entre los distintos tipos de fuentes, creando una infraestructura que no solo facilitara el

³¹ Entre 2003 y 2006, el *TLG* introdujo información sobre los lemas de todas las palabras contenidas en su base de datos, gracias a lo cual elaboró uno de los más potentes motores de búsqueda léxica sobre textos griegos (cf. Johnson 1994). Ello es una muestra de las posibilidades de *RI* que ofrece un uso inteligente de las *expresiones regulares* (cf. *supra*, II.2.3. *Expresiones regulares*).

³² De aquí en adelante nos referimos a la biblioteca digital *Perseus* simplemente mediante el término *Perseus*.

acceso a la información sino que al mismo tiempo reflejara la visión global que el proyecto tenía de la cultura antigua, como un complejo fenómeno que únicamente puede comprenderse estableciendo relaciones entre los testimonios que nos han llegado (*cf.* Crane *et al.* 2009).

Resulta claro que un repositorio de textos del tipo que pretender ser necesita algún tipo de regulación de acuerdo con un estándar que facilite el intercambio efectivo de información, la claridad y la interoperabilidad. El modo más fácil para conseguirlo es crear documentos conforme a los esquemas definidos en el estándar *TEI*³³. En efecto, *Perseus* hace un uso extensivo del lenguaje *XML* para la codificación de los textos que incluye; sin embargo, como veremos, todavía no termina de aprovechar todas las posibilidades que la normativa *TEI* ofrece³⁴. En nuestro estudio hemos utilizado el texto de las *Historias* disponible en *Perseus*; aunque es una digitalización del texto de Godley, que no es una edición crítica, es el único texto disponible libre de derechos de autor³⁵.

III.3.2. LA NORMATIVA *TEI* APLICADA A LAS *HISTORIAS*

Dee 2014 contiene un breve resumen de los recursos actualmente disponibles para aprender a entender y usar la normativa *TEI*³⁶. Aunque la mayoría de los usuarios de la

³³ El proyecto *Perseids: a collaborative editing platform for source documents in Classics* (<<http://sites.tufts.edu/perseids>>) es una iniciativa asociada a la *Perseus Digital Library*, centrada en la aplicación a la docencia de los recursos recopilados por esta última. Esta iniciativa defiende la participación activa del público para su desarrollo, para lo cual incluye un sistema de revisión por expertos que garantice el cumplimiento de los estándares como el *TEI*.

³⁴ Desde *Perseus* el esfuerzo por mantenerse actualizados con respecto a los avances de la tecnología es constante. En particular, el interés por la codificación de información sintáctica ha sido creciente en los últimos años (*cf.* Bamman & Crane 2009, Bamman *et al.* 2010 y Almas *et al.* 2013).

³⁵ En su momento nos planteamos modificar nosotros mismos la codificación *XML* de la versión de *Perseus*, introduciendo las modificaciones pertinentes con respecto a la edición más reciente de Wilson. La tarea se reveló mucho más compleja de lo que parecía a primera vista, de modo que nos hemos limitado a comparar el uso efectivo que hace *Perseus* del lenguaje *XML*, proponiendo mejoras en el sistema de etiquetado actual del texto.

³⁶ En esencia, el artículo de Dee 2014 es una reflexión sobre los resultados de una encuesta destinada a mejorar el apoyo del Consorcio *TEI* a usuarios noveles. Lo que sigue es un resumen de lo expuesto por esta investigadora.

normativa reciben una formación en persona, existen a disposición de cualquier interesado recursos para aprender a utilizarla, fundamentalmente páginas web³⁷.

Con respecto a los resultados de la encuesta, lo primero que salta a la vista son las críticas hacia la normativa *TEI*, en particular a su versión digitalizada, en áreas como la facilidad de búsqueda de información, la falta de contextualización y la interfaz. Estos mismos problemas, salvo el último caso, afectan a la versión «impresa», si acaso de una manera mucho más grave, por las limitaciones que el medio físico impreso impone de por sí. En general resulta difícil encontrar lo que uno necesita, especialmente si no se está seguro de cuáles son las necesidades específicas. Por otro lado, los ejemplos concretos de uso de esta normativa son escasos, y a menudo están desvinculados de su trasfondo teórico. Con todo, sigue siendo el mejor recurso disponible actualmente en términos de exhaustividad.

Los problemas de dificultad en la búsqueda de contenido específico dentro de las guías pueden deberse en parte a un enfoque demasiado «egocéntrico» por parte del Consorcio *TEI*. Con esto queremos decir que la guías han sido diseñadas para poner de relieve con su propia estructura algunas de las características de los documentos *XML*, en detrimento de otras técnicas de *RI*, quizá más antiguas, pero indudablemente efectivas. En efecto, la versión *online* de las guías consiste en una serie de documentos *XML*, estructurados de forma que sea patente el característico modelo *OHCO* del texto, con un índice al margen que permite saltar de una sección a otra. El contenido se divide en bloques anidados uno dentro de otro; la lógica es impecable, pero en términos de claridad descriptiva deja mucho que desear. Por otro lado, la única herramienta de búsqueda de información además de este índice es una lista de los atributos y elementos reseñados a lo largo de las guías, al final de la edición impresa. Como mínimo, debería ser posible consultar las guías de acuerdo con el género literario y el idioma (*cf.* Dee 2014 párr. 46)³⁸.

³⁷ La lista de recursos didácticos sobre el uso de normativa *TEI* es enorme: además del material disponible en la propia página web de la *TEI* (<<http://www.tei-c.org/Support/Learn/tutorials.xml>>), y de obras genéricas como la de Burnard 2014, podemos citar, entre otros, colecciones de material publicado por instituciones académicas como la Universidad de Brown (<<http://listserv.brown.edu/archives/>>) o la Universidad de Oxford (<<http://tei.oucs.ox.ac.uk/GettingStarted/html/>>); compitiendo con los anteriores en exhaustividad debemos mencionar a Van den Branden *et al.* 2010, así como otros recursos como Almas 2012, Mandell 2013, Chesley 2012, Spiro 2010. Algunas publicaciones se centran en la utilidad de la normativa para un público específico, como Halsell 2013 o Roueché & Flanders 2014.

³⁸ Burghart 2013 ha investigado formas alternativas de «navegar» a través de las guías para humanistas familiarizados con el propósito teórico de las guías y que simplemente busquen la etiqueta apropiada para una

Con respecto a la falta de ejemplos, los participantes en la encuesta mostraron interés por la creación de un repositorio de textos etiquetados que sirva de material didáctico. Una búsqueda en Internet a través de cualquier motor de búsqueda como Google proporciona rápidamente ejemplos de etiquetado *XML*. Por desgracia, desde su orígenes como una herramienta al servicio del lingüista y el estudioso de la literatura, las aplicaciones del lenguaje *XML* se han extendido a multitud de otros campos; en consecuencia, una herramienta de *RI* generalista como los motores de búsqueda mencionados, movidos por la fuerza de la mayoría, resulta poco útil para encontrar material producido por filólogos.

La necesidad de la que se hace eco esta encuesta, y que nuestra propia experiencia confirma, es la de crear un directorio ordenado de recursos específicos donde el interesado pueda definir sus necesidades, encontrar ejemplos de uso creados por usuarios con un objetivo similar, y expandir sus conocimientos a partir del material dado (*cf.* Dee 2014, parr. 48); en este sentido, aunque una de las aspiraciones de *Perseus* es precisamente convertirse en un «repositorio» de material codificado al alcance de cualquiera, su uso de la normativa *TEI* es bastante irregular y las referencias directas a la normativa *TEI*, inexistentes³⁹.

III.3.2.1. Objetivo y utilización de la normativa. Un marco abierto a la personalización

La normativa proporciona un esquema de codificación adecuado para describir una amplia gama de textos. Con este propósito, un esquema *TEI* admite diferentes y variados métodos para resolver problemas similares, y por ello define un conjunto de elementos mucho más grande de los que probablemente serían necesarios para un proyecto en particular. Además de estos, un esquema *TEI* puede ampliarse en el caso de que las herramientas recogidas en esta normativa no resulten adecuadas o suficientes para un caso dado⁴⁰.

situación concreta. La única sección realmente completa de este resumen está dedicada a la anotación de diversos elementos propios de un aparato crítico, y puede consultarse en <<http://marjorie.burghart.online.fr>>.

³⁹ En *Perseus* coexisten documentos conformes y no-conformes con los esquemas definidos por las guías, e incluso documentos cuya estructura es más similar a la de un documento *SGML*.

⁴⁰ Basamos nuestra exposición en lo recogido dentro de los epígrafes 23.3, 23.4 de la normativa (*cf.* *TEI-P5*, 700-714) con respecto a las modificaciones pertinentes, así como en iv.1 y la práctica totalidad del epígrafe 1 para las convenciones establecidas por la normativa *TEI* para la creación y manejo de esquemas *XML* (*cf.* *TEI-P5*, xvi-xxii; 1-17).

Estrictamente hablando la normativa *TEI* proporciona un juego de elementos y atributos, las reglas sintácticas según las cuales pueden combinarse, así como algunas recomendaciones sobre la interpretación que debe darse a las expresiones resultantes. En este sentido, constituyen tanto una *definición del tipo de documento* como una *declaración del tipo de documento* (cf. *supra* II.2.4.3. *Extensible Markup Language*). Podemos decir para especificar aún más lo ya dicho, que un esquema *TEI* o un esquema conforme a la normativa *TEI* es una lista de reglas combinatorias y de limitaciones que refleja un modelo abstracto del texto, o al menos lo intenta.

La normativa *TEI* se comporta como un marco que puede ser modificado según las necesidades del usuario. En algunos casos la descripción de un elemento es lo suficientemente amplia como para permitir la introducción de variantes sin necesidad de una modificación explícita del modelo utilizado⁴¹. Sin embargo, suele ser más frecuente establecer un margen de posibilidades limitado y bien definido antes de realizar el marcado. Para ello puede resultar necesario personalizar el esquema básico suprimiendo o añadiendo elementos, modificando clases de elementos, y cambiando su nombre⁴².

El esquema *TEI* está compuesto de un conjunto de declaraciones de clase y marcos, así como de un número determinado de módulos. Cada módulo contiene declaraciones sobre elementos y sus atributos (cf. TEI-P5, 701). La combinación de diferentes módulos produce un esquema. El usuario puede modificar el contenido de estos módulos para generar un esquema acorde con el modelo de texto en el que se basa su análisis de múltiples formas, pero para que el resultado pueda ser validado por la normativa *TEI* se contemplan estrictamente las siguientes posibilidades:

1. Eliminación de elementos concretos, de atributos o de valores posibles para uno o varios atributos.
2. Alteración del nombre o identificador genérico asociado a un elemento, sin que se pierdan por ello sus propiedades sintácticas o su valor semántico.

⁴¹ El elemento <note> posee un atributo llamado *type* que admite cualquier valor, lo que en la práctica se utiliza para generar diferentes tipos de notas (cf. TEI-P5, 700).

⁴² La normativa *TEI* prevé métodos para que un documento que utilice un esquema modificado pueda seguir siendo válido (cf. *supra* la definición de «validez» en el uso del lenguaje *XML* que vimos en el capítulo II.2.4.3. *Extensible Markup Language*; también TEI-P5, 701). En nuestro análisis no hemos descendido hasta este nivel de detalle técnico, sin embargo. Nos limitamos a ofrecer unas consideraciones teóricas sobre la dirección que deberían seguir los cambios para adaptarse mejor a los rasgos específicos de nuestro objeto de estudio.

3. Adición de nuevos elementos a una clase existente.
4. Definición de nuevos atributos o valores para atributos existentes, tanto para elementos particulares como para clases de elementos.

La piedra angular de la arquitectura conceptual del esquema *TEI* son los módulos. Cada uno de estos sirve para definir determinados elementos *XML*, así como sus atributos; a su vez, cada elemento puede ser asignado a una o más clases de elementos. Se trata de un sistema de clasificación doble que busca facilitar el proceso de desarrollo de un esquema de etiquetado *XML*, puesto que permite un amplio abanico de posibilidades combinatorias.

A pesar de esta flexibilidad, ciertos módulos son especialmente importantes y deben estar presentes más allá de toda modificación posible. Tal es el caso del módulo *{tei}*, que define las clases, macros y tipos de datos utilizados por el resto de módulos; el módulo *{core}*, que define elementos y atributos de uso general; el módulo *{header}*, necesario para que el documento final sea conforme a la normativa *TEI*; por último, el módulo *{textstructure}*, donde se definen elementos básicos para codificar la mayoría de objetos textuales o semejantes. Estos cuatro módulos son, por lo tanto, prácticamente omnipresentes, y así los mantendríamos en nuestro esquema «virtual».

La definición de un esquema *TEI* se realiza por medio de un documento *XML*, utilizando los elementos previstos para ello (cf. *TEI-P5*, §22 «Documentation Elements»). Este documento recibe de manera informal el apelativo *ODD*, a raíz de su objetivo original, «One Document Does it all». De cara a un uso profesional, y especialmente para facilitar su integración con algún tipo de aplicación informática, conviene «traducir» un documento *ODD* a algún lenguaje específico para la definición de esquemas utilizados hoy⁴³.

III.3.2.2. Nuestra propuesta: descripción de *estructuras* y *entidades*

El uso que hemos hecho de la información recogida por la normativa *TEI* viene marcado naturalmente por la tipología y características de aquellos rasgos que deseamos describir utilizando el lenguaje *XML*⁴⁴. Estos se pueden agrupar en los siguientes grupos:

⁴³ Los tres lenguajes de esquema más habituales hoy son el lenguaje *DTD XML*, el lenguaje *W3C Schema*, y el lenguaje *ISO RELAX NG*. En general, la terminología empleada por la normativa *TEI* está basada en este último.

⁴⁴ Seguimos las mismas convenciones tipográficas utilizadas por la normativa *TEI* para transcribir los nombres de los diferentes códigos *XML*: corchetes angulares para los elementos (*<q>*, *<hi>*); la arroba delante del nombre de un atributo (*@direct*, *@source*).

a) segmentación del texto, b) estructura argumental, c) entidades (personas (c1); lugares (c2); «objetos» [físicos y conceptuales]).

III.3.3. DESCRIPCIÓN DE *ESTRUCTURAS*

A continuación nos centramos en los recursos para describir relaciones entre los datos, así como la organización de los mismos en el texto, es decir, su *estructura*. Tras un breve preámbulo sobre la distribución del contenido de las *Historias*, dividimos nuestro resumen de las herramientas que ofrece la normativa *TEI* en dos partes, centrándonos en primer lugar en la codificación de la segmentación formal del texto, así como en los recursos para insertar notas, índices y referencias, bajo el epígrafe «descripción de estructuras simples»; en segundo lugar, bajo la denominación «estructuras complejas», nos detenemos en formas de codificar relaciones complejas entre secciones del texto y marcar referencias internas.

III.3.3.1. Estructura y contenido de las *Historias*

No está claro si la obra nos han llegado en un estado que el autor hubiera considerado definitivo, o si bien algunas partes se publicaron sin haber recibido una última revisión. En efecto, la composición de las *Historias* es anómala, hecho que ha llevado a algunos estudiosos a considerar que se trata de un agregado de diferentes obras escritas por separado y en diferentes momentos de la vida del autor. De este modo explican estos estudiosos el desajuste entre la primera parte de la obra, donde abundan los excursos etnográficos, y el relato más continuo y coherente de la segunda guerra médica en los últimos tres libros, unido a varias incoherencias argumentales y pasajes donde el autor promete o deja entrever una continuación que finalmente no se hace realidad. El principal sostén de esta corriente analítica es Jacoby 1913, que recoge los argumentos sobre los que se apoya⁴⁵: Heródoto habría sido originalmente un geógrafo o un etnógrafo del estilo de Hecateo de Mileto (cf. Riemann 1967), y como tal habría compuesto una serie de *λόγοι* etnográficos independientes, fruto del conocimiento adquirido en sus viajes. La estancia en Atenas, sin embargo, y el contacto con el círculo de Pericles le habría permitido desarrollar un concepto de historia

⁴⁵ Cf. Jacoby 1913, 152 (= col. 376), para quien resulta incoherente que Persia continúe siendo fuerte al final de la obra. Sobre la cuestión remitimos a Oswald 1995, esp. 47ss., que contiene un análisis más reciente de la cuestión, así como a Lendle 1992, 36-68.

global con el que unir aquellos y otros escritos de juventud en un único relato en torno a las guerras médicas⁴⁶.

Sin embargo, ni esta teoría ni la propuesta de Aly 1969 de considerar las *Historias* como una *Rahmenerzählung*⁴⁷ explican con suficiente firmeza las evidentes relaciones que se establecen entre las diferentes partes. Por el contrario, una lectura atenta revela la existencia de un propósito claro desde el principio por parte de Heródoto, y que por lo tanto la forma y estructura de la obra tal y como se ha conservado responde a la voluntad del autor⁴⁸. En efecto, las Guerras Médicas sirven de hilo conductor al resto de relatos, que establecen entre sí un complejo sistema de relaciones internas, de forma que unos sirven de introducción a la acción que va a producirse, o de preparación para una cierta interpretación de estos hechos⁴⁹. Así, los discursos y relatos destacan puntos decisivos para la trama principal o explican alguna cuestión.

En el proemio se exponen los elementos básicos que mueven la obra: la culpa o la responsabilidad (αἰτία) de una serie de «actos humanos» (τὰ γινόμενα ἐξ ἀνθρώπων) son los objetivos principales de la «investigación» (ἱστορίη) herodotea; no obstante, los criterios que sigue la «exposición» de los resultados (ἀπόδειξις), es decir, la narración propiamente dicha, quedan en el aire. Heródoto ofrece una concepción humana, casi personal, de la historia, donde los pensamientos de ciertas personas, y, sobre todo, las acciones que se derivan de aquellos tienen una importancia capital. En este sentido es la ὕβρις de los reyes la principal αἰτία del enfrentamiento entre griegos y bárbaros, y el hilo conductor de los seis arcos argumentales principales. El primer arco es breve, y representa una especie de introducción donde se mencionan las disputas entre griegos y extranjeros en época mitológica (cf. 1, 1-5). Las acciones del rey lidio Creso, πρῶτος τῶν ἡμεῖς ἴδμεν τοῦς μὲν κατεστρέψατο Ἑλλήνων ἐς φόρου ἀπαγωγῆν (cf. 1, 6, 2), representan el segundo

⁴⁶ Cf. Latte 1958, cuya teoría no deja de ser una expansión imaginativa de la hipótesis analítica básica. Una versión más simple, propuesta por De Sanctis 1926 y apoyada por Powell 1939 considera que el núcleo central de las *Historias* fue un relato logográfico centrado en Persia, unas Περιοδικά. Rodríguez Adrados ofrece un escueto resumen de los aciertos de esta idea (cf. Schrader 1992, 37-38).

⁴⁷ El término *Rahmenerzählung* se refiere a un «marco compositivo» para un conjunto de historias novelescas, originalmente separadas. Howald 1944, 11-45, se adhiere parcialmente a esta idea.

⁴⁸ Sobre el punto de vista unitario, cf. Cobet 1971 y Meister 1990, 32 ss.

⁴⁹ El iniciador de la corriente «unitaria» de análisis de las *Historias* fue Pohlenz 1961; otros investigadores alemanes, como Bornitz 1968 y Gottlieb 1963 han contribuido a la solidez de esta hipótesis con un exhaustivo análisis de las conexiones internas de la obra.

arco (1, 6-94), así como la primera α ι τ η «histórica», de la que derivan las acciones tomadas por sucesivos reyes bárbaros, protagonistas de los cuatro arcos argumentales restantes: Ciro (1, 141-214), Cambises (2, 1-3; 70), Darío (3, 61-7, 4), y Jerjes (desde 7, 5-8 hasta el final de la obra)⁵⁰.

Entretejidos en las líneas argumentales que acabamos de enumerar, Heródoto inserta numerosos relatos, noticias y pensamientos de muy diversa índole: geográficos, etnográficos, históricos, míticos o simplemente anecdóticos. Estos hechos «asombrosos» (θ ω μ α σ τ $\acute{\alpha}$), son igualmente fruto de su investigación y constituyen en núcleo de las numerosas digresiones que salpican sin cesar el texto de las *Historias*. Heródoto sigue una progresión narrativa *hodológica*; es decir, el hilo argumental principal funciona a modo de camino, donde los personajes, lugares o acontecimientos son una oportunidad para «dar un rodeo» y descansar o disfrutar de otro tipo de información⁵¹. A modo de ejemplo, podemos mencionar aquí que las acciones militares de Ciro sirven de excusa para hablar de los babilonios (*cf.* 1, 178-200) y los masagetas (*cf.* 1, 201-216); el deseo de conquista de Cambises brinda la ocasión de extenderse en un relato sobre Egipto (2, 2-182) y los etíopes (3, 20-24), y la campaña de Darío da pie al último gran excursus sobre pueblos bárbaros, dedicado a los escitas (4, 5-82).

Lidios, medos, persas y bárbaros en general no son los únicos protagonistas de la narración. La región de Jonia y sus habitantes son objeto de su propio relato (*cf.* 1, 142-151), así como la región de Samos (3, 39-60; 120-149). La historia de Atenas (1, 59-64; 5, 55-96; 6, 121 ss.), de Esparta (1, 65-68; 5, 39-48; 6, 51-84) se alternan con la acción en el bando persa, hasta la llamada «rebelión jonia» (5, 28 ss.), en la que convergen la narrativa griega y persa. En el relato de la segunda guerra médica, sin embargo, Heródoto vuelve a exponer los acontecimientos de forma paralela, alternando un bando y otro, y siguiendo el orden cronológico natural. Al fracaso de Darío en Maratón (6, 102 ss.) le sigue la campaña de Jerjes, desde la decisión de ir a la guerra de nuevo contra Grecia (5, 5 ss.), pasando por la leva de su enorme ejército (5, 59 ss.), las batallas de las Termópilas (7, 198-239), de Artemisio (8, 1-23) y Salamina (8, 40-96), hasta las victorias griegas en Platea (9, 19-89) y Micala (9, 90-107), que, junto con la captura de Sesto, marcan el final del enfrentamiento.

⁵⁰ Sobre la estructura de la obra remitimos a Jacoby 1913, 288ss., Wood 1972 y Waters 1985, 47 ss.

⁵¹ La *Ilíada* presenta un modelo de narración semejante: la exposición del conflicto entre Agamenón y Aquiles se ve interrumpida aquí y allá con digresiones de diferente clase y que, al final, sirven de apoyo y complemento a la historia principal (*cf.* Schrader 1992, 31-32; 38-39).

El interés moderno en las *Historias* no descansa en la exactitud o la veracidad histórica de la información que proporciona sobre egipcios, persas o escitas sino en la batalla conceptual subyacente⁵². La libertad y la esclavitud, la democracia y la tiranía, la medida y el deseo sin control se enfrentan a través de sus «avatares» humanos: Jerjes, Demarato⁵³, Creso, Leónidas, Temístocles etc. Un segundo *leitmotif* que atraviesa la obra es la naturaleza transitoria de todas las cosas, idea que se expresa con especial claridad al comienzo (1, 5; 1, 207, en la conversación entre Solón y Creso). Heródoto es un hombre de fuertes convicciones religiosas (cf. Legrand 1932, 131-139), que cree en la influencia de una divinidad (τὸ θεόν, ὁ θεός) movida por la «envidia». Esta se manifiesta castigando la ὕβρις humana, encarnada en la figura de Creso (1, 30-33), Polícrates de Samos (3, 39, ss.) o Jerjes (7, 35), o bien a través de vaticinios, sueños y oráculos. Sin embargo, el carácter agónico típicamente griego impulsa al historiador a destacar el papel de las acciones y pensamientos humanos en el desarrollo de las *Historias*, más allá de su creencia en un «orden divino» de las cosas (cf. Schadewalt 1982, 158).

Por otra parte, aunque represente un conflicto entre Grecia y el resto del mundo «no-griego», Heródoto muestra una marcada ecuanimidad, a la hora de hablar de los participantes de uno y otro bando. Si acaso, en ocasiones se revela más a favor del partido bárbaro que al contrario, especialmente en el caso de Egipto (2, 4; 32; 50; 58; 77; 82). No obstante, su favoritismo es sutil: así, reconoce que las campañas de Darío, Jerjes y Artajerjes fueron más nocivas para Grecia que todas las guerras de las veinte generaciones anteriores (6, 98). A su vez, denuncia las carencias más graves de los griegos, especialmente su falta de unidad, el gusto por la rivalidad entre particulares y entre ciudades, que llevó a muchas polis a unirse al bando persa en contra de sus propios compatriotas, algo especialmente patente durante la «rebelión jonia» (cf. 2, 565 ss.). Al mismo tiempo, expresa su admiración por el νόμος, por la tradición y las leyes antiguas, de las que tanto se enorgullecían los espartanos, cuya valentía alaba con frecuencia (cf. 7, 101-104, por ejemplo). Finalmente, sin embargo, los salvadores de Grecia fueron los atenienses, una opinión que el mismo reconoce era ἐπίφθονον μὲν πρὸς τῶν πλεόνων ἀνθρώπων, «odiosa para la mayoría». La

⁵² Al respecto, cf. Nenci 1988 y Rollinger 1993. En otro orden de cosas, la contribución de Heródoto al campo de la cronología es importante, como señalan Strasburger 1982b.

⁵³ Schadewalt 1982, 215, analiza el diálogo entre ambos personajes en 7, 101-104.

filiación política de Heródoto ha sido un objeto de interés importante, paralelo al estudio de las formas en las que el historiador expresaba su opinión personal⁵⁴.

A pesar de sus esfuerzos por mantenerse imparcial, su hostilidad hacia la tiranía es patente, como se puede apreciar en el largo discurso pronunciado por el corintio Socles a propósito de la intención de los lacedemonios de controlar Atenas mediante la imposición de un gobierno tiránico (5, 92), o en la relación que establece entre un gobierno próspero y la ausencia de tiranía (5, 78: ἀπαλλαχθέντες δὲ τυράννων μακροῦ πρότοι ἐγένοντο)⁵⁵. Excepciones a esta regla son Pisístrato, en el caso de Atenas, o Polícrates de Samos.

Rodríguez Adrados (en Schrader 1992, 38-43) describe algunos de los recursos empleados por Heródoto para la composición de su obra. Como Homero, el historiador utiliza una «composición abierta», es decir, que las unidades que componen el conjunto no están constreñidas por limitaciones en cuanto a su extensión, su orden o su temática⁵⁶. En este modelo de trabajo, las historias paralelas, las expansiones y las digresiones se utilizan para hacer énfasis en los aspectos que el autor desea destacar. Por otra parte, esta forma de componer puede reflejar las propias ideas del historiador sobre el funcionamiento de las *Historias*, que concibe como un cruce de voluntades, individuales y grupales. En lugar de exponer por separado y de modo racional las causas de los hechos, recurre al paralelismo entre personajes y relatos para expresarlas. Es una técnica propia de la composición literaria oral, basada en la asociación cognitiva y destinada a reforzar la memoria de lo que se cuenta⁵⁷. Un ejemplo de esto es la historia del anillo de Polícrates (3, 39-60; 120-125), insertada a modo de presagio entre la exposición de los crímenes de Cambises y su muerte (3, 64; cf. Rodríguez Adrados, en Schrader 1992, 40-41). Dentro de esta digresión se

⁵⁴ Strasburger 1982 fue de los primeros en demostrar que Heródoto no era abiertamente pro-ateniense (cf. Ostwald 1991).

⁵⁵ El tirano expulsado aquí es Pisístrato; otro pasaje donde el desprecio a la figura del tirano es patente se encuentra en boca de Ótanes, a propósito del poder unipersonal: ἕνα μὲν ἡμέων μούναρχον μηκέτι γενέσθαι· οὔτε γὰρ ἡδὺ οὔτε ἀγαθόν (3, 80).

⁵⁶ El tipo de composición opuesta a la «composición abierta» es la «composición cerrada», característica de la lírica o el teatro, géneros donde existen unas reglas claras que limitan el espacio o el orden de los elementos (cf. Beck 1971).

⁵⁷ Los recursos empleados por Heródoto para mantener la cohesión temática derivan del carácter oral del método de composición de la obra: repeticiones y paralelismos (cf. *supra* 1.2.3.6. *Las huellas de la μνήμη en el texto*).

enmarcan otras, creando una serie de composiciones en anillo sucesivas, que podemos describir utilizando una convención similar a la empleada por en lenguaje *XML*⁵⁸:

```
<LINEA PRINCIPAL: Cambises>
<DIGR.: Polícrates>
<DIGR.: Expedición lacedemonia>
<ALTR.: (versión samia) + (versión espartana)
<DIGR.: Periandro> (3, 48-54)
<FIN DIGR.: /Expedición lacedemonia>
</Polícrates> <APÉND.: Θώματα de Samos>
<FIN LÍNEA PRINCIPAL/Cambises> (3, 64)
<FIN DIGR.: /Polícrates> (3, 120-125)
```

Veamos ahora de forma expositiva qué representa este esquema. Después de que Polícrates recibiera una carta del faraón Amasis advirtiéndole contra la prosperidad excesiva, y la escena del pescador y el anillo, Heródoto vuelve atrás para contar un suceso anterior relacionado con el tirano: su enfrentamiento con los samios y la ayuda que estos recibieron de Esparta. Tras dar las dos versiones de las causas de esta ayuda, aprovecha la mención al resentimiento de los espartanos con Polícrates para contar la historia de Periandro, un cuento sobre las desgraciadas consecuencias del resentimiento. Tras mencionar en pocas líneas la derrota espartana, nos cuenta, a modo de apéndice, el final de aquellos samios rebelados, transformados en piratas y esclavizados finalmente por eginetas, con quienes tenían una cuenta pendiente. La serie de digresiones se interrumpe aquí para volver a la línea argumental principal, y se describe el final de Cambises. Heródoto retrasa el final de Polícrates, cuya historia había quedado inconclusa, para aprovechar la tensión del público por conocer el final y utilizarla para reforzar el papel ejemplar de la historia del tirano con respecto a la de Cambises. La ὕβρις como causa de un mal, que exige retribución, en la forma de un final perjudicial para el infractor es el tema común a todas estas historias, reforzado por la relación establecida entre ellas. El hecho de que se interrumpan entre sí contribuye, paradójicamente, a conseguir este efecto, aprovechando la «tensión no resuelta» generada en el público, que constantemente ansía conocer un final que tarda en llegar.

⁵⁸ Nótese la ruptura del encapsulamiento de las etiquetas que se produce tras la digresión sobre la expedición lacedemonia. Sobre los problemas que ocasionan este tipo de alteraciones para la aplicación de un sistema de etiquetado descriptivo, cf. *supra* II.2.4. *Lenguajes descriptivos*.

III.3.3.3. Estructuras simples

III.3.3.3.a. El sistema CTS-URN

El *Canonical Text Service Protocol (CTS)* es un sistema de notación de referencias textuales diseñado para mantenerse constante y funcional a pesar de la continua evolución de la tecnología. La idea surgió durante el desarrollo del proyecto *Homer Multitext*⁵⁹, como parte de la arquitectura *CITE*, una forma de organizar los datos del texto concebida por los técnicos informáticos del proyecto, Christopher Blackwell y Neel Smith⁶⁰. Ambos se dieron cuenta de la carencia de un modelo canónico de notación textual entre los humanistas digitales, que se veían obligados a utilizar otros sistemas, concebidos con el texto impreso en mente, pero que no satisfacen completamente la necesidad de precisión y la falta de ambigüedad que el medio digital exige.

Uniform Resource Names

El sistema *CTS-URN* se apoya en el uso de *Uniform Resource Names (URN)*, una forma de notación propuesta por Smith & Weaver 2009, que aspira a ofrecer referencias textuales únicas y persistentes, inspirado en el concepto análogo de *Uniform Resource Locator (URL)* que rige la asignación de referencias textuales a páginas web en Internet, algo que comúnmente se conoce como *direcciones web*. La segunda fuente de inspiración para este sistema es el modelo *OHCO* del texto que ya describimos al hablar del *XML* (cf. *supra* II.2.4. *Lenguajes descriptivos*).

Toda referencia *CITE-URN* localiza un pasaje dentro de dos jerarquías diferentes y complementarias. Dentro de la primera jerarquía se identifica el texto en cuestión dentro de un grupo de textos mayor que lo contenga. El criterio de agrupación de los textos puede ser la autoría (las obras de Cicerón o Plutarco), u otro rasgo en común. Un nivel por debajo del «grupo de textos» se encuentra el nivel de la «versión», un concepto que engloba las traducciones a otros idiomas de la obra o el pasaje, así como las distintas ediciones críticas disponibles, que para el sistema no serían más que una «versión en el mismo idioma».

⁵⁹ El *Homer Multitext* es una iniciativa promovida por Casey Dué y Mary Ebbott, y supervisada por G. Crane (cf. Crane 2014), que trata de representar de forma interactiva y visual la transmisión textual de la *Ilíada* y la *Odisea*, aprovechando la versatilidad del medio digital (cf. <<http://www.homermultitext.org/>>).

⁶⁰ Lo que sigue es un resumen de lo expuesto en la propia página web de los desarrolladores de la arquitectura *CITE* (<<http://cite-architecture.github.io/about/>>).

La segunda jerarquía identifica el pasaje dentro de la estructura lógica y formal de la obra: libros, capítulos, secciones, versos, estrofas etc. El sistema *CITE-URN* permite descender aún más e identificar una sección de texto inferior a estas unidades, o incluso una única palabra, si bien esto sólo es posible si la referencia incluye una mención a la *versión* (edición o traducción).

Veamos, por ejemplo, cómo se trasladarían una serie de referencias al texto de Heródoto desde el sistema que utilizamos en este trabajo, basado en las recomendaciones del *DGE*, y en el sistema *CTS-URN*:

```
Hdt. > urn:cts:greekLit:tlg0016
Hdt. 1 > urn:cts:greekLit:tlg0016.1
Hdt. 1, 16 > urn:cts:greekLit:tlg0016.1.16
Hdt. 1, 16-21 > urn:cts:greekLit:tlg0016.1.16-1.21
Hdt. 1, 16, 1, «Ἀρδύοις» > urn:cts:greekLit:tlg0016.1.16.1:@Ἀρδύοις
```

Una «dirección» *URN* se compone de una serie de campos, separados por dos puntos, en los que se ubica el pasaje desde el elemento mayor en la jerarquía hasta el menor. Los dos primeros campos son fijos y sirven para *declarar* que se trata de una *URN*, que sigue el protocolo *CTS*; el tercero identifica el pasaje dentro de un supergrupo de textos, en nuestro caso la literatura griega (*greekLit*); en cuarto lugar se identifica la obra y la versión, que en este caso es la edición digitalizada por el *TLG*, y que se corresponde con el número identificativo dentro de aquel corpus (*tlg0016*); se siguen, por último, las referencias al libro, sección y línea, separados estas tres últimas por punto. Toda referencia a una sección de texto o palabra concreta va precedida de una arroba (@).

III.3.3.3.b. Recursos XML para describir estructuras simples

α. Segmentación del texto

La normativa *TEI* describe elementos que pueden aparecer en prácticamente cualquier tipo de texto, que no poseen una estructura externa o interna demasiado elaborada y que habitualmente suelen identificarse o marcarse mediante convenciones tipográficas como alternancia entre mayúscula o minúscula, signos de puntuación o cambios en la disposición del texto (*cf.* *TEI-P5*, §3).

Párrafos

El párrafo es la unidad de organización visual básica de los textos en prosa modernos. El párrafo se comporta como la unidad jerárquica superior dentro de la cual se insertan el resto de elementos básicos descritos en la sección tercera de la normativa. Al respecto, esta distingue elementos a nivel de frase, que solo pueden aparecer en el interior de un párrafo y completamente contenidos en él (fechas, nombres, citas), de los *chunks*, secciones de texto que pueden aparecer entre párrafos, pero no en su interior (el propio párrafo, o el elemento `<ab>` [*anonymous block*]), que en determinados casos se utiliza como alternativa al párrafo, *cf.* TEI-P5, §16.3), y finalmente de elementos intermedios, que pueden aparecer dentro de un párrafo o entre dos párrafos (citas bibliográficas, notas, listas) (*cf.* TEI-P5, §3.1).

El párrafo se marca con el elemento `<p>`. Si se desea establecer algún tipo de subdivisión interna, pueden emplearse los elementos `<s>` o `<seg>` («segmentos»; *cf.* TEI-P5, §16 y §17, respectivamente), aunque se recomienda sólo en casos donde este tipo de división sea regular.

Puntuación

La normativa *TEI* contempla dos posibles causas de problemas para el marcado de los signos de puntuación. El uso de un conjunto de caracteres distinto del alfabeto latino empleado para el etiquetado es el más evidente; en el caso de un texto griego como el que nos ocupa, Unicode proporciona códigos de referencia que cubren nuestras necesidades. Asimismo, *Unicode* recoge bajo la categoría «glifos» algunos signos de puntuación con un sentido bastante amplio, tales como el punto, la coma o el guión (*cf.* TEI-P5, §3.2).

A la hora de codificar los signos de puntuación, resulta más interesante reflejar su función, en lugar de una simple declaración de su presencia. Los métodos para ello varían dependiendo del sentido particular de estos signos que se desee describir; por ejemplo, el epígrafe §16.3 incluye elementos útiles para indicar usos de la puntuación como delimitador de segmentos textuales, mientras que para ciertos usos del signo de exclamación o interrogación que no indican una separación de unidades textuales, sino la simple adición de un matiz semántico a un término completo («hoy comeremos ¿maíz?»), es mejor utilizar el elemento `<pc>`, *punctuation character* (*cf.* TEI-P5, §17.1).

Texto resaltado y citas

La definición de texto «resaltado» contemplada en la normativa *TEI* como «the use of any combination of typographic features (font, size, hue, etc.) in a printed or written text in order to distinguish some passage of a text from its surroundings» plantea algún problema cuando se trabaja con una edición de un texto antiguo (cf. *TEI-P5*, §3.3). En efecto, el deseo de abrazar el mayor número de circunstancias posibles que impregna la normativa *TEI* conlleva una cierta esclavitud con respecto a la «mayoría» de textos disponibles para el marcado *XML*: los textos transmitidos de la Antigüedad grecorromana representan solo una fracción del conjunto al que va dirigida la normativa, de modo que algunas de las características especiales que los definen no encuentran un reflejo inmediato en ella. Así, la mayoría de elementos recogidos en esta subsección del módulo §3 se refieren a métodos gráficos de resaltado, insistiendo mucho menos en otras formas de llamar la atención del lector sobre algún elemento del texto⁶¹.

Los textos antiguos, y especialmente el de Heródoto, inmerso en un cultura fuertemente oral, recurren a procedimientos retóricos, lingüísticos con la misma función de resaltado que las alteraciones tipográficas de nuestros textos actuales. Para etiquetar este tipo de rasgos, disponemos del elemento `<emph>`, *emphasized* (cf. *TEI-P5*, §3.3.2.2)⁶². Este elemento simplemente delimita una sección. Si se desea especificar más, las opciones recogidas por la normativa se limitan a la posibilidad de especificar algún tipo de cambio tipográfico en el texto así etiquetado, es decir, funciones procedimentales y no descriptivas.

Por otra parte, la normativa *TEI* dedica un apartado a secciones del texto atribuidas por el autor a una fuente o autor diferente, denominadas de forma genérica «citas» (cf. *TEI-P5*, §3.3.3). Para ello dispone una serie de elementos que podemos aplicar al texto de Heródoto sin modificación alguna, en especial dos: `<said>`, que junto con el atributo `@direct` puede servir para delimitar pasajes en estilo directo⁶³; `<quote>`, *quotation*,

⁶¹ La descripción del elemento `<hi>`, *highlighted*, se refiere explícitamente a elementos del texto diferenciados gráficamente del resto.

⁶² Relacionados con el elemento `<emph>`, los elementos `<foreign>` y `<distinct>`, denotan términos o expresiones diferentes del texto circundante en un sentido lingüístico (cf. *TEI-P5*, §3.3.2.3). En el caso de las *Historias* su uso es bastante limitado, aunque podrían resultar interesantes para un estudio lexicográfico del texto.

⁶³ Para el etiquetado de un texto en griego antiguo como el de las *Historias* excluimos dos elementos más recogidos en este módulo de la normativa *TEI*: el elemento `<q>`, *quoted*, (cf. *TEI-P5*, 75) no nos interesa porque describe secciones del texto que ya están delimitadas gráficamente como citas o estilo directo

enmarca un pasaje procedente de otra fuente, que puede explicitarse mediante el atributo *@source*. Pueden ser interesantes en menor medida, *<mentioned>* y *<soCalled>*⁶⁴. El primero resulta apropiado en un contexto de análisis metalingüístico, donde una palabra o palabras presentes en el texto son objeto de una reflexión, en lugar de cumplir su función lingüística propia (cf. *TEI-P5*, 76). El elemento *<soCalled>* designa pasajes de cuya responsabilidad se distancia el autor del texto, sin indicar su procedencia; puede servir para codificar denominaciones alternativas con las que Heródoto no se muestra conforme o, con más frecuencia, expresiones irónicas.

β. Cambios editoriales

Los elementos recogidos en este módulo permiten codificar aspectos del texto relacionados con la actividad del editor, tales como lagunas (*<gap>*, TEI 2016, 86), adiciones (*<add>*), eliminaciones (**) o pasajes de lectura incierta (*<unclear>*). De modo semejante incluye una serie de elementos que representan varios tipos de «elecciones» editoriales, tales como correcciones (*<corr>*) o regularizaciones de una forma anómala (*<reg>*); estas pueden agruparse mediante el elemento *<choice>* (TEI 2016, 83) para una mayor claridad (cf. *TEI-P5*, §3.4).

γ. Nombres, números, fechas y abreviaturas

Las recomendaciones recogidas en este módulo están concebidas especialmente para el etiquetado de un texto moderno. Permiten distinguir nombres, números o abreviaturas de la masa de texto circundante, pero se basan en la existencia de cierta regularidad y sistematicidad en la expresión de estos datos. Esto es especialmente patente en las direcciones y fechas. Es de sobra conocida la ausencia de un único calendario en la Grecia antigua, y las abreviaturas no son muy frecuentes en un texto antiguo. Los elementos *<date>* y *<time>* presuponen la existencia de algún tipo de sistema de referencia estandarizado que regule la expresión lingüística del tiempo⁶⁵, y requerirían una adaptación especial para los textos históricos antiguos (cf. *TEI-P5*, §3.5).

mediante guiones, sangría, etc. La presencia de esta clase de recursos visuales en un texto antiguo es obra de algún editor y, como tal, producto de un acto interpretativo o descriptivo, idéntico al que motiva el uso de un elemento *XML*; por lo tanto, su aplicación nos parece redundante. La ausencia general de un sistema de notación bibliográfica definido formalmente en la Antigüedad nos lleva a ignorar el elemento *<cit>*, *citation*, apropiado para citas bibliográficas modernas.

⁶⁴ Mediante el atributo *@who* es posible indicar quién está hablando en un pasaje en estilo directo.

⁶⁵ Cf. los ejemplos aducidos en TEI 2016, 97-98.

Con respecto a los nombres, el elemento `<name>` descrito aquí tampoco resulta adecuado para el texto de las *Historias*. En su lugar, resultan más interesantes los elementos descritos en el módulo §13, puesto que permiten un grado más profundo de análisis al tratar los nombres como la representación de entidades físicas y no como simples palabras⁶⁶.

δ. Anotaciones e índices

El elemento `<note>` permite la inserción de comentarios de diversa clase, tanto si se trata de codificar algo presente en el texto, como si se trata de un añadido posterior del etiquetador. Este tipo de elementos suele ir asociado a un término o sección del texto, denominado «punto de unión» (*point of attachment*, cf. TEI 2016, 108). Si el punto de unión es muy extenso, la normativa recomienda ubicar el elemento `<note>` en un párrafo separado, y utilizar el atributo `@target` para especificar el lugar exacto al que se refiere (cf. TEI-P5, §3.8).

En la normativa *TEI* encontramos recomendaciones para facilitar la creación de índices y sistemas *RI* básicos semejantes mediante etiquetas *XML*. El proceso puede realizarse manualmente, por medio de elementos de carácter visual-estructural como `<div>` o `<list>`, que sirven para describir una sección de texto como un bloque separado, dentro del cual pueden marcarse las entradas del índice mediante `<item>`, así como codificar el vínculo correspondiente con otras secciones mediante los elementos `<ptr>` y `<ref>`. TEI-P5, 109, incluye un ejemplo de uso de estos elementos:

```
<div type="index">
<!--...-->
<list type="index">
<item>Women, how cause of mel. <ref>193</ref>; their vanity in
apparell taxed, <ref>527</ref>; their counterfeit tears
<ref>547</ref>; their vices <ref>601</ref>, commended,
<ref>624</ref>.
</item>
<item>Wormwood, good against mel. <ref>443</ref>
</item>
<item>World taxed, <ref>181</ref>
</item>
<item>Writers of the cure of mel. 295</item>
<!--...-->
</list>
</div>
```

⁶⁶ Sobre la importancia de los nombres en las *Historias* de Heródoto remitimos a Campos Daroca 1992.

Además de esta posibilidad, es posible facilitar la generación automática de un índice por medio de un programa informático con el elemento `<index>` rodeando aquellos pasajes que deben ser recogidos⁶⁷; dentro del texto así etiquetado, el elemento `<term>` nos permite especificar una o varias palabras clave como texto de entrada para el pasaje en el índice resultante. Por ejemplo (cf. TEI-P5, 110):

```
<p>The students understand procedures for Arabic lemmatisation
<index><term>Lemmatization, Arabic</term></index>and are beginning
to build parsers.</p>
```

Asimismo es posible estructurar la información recogida en el índice, creando diferentes subíndices con información relacionada mediante el atributo `@indexName`, un proceso denominado «indización multi-nivel»; por ejemplo (cf. TEI-P5, 112):

```
<p>Sir      John      Ashford      <index      indexName="INDEX-
PERSONS"><term>Ashford, John</term></index> was, coincidentally,
born      in      <index      indexName="INDEX-PLACES"><term>Ashford
(Kent)</term></index>Ashford...</p>
```

ε. Sistemas de referencia

Dentro de este módulo se recogen diferentes métodos para representar sistemas de conexión intertextual, es decir, formas de codificar determinadas referencias a otras obras. La normativa *TEI* distingue entre sistemas de referencia basados en unidades estructurales (capítulos, párrafos, oraciones; estrofas o versos), en unidades tipográficas (números de página y línea), o en divisiones creadas específicamente para construir un sistema de referencia (capítulo y versículo en la *Biblia*; libros y secciones en las *Historias*) (cf. TEI-P5, §3.10).

Si el sistema de referencia que se desea representar se basa en la misma estructura lógica utilizada para codificar el resto del texto, el valor concreto de las referencias puede

⁶⁷ Si el texto es muy extenso puede ser más conveniente utilizar un elemento vacío `<anchor>` para delimitar el comienzo y el final del pasaje en cuestión, y remitir el elemento `<index>` a alguno de estos puntos (cf. TEI-P5, 111):

```
<p>We now turn to the topic of Arabic lemmatisation <index
spanTo="#ALAMEND"><term>Lemmatization, Arabic</term></index> concerning which it
is important to note [...] <!-- much learned material omitted here --> and now we
can build our parser.<anchor xml:id="ALAMEND"/></p>
```

indicarse mediante los atributos *@xml:id* y *@n*. Estos atributos pueden aplicarse a un serie de elementos (cf. TEI-P5, §3.10.1), como por ejemplo, el delimitador de bloques *<div>* y derivados (*<div1>*, *<div2>*, *<div3>*...). El valor del atributo *@n* sirve como identificador numérico para cada una de las unidades estructurales en las que se divide un texto, como puede apreciarse en la versión XML de las *Historias* alojada en *Perseus*:

```
<text n="urn:cts:greekLit:tlg0016.tlg001.perseus-grc1">
<body>
<div1 type="book" n="1">
<div2 type="chapter" n="1">
<div3 type="section" n="0">
<p> <milestone unit="para"/>Ἡροδότου Ἐλικαρνησέος ἱστορίας
ἀπόδεξις ἦδε, ὡς μήτε τὰ γενόμενα ἐξ ἀνθρώπων τῷ χρόνῳ ἐξίτηλα
γένηται, μήτε ἔργα μεγάλα τε καὶ θωμαστά, τὰ μὲν Ἕλλησι τὰ δὲ
βαρβάροισι ἀποδεχθέντα, ἀκλεᾶ γένηται, τὰ τε ἄλλα καὶ δι' ἣν
αἰτίην ἐπολέμησαν ἀλλήλοισι.
</p>
```

Aquí observamos dos tipos de valores posibles para el atributo *@n*. Dentro del elemento *<text>*, *@n* indica una referencia estándar al documento completo de las *Historias* siguiendo la arquitectura *CITE-URN*; en los tres elementos siguientes el valor de *@n* especifica la numeración de cada uno de las unidades tradicionales utilizadas para dividir la obra de Heródoto: libro, capítulo y sección.

El atributo *@xml:id* sirve para definir identificadores únicos para un determinado elemento del texto; estos no pueden contener espacios en blanco, deben comenzar por una letra o un guión bajo, y en general sólo puede contener letras, dígitos, guiones, guiones bajos, puntos y caracteres extendido o de combinación definidos en la especificación general del lenguaje *XML*. Como ejemplo veamos cómo sería el pasaje anterior de las *Historias* sustituyendo *@n* por *@xml:id*:

```
<div1 type="book" xml:id="Hdt.1">
<div2 type="chapter" xml:id="Hdt.1.1">
<div3 type="section" xml:id="Hdt.1.1.0">
<p> <milestone unit="para"/>Ἡροδότου Ἐλικαρνησέος ἱστορίας
ἀπόδεξις ἦδε, ὡς μήτε τὰ γενόμενα ἐξ ἀνθρώπων τῷ χρόνῳ ἐξίτηλα
γένηται, μήτε ἔργα μεγάλα τε καὶ θωμαστά, τὰ μὲν Ἕλλησι τὰ δὲ
βαρβάροισι ἀποδεχθέντα, ἀκλεᾶ γένηται, τὰ τε ἄλλα καὶ δι' ἣν
αἰτίην ἐπολέμησαν ἀλλήλοισι.
</p>
```

Puede ocurrir, sin embargo, que el texto que se debe codificar utilice otros elementos como base para construir el sistema de referencia, tales como número de página en una edición determinada; en estos casos, el módulo §20 recoge algunas posibles soluciones⁶⁸. En último lugar, el elemento `<milestone>` sirve para marcar lugares en el texto donde el valor del sistema de referencia cambia (cf. TEI-P5, 115; §3.10.3); este último es la solución más simple cuando el sistema de referencia no se corresponde con una división jerárquica clara, o cuando coexisten múltiples estructuras de este tipo⁶⁹. La biblioteca *Perseus* sigue uno de los usos de este elemento recomendados por la normativa *TEI*, donde los capítulos, libros y secciones vienen marcados con elementos `<div>` (estructura visual-lógica), mientras que elementos `<milestone>` indican el final de las diferentes líneas argumentales:

```
<div3 type="section" n="1">
<p>
<milestone unit="para"/>Περσέων μὲν νῦν οἱ λόγιοι Φοίνικας αἰτίους
φασὶ γενέσθαι τῆς διαφορῆς. τούτους γὰρ ἀπὸ τῆς Ἐρυθρῆς καλεομένης
θαλάσσης ἀπικομένους ἐπὶ τήνδε τὴν θάλασσαν, καὶ οἰκῆσαντας τοῦτον
τὸν χῶρον τὸν καὶ νῦν οἰκέουσι, αὐτίκα νῦν αὐτιλίησι μακροῖσι
ἐπιθέσθαι, ἀπαγινέοντας δὲ φορτία Αἰγύπτια τε καὶ Ἀσσύρια τῆ τε
ἄλλῃ ἐσαπικνέεσθαι καὶ δὴ καὶ ἐς Ἄργος.
</p>
</div3>
```

Desde un punto de vista formal, el método específico elegido para codificar el sistema de referencias del texto debe aparecer declarado en la cabecera del documento *XML* (cf. TEI-P5, §3.10.4; §16.2.5), dentro del elemento `<refsDecl>` (cf. §2.3.6). La declaración puede consistir en una descripción en prosa o bien en una definición más formal utilizando el elemento `<cRefPattern>`⁷⁰. Esta última tiene la ventaja de permitir el procesamiento

⁶⁸ Es posible codificar segmentaciones alternativas utilizadas en diferentes ediciones por medio de los atributos `@ed` y `@unit` para indicar que un pasaje constituye una unidad determinada en una edición específica (cf. TEI-P5, 121). Encontramos un uso similar en el ejemplo citado (Hdt. 1, 1, 1), donde el valor de `@unit`, «para», indica que el texto marcado con el elemento `<milestone>` es un párrafo.

⁶⁹ Cf. TEI-P5, 119: «`<milestone>` marks a boundary point separating any kind of section of a text, typically but not necessarily indicating a point at which some part of a standard reference system changes, where the change is not represented by a structural element». Otros elementos dentro de la misma categoría son `<gb>`, *gathering begins*; `<pb>`, *page break*; `<lb>`, *line break*; `<cb>`, *column break*.

⁷⁰ Remitimos a TEI-P5, 123 para una muestra de cómo se define formalmente un sistema de referencia mediante `<cRefPattern>`.

posterior por algún tipo de programa informático, pero como todo, la elección depende de las aplicaciones posteriores que pueda tener el texto así etiquetado.

La versión de *Perseus* sigue la segunda recomendación, y define el sistema de referencia de la siguiente manera:

```
<encodingDesc>
<refsDecl n="CTS">
<cRefPattern matchPattern="(\\w+).(\\w+).(\\w+)"
replacementPattern="#xpath(/tei:TEI.2/tei:text/tei:body/tei:div1[@
n=\\'$1\\']/tei:div2[@n=\\'$2\\']/tei:div3[@n=\\'$3\\'])">
<p>This pointer pattern extracts Book and Chapter and Section</p>
</cRefPattern>
<cRefPattern matchPattern="(\\w+).(\\w+)"
replacementPattern="#xpath(/tei:TEI.2/tei:text/tei:body/tei:div1[@
n=\\'$1\\']/tei:div2[@n=\\'$2\\'])">
<p>This pointer pattern extracts Book and Chapter</p>
</cRefPattern>
<cRefPattern matchPattern="(\\w+)"
replacementPattern="#xpath(/tei:TEI.2/tei:text/tei:body/tei:div1[@n=
\\'$1\\'])">
<p>This pointer pattern extracts Book</p>
</cRefPattern>
</refsDecl>
<refsDecl doctype="TEI.2">
<state delim="." unit="book"/>
<state delim="." unit="chapter"/>
<state unit="section"/>
</refsDecl>
</encodingDesc>
```

Se trata de un extracto de la cabecera del documento *XML* de las *Historias*, donde se describen formalmente dos sistemas de referencia: «CTS» y «TEI.2», por medio de dos elementos *<refsDecl>*. Sólo en el caso del sistema «CTS» se utiliza el elemento *<cRefPattern>*; en este caso, su propósito no es tanto describir el sistema en sí como recoger un método para extraer secciones específicas del texto mediante la herramienta

informática *xPath*⁷¹. Se trata de un método de *RI* básico (*reducción*), que funciona de modo parecido a la lectura manual del libro. El valor de los atributos *@n* representa la numeración y los diferentes elementos *<div>* representan cada una de las tres unidades estructurales en las que se divide la obra de forma tradicional: libro, capítulo y sección. Sustituyendo el valor de *@n* en la fórmula indicada dentro del elemento *<cRefPattern>* es posible «reducir» el texto únicamente a una sección determinada; por ejemplo, para seleccionar Hdt. 1, 123, 2 la fórmula sería la siguiente:

```
#xpath(/tei:TEI.2/tei:text/tei:body/tei:div1[@n=\ '$1\ ']/tei:div2[@n=\ '$123\ ']/tei:div3[@n=\ '$2\ '])
```

III.3.3.4. Estructuras complejas

Una simple lectura del texto revela que la separación en libros no respeta la segmentación argumental de la obra, de modo que un bloque argumental queda interrumpido abruptamente por el corte entre uno y otro. Sin embargo, el peso de la tradición, y el elevado grado de interrelación entre las diferentes partes de la obra hacen inviable un cambio en el sistema de referencias actual (*cf.* Legrand 1932, 227). En consecuencia, la búsqueda de una división «lógica» de las *Historias* se ha plasmado en agrupaciones alternativas de las secciones del texto, de acuerdo con su coherencia argumental (*cf.* Jacoby 1913). A continuación vemos qué métodos ofrece la normativa *TEI* para describir este tipo de datos.

III.3.3.4.a Recursos XML para describir estructuras complejas

Uno de los puntos fuertes del lenguaje XML es su enfoque estructuralista: el texto es concebido como un sistema de elementos que cumplen diferentes funciones y que se relacionan entre sí de manera ordenada, conformando una estructura jerárquica (*cf. supra*, II.2.4.3. «El modelo OHCO del texto»). Por esta razón la normativa *TEI* abunda en recursos destinados a representar todo tipo de rasgos del texto relacionados con su ordenación (*cf.* TEI-P5, §VI. 2.2., «Estructura argumental»).

Por defecto un documento XML conforme a la normativa *TEI*, también llamado «documento *TEI*», incluye una serie de elementos básicos que definen su estructura a gran escala (*cf.* TEI-P5, §4). Así, un documento *TEI* consiste en un elemento *<TEI>* que con-

⁷¹ *xPath* es un lenguaje procedimental diseñado para facilitar la búsqueda de información en documentos XML (*cf.* <https://www.w3.org/TR/xpath>).

tiene a su vez un elemento `<teiheader>` y un elemento `<text>`. El elemento `<teiheader>` actúa como una cabecera al cuerpo principal del texto, donde se describen fundamentalmente el método y los módulos utilizados para etiquetar los datos (cf. TEI-P5, 156). En nuestro caso particular:

```
<TEI.2>
<teiHeader type="text">
...
</teiHeader>
<text n="urn:cts:greekLit:tlg0016.tlg001.perseus-grc1">
...
</text>
</TEI.2>
```

La mayoría de elementos que venimos describiendo aparecen anidados dentro del elemento `<text>`. Un texto *TEI* puede concebirse como un todo unitario o bien como algo compuesto o complejo, formado a partir de varios componentes independientes unos de otros en algún rasgo importante (cf. TEI-P5, 156). En ambos casos el texto contenido dentro de un elemento `<text>` suele estructurarse en tres bloques. En primer lugar la materia preliminar, sea una intruducción o preludeo, definida formal o semánticamente, se delimita mediante un elemento `<front>`; a continuación, una etiqueta `<body>` enmarca el cuerpo principal del texto, o bien la etiqueta `<group>`, si se trata de un texto compuesto; por último, un elemento `<back>` sirve para describir aquella sección del texto que suele incluirse al final de la obra, tal como índices, apéndices etc. Para completar las herramientas descriptivas de la estructura del texto, la normativa *TEI* contempla un elemento `<floatingText>` para delimitar secciones de texto insertas o anidadas dentro del cuerpo principal, pero independientes de su organización jerárquica; esto resulta útil cuando se entrecruzan múltiples líneas narrativas (cf. TEI-P5, 157).

El texto de las *Historias* etiquetado por *Perseus* únicamente contiene un elemento `<body>`, sin `<front>` o `<back>`, ni referencia al carácter unitario o compuesto del texto.

α. Divisiones del cuerpo del texto

Los párrafos o listas representan un tipo de estructura superficial, visual, del texto, y la normativa *TEI* denomina a los elementos que describen este tipo de segmentaciones como componentes o elementos del nivel de composición (cf. TEI-P5, 158). La agrupación secuencial y jerárquica de este tipo de estructuras da lugar a varios tipos de divisiones del texto, como secciones, capítulos o «libros» (cf. TEI-P5, §4.1). En lugar de estos nombres tradicionales la versión de *Perseus* utiliza para el etiquetado de estas divisiones un mismo elemento, distinguiéndolas entre sí por el valor del atributo `@type` así como por un sufijo

numérico que indica la profundidad del elemento en cuestión dentro de la estructura jerárquica. Por ejemplo, el comienzo del libro segundo:

```
<div1 type="book" n="2">
<div2 type="chapter" n="1">
<div3 type="section" n="1">
<p>
<milestone unit="para"/>τελευτήσαντος δὲ ἑ Κύρου παρέλαβε ἦν
βασιληίην Καμβύσης, Κύρου ἐὼν παῖς καὶ Κασσανδάνης τῆς Φαρνάσπεω
θυγατρὸς, τῆς προαποθανούσης Κῦρος αὐτός τε μέγα πένθος ἐποιήσατο
καὶ τοῖσι ἄλλοισι προεῖπε πᾶσι τῶν ἦρχε πένθος ποιέεσθαι.
</p>
</div3>
```

Por su uso de la numeración, esta solución recibe el apelativo de «división numerada» (cf. TEI-P5, 159) y su claridad resulta más que evidente.

β. Elementos comunes

Para garantizar el soporte más amplio posible a todo tipo de divisiones textuales, la normativa *TEI* agrupa los elementos utilizados para este cometido en cinco clases, según su ubicación en el texto⁷². Lo que sigue procede de TEI-P5, §4.2:

- a) *model.divTop*: agrupa elementos que aparecen al principio de una división textual (incluye *c* y *e*).
- b) *model.divBottom*: para elementos que aparecen al final de una división textual (incluye *d* y *e*).
- c) *model.divTopPart*: para elementos que solo pueden aparecer al principio de una división textual. Incluye el elemento *<head>*, *<opener>* y *<signed>*⁷³.

El elemento *<opener>* se utiliza para agrupar varios elementos que describen fórmulas y expresiones típicas o particulares del preludio a una obra, tales como la fecha (*<dateline>*) o la autoría (*<byline>*). En nuestro texto de ejemplo, Hdt. 1, 1, 0 representa el proemio de la obra, pero no se ha distinguido del resto del texto mediante ninguno de estos procedimientos, aunque en algunas ediciones se indica su carácter mediante el sufijo «pr.» añadido a la citación, como en la versión codificada por el *TLG*.

⁷² Cf. TEI-P5, 164-165, y §1.2 para una definición de «clase de elemento».

⁷³ El elemento *<head>* es el único miembro de una clase propia, *model.headLike*, creada para facilitar la modificación del esquema *TEI* en lo referente a descripción de elementos similares (cf. TEI-P5, 165).

- d) *model.divBottomPart*: agrupa elementos que solo pueden aparecer al final de una división textual. Incluye los elementos *<closer>*, *<postscript>*, *<signed>* y *<trailer>*.

El elemento *<closer>* agrupa elementos que describen fórmulas y expresiones que suelen aparecer al final de una división, por ejemplo, la despedida de una carta. Aunque semejante a *<opener>*, no encontramos en las *Historias* ocasión donde resulte apropiada su utilización. En cambio, sí que puede ser interesante el elemento *<trailer>*, que describe una sección del texto que cumple una función semejante a la de una cabecera (elemento *<head>*), pero situado al final de una división, tal como hace Heródoto al repetir ciertos sintagmas para marcar el final de un *λόγος* y el principio de otro.

- e) *model.divWrapper*: para elementos que pueden ubicarse tanto al comienzo como al final de una división textual. Incluye los elementos *<argument>*, *<byline>*, *<dateline>*, *<docAuthor>*, *<docDate>*, *<epigraph>*, *<meeting>* y *<salute>*. Estos elementos se apoyan en convencionalismos formales y temáticos que no están presentes en las *Historias* y cuya aplicación resulta, por lo tanto, inviable.

γ. Textos agrupados y «flotantes»

Según su estructura interna, la normativa *TEI* distingue entre textos «unitarios» y «compuestos». Los elementos descritos hasta ahora permiten describir sin mayor problema los rasgos estructurales más importantes de un texto del primer grupo, pero, en el segundo caso, la normativa nos ofrece dos posibilidades: *<group>* y *<floatingText>* (cf. *TEI-P5*, §4.3). El elemento *<group>* «contains the body of a composite text, grouping together a sequence of distinct texts (or groups of such texts) which are regarded as a unit for some purpose», y aduce como ejemplo una antología de obras de un autor, o una colección de cartas (cf. *TEI-P5*, 170, y §4.3.1). El adjetivo *distinct* de la definición insiste en que las piezas que forman parte de un texto compuesto así descrito son relativamente independientes entre sí y con respecto al texto principal. Aunque no se especifica, conocido el funcionamiento del lenguaje XML, esta independencia a la que se refiere la normativa debe ser fundamentalmente formal (cf. *TEI-P5*, 174).

Sin embargo, no todos los textos ofrecen una división clara y concisa de sus elementos constituyentes, con unos rasgos formales claros y constantes que separen secciones según su argumento o algún otro criterio. Las *Historias* de Heródoto son, en este sentido, una colección de diferentes relatos unidos por un hilo argumental, pero distan mucho de

ofrecer un aspecto fácilmente divisible: unas tramas se insertan en otras, interrumpiendo con frecuencia el relato principal y formando con ello las múltiples digresiones que caracterizan esta obra. Si queremos reconocer y describir este carácter complejo del texto de Heródoto, el elemento `<group>` resulta inadecuado. En atención a casos similares al que nos ocupa, la normativa *TEI* propone usar `<floatingText>`, un elemento que permite el etiquetado de secciones del texto insertas dentro del texto principal, sin implicar una separación estricta del mismo, y sin implicar un cambio en la estructura (cf. TEI-P5, 178-179).

δ. «Front Matter» y «Back Matter»

Las expresiones *front matter* y *back matter* se refieren al material diferente del cuerpo del texto que suele aparecer al principio y al final de una obra impresa moderna. Así, mediante el elemento `<back>` se describen secciones del texto que habitualmente se sitúan al final, como índices, glosarios, apéndices, notas o bibliografía; por otra parte, el elemento `<front>` codifica material de tipo introductorio (cf. TEI-P5, §4.5 y §4.7). Mientras que `<back>` resulta poco útil para la codificación de las *Historias*, por la ausencia en ellas del tipo de secciones textuales recogidas por su definición (cf. TEI-P5, 184-185), la definición de `<front>` permite su utilización en secciones de texto que no estén al comienzo de la obra (cf. TEI-P5, 180-182), siempre y cuando tengan un carácter introductorio⁷⁴; por ejemplo, en las secciones que presentan el comienzo de una digresión:

```
<text>
<body>
<front></front>
<floatingText>
<body>
</body>
</floatingText>
</body>
</text>
```

ε. Enlaces, segmentación y alineación del texto

En relación con las ideas que el módulo §4 aporta para codificar la estructura del texto, añadimos a continuación lo expuesto en TEI-P5, §16, donde se ofrecen varias soluciones para aquellos casos donde la estructura que se desea describir no es lineal ni jerár-

⁷⁴ Remitimos a la sugerencia que hace la normativa *TEI* en este sentido para los *Cuentos de Canterbury*, de Chaucer (cf. TEI-P5, 179).

quica y, en consecuencia, plantea problemas. Todas las técnicas descritas en este módulo se basan en el marco *W3C XPointer* (cf. Grosso *et al.* 2003); en particular, la normativa recomienda las estrategias basadas en el uso del atributo *global @xml:id* (cf. TEI-P5, §1.3.1.1, y §V.5.2).

ζ. Enlaces

Se considera que un elemento apunta a otros si contiene un atributo cuyo valor sea una referencia a los demás; este elemento recibe el nombre de «elemento apuntador» (*pointer element*) o simplemente «apuntador» (*pointer*) (cf. TEI-P5, §16.1). Los elementos `<note>`, `<ref>` y `<ptr>` son «apuntadores», puesto que codifican una asociación entre un lugar del documento y uno o más elementos, especificados por el valor del atributo `@target`. Un enlace o elemento `<link>` es una variante del elemento apuntador, que especifica no solo el destino de la asociación, sino también la procedencia por medio del atributo `@target`. Al igual que `<ref>` y `<ptr>` admiten los atributos `@type` y `@subtype` como método de clasificación. La diferencia entre un enlace y un elemento `<ref>` o `<ptr>` es que contiene una referencia a dos lugares o elementos del texto, sin que ninguno de ellos tenga que ser la localización del propio elemento `<link>`⁷⁵:

```
<link target="#sa-p1 #sa-p2"/>
```

Que es lo mismo que:

```
<ptr xml:id="sa-p1" target="#sa-p2"/><ptr xml:id="sa-p2"
target="#sa-p1"/>
```

El hecho de que el elemento `<link>` no utilice su propia ubicación en el texto para describir una conexión amplía el rango de sus posibles usos, uno de los cuales puede ser agrupar pasajes o elementos interconectados en alguna parte del documento *XML*. Si se hace un uso adecuado del atributo `@type` es posible obtener una lista ingente de enlaces al final del etiquetado; estos pueden agruparse fácilmente mediante el elemento `<linkGrp>`, formando lo que se denomina una *red*. Este elemento forma parte de la clase *att.pointing.group*, es decir, de los elementos utilizados para agrupar y establecer conexiones entre otros elementos, y por lo tanto admite los atributos propios de esta clase: `@domains`, para indicar el elemento dentro del cual están englobados el resto de elementos

⁷⁵ Cf. la descripción más extensa que se hace de las diferencias entre el uso de uno y otro método en TEI-P5, 515-518.

indicados por `<linkGrp>`; `@targFunc`, *target function*, permite extender la descripción del valor del atributo `@target` de cada enlace contenido dentro de `<linkGrp>`, explicitando una función definida por el usuario⁷⁶. El elemento `<linkGrp>` no solo permite agrupar enlaces presentes en otras partes del documento, sino que también facilita la clasificación de los mismos, puesto que se puede establecer el valor del atributo `@type` de una sola vez, en lugar de repetirlo en cada elemento `<link>` por separado (cf. TEI-P5, 520). La red resultante del uso del elemento `<linkGrp>` puede ser procesada con posterioridad por un programa informático, otorgando al documento XML algunas funcionalidades propias de una base de datos (cf. TEI-P5, 519).

η. Mecanismos de referencia o «pointing»

La normativa *TEI* proporciona métodos para apuntar *a)* a documentos diferentes de otros documentos; *b)* a elementos individuales presentes en otros documentos mediante su `@xml:id`; *c)* a elementos individuales a partir de su posición dentro de la estructura del documento XML; *d)* a contenido de cualquier tipo y extensión mediante esquemas *Xpointer* definidos en TEI-P5, 521 (cf. TEI-P5, §16.2).

Todos los atributos descritos por la normativa *TEI* que sirven para indicar una referencia a otra cosa están definidos como referencias *URI* (*Universal Resource Indicator*). Este tipo de referencias permiten establecer una asociación con un documento completo, pero lo habitual es que se utilicen para indicar una sección dentro de un documento; para ello no es necesario proporcionar la *URI* en toda su extensión, sino tan sólo la parte que describe aquella sección que nos interesa, lo que se conoce como un identificador de fragmento. Técnicamente hablando, este identificador es la sección de una referencia *URI* que va después del primer signo #.

En el caso de la versión del texto de Heródoto digitalizada por la biblioteca *Perseus*, las referencias se limitan a otros pasajes del texto y están codificadas siguiendo el esquema *CITE-URN*, que en esencia es una extensión del mecanismo fundamental de las referencias *URI*⁷⁷.

⁷⁶ En el caso de que todos los enlaces recopilados dentro de `<linkGrp>` se encuentren dentro del elemento estructural `<body>`, este será el valor del atributo `@domains`. Los posibles valores del atributo `@targFunc` deben ser convenientemente recogidos en la sección de la cabecera dedicada a describir el método de codificación (`<encodingDesc>`; cf. TEI-P5, 520, n. 93).

⁷⁷ Cf. TEI-P5, 525-526 para instrucciones con las que personalizar un esquema *URI*.

Uno de los esquemas *XPointer* más importantes, definido también por la *W3C* es el esquema *xpath()*, que permite seleccionar cualquier parte de la estructura *XML* utilizando la sintaxis definidas por la especificación *XPath* (cf. TEI-P5, §16.2.4.2). A menudo se necesitan métodos para establecer conexiones no sólo con elementos aislados, sino también con puntos intermedios entre elementos o secciones que se extiendan a lo largo del documento. La descripción de la mayoría de las digresiones presentes en las *Historias* supone un desafío de este tipo: desde el punto de vista de la mecánica subyacente al lenguaje *XML*, el relato de Candaules (cf. Hdt. 1, 7, 2-20) es una unidad argumental o semántica que se superpone a otras unidades estructurales (párrafos, secciones y capítulos)⁷⁸.

Algunos de los elementos descritos hasta ahora (*<ref>*, *<ptr>*, *<link>*, *<note>*) describen una asociación entre dos puntos o «nodos» de un documento *XML*, es decir, entre elementos y atributos. Sin embargo, como demuestra el ejemplo anterior, se hace necesario algún método para referirse a puntos intermedios entre elementos, o a pasajes que se extienden a lo largo de varios elementos. Ahora bien, desde la perspectiva de un esquema de referencia como el que estamos describiendo un documento *TEI* es una estructura en árbol sobrepuesta a un flujo de caracteres; esto significa que cualquier búsqueda de información se realiza en dos planos: *a*) localizando elementos situados en algún punto de la estructura de árbol, y *b*) seleccionando rangos de caracteres más o menos extensos dentro del flujo del texto. Tanto los métodos basados en el lenguaje *xPath* como las referencias que puedan darse a una sección del texto basadas en el valor de su atributo *@xml:id* solo pueden encontrar secciones del texto contenidas dentro de un elemento; por otro lado, estos métodos no pueden localizar una sección que se extienda a lo largo de varios elementos, o que solo ocupe una parte del mismo. Búsqueda por caracteres y búsqueda por elementos son técnicas de *RI* diferentes, y estrictamente hablando, incompatibles. Cuando alguno de los esquemas de referencia *TEI* trata de combinarlos, en realidad está realizando dos tipos de búsqueda por separado, y el resultado será, o bien una secuencia de caracteres, o bien un elemento.

⁷⁸ La versión de *Perseus* inserta un elemento *<milestone>* en Hdt. 1, 8, 1 para indicar que ahí comienza una digresión. Sin embargo, su sentido es poco claro. ¿Por qué no sitúa el inicio de la historia en la primera mención de Candaules (1, 7, 2)? ¿Significa que considera que la sección entre 1, 7, 2 y 1, 8, 1 es una introducción al relato y por lo tanto no forma parte de la digresión? En tal caso podría haberse etiquetado esta sección preliminar mediante un elemento *<front>*.

Un esquema de referencia *TEI* distingue cuatro tipos de «objetos» susceptibles de ser localizados mediante diferentes técnicas de *RI* como las que hemos referido ya (*cf.* TEI-P5, 525-526):

- a) **Nodo**: representa un único ítem dentro del conjunto de datos que contiene un documento *XML*; según el tipo de contenido al que hacen referencia podemos hablar de nodos textuales, nodos de elemento y nodos de atributo⁷⁹.
- b) **Secuencia**: es un grupo ordenado de uno o más ítems, en la que un ítem puede ser bien un nodo o parte de un nodo textual.
- c) **Flujo de texto**: es la concatenación de las diferentes secuencias de caracteres contenidas en los nodos textuales de un documento. Un flujo de texto comienza en un nodo de referencia y engloba todo el texto dentro de ese nodo junto con el resto del texto a continuación en el orden en que se presenta dentro del documento, ignorando las etiquetas *XML* presentes en él⁸⁰.
- d) **Punto**: representa un lugar entre nodos o caracteres dentro de un documento. Todo punto es adyacente a caracteres o elementos, pero nunca a otro punto; por lo tanto, solo puede hacerse referencia a un punto mediante su relación con un elemento o un nodo textual, es decir, con una parte del texto que pueda localizarse mediante *XPath* o un identificador de fragmento. No se pueden utilizar puntos para indicar el comienzo de una selección de caracteres.

Para la detección de estos cuatro tipos, la normativa recomienda seis esquemas de referencia (*cf.* TEI-P5, 527):

1. *xpath()*: hace referencia a un nodo o conjunto de nodos utilizando la sintaxis propia de *XPath* (*cf.* §16.2.4.2). *XPath* es un lenguaje procedimental que permite hacer búsquedas dentro de un documento *XML*; consiste en una serie de términos clave, con los que se dan órdenes a cualquier tipo de programa informático capaz de manejarlo. Su capacidad para utilizar nombres y valores de

⁷⁹ Solo mencionamos los tres nodos habituales en una búsqueda de *RI* sobre un documento *XML*. Para una definición formal de nodo y todos sus tipos remitimos a Clark & DeRose 1999.

⁸⁰ Los esquemas de referencia que localizan flujos de texto requieren como dato necesario para la búsqueda un número de caracteres; dado que este objeto ignora las etiquetas *XML*, el número de caracteres que formen parte de estas etiquetas no se tiene en cuenta.

elementos y atributos para seleccionar nodos en el documento lo hace uno de los sistemas de *RI* aplicables a *XML* más robustos que existen; no se ve afectado por cambios en la secuencia de caracteres que puedan producirse en el texto, y no requiere la adición al documento de identificadores únicos mediante *@xml:id*, tarea que puede ser bastante laboriosa. Por estas razones es el esquema de referencia recomendando por la versión de *Perseus*, como puede verse en la cabecera del documento:

```
<refsDecl n="CTS">
<RefPtrPattern matchPattern="(\\w+).(\\w+).(\\w+)"
replacementPattern="#xpath(/tei:TEI.2/tei:text/tei:body/tei:div1[@
n='§1\']/tei:div2[@n='§2\']/tei:div3[@n='§3\'])">
<p>This pointer pattern extracts Book and Chapter and Section</p>
</RefPtrPattern>
<RefPtrPattern matchPattern="(\\w+).(\\w+)"
replacementPattern="#xpath(/tei:TEI.2/tei:text/tei:body/tei:div1[@
n='§1\']/tei:div2[@n='§2\'])">
<p>This pointer pattern extracts Book and Chapter</p>
</RefPtrPattern>
<RefPtrPattern matchPattern="(\\w+)"
replacementPattern="#xpath(/tei:TEI.2/tei:text/tei:body/tei:div1[@
n='§1\'])">
<p>This pointer pattern extracts Book</p>
</RefPtrPattern>
```

El lenguaje *XPath* puede utilizarse igualmente para describir la localización de un objeto determinado en el texto. En consecuencia, los esquemas de referencia que mencionamos a continuación pueden utilizar esta «dirección» como un parámetro de búsqueda, al que nos referimos como *XPATH*⁸¹. Mantenemos la fórmula que sigue la normativa *TEI* para describir brevemente los esquemas⁸²:

⁸¹ Nótese las diferentes convenciones tipográficas utilizadas para distinguir entre el lenguaje procedimental (*XPath*), la referencia a una sección del documento utilizando la sintaxis propia de este lenguaje (*XPATH*) y la expresión con la que se indica una referencia en este lenguaje dentro de un documento *XML* (*xpath()*).

⁸² La barra vertical dentro del paréntesis indica una alternativa entre las opciones indicadas a uno y otro lado; la coma separa parámetros que deben aparecer necesariamente.

«tipo de objeto» comando(PARÁMETRO 1a | PARÁMETRO 1b, PARÁMETRO 2)

2. **left(IDREF | XPATH)** y **right(IDREF | XPATH)**: hace referencia a un punto anterior o posterior de un nodo o conjunto de nodos (cf. §16.2.4.3; §16.2.4.4), mediante una dirección *XPATH* o mediante el valor de un atributo *@xml:id* (el argumento *IDREF*).
3. **string-index(IDREF | XPATH, OFFSET)**: selecciona un punto dentro de un nodo textual (§16.2.4.5), basándose en la posición de un carácter relativa al nodo identificado por el parámetro *IDREF* o *XPATH*. El parámetro *OFFSET* es un número entero que especifica la posición relativa del punto buscado; un valor 0 para *OFFSET* representa una posición inmediatamente anterior al primer carácter que se encuentre en el nodo indicado mediante *IDREF* o *XPATH*⁸³.
4. **range(APUNTADOR, APUNTADOR[, APUNTADOR, APUNTADOR ...])**: hace referencia a un rango entre dos puntos (cf. §16.2.4.6), y localiza una secuencia que comienza en el lugar indicado por el valor del primer elemento *APUNTADOR*. Con *range()* es posible localizar secuencias de nodos no contiguos, es decir, separados por algún otro nodo, añadiendo sucesivas parejas de apuntadores a la lista de parámetros (cf. TEI-P5, 528).
5. **string-range(IDREF | XPATH, OFFSET, LENGTH[, OFFSET, LENGTH ...])**: selecciona una secuencia basándose en la posición de los caracteres relativa a un nodo identificado mediante el primer parámetro (cf. TEI-P5, §16.2.4.7), del mismo modo que ocurría con el comando *string-index()*. El parámetro *LENGTH* es un número entero positivo que indica la longitud en caracteres del flujo de texto incluido en la secuencia.
6. **match(IDREF | XPATH, 'REGEX' [, INDEX])**: selecciona una secuencia basándose en una expresión regular que define una parte del flujo del texto

⁸³ El valor de *OFFSET* se mide en número de caracteres; cf. *Unicode Normalizations Forms* sobre las precauciones que deben tomarse a la hora de definir qué combinaciones de caracteres se consideran un único carácter, o dos. En el caso del griego, la duda recae en los acentos y diacríticos (cf. TEI-P5, 528).

relativa al nodo identificado mediante *IDREF* o *XPATH* (cf. §16.2.4.8)⁸⁴. El parámetro *INDEX* es opcional y permite indicar mediante un valor numérico qué coincidencia debe elegirse en el caso de que la expresión regular coincida con más de una secuencia.

θ. Bloques, segmentos y «anclas»

Es posible utilizar elementos para describir tanto una secuencia como un punto dentro del flujo del texto. Su principal ventaja es que pueden recibir un atributo *@xml:id* y facilitar con ello las tareas de *RI*. Por ejemplo, el elemento *<anchor>*, punto de anclaje, permite añadir un identificador a cualquier punto dentro del texto, sea o no un elemento textual (cf. TEI-P5, §16.3). El uso del elemento *<anchor>* es similar al de *<milestone>*, por cuanto permite la superposición de una estructura del texto sobre otra ya existente, como es la división tradicional del texto de Heródoto en libros, partes y secciones. A diferencia de *<milestone>*, no se recomienda su aplicación para realizar una segmentación alternativa del texto, sino para crear puntos de referencia útiles para otros elementos (cf. TEI-P5, 533).

En segundo lugar, el elemento *<seg>*, segmento arbitrario, representa cualquier tipo de segmentación del texto por debajo del nivel del *chunk*. El elemento *<seg>* no posee un uso específico definido, más allá de extender los elementos descriptivos disponibles para otras unidades del texto no contempladas por la normativa⁸⁵.

Por último, el elemento *<ab>*, bloque anónimo, contiene cualquier unidad de texto equivalente a un párrafo, pero sin la carga semántica de este (cf. TEI-P5, 532). Además de los atributos *@type* y *@subtype*, los elementos *<ab>* y *<seg>* poseen un atributo *@part* gracias al cual es posible unificar una unidad de texto que se halla fragmentada.

ι. Agrupamiento de secciones no continuas

La estructura de un texto real no es necesariamente jerárquica, y las unidades argumentales, sintácticas o gráficas que lo componen suelen verse interrumpidas con frecuencia para continuar en otro lugar del texto. Cuando no es posible englobar dentro de un ele-

⁸⁴ Puesto que el valor del parámetro *REGEX* está delimitado por apóstrofes, cualquier apóstrofo que forme parte de la expresión regular debe ser «escapado» (cf. *supra* II.2.3. *Expresiones regulares*).

⁸⁵ Si se usa para delimitar oraciones, es mejor utilizar los elementos recogidos en el módulo §17.1, tal como *<s>*, que describe una oración gráficamente, como una secuencia de caracteres delimitada por signos de puntuación.

mento todas las partes de una unidad, es posible utilizar algunos elementos para representar el vínculo existente entre las partes. El resultado es un elemento «virtual», puesto que la unión de las partes que forman el elemento se produce fuera de la forma física del texto original (*cf.* TEI-P5, §16.7).

Etiquetar un elemento virtual consiste en definir una relación entre dos o más lugares del texto y definirlos como un grupo de unidades que conforman un único elemento. Esto se puede conseguir mediante el elemento `<link>`, ya descrito, añadiendo un par de atributos simples que permiten describir el orden de cada parte dentro del contenido del elemento virtual; `@next` hace referencia al siguiente elemento del grupo, mientras que `@prev` hace lo propio con el elemento anterior. Sin embargo, el elemento recomendado para esta tarea es `<join>` (*cf.* TEI-P5, 551), ya que permite describir a qué elemento se correspondería el grupo virtual, mediante el atributo `@result`. Al igual que pasaba con `<link>`, es posible reunir varios elementos del mismo tipo dentro de un elemento `<joinGrp>`, para facilitar el análisis posterior del texto.

z. Inserción de comentarios sobre la estructura

Este módulo describe formas en las que añadir información que podemos clasificar como fruto de la interpretación particular de la persona o personas responsables del etiquetado, así como vincularla explícitamente con elementos textuales presentes en el documento (*cf.* TEI-P5, §17, «Mecanismos de análisis simple»). La normativa *TEI* incluye en este ámbito elementos y atributos que describen la estructura lingüística del texto (*cf.* §17.1), con los que indicar que el elemento utilizado para caracterizar una sección es fruto de una interpretación propia (*cf.* §17.2 y §17.3), con algunos ejemplos (*cf.* §17.4).

Mediante etiquetas *XML* es posible delimitar las unidades lingüísticas presentes en un texto, sin importar la definición exacta del valor que otorguemos a cada una de ellas⁸⁶. Las palabras aisladas se delimitan con un elemento `<w>`, independientemente de que gráficamente se encuentre separada del resto. Por encima de este nivel, tres elementos recogen distintas perspectivas de análisis sobre las unidades en las que pueden agruparse las palabras: `<s>` (*s-unit*) describe una división del texto semejante o equivalente a la ora-

⁸⁶ La definición exacta de las unidades de análisis representadas por estos elementos debe aparecer convenientemente recogida en la sección correspondiente de la cabecera del documento *XML*.

ción⁸⁷; <cl> (*cláusula*) y <phr> (*frase*) representan un rango bastante amplio de divisiones posibles inferiores a la oración. Estos cuatro elementos admiten un atributo @function, además de los ya mencionados @type y @subtype.

Más allá del ámbito exclusivamente lingüístico, la forma más directa y simple de añadir comentarios de tipo analítico en algún pasaje del texto es utilizar los elementos , <interp>, junto con sus respectivos elementos aglutinadores <spanGrp> e <interpGrp> (cf. *supra* el comportamiento de este tipo de elementos en el caso de <linkGrp>). El elemento permite asociar un comentario directamente a una sección del texto, que puede indicarse mediante los atributos @from y @to o mediante el mismo atributo @target común a otros elementos de enlace⁸⁸.

Este método resulta especialmente interesante para describir unidades narrativas que no coinciden con la división gráfica o tradicional del texto⁸⁹. El ejemplo aducido en la normativa, de la épica germánica (cf. TEI-P5, 577), nos ofrece un buen precedente de lo que puede hacerse en el texto de Heródoto:

```
<p xml:id="P1">
  <s xml:id="S1">Sigmund ... was a king in Frankish country.</s>
  <s xml:id="S2">Sinfiotli was the eldest of his sons.</s>
  <s xml:id="S3">Borghild, Sigmund's wife, had a brother ...</s>
  <s xml:id="S4A">But Sinfiotli ... wooed the same woman</s>
  <s xml:id="S4B">and Sinfiotli killed him over it.</s>
  <s xml:id="S5">And when he came home, ... she was obliged to
  accept it.</s>
  <s xml:id="S6">At the funeral feast Borghild was serving beer.</s>
  <s xml:id="S7">She took poison ... and brought it to
  Sinfiotli.</s>
```

⁸⁷ El elemento <s> tiene algunas limitaciones de uso importantes: sólo puede contener datos textuales a nivel de frase y no puede contener otros elementos <s> en su interior; los elementos <cl> y <phr> permiten etiquetar unidades inferiores dentro de un elemento <s>. Se distingue así de <seg>, que describe de forma genérica cualquier tipo de división textual, especificada mediante el atributo @type (cf. TEI-P5, 567).

⁸⁸ Los atributos @from y @to deben utilizarse obligatoriamente en conjunto, y solo permiten seleccionar una extensión de texto ininterrumpida entre los dos marcadores; si la sección de texto no es continua debe emplearse @target para indicar la referencia (cf. TEI-P5, 574-575).

⁸⁹ Cf. TEI-P5, 579, sobre las diferencias de uso de <interp> y a la hora de comentar la estructura narrativa: el primero requiere la adición de etiquetas específicas en el texto, lo que supone una complicación con respecto al uso inmediato que puede hacerse de .

```

<s xml:id="S17">Sinfiotli drank it off and at once fell dead.</s>
<anchor xml:id="EOS17"/>
</p>
<p xml:id="P2">Sigmund carried him a long way in his arms ... </p>
<p xml:id="P3">King Sigmund lived a long time in Denmark ... </p>
<p xml:id="P4">Sigmund and all his sons were tall ... </p>
<spanGrp resp="#TMA" type="narrative-structure">
<span from="#S1" to="#S3">introduction</span>
<span from="#S4A">conflict</span>
<span from="#S4B">climax</span>
<span from="#S5" to="#S17">revenge</span>
<span from="#EOS17">reconciliation</span>
<span from="#P2" to="#P4">aftermath</span>
</spanGrp>

```

Otros datos relativos al análisis pueden añadirse, como el grado de seguridad otorgado a la información que se describe (*@cert*), o la persona o entidad responsable de los datos añadidos (*@resp*).

λ. Descripción de estructuras no jerárquicas

El lenguaje *XML* se basa en un modelo jerárquico del texto, y aunque existen diversas soluciones al problema de la descripción de rasgos o estructuras no jerárquicas, a día de hoy no existe una única solución que cumpla las condiciones deseables de simplicidad, capacidad de representación de todas las circunstancias posibles y facilidad para su validación formal o mecánica (*cf.* TEI-P5, 645). Así pues, tenemos cuatro métodos para rodear o solventar parcialmente el problema (*cf.* TEI-P5, §20):

- 1) codificar la misma información de varias maneras;
- 2) utilizar elementos vacíos para delimitar estructuras no anidadas⁹⁰;
- 3) dividir artificialmente una unidad del texto en secciones que puedan anidarse correctamente, y etiquetar la unidad por medio de un elemento virtual;
- 4) separar el marcado descriptivo del texto descrito.

El primer método requiere crear tantas copias del texto etiquetado como posibles codificaciones existan (*cf.* §20.1). Es el método más simple, pero no permite insertar en un todo interrelacionado diferentes visiones del texto; es decir, no se pueden aplicar todas las

⁹⁰ Esta parece haber sido la opción escogida por *Perseus*.

etiquetas en una única copia del texto. Si queremos describir la estructura lingüística, utilizaremos `<s>`, `<cl>` y `<phr>`, y, si nos interesa la organización argumental de la obra, `<div>` (y derivados), en combinación con `<seg>`, `<milestone>` u otros elementos similares; sin embargo, no siempre es posible combinar todas las posibilidades.

El segundo método se centra en el problema particular del solapamiento de unidades en el texto y de la vulneración de la regla de anidamiento que conlleva su etiquetado normal. Mediante elementos vacíos del tipo `<milestone>`, `<anchor>`, `<seg>`, o algún otro creado o modificado a propósito para este cometido, es posible indicar el comienzo o los límites del segmento de texto que representa el rasgo que se va a describir, sin alterar la rígida estructura jerárquica del documento *XML*⁹¹.

El tercer método es una extensión del anterior, en la que, con posterioridad a la subdivisión del texto, los segmentos resultantes son recombinados formando unidades no presentes literalmente en el texto, es decir, *virtuales*. Este último paso permite superar las limitaciones de la simple fragmentación del texto, aunque las herramientas proporcionadas por la normativa *TEI* pueden quedarse cortas y requerir de cierto nivel de modificación⁹².

La última solución es similar a la primera, y consiste en separar el marcado del propio texto, como ya comentamos antes (*cf.* TEI-P5, §16.9).

Soluciones no basadas en XML

El problema a la hora de describir diferentes organizaciones estructurales simultáneas en el texto con *XML* se encuentra en los fundamentos que sustentan el funcionamiento de este lenguaje descriptivo. Los cuatro métodos anteriores no dejan de representar un rodeo, una forma de ignorar el problema en lugar de darle una auténtica solución. Muchos piensan que para ello es necesario modificar los fundamentos básicos del lenguaje *XML*, implementando características o mecanismos nuevos o existentes en otras herramientas descriptivas

⁹¹ DeRose 2004 describe en detalle varias técnicas basadas en la inserción de este tipo de elementos y los nombres que reciben: *Trojan milestons*, *HORSE markup*, *CLIX* o *COLT*; en esencia, todas ellas subdividen los elementos originales, sobrecargando en ocasiones el texto. No sólo es un método un tanto incómodo, sino que puede ocasionar problemas de validación del documento *XML* resultante si las modificaciones no se insertan debidamente según lo previsto por la normativa *TEI* (*cf.* TEI-P5, 650 y n. 119).

⁹² Además de `<seg>`, otros elementos pertenecientes a la clase *att.segLike*, como `<ab>`, o `<div>`, pueden utilizarse para fragmentar otras unidades, mediante el atributo `@part`; por desgracia, este atributo sólo admite tres valores, I (inicial), M (media), F (final), pero no permite especificar el orden en que deben combinarse (*cf.* TEI-P5, 652).

(cf. TEI-P5, §20.5). La normativa *TEI* se limita a mencionar las propuestas más importantes en este sentido, remitiendo a la bibliografía correspondiente (cf. TEI-P5, 655). Una de las opciones más sencillas es recuperar el rasgo *CONCUR*, presente en el lenguaje *SGML*, y que permitía la coexistencia de varias opciones de marcado para un mismo elemento (cf. Hilbert *et al.* 2005). Otras opciones tratan de expandir las posibilidades del lenguaje *XML* alterando lo menos posible su funcionamiento habitual, ya sea combinando varios documentos *XML* en uno solo⁹³, procesando los datos codificados de un modo diferente⁹⁴ o proponiendo nuevos mecanismos de anotación que sustituyan en mayor o menor medida al actual lenguaje de marcado *XML*⁹⁵.

III.3.4. DESCRIPCIÓN DE ENTIDADES

En último lugar, dentro de nuestro repaso de las opciones de etiquetado descriptivo que ofrece la normativa *TEI*, nos detenemos a considerar qué métodos existen para describir «entidades» presentes en el texto. Mediante este término nos referimos a personas, objetos o conceptos que deseamos describir no tanto como una secuencia de caracteres (<*w*>), ni como unidades lingüísticas de cierta clase (atributo *@type* añadido al elemento <*w*>), sino como realidades externas en cierto modo a la forma concreta en la que se nos presenta el texto. Los elementos descritos en esta sección describen palabras o conjuntos de palabras que hacen referencia a una misma realidad, y nos permiten contabilizar la presencia o ausencia de estas a lo largo de la obra.

En el capítulo §3.5 de la normativa *TEI* se recogen los métodos básicos para indicar que un segmento del texto es un nombre propio o un cadena de referencia⁹⁶ de un tipo concreto, especificado mediante el atributo *@type*. El capítulo §13, *Nombres, fechas, personas y lugares*, continúa donde lo deja el anterior y describe en mayor detalle una serie de ele-

⁹³ Como el llamado «XML coloreado» de Jagdish *et al.* 2004, y el modelo teórico de múltiples árboles estructurales desarrollado por Dekhtyar & Iacob 2005.

⁹⁴ La propuesta de los *Just-In-Time-Trees* de Durasau & O'Donnell 2002 propone combinar los datos de un documento *XML* estándar de un modo alternativo para obtener diferentes estructuras.

⁹⁵ La llamada propuesta *LMNL* (*Layered Markup and Annotation Language*), basada en la denominada *core range algebra* de Tennison & Piez 2002, así como la propuesta *MLCD* (*Markup Languages for Complex Documents*) de Huitfeldt & Sperberg-McQueen 2001 plantean un sistema de notación y un concepto de estructura completamente nuevo y diferente del representado por *XML*.

⁹⁶ Una cadena de referencia o *referring string* es una expresión que representa otra cosa, por ejemplo, una persona, un objeto o un lugar (cf. TEI-P5, 88).

mentos que codifican específicamente tres tipos de cadenas de referencia: nombres, lugares, «organizaciones» y fechas (*cf.* TEI-P5, 437)⁹⁷.

Los elementos del módulo §13, «names, dates, people, and places», admiten varios atributos que permiten establecer un código de identificación para cada entidad (*@key*), o proporcionar una referencia externa con idéntica función, bien sea mediante un identificador URI (*@ref*) o mediante la forma considerada canónica (*@nymRef*). Además de estos métodos, es posible crear una clasificación propia de las entidades etiquetadas mediante el atributo *@role*, que admite una serie de valores separados por espacios en blanco (*cf.* TEI-P5, 437). Por otra parte, a la hora de identificar un segmento de texto como una referencia a una entidad, es importante codificar quién es el responsable de hacerlo, así como el grado de certeza que se posee sobre la interpretación. Para ello, la normativa incluye los atributos *@resp* y *@cert*: el primero indica la persona o institución que realizar la interpretación (en nuestro caso, habitualmente es un editor, como Wilson); el segundo contiene un valor numérico que representa el grado de certeza estimado⁹⁸.

III.3.4.1. Personas

Los elementos *<rs>* y *<name>* descritos en el capítulo §3.5 resultan insuficientes para describir el papel que juegan las palabras que hacen referencia a entidades dentro de la estructura del texto. Para ello la normativa incluye una serie de elementos concebidos específicamente para representar este tipo de palabras (*cf.* TEI-P5, §13.2, 441-452).

III.3.4.1.a. Nombres propios

Para codificar que una palabra es un nombre propio o una expresión referida a una persona, el elemento específico más adecuado es *<persName>*. Los nombres griegos carecen de una estructura fija semejante al binomio nombre-apellidos que idiomas modernos

⁹⁷ Este módulo contiene instrucciones para el etiquetado de segmentos de texto que indiquen fechas. TEI-P5, 439, recoge un ejemplo práctico de uso del elemento *<taxonomy>* para crear una serie de valores que definan periodos de tiempo concretos, que pueden emplearse con valor del atributo *@period*, y así salvar las diferencias entre el sistema de notación actual y los sistemas antiguos.

⁹⁸ Concretar el grado de certeza, que es por definición una magnitud incontable, en una escala numérica puede resultar complejo, y sólo es recomendable si es posible establecer unos criterios claros; es habitual utilizar una escala reducida de valores, definidos por una expresión y un número, por ejemplo: seguro (1), muy seguro (2), poco seguro (3), poco probable (4), imposible (5).

como el español y otros idiomas europeos han heredado del latín, razón por la cual elementos *XML* que describen este tipo de nombres (<surname>, <forename>) podrían descartarse a primera vista de nuestra selección⁹⁹. Otros elementos relacionados con <persName> y recogidos dentro de este mismo apartado pueden resultar de cierta utilidad para la creación de algún tipo de índice automático o de base de datos: <roleName> permite identificar un nombre o sintagma referido a la función o profesión de una persona y utilizado como forma de identificación del mismo; <addName> describe un nombre o expresión utilizada como forma de identificación alternativa, es decir, como un apodo o epíteto; en último lugar, según la normativa *TEI*, el elemento <genName> «contains a name component used to distinguish otherwise similar names on the basis of the relative ages or generations of the persons named» (cf. *TEI-P5*, 441), como es el caso del adjetivo *Vetus* en latín, o las indicaciones numéricas asociadas a miembros homónimos de una casa real (Felipe I, Carlos III). Los escritores griegos prefieren utilizar otros métodos identificativos, como apodos o epítetos basados en algún rasgo físico o logro destacado del personaje, por lo que no es un recurso demasiado útil en nuestro caso.

III.3.4.1.b. Nombres de «organizaciones»

El término «organización» indica aquí «any named collection of people regarded as a single unit» (cf. *TEI-P5*, 447), ya se trate de instituciones o grupos raciales o políticos, y puede etiquetarse mediante el elemento *orgName*. La normativa *TEI* recomienda el uso del atributo *@type* para indicar propiedades del nombre o sintagma en cuestión, en lugar de para describir la organización referida (una frase, un acrónimo etc.).

III.3.4.1.c. Datos biográficos y prosopográficos

Una de las ventajas del etiquetado *XML* es que facilita la elaboración de bases de datos a partir del texto. En este sentido el papel de este lenguaje es no sólo describir los elementos lingüísticos formales presentes en el texto, sino también explicitar la información que proporciona el texto añadiendo datos no presentes en el mismo. En el caso del elemento <persName>, referido a nombres de persona, es posible que los atributos básicos recogidos en la normativa se queden cortos para nuestro etiquetado descriptivo. La misma normativa contempla esta posibilidad y ofrece varios elementos que añaden información de

⁹⁹ En lenguas donde el patronímico reviste cierta sistematicidad, la normativa *TEI* también admite el uso de <surname> para su codificación (cf. *TEI-P5* 2016, 444).

tipo biográfico o prosopográfico: *<trait>*, *<state>* y *<event>* (cf. TEI-P5, §13.3, «Datos biográficos y prosopográficos»).

Estos elementos añaden tres tipos de datos a un elemento etiquetado como nombre de persona: características que no cambian con el tiempo («rasgos» o *traits*); características o rasgos válidos únicamente durante o en un momento específico en el tiempo («estados» o *states*); por último, «incidentes» (*events*) relacionados con un cambio de «estado» o la adquisición de un determinado «rasgo». Los elementos *<state>*, *<trait>* y *<event>* son subelementos dependientes de *<persName>*, es decir, que deben anidarse dentro de un elemento *<persName>* y no pueden usarse independientemente (cf. TEI-P5, 453).

III.3.4.2. Lugares

Al tratarse de una obra histórica, las *Historias* de Heródoto están llenas de datos sobre lugares tanto naturales como producto del trabajo del hombre. La normativa *TEI* recoge tres modos de describir una palabra o sintagma como una referencia a un lugar. El elemento *<placeName>* contiene uno o más nombres que indican una unidad dentro de un sistema geopolítico o administrativo, tales como provincias, países, comarcas etc. En segundo lugar, *<geogName>* describe otro tipo de referencias a lugares basadas en sus características geofísicas, como montañas o ríos. En último lugar, la normativa contempla la posibilidad de interpretar una expresión que indique cierta relación entre otros nombres de lugar como una referencia de lugar relativa, y proporciona algunas formas de codificarla de forma especial (cf. §13.2.3.3).

Además de los atributos *@key* y *@ref*, el elemento *<placeName>* admite varios subelementos con los que se describen una organización geopolítica compleja; en orden ascendente de acuerdo con el tamaño o importancia de la división (cf. TEI-P5, 450), son:

- *<district>*: válido para cualquier clase de unidad inferior a un asentamiento como un barrio, un puesto de guardia etc.;
- *<settlement>*: contiene el nombre de un asentamiento; puede ser una ciudad, un pueblo o una aldea claramente separada de otras unidades similares;
- *<region>*: se utiliza para agrupaciones de asentamientos con una entidad inferior a la de un país;
- *<country>*: describe una unidad geopolítica coincidente con lo que tradicionalmente se denomina «país» (¿una unidad política independiente?), superior

a la región e inferior a la unidad mayor recogida por la normativa *TEI*, el «bloque»;

- *<bloc>*: un «bloque» consiste en la agrupación de varios estados o países etiquetados con el elemento *<country>*, siguiendo cualquier tipo de criterio.

En lo que respecta a las divisiones geográficas descritas por *<geogName>*, es posible etiquetar por separado partes de la referencia, como puede ser un nombre común que identifique el rasgo geofísico concreto utilizado como referencia (el monte Atos):

```
<geogName type="río">
<name>Ευφρότης</name>1
<geogFeat>ποταμός</geogFeat>
</geogName>
```

Las referencias descritas mediante *<placeName>* y *<geogName>* suelen ser absolutas, en cuanto no necesitan ninguna otra especificación para tener sentido. Sin embargo, no es infrecuente la mención a un lugar por su ubicación con respecto a otro u otros lugares. En estos casos, pueden emplearse los elementos *<offset>* y *<measure>* dentro del correspondiente *<placeName>* u *<orgName>* para especificar el tipo de relación (*<offset>*; «cerca de», «a los pies de») entre los lugares que forman parte de la referencia, así como la distancia que los separa, de estar especificada (*<measure>*).

III.3.4.3. Relaciones entre entidades

La adición de todos estos datos y el propio uso de elementos puede saturar un documento *XML* volviéndolo ilegible para un ser humano, e incluso puede dificultar el procesamiento posterior por ordenador, si ese es el objetivo del etiquetado¹⁰⁰. Resulta interesante poseer y utilizar un sistema de referencias que nos permita separar con facilidad los datos añadidos mediante el etiquetado del propio texto, reduciendo la presencia de lenguaje *XML* en el texto al mínimo necesario. Hemos mencionado varios atributos que cumplen precisamente esta función, tal como *@ref*, *@xml:id*, *@key* o *@nym*. Cada uno de ellos proporciona una identificación única para cada entidad nombrada en el texto, independientemente de las palabras exactas que constituyen la referencia. El uso de estos métodos no elimina la necesidad de definir cada entidad etiquetada mediante una serie de datos, sino que limita el

¹⁰⁰ Recordemos que un ordenador debe «leer» todo el documento para procesar los datos, independientemente de que sólo resulte interesante para la búsqueda una sección concreta (*cf. supra, II.2.1.2. Funcionamiento:software*).

esfuerzo de introducir estos datos a un sólo lugar, al que el futuro lector o programa informático que procese el texto será remitido después¹⁰¹.

Los posibles valores de estos atributos, es decir, de las referencias utilizadas, se dividen en dos clases: referencias externas al texto y referencias internas, incluidas dentro del mismo documento. En el primer caso la definición de las entidades a las que dirige la referencia se encuentra fuera del documento *XML* y su uso puede implicar una ausencia de responsabilidad por parte del etiquetador, porque se trate de un estándar dentro de su disciplina. Un ejemplo de este tipo es el sistema de referencias *CITE-URN*.

Si el etiquetado del texto está reflejando una interpretación particular, si la identificación de las entidades mencionadas en el texto es un tema sujeto a controversia, o si se trata de un área de conocimiento poco estudiada, encontrar un estándar ya definido que utilizar como referencia es una tarea mucho más difícil. En estos casos resulta conveniente codificar dentro del mismo documento *XML* una definición de las entidades a las que sea necesario referirse en un lugar concreto del documento, separado del cuerpo del texto¹⁰². Los elementos que analizamos a continuación sirven a este propósito.

Los elementos *<listPerson>* (TEI-P5, 454), *<listRelation>* (TEI-P5, 461), *<listOrg>* (TEI-P5, 462), *<listPlace>* (TEI-P5, 464), funcionan como contenedores pensados para reunir definiciones de personas, relaciones, organizaciones o lugares mencionados en el texto y etiquetados con su correspondiente elemento (*cf. supra*). Cuando hablamos de definición aquí nos referimos a una serie de datos marcados mediante un elemento apropiado, y no a una descripción narrada del tipo que podamos encontrar en un diccionario. No olvidemos que en la práctica un documento *XML* es una herramienta para facilitar el procesamiento y recuperación de información por ordenador, y que el funcionamiento de este último condiciona la forma en la que se presentan los datos (*cf. supra, II.2.1. El ordenador*).

La información relativa a personas se codifica mediante un elemento *<person>*, o *<personGrp>*, en el caso de grupos de individuos tratados como una única persona¹⁰³. Una

¹⁰¹ Lo que sigue es un resumen de TEI-P5, 206, 453-473.

¹⁰² La definición de las entidades puede aparecer dentro de la cabecera, por ejemplo, o anidada en elemento *<particDesc>*, «descripción de los participantes» (*cf. TEI-P5, 454*).

¹⁰³ La normativa recomienda utilizar el elemento *<personGrp>* para codificar agrupaciones contextuales, no estables de individuos. Cuando un grupo de individuos posee una identidad estable, marcada por la existencia de un nombre propio que los identifique (los escitas, los atenienses), hablamos de una «organi-

forma simple de utilizarlos es asignar un código de identificación a cada uno mediante el atributo *@xml:id*, y consignar dentro de ellos las diferentes formas presentes en el texto en las que aparece mencionada la entidad:

```
<listPerson>
  <person xml:id="Darius.Histaspes">
    <persName>Δαρεΐος ὁ ...</persName>
  </person>
  <person xml:id="Darius">
    <persName>Δαρεΐος</persName>
  </person>
</listPerson>
```

Además de los atributos típicos de la clase de elementos a la que pertenece (*@type*, *@cert*, *@resp*, *@source*), *<person>* y *<personGrp>* admiten tres elementos propios: *@role*, *@sex* y *@age*. Su uso está limitado a datos simples. Para una definición más compleja se recomienda el uso de los elementos reseñados al hablar de los datos biográficos y prosopográficos (*cf.* TEI-P5, 455-461). Estos mismos elementos y atributos pueden utilizarse para describir organizaciones mediante el elemento *<org>*¹⁰⁴.

Naturalmente, los nombres de lugar requieren atributos y subelementos específicos y diferentes de los ya mencionados para su correcta definición. Para ello se utiliza el elemento *<location>*, cuyo valor puede tener varias formas (*cf.* TEI-P5, 465). En el caso de Heródoto, la opción final elegida dependerá de la aplicación práctica final del texto etiquetado. Mientras que un proyecto como *Hellespont* seguramente se decante por proporcionar una serie de coordenadas para cada lugar (mediante un elemento *<geo>*), otros investigadores con un interés más literario y menos histórico pueden contentarse con una descripción textual (elemento *<desc>*) o con una ubicación genérica mediante los elementos mencionados a la hora de hablar de *<placeName>*.

Además de clasificar las definiciones de las entidades personales, organizaciones y lugares mencionadas en el texto, es posible y conveniente explicitar las relaciones exis-

zación», definida mediante el elemento *<org>* (*cf.* TEI-P5, 455). Sobre la extraordinaria importancia de los nombres en Heródoto, *cf.* Campos Daroca 1992, 87-141.

¹⁰⁴ El elemento *<org>*, adecuadamente ubicado dentro de una lista (*<listOrg>*) resulta útil no sólo para relacionar diferentes denominaciones de la misma entidad colectiva, sino para introducir interpretaciones propias sobre la razón de dichos cambios, como pueda ser la asignación de una denominación a un periodo histórico concreto (que la liga de Delos recibiera otro nombre antes etc.; *cf.* TEI-P5, 463 para un ejemplo).

tentes entre las entidades mediante un elemento *<listRelation>* y *<relation>* (cf. TEI-P5, 461-462). Para la normativa *TEI*, una relación es «any kind of describable link between specified participants» (cf. TEI-P5, 461), y puede definirse formalmente mediante una sucesión de datos relativos a los participantes (atributo *@name*), y al tipo de relación que establecen entre sí. La visión de la normativa con respecto a la tipología de las relaciones entre los participantes es estrictamente jerárquica: o bien todos los participantes se encuentran al mismo nivel, o uno de ellos ejerce un papel dominante. En el primer caso, la relación se describe con el atributo *@mutual*; en el segundo, debemos consignar qué participantes ejercen un papel activo (atributo *@active*) y quiénes un papel pasivo (atributo *@passive*).

III.4. TOPIC MODELS

III.4.1. TOPIC MODELING EN HUMANIDADES

El estudio de Templeton 2011 sobre la aplicación de la técnica de *topic modeling* en las Humanidades divide los trabajos publicados hasta entonces en enfoques sincrónicos y diacrónicos. Los enfoques sincrónicos identifican temas en un momento determinado: Jockers 2010, por ejemplo, usó *topic models* para clasificar por tema los escritos de 177 autores de blogs que escribieron durante el «Día de las Humanidades Digitales» en 2010; por otro lado, Meeks 2011 utilizó esta técnica para identificar las definiciones que diversos estudiosos del campo de las Humanidades y de las Humanidades Digitales daban de su propia disciplina, utilizando una muestra bastante pequeña (50 textos), y a partir de los datos obtenidos fue capaz de crear una visualización de los resultados en forma de redes de investigadores y temas; similar a este es el examen de Proust realizado por Drouin 2011, para identificar temas y mostrarlos como una red gráfica.

En segundo lugar, los enfoques diacrónicos estudian la evolución del tratamiento de unos temas a lo largo del tiempo. Este enfoque es más complejo a nuestro parecer, puesto que, además de la detección de los temas en cada uno de los textos de muestra, trata de establecer relaciones entre diversos autores. Para reducir al mínimo los posibles errores debidos a diferencias de estilo o temática ignoradas por el algoritmo, el enfoque diacrónico suele aplicarse a corpus de textos que mantienen una coherencia conocida y comprobada a lo largo de un periodo determinado, como por ejemplo publicaciones periódicas. Abundan en este sentido los estudios de periódicos publicados en los tres últimos siglos, como el trabajo de Block & Newman 2006 sobre la evolución de los temas tratados en el periódico *Pennsylvania Gazette* entre 1728 y 1800. En el ámbito científico, Griffiths & Steyvers 2004 extrajeron con éxito *topic models* de los *abstracts* publicados en la revista *PNAS* entre 1991 y 2001, con los que analizaron la evolución de los intereses de la revista en este periodo. Dentro de la Filología, debemos a Mimno 2012 un estudio similar con 24 revistas de Filología Clásica y Arqueología a lo largo de los últimos 150 años, en el que se aprecia la evolución de los temas tratados y los puntos que separan o unen entre sí a las publicaciones.

III.4.2. TOPIC MODELING DE TUCÍDIDES

Nuestro punto de partida es el análisis de Tucídides mediante *topic modeling* realizado por Thomas Köntges¹⁰⁵. Köntges empezó a trabajar en una interfaz para el algoritmo de *topic modeling* basado en *LDA* planteado por Blei 2012. El resultado es un programa que se ejecuta a través del lenguaje procedimental *R*, capaz de generar *topic models* de un texto de forma relativamente sencilla; los resultados adoptan la forma de una tabla de datos tradicional, así como una visualización interactiva en forma de gráficos interactivos que utilizan la distancia y el tamaño relativos de una serie de círculos de color en un plano cartesiano para representar la frecuencia y relación entre los temas detectados. Köntges realizó una prueba con algunos pasajes de Tucídides: el programa detectaba una serie de temas, y la revisión de los términos incluidos en alguno de ellos permitía incluso ponerles «nombre» (acciones de los atenienses, acciones de los espartanos, acciones navales). A día de hoy, sin embargo, no se ha publicado ningún estudio científico que desarrolle estos resultados preliminares que, por lo tanto, mantienen el estatus de una prueba preliminar.

A partir de estos primeros resultados, Köntges mejoró aquella primera herramienta y la publicó bajo el nombre de *(Meletē)ToPān*, o simplemente, *ToPan*¹⁰⁶. Movidos por el atractivo de esta nueva herramienta nos dispusimos a intentar emular lo conseguido con Tucídides para el texto de su predecesor Heródoto. No sólo el tamaño de la muestra era mayor, sino que el dialecto particular en el que se escribieron las *Historias* debía ser adaptado antes de llevar a cabo el análisis propiamente dicho.

III.4.3. TOPIC MODELING DE LAS HISTORIAS

III.4.3.1. Preparación del texto

ToPan incorpora un tutorial no muy extenso pero suficiente para comprender las fases fundamentales del proceso, a saber:

- a) *Introducción de los datos*. *ToPan* permite utilizar un texto alojado en la base de datos del proyecto *Perseids*, así como utilizar un archivo modificado y alojado en el mismo ordenador desde el que se ejecuta el programa. En el segun-

¹⁰⁵ Los resultados descritos están disponibles en <<https://github.com/ThomasK81/TopicModellingR/>>.

¹⁰⁶ Cf. Köntges 2016b; el programa en sí está disponible en <https://github.com/ThomasK81/ToPan>.

do caso, las opciones representan diferentes formas de modificación del texto con el propósito de resaltar diversos datos y facilitar su procesamiento por ordenador.

- b) *Normalización morfológica del texto*. El método concreto utilizado depende del tipo de archivo que se utilice en el primer caso. Para nuestro estudio nos decantamos por utilizar la versión del texto disponible en el proyecto *Perseids*, de modo que la normalización del texto corre a cargo de la herramienta de análisis morfológico *Morpheus*, desarrollada para *Perseus*.
- c) *Fijación de los valores para la generación de topic models*. Estos valores incluyen el número de temas estimado para el texto, el número de palabras que deben ser excluidas del análisis (*stopwords*) etc. Aquí resulta particularmente frustrante la ausencia de algún tipo de explicación sobre los diferentes ajustes que pueden hacerse antes de obtener los resultados.
- d) *Análisis de los resultados*. *ToPan* ofrece dos tipos de resultados: una pareja de tablas, denominadas ϕ y θ , donde se indican las relaciones de frecuencia entre temas y pasajes del texto.

La preparación previa del texto suele denominarse «normalización», y consiste en la adaptación de la forma lingüística del texto para facilitar el procesamiento informatizado del mismo. Como el nombre indica, normalizar un texto implica modificar cualquier forma anómala o extraña y sustituirla por otra «canónica» o «normalizada» equivalente. Para esta tarea recurrimos en un primer momento a un conjunto de herramientas denominadas *Classical Language ToolKit (CLTK)*, ejecutadas a través del lenguaje procedimental *Python*¹⁰⁷. De este modo fuimos capaces de crear una versión de cada libro en la que cada término era sustituido por la forma principal de su lema. Sin embargo, como ya hemos comentado, *ToPan* proporciona su propio método de lematización; este procedimiento reveló ser más preciso a la hora de asignar lemas a las formas del texto que el *CLTK*, lo que nos llevó a abandonar este conjunto de herramientas en favor de *Morpheus*. La segmentación del texto se realiza también automáticamente gracias a la estructura *CITE-URN* que ya incorpora la versión etiquetada de *Perseus*. Esto permite preservar la división canónica del texto, lo que

¹⁰⁷ El *CLTK* es un conjunto especializado de herramientas de procesamiento y análisis lingüístico por ordenador, dirigido a filólogos, y basado en otro grupo de herramientas denominado *NLTK (Natural Language ToolKit)*, desarrollado para lingüistas en general. La documentación está disponible en <<http://docs.cltk.org/en/latest/greek.html>>.

facilita el análisis posterior de los resultados. Cada una de estas secciones se comportará como una «unidad semántica mínima», que será comparada con las demás en cuanto a la similitud de lemas que aparecen en ellas y a su distribución.

III.4.3.2. Resultados

Aunque los defensores de este tipo de técnicas recurren con frecuencia al atractivo visual de las representaciones gráficas, los *topic models* generados por *ToPan* pueden descargarse en forma de una tabla de datos. Este último modo de representación facilita la correlación entre los temas localizados y las secciones «canónicas» del texto, compensando la pérdida de una visión de conjunto con un desglose más detallado de los datos¹⁰⁸. En anexo incluimos un resumen de la información obtenida para cada uno de temas, dividido en dos bloques: en primer lugar se indican las palabras que componen el «tema» y se incluyen todas las observaciones relativas a estas¹⁰⁹.

Como ya mencionamos al hablar del funcionamiento del ordenador, y más en concreto, de los algoritmos de *ASL* en los que se basa la técnica del *topic modeling* (cf. *supra*, II.3.1. *Análisis semántico latente*), la interpretación de los datos obtenidos, esto es, la generación de información, es una tarea exclusiva del usuario humano. En consecuencia, los temas obtenidos mediante *ToPan* se identifican simplemente mediante una referencia numérica. A pesar de nuestras expectativas, aún es necesario mucho trabajo para convertir la técnica del *topic modeling* en una herramienta provechosa de recuperación de información de textos en griego antiguo. Como afirma Murdock & Allen 2015, «topic models remain a black box both for modelers and for end users [...] many decisions [...] lack clear rationales [...] the results are hard to analyze, summarize, and visualize».

¹⁰⁸ Después de seleccionar *stopwords* y ejecutar el programa, obtuvimos quince *topic models* y una lista de los veinticinco términos más sobresalientes en todo el texto.

¹⁰⁹ Debido al modo en que está diseñado *ToPan*, los errores de lematización no son infrecuentes y es necesario explicar sus causas y soluciones.

IV. CONCLUSIONES

A lo largo de este trabajo hemos intentado ofrecer un panorama de las técnicas de recuperación automática de información desarrolladas dentro y fuera del ámbito de la Filología Clásica más importantes, dirigidos por el deseo de mejorar su utilización actual. De todas ellas, el etiquetado *XML* resulta el método *RI* más transparente, que otorga al investigador un poderoso instrumento de codificación. En efecto, si bien no permite un análisis *a priori* de los datos, el lenguaje *XML* posibilita la expresión de un número indefinido de rasgos textuales, de un modo que no sólo resulta comprensible para un lector humano, sino para un programa informático. Las únicas limitaciones al tipo de rasgos codificables son pocas, además de una consecuencia de la naturaleza lógica y matemática de cualquier mecanismo informático digital: *a)* los datos deben tener un principio y un final claro y *b)* los tipos de datos y las relaciones entre los mismos deben estar perfectamente definidas. De la primera condición se deriva la regla del anidamiento que comentamos antes (*cf. supra*, II.2.4. *Lenguajes descriptivos*), y la característica estructura jerárquica de un documento *XML*. Al margen de estas observaciones, la expresividad del lenguaje *XML* y el control que ofrece sobre los datos lo convierten en una herramienta muy apropiada para el procesamiento de texto por ordenador. Por estas razones, a partir del análisis del procedimiento de generación de *topic models*, dirigimos la mirada a los lenguajes descriptivos como método de *RI* alternativo. En este caso, nos hemos centrado en la optimización del uso del sistema de etiquetado *XML* para codificar la información detectada en el texto.

En efecto, hemos tratado de mostrar que el uso del lenguaje *XML* hecho por *Perseus* resulta pobre en comparación con las posibilidades que ofrece, recogidas en la teoría y con ejemplos en la normativa *TEI*. El texto de Heródoto, con su riqueza estructural y temática, derivada de la naturaleza oral del texto transmitido, es un campo de trabajo perfecto para demostrar la versatilidad de esta herramienta, y esperamos que lo expuesto anime a otros a aplicarla a otros autores.

Por otra parte, aunque las herramientas de *ASL* son prometedoras, aún están lejos de proporcionar unos resultados equiparables a los que consigue un ser humano mediante una lectura atenta. La aportación de la técnica de *topic modeling* se reduce a una confirmación estadística de las relaciones entre varios grupos de palabras presentes en el texto de las *Historias*, pero la atribución de un significado más profundo a estos conjuntos de términos ha resultado delicada, e infructuosa en líneas generales. Los defensores del empleo de este tipo de técnicas mencionan su utilidad para observar la presencia continua de estos conjuntos dentro de un corpus en el que existe una cierta progresión. Sin embargo, al respecto

sigue resultando mucho más útil un índice de palabras clave, tal como el *index nominum* o el *index rerum* que suele acompañar tradicionalmente a cualquier edición crítica. Estos índices son fruto de un estudio largo y riguroso, y el hecho de que sean confeccionados a mano no les resta fiabilidad, ni mucho menos el nivel de exhaustividad que viene asociado al empleo de un ordenador¹.

¹ El léxico de las *Historias* realizado por Powell 1939 es un ejemplo del gran nivel de precisión que puede alcanzarse sin necesidad de procedimiento mecánico alguno. Lo mismo podemos decir de otras obras lexicográficas como el *DGE*, en las que la introducción del ordenador ha sido tardía (*cf.* Rodríguez Adrados & Rodríguez Somolinos 1994).

LISTA DE ABREVIATURAS

Para los nombres de los autores antiguos y los títulos de sus obras seguimos las abreviaturas del *Diccionario Griego-Español* y del *Thesaurus Linguae Latinae*. En el caso de las revistas, hemos empleado las abreviaturas recogidas en el listado de *L'Année Philologique*. Allí donde se trata de revistas no presentes en el mismo, damos el enunciado completo del título y, en su caso, del subtítulo.

- FGH* Jacoby, F. (ed.), (1923-1958), *Die Fragmente der griechischen Historiker*, vol. I-III C 2, Berlin/Leiden, Brill («vermehrte Neudrucke», Leiden, Brill, 1954-).
- LSAG* Jeffery, L. H. (1961), *The local scripts of archaic Greece*, Oxford, Clarendon.
- LIMC* *Lexicon Iconographicum Mythologiae Classicae* (1981-2009), 8 vols.; *Indices*, 2 vols.; *Supplements*, 2 vols., Zurich, Artemis.
- TEI-P5* TEI Consortium (eds.), (2016), *TEI P5: Guidelines for electronic text encoding and interchange. Version 3.0.0. Last updated on 29th March 2016, revision 89ba24e*, TEI Consortium.

BIBLIOGRAFÍA

- Adam, J.-M. (1992), *Les textes: types et prototypes*, Paris, Nathan.
- Adam, J.-M. (1999), *Linguistique textuelle. Des genres des discours aux textes*, Paris, Nathan.
- Adams, M. J. (1990), *Beginning to read: thinking and learning about print*, Cambridge, MIT Press.
- Ahl, F. M. (1967), «Cadmus and the palm-leaf tablets», *AJPh*, 88, 188-194.
- Alam, F. *et al.* (2015), «Resource description framework and topic maps: complementary or competitive?», *Proceedings of the Pakistan Academy of Sciences*, 52 (1), 1-14.
- Alcaraz Varó, E. & Martínez Linares, M. A. (1997), *Diccionario de lingüística moderna*, Barcelona, Ariel.
- Allen, T. M. (1889), *Notes on abbreviations in Greek manuscripts*, London, Henry Frowde.
- Almas, B. (2012), *NEH Institute Working with Text in a Digital Age: digital edition demonstration and sample code*, [página web], consultado el 4-4-2017, <<https://github.com/balmas/tei-digital-age>>.

- Almas, B. *et al.* (2013), «The linked fragment: TEI and the encoding of text re-use of lost authors» en *The Linked TEI: Text Encoding in the Web. TEI Conference and Members Meeting 2013*, Roma, DIGILAB Università della Sapienza & TEI Consortium.
- Alsina, J. (1963), «Notas para una historia de la Filología Clásica», *Convivium*, 15-16, 169-191.
- Aly, W. (1969), *Volksmärchen, Sage und Novelle bei Herodot und seinen Zeitgenossen. Eine Untersuchung über die volkstümlichen Elemente der altgriechischen Prosaerzählung*, Göttingen, Vandenhoeck & Ruprecht.
- Amadasi, M. G. (1998), «Sulla formazione e diffusione dell' alfabeto», en G. Bagnasco & F. Cordano (eds.), *Scritture mediterranee fra IX e VIII sec. a.C. Atti del Seminario: Università degli Studi di Milano, Istituto di Storia Antica, 23-24 febbraio 1998*, Milano, Edizioni Et, 27-51.
- Ananda, A. L. & Srinivasan, B. (1984), «An extensive bibliography on computer networks», *ACM SIGCOMM Computer Communication Review*, 13-14 (5-1), 78-98.
- Andersen, O. (1987), «Mündlichkeit und Schriftlichkeit im frühen Griechentum», *Antike und Abendland*, 33, 29-44.
- Andersen, O. (1989), «The significance of writing in early Greece: a critical appraisal», en K. Schousboe & M. T. Larsen (eds.), *Literacy and society*, Copenhagen, Akademisk Forlag, 73-90.
- Anderson, J. R. (2015⁸), *Cognitive psychology and its implications*, New York, Worth.
- Armayer, O. C. (1985), *Herodotus' autopsy of the Fayoum: lake Moeris and the labyrinth of Egypt*, Amsterdam, J. C. Gieben.
- Aschoff, V. (1983), «The early history of the binary code», *IEEE Communications Magazine*, 21 (1), 4-10.
- Astarita, M. L. (1993), *La cultura nelle Noctes Atticae*, Catania, Università di Catania.
- Astle, D. E. & Scerif, G. (2011), «Interactions between attention and visual short-term memory (VSTM): What can be learnt from individual and developmental differences?», *Neuropsychologia*, 49 6, 1435-1445.

- Astrid Ciro, L. (2014), *El léxico de la informática e Internet en el ámbito hispanohablante*, Lleida, Universidad de Lleida.
- Asunción, A. *et al.* (2009), «On smoothing and inference for topic models», en J. Bilmes & A. Ng (eds.), *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI) 2009. Montreal, Canada, 18-21 June 2009*, Corvallis, AUAI Press, 27-34.
- Austin, J. (1962), *How to do things with words*, Oxford, Clarendon.
- Baddeley, A. (1966), «Short-term memory for word sequences as a function of acoustic, semantic and formal similarity», *Quarterly Journal of Experimental Psychology*, 18 4, 362-365.
- Baddeley, A. & Hitch, G. (1974), «Working memory», en Gordon, H. B. (eds.), *Psychology of learning and motivation*, New York, Academic Press, 47-89.
- Baddeley, A. *et al.* (1975), «Word length and the structure of short-term memory», *Journal of Verbal Learning and Verbal Behaviour*, 14, 575-589.
- Baddeley, A. (1983), *Your memory: a user's guide*, Harmondsworth, Penguin Books.
- Baddeley, A. (2000), «The episodic buffer: a new component of working memory», *Trends in Cognitive Science*, 4 (11), 417-423.
- Bakker, E. J. (2008), «Epic remembering» en Mackay, E. A. (eds.), *Orality, literacy, memory in the ancient Greek and Roman world*, Leiden, Brill, 65-77.
- Bakker, E. J., De Jong, I. J. F. & Van Wees, Hans (eds.) (2002), *Brill's companion to Herodotus*, Leiden, Brill.
- Balota, D. A. & Marsh, E. J. (2004), *Cognitive psychology: key readings*, New York, Psychology Press.
- Bamman, D. & Crane, G. (2009), «Structured Knowledge for Low-Resource Languages: The Latin and Ancient Greek Dependency Treebanks», en G. Heyer (ed.), *Text Mining Services – Building and applying text mining based service infrastructures in research and industry, Proceedings of the Conference on Text Mining Services – TMS 2009 at Leipzig University*, Leipzig, Springer.

- Bamman, D. *et al.* (2010), «Corpus linguistics, treebanks and the reinvention of philology», en W. Abramowicz *et al.* (eds.), *Informatik 2010. Business process and service science. Proceedings of ISSS and BPSC, september 27 – october 1, 2010, Leipzig, Germany*, Bonn, Köllen Druck, 542-551.
- Barbier, F. (2005), *Historia del libro*, Madrid, Alianza.
- Barnard, D. *et al.* (1988), «SGML-based markup for literary texts: two problems and some solutions», *Computers and the Humanities*, 22, 265-276.
- Bartolomé, J., Cruz González, M. & Quijada, M. (ed.) (2004), *La escritura y el libro en la Antigüedad*, Madrid, Ediciones Clásicas.
- Bassi, K. (1998), *Acting like men: gender, drama, and nostalgia in ancient Greece*, Ann Arbor, University of Michigan Press.
- Beazley, J. D. (1948), «Hymn to Hermes», *AJA*, 52 (3), 336-340.
- Beck, I. (1971), *Die Ringkomposition bei Herodot und ihre Bedeutung für die Beweisstechnik*, Hildesheim, Weidmann.
- Benveniste, E. (1966), *Problèmes de linguistique générale*, Paris, Gallimard.
- Berenguer Amenós, J. (1960), *Heródoto, Antología de las Historias*, Barcelona, Bosch.
- Berenguer Amenós, J. (ed. y trad.) (1960b), *Heródoto. Historias. Libro I*, Madrid, CSIC.
- Berenguer Amenós, J. (ed. y trad.) (1971), *Heródoto. Historias. Libro II*, Madrid, CSIC.
- Bernabé Pajares, A. & Hernández Muñoz, F. G. (2010²), *Manual de crítica textual y edición de textos griegos*, Madrid, Akal.
- Berners-Lee, T., Hendler, J. & Lassila, O. (2001), «The semantic web», *Scientific American*, 284 (5), 28-37.
- Berners-Lee, T. & Fischetti, M. (1999), *Weaving the web*, San Francisco, Harper-Collins.
- Berners-Lee, T. & Hendler, J. (2001), «Publishing on the semantic web», *Nature*, 410 6832, 1023-1024.

- Blei, D. M. (2012), «Probabilistic topic models», *Communications of the ACM*, 55 (4), 77-87.
- Blei, D. M. & Jordan, M. (2003), «Latent Dirichlet allocation», *Journal of Mathematical Research*, 3, 993-1022.
- Block, S. & Newman, D. (2006), «Probabilistic topic decomposition of an eighteenth-century newspaper», *Journal of the American Society for Information Science and Technology*, 57 5.
- Boardman, J. (ed.) (1981-2009), *Lexicon iconographicum mythologiae classicae*, Zurich, Artemis & Winkler Verlag.
- Boardman, J. (1992), «Linos», *LIMC*, 6 (1), 290.
- Bonifazi, A. (2008), «Memory and visualization in homeric discourse markers» en Mackay, E. A. (eds.), *Orality, literacy, memory in the ancient Greek and Roman world*, Leiden: Brill, 35-65.
- Borgeaud, P. (ed.) (1988), *Le mémoire des religions*, Genève, Labor et Fides.
- Börner, K. et al. (2007), «Network science», *Annual Review of Information Science and Technology*, 41, 537-607.
- Bornitz, H.-F. (1968), *Herodot-Studien. Beiträge zum Verständnis der Einheit des Geschichtswerkes*, Berlin, De Gruyter.
- Bouvier, D. (1988), «L'aède et l'aventure de mémoire. Remarques sur le problème d'une dimension religieuse de la mémoire dans l'*Iliade* et l'*Odyssee*», en P. Borgeaud (ed.), *Le mémoire des religions*, Genève: Labor et Fides, 63-79.
- Bowman, A. & Woolf, G. (ed.) (1994), *Literacy and power in the ancient world*, Cambridge, Cambridge University Press.
- Bradley, N. (2002), *The XML companion*, Boston, Addison-Wesley.
- Bremer, J. M., De Jong, I. J. F. & Kalff, J. (ed.) (1987), *Homer: beyond oral poetry: recent trends in Homeric interpretation*, Amsterdam, Grüner.
- Bremmer, J. N. (1999), «Linos», *Neue Pauly*, 7, 252-253.

- Brisson, L. (1982), *Platon, les mots et les mythes*, Paris, Maspero
- Brown, R. & Kulik, J. (1977), «Flashbulb memories», *Cognition*, 5 (1), 73-99.
- Brunova, E. (2013), «Parametrical words in the sentiment lexicon», *International Journal of Cognitive Research in Science, Engineering and Education*, 1 (2), 57-64.
- Buntine, W. (2002), «Variational extensions to EM and multinomial PCA» en T. Elomaa *et al.* (eds.), *Proceedings of the 13th European Conference on Machine Learning, Helsinki, Finland, August 19-23, 2002*, Berlin, Springer, 23-34.
- Buntine, W. & Jakulin, A. (2006), «Discrete component analysis», en C. Saunders *et al.* (eds.), *Subspace, latent structure and feature selection*, Berlin, Springer, 1-33.
- Burghart, M. (2013), *TEI cheatsheets*, [página web], consultado el 4-4-2017, <http://wiki.tei-c.org/index.php/TEI_Cheatsheets>.
- Burnard, D. (2014), *What is the text encoding initiative? How to add intelligent markup to digital resources*, Marseille, OpenEdition Books.
- Busa, R. (1975), *Index Thomisticus: Sancti Thomae Aquinatis operum omnium indices et concordantiae (Indices)*, Stuttgart, Frommann-Holzboog.
- Buzzetti, D. (2002), «Digital representation and the text model», *New Literary History*, 33 (1), 61-88.
- Calame, C. (1993), «Rythme, voix et mémoire de l'écriture en Grèce classique», en R. Pretagostini (ed.), *Tradizione e innovazione nella cultura greca da Omero all'età ellenistica: scritti in onore di Bruno Gentili*, 3 vols., Rome, Gruppo Editoriale Internazionale, II, 785-799.
- Caleme, C. (2000), «Orphik, Orphische Dichtung», *Neue Pauly*, 9, 58-70.
- Calinescu, M. (1993), «Orality in literacy: some historical paradoxes of reading», *YJC*, 6, 175-190.
- Campbell-Kelly, M. (2009), «Origin of computing», *Scientific American*, 301 (3), 62-69.
- Campbell-Kelly, M. *et al.* (2004³), *Computer: a history of the information machine*, Boulder, Westview.

- Campos Daroca, J. (1991), «La envidia de los griegos: aspectos de la memoria en la primera historiografía», *Florilib*, 2, 93-98.
- Campos Daroca, J. (1991b), «Saber y poder en las *Historias* de Heródoto», *Florilib*, 3, 101-115.
- Campos Daroca, J. (1992), *Experiencias del lenguaje en las Historias de Heródoto*, Almería, Instituto de Estudios Almerienses.
- Campos Daroca, J., Fuentes González, P. P. & López Cruces, J. L. (1993), «La *Biblia* y Heródoto», en *Biblia y culturas. IV Simposio Bíblico Español (I Simposio Bíblico Ibero-Americano)*, Granada, 7-10 sept. 1992, Valencia / Granada, Fundación Bíblica Española, 245-251.
- Campos Daroca, J. & López Cruces, J. L. (1994), «Investigación lingüística e historiográfica en las *Historias* de Heródoto», en *Actas del VIII Congreso Español de Estudios Clásicos (Madrid, 23.28 de septiembre de 1991)*, vol. I, Madrid, Ed. Clásicas, 79-86.
- Canet Vallés, J. L. (2014), «Reflexiones sobre las humanidades digitales», en S. López Poza & N. Pena Sueiro (eds.), *Humanidades digitales: desafíos, logros y perspectivas de futuro*, A Coruña, Universidade da Coruña, SIELAE, 11-20.
- Canettieri, P. et al. (2008), «Philology and Information Theory», *Cognitive Philology*, 1.
- Carmona Vázquez, A. (2000), «Estudio de campos semánticos en griego: presupuestos lingüísticos», *ExcPhilol*, 10, 47-62.
- Carrasco Serrano, G. & Oliva Mompeán, J. C. (ed.) (2010), *El mediterráneo antiguo: lenguas y escrituras*, Madrid, Centro de Estudios del Próximo Oriente.
- Cartledge, P. A. (2002²), *Sparta and Lakonia: a regional history, 1300-362 b. C.*, London, Routledge.
- Castillo Blasco, L. (2006), *Elaboración de un tesoro de información de actualidad y conversión en red semántica para su empleo en un sistema de recuperación periodístico*, Valencia, Universidad de Valencia.
- Cavallo, G. (2004), «Entre el *volumen* y el *codex*: la lectura en el mundo romano», en G. Cavallo & R. Chartier (eds.), *Historia de la lectura en el mundo occidental*, Madrid, Santillana, 109-317.

- Cavallo, G. & Chartier, R. (ed.) (2004), *Historia de la lectura en el mundo occidental*, Madrid, Santillana.
- Ceccarelli, P. (2013), *Ancient Greek letter writing: a cultural history (600 BC – 150 BC)*, Oxford, Oxford University Press.
- Chabert, J.-L. (1999), *A history of algorithms: from the pebble to the microchip*, Berlin, Springer.
- Chacón Beltrán, M. R. (2008), «El uso de *expresiones regulares* en la detección de errores escritos: implicaciones para el diseño de un corrector gramatical», en A. Moreno Sandoval (ed.), *El valor de la diversidad (meta)lingüística. Actas del VIII Congreso de Lingüística General*, Madrid, UAM, 27-34.
- Chadwick, J. (1973²), *Documents in Mycenaean Greek*, Cambridge, Cambridge University Press.
- Chambers, M. *et al.* (1981), *The Oxyrrhyncus papyri, XLVIII, edited with translations and notes*, London, Egypt Exploration Society.
- Chantraine, P. (1950), «Les verbes grecs signifiant “lire”», en *Παγκραπεία. Mélanges Henri Grégoire*, Brussels, Éditions de l’Institut de Philologie et d’Histoire Orientales et Slaves, 115-126.
- Charaudeau, P. & Maingueneau, D. (2002), *Diccionario de análisis del discurso*, Buenos Aires / Madrid, Amorrortu.
- Chartier, R. (2004), «Books, and reading from the printed word to the digital text», *Critical Inquiry*, 311, 133-152.
- Chassant, L. A. (1884), *Dictionnaire des abréviations latines et françaises usitées dans les inscriptions lapidaires et métalliques, les manuscrits et les chartes du Moyen Âge*, Paris, Jules Martin.
- Chesley, A. (2012), *TEI: an overview*, [video], consultado el 4-4-2017, <<https://www.youtube.com/watch?v=R6iiIFrWvmU>>.
- Chica, A. B. *et al.* (2013), «Two cognitive and neural systems for endogenous and exogenous spatial attention», *Behavioural Brain Research*, 237, 107-123.

- Cobet, J. (1971), *Herodots Exkurse und die Frage der Einheit seines Werkes*, Wiesbaden, Franz Steiner.
- Cogan, M. (1981), *The human thing: the speeches and principles of Thucydides' History*, Chicago, University of Chicago Press.
- Cole, M. & Scribner, S. (1981), *The psychology of literacy*, Cambridge, Harvard University Press.
- Coleman, J. (1992), *Ancient and medieval memories*, Cambridge, Cambridge University Press.
- Connors, R. J. (1993), «*Actio: a rhetoric of written delivery (iteration two)*», en J. F. Reynolds (ed.), *Rhetorical memory and delivery: classical concepts for contemporary composition and communication*, Hillsdale, NJ, Routledge, 65-77.
- Conrad, R. & Hull, A. J. (1964), «Information, acoustic confusion and memory span», *British Journal of Psychology*, 55, 429-432.
- Coombs, J. H. et al. (1987), «Markup systems and the future of scholarly text processing», *Communications of the ACM*, 30, 933-947.
- Corcella, A. (1986), «Dionisio Skytobrachion, i *phoinikeia*, e l'«alfabeto pelasgico»: per una corretta interpretazione di Diodoro III.67.1 (FGrHist 32 F 8)», *AAT*, 120, 41-82.
- Cornellá, A. (2004²), *Infoxicación: buscando un orden en la información*, Barcelona, Informía.
- Crane, G. (2014), *Opening up Classics and the Humanities: computation, the Homer Multitext project and citizen science*, *Perseus Digital Library Updates*, [artículo de blog], consultado el 4-4-2017, <<https://sites.tufts.edu/perseusupdates/>>.
- Crane, G. (2004), «Classics and the computer: an end of the history», en S. Schreibman et al. (eds.), *A companion to Digital Humanities*, Oxford, Blackwell, 46-56.
- Crane, G. et al. (2009), «Classics in the million book library», *Digital Humanities Quarterly*, 3 (1).
- Crookes, G. (1990), «The utterance, and other basic units for second language discourse analysis», *Applied Linguistics*, 11 (2), 183-199.

- D'Angour, A. J. (1999), «Archinus, Eucleides and the reform of the Athenian alphabet», *BICS*, 43, 109-130.
- Damasio, A. R. (1995), *Descartes' error. Emotion, reason, and the human brain*, New York, Avon.
- De Sanctis, G. (1926), «La composizione della storia di Erodoto», *Rivista di Filosofia*, 54, 289-309.
- Dee, S. (2014), «Learning the TEI in a digital environment», *Journal of the Text Encoding Initiative*, 7.
- Deerwester, S. et al. (1990), «Indexing by latent semantic analysis», *Journal of the American Society of Information Science*, 41 (6), 391-407.
- Dekhtyar, A. & Iacob, I. E. (2005), «A framework for management of concurrent XML», *Data & Knowledge Engineering*, 52 (2), 185-208.
- Delcourt, C. (1989), «Where have all the key words gone?», *CHum*, 23 (4), 285-291.
- DeRose, S.-J. (2004), «Markup overlap: a review and a horse», *Proceedings of Extreme Markup Languages*.
- DeRose, S.-J. et al. (1990), «What is text, really?», *Journal of Computing in Higher Education*, 1 (2), 3-26.
- Derrida, J. (1981), *Dissemination*, Chicago, University of Chicago Press.
- Detienne, M. (1967), *Les maîtres de vérité dans la Grèce archaïque*, Paris, Maspero.
- Detienne, M. (ed.) (1988), *Les savoirs de l'écriture*, Villeneuve d'Ascq, Presses Universitaires de Lille.
- Detienne, M. (1990), *La escritura de Orfeo*, Barcelona, Peninsula.
- Dewald C. & Marincola, J. (eds.) (2006), *The Cambridge companion to Herodotus*, Cambridge/New York, Cambridge University Pr.
- Dickey, E. (2007), *Ancient Greek Scholarship: a guide to finding, reading, and understanding scholia, commentaries, lexica, and grammatical treatises, from their beginnings to the Byzantine period*, Oxford, Oxford University Press.

- Diderot, D. & Alembert, J. Le Rond D' (1751), *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, vol. V, Paris, Briasson.
- Dijkstra, E. W. (1989), «On the cruelty of teaching computer science», *Communications of the Association for Computing Machinery*, 32 (12), 27-35.
- Domingo Baguer, I. (2013), *Para qué han servido los libros*, Zaragoza, Universidad de Zaragoza.
- Dorandi, T. (2007), *Nell'officina dei classici: come lavoravano gli autori antichi*, Roma, Carocci.
- Dorandi, T. (2014), «Ancient ἐκδόσεις: further lexical observations on some Galen's evidences», *Lexicon Philosophicum*, 2, 1-23.
- Dorandi, T. (2016), «Notebooks and collections of excerpts: moments of *ars excerpendi* in the Greco-Roman world», en A. Cevoloni (ed.), *Forgetting machines: knowledge management evolution in early modern Europe*, Leiden, Brill, 37-57.
- Dorjahn, A. P. (1935), «Anticipation of arguments in Athenian courts», *TAPhA*, 66 , 274-295.
- Dover, K. J. (1973), *Thucydides*, Oxford, Clarendon.
- Driver, G. R. (1954), *Aramaic documents of the fifth century BC*, Oxford, Clarendon.
- Drouin, J. (2011), *Foray into topic modeling ecclesiastic Proust archive*, [página web], consultado el 10-3-2016, <<http://www.proustarchive.org/>>.
- Ducrot, O. & Schaeffer, J.-M. (1995), *Nouveau dictionnaire encyclopédique des sciences du langage*, Paris, Seuil.
- Durasau, P. & O'Donnell, M. B. (2002), «Coming down from the trees: next step in the evolution of markup?», *Proceedings of Extreme Markup Languages*.
- Edmunds, A. & Morris, A. (2000), «The problem of information overload in business organisations: a review of the literature», *International Journal of Information Management*, 20, 17-28.

- Eito Brun, R. (2003), «Topic maps y la indización de recursos electrónicos en la web», *El profesional de la información*, 12 (2), 141-148.
- Eppler, M. J. & Mengis, J. (2004), «The concept of information overload: a review of literature from Organization Science, Accounting, Marketing, MIS, and related disciplines», *The Information Society*, 20, 325-344.
- Esteban Ortega, J. (1997), *Emilio Lledó: una filosofía de la memoria*, Salamanca, San Esteban.
- Evans, J. A. S. (1982), *Herodotus*, Boston, Hall & Company.
- Fabrikant, S. I. & Battenfield, B. P. (2001), «Formalizing semantic spaces for information access», *Annals of the Association of American Geographers*, 91 (2), 263-280.
- Fehling, D. (1971), *Die Quellenangaben bei Herodot*, Berlin, De Gruyter.
- Ferlauto, F. (1990), «Prometeo e le grammaton suntheseis», *BollClass*, 11, 164-177.
- Fernández Galiano, M. (trad.) (1951), *Heródoto, nueva versión directa*, Barcelona, Labor.
- Ferreiro, E. (2007), «Las unidades de lo oral y las unidades de lo escrito», *Archivos de Ciencias de la Educación*, 1 (1), 195-230.
- Finley, M. I. (1975), *The use and abuse of history*, London, Chatto and Windus.
- Finley, M. I. (1985), *Ancient history, evidence and models*, London, Chatto and Windus.
- Finnegan, R. (1977), *Oral poetry. Its nature, significance and social content*, Cambridge, Cambridge University Press.
- Finnegan, R. (2011), *Why do we quote? The culture and history of quotation*, Cambridge, OpenBook Publishers.
- Fischer-Starcke, B. (2010), *Corpus linguistics in literary analysis: Jane Austen and her contemporaries*, London, Continuum.
- Fitzgerald, M. (2012), *Introducing regular expressions*, Beijing, O'Reilly

- Foley, J. M. (2002), «Editing and translating traditional oral epic: the south slavic songs and Homer», en I. Worthington & J. M. Foley (eds.), *Epea and grammata: oral and written communication in ancient Greece*, Leiden, Brill, 3-27.
- Ford, A. (2002), *The origins of criticism, literary culture and poetic theory in classical Greece*, Princeton, Princeton University Press.
- Foxhall, L. & Davies, J. K. (ed.) (1989), *The Trojan War: its historicity and context. Papers of the First Greenbank Colloquium, Liverpool 1981*, Bristol, Bristol Classical Press.
- Fraser, P. M. (1972), *Ptolemaic Alexandria*, vol. I: *Text*, Oxford, Oxford University Press.
- Fredouille, J.-C. et al. (ed.) (1997), *Titres et articulations du texte dans les œuvres antiques: actes du colloque international de Chantilly, 13-15 décembre 1994*, Turnhout, Brepols.
- Froger, D. (1968), *La critique de textes et son automatisaton*, Paris, Dunod.
- Fuchs, C. (ed.) (1985), *Aspects de l'ambigüité et de la paraphrase dans les langues naturelles*, Berna, Peter Lang.
- Fuentes González, P. P. (1989), «Magisterio y literatura en el estudio de la diatriba del cínico Teles», en *Actas del VII Congreso Español de Estudios Clásicos*, Madrid, Universidad Complutense de Madrid, 183-188.
- Fuentes González, P. P. (1993), «“Lexis” de Richard Goulet: un programa para la elaboración automática de léxicos de textos griegos y latinos en Macintosh™ (Apple)», *EClás*, 35, nº 104, 83-96.
- Fuentes González, P. P. (1998), *Les diatribes de Télès: Introduction, texte revu et commentaire des fragments, avec en appendice une traduction espagnole*, Paris, J. Vrin.
- Fuentes González, P. P. (2016), «Théodore», en R. Goulet (ed.), *Dictionnaire des Philosophes Antiques*, vol. VI, Paris, CNRS, 914-924.
- Gallagher, W. (1993), *The power of place: how our surroundings shape our thoughts, emotions and actions*, New York, Poseidon Press.

- Gangutia Elícegui, E. (1977), «Teorías semánticas en la Antigüedad», en F. Rodríguez Adrados *et al.* (eds.), *Introducción a la lexicografía griega*, Madrid, Instituto Antonio de Nebrija, 3-35.
- Gangutia Elícegui, E. (2007), «El Diccionario Griego-Español», *Arbor Ciencia, Pensamiento y Cultura*, 727, 749-769.
- García Martul, D., & Franco-Álvarez, G. (2009), «¿Tesauros, mapas conceptuales o *topic maps* para la Biblioteca Digital Augusto Millares?», *Boletín Millares Carlo*, 28, 133-155.
- García, J. F. (2002), «Ritual speech in early Greek song», en I. Worthington & J. M. Foley (eds.), *Epea and grammata: oral and written communication in ancient Greece*, Leiden, Brill, 29-55.
- Gardiner, A. H. (1989), *Langage et acte de langage: aux sources de la pragmatique*, Paris, Presses Universitaires de Lille.
- Gardthausen, V. (1911²), *Griechische Paläographie*. Erster Band: *Das Buchwesen im Altertum und im byzantinischen Mittelalter*, Leipzig, Verlag von Veit & Comp.
- Garrido Arilla, M. R. (1996), «Fundamentos del análisis documental», en J. López Yepes (ed.), *Manual de información y documentación*, Madrid, Pirámide, 229-250.
- Garshol, L. M. (2004), «Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all», *Journal of Information Science*, 30, 378-391.
- Gartner, R. (2016), *Metadata: shaping knowledge from Antiquity to the semantic web*, Berlin, Springer.
- Gavrilov, A. K. (2007), «Techniques of reading in classical Antiquity», *CQ*, 47 (1), 56-73.
- Gera, D. L. (2003), *Ancient Greek ideas on speech, language and civilization*, Oxford, Oxford University Press.
- Gibaldi, J. (ed.) (1992), *Introduction to scholarship in modern languages and literatures*, New York, MLA Press.
- Gil Urdiciáin, B. (1996a), «Lenguajes documentales (I)», en J. López Yepes (ed.), *Manual de información y documentación*, Madrid, Pirámide, 324-354.

- Gil Urdiciaín, B. (1996b), «Lenguajes documentales (II)» en J. López Yepes (ed.), *Manual de información y documentación*, Madrid, Pirámide, 354-373.
- Gil, L. (ed.) (1963), *Introducción a Homero*, Madrid, Guadarrama.
- Gil, L. (1995), *La palabra y su imagen: la valoración de la obra escrita en la Antigüedad: discurso leído en la solemne apertura del curso académico 1995-1996*, Madrid, Universidad Complutense de Madrid.
- Gilbert, P. (1973), «Automatic collation: a technique for medieval texts», *Computers and the Humanities*, 7 (3), 139-145.
- Gleitman, L. et al. (1989), «Great expectations», en A. M. Galaburda (ed.), *From reading to neurons*, Cambridge, MIT Press, 91-132.
- Godley, A. D. (trad.) (1922), *Herodotus, with an english translation*, Cambridge, Cambridge University Press.
- Goldfarb, C. (1997), «SGML: the reason why and the first published hint», *Journal of the American Society for Information Science*, 48 (7), 656-661.
- Gómez y Méndez, J. M. & Leal Gil, E. (2001), «Delimitación del vocablo *cibernética* y otras voces tecnológicas en periodismo», *Estudios sobre el Mensaje Periodístico*, 7, 97-107.
- Goody, J. (1968), *Literacy in traditional societies*, Cambridge, Cambridge University Press.
- Goody, J. (1977), *The domestication of the savage mind*, Cambridge, Cambridge University Press.
- Goody, J. (1981), *The interface between the written and the oral*, Cambridge, Cambridge University Press.
- Gordon, H. B. (ed.) (1974), *Psychology of learning and motivation*, New York, Academic Press.
- Gottlieb, G. (1963), *Das Verhältnis der Ausserherodoteischen Überlieferung zu Herodot*, Bonn, Habelt.

- Goulet-Cazé, M.-O. *et al.* (ed.) (2000), *Le commentaire entre tradition et innovation: actes du colloque international de l'Institut des traditions textuelles (Paris et Villejuif, 22-25 septembre 1999)*, Paris, Vrin.
- Goulet, R. (1990), «Lexis: un logiciel d'indexation et de lexicalisation automatiques pour Macintosh™ (Apple)», *RISSH*, 26, 93-110.
- Graf, F. & Johnston, S. I. (2007), *Ritual texts for the afterlife: Orpheus and the bacchic gold tablets*, London, Routledge.
- Grassl, H. (1972), «Herodot und die griechische Schrift», *Hermes*, 100, 169-175.
- Graziosi, B. & Haubold, J. (ed.) (2010), *Homer: Iliad, book VI*, Cambridge, Cambridge University Press.
- Greetham, D. (1992), «Textual scholarship», en Gibaldi, J. (eds.), *Introduction to scholarship in modern languages and literatures*, New York, MLA Press, 103-137.
- Greetham, D. (2007), «What is textual scholarship?», en S. Eliot & J. Rose (eds.), *A companion to the history of the book*, Malden / Oxford / Victoria, Blackwell, 21-33.
- Griffin, J. (1986), «Homeric words and speakers», *JHS*, 106, 36-57.
- Griffiths, T. & Steyvers, M. (2004), «Finding scientific topics», *Proceedings of the National Academy of Sciences*, 101 Suppl. I, 5228-5285.
- Grosso, P. *et al.* (2003), *XPointer framework*, [página web], consultado el 4-4-2017, <<http://www.w3.org/TR/xptr-framework/>>.
- Guespin, E. (1971), «Problématique des travaux aux discours politiques», *Langages*, 23, 3-24.
- Guillaume, G. (1973), *Principes de linguistique théorique de Gustave Guillaume*, Paris, Klincksieck.
- Gutiérrez Ordóñez, S. (1997), *La oración y sus funciones*, Madrid, Arco.
- Hadas, M. (1954), *Ancilla to classical reading*, New York, Columbia University Press.
- Hainsworth 1968 (1968), *The flexibility of the homeric formula*, Oxford, Clarendon.

- Hall, E. (1991), *Inventing the barbarian: Greek self-definition through tragedy*, Oxford, Clarendon.
- Halsell, L. A. (2013), «TEI and so can you: corpus linguistics and the Text Encoding Initiative», *Pacific Northwest Library Association Quarterly*, 77 3, 63-70.
- Halverson, J. (1992), «Havelock on Greek orality and literacy», *JHS*, 53, 148-163.
- Hampe, B. (ed.) (2005), *From perception to meaning: image schemas in cognitive linguistics*, Berlin, Mouton de Gruyter.
- Harrison, S. J. (ed.) (2001), *Texts, ideas and the classics: scholarship, theory, and classical literature*, Oxford, Oxford University Press.
- Hartog, F. (1988), *The mirror of Herodotus: the representation of the other in the writing of history*, Berkeley, University of California Press.
- Hauvette, A. (1894), *Hérodote historien des Guerres médiques*, Paris, Hachette.
- Havelock, E. & Hershbell, J. P. (ed.) (1978), *Communication arts in the ancient world*, New York, Hastings.
- Havelock, E. A. (1982), *The literate revolution in Greece and its cultural consequences*, Princeton, Princeton University Press.
- Havelock, E. A. (1986), *The muse learns to write. Reflexions on orality and literacy from Antiquity to the present*, New Haven / London, Yale University Press.
- Heath, F. G. (1972), «Origins of the binary code», *Scientific American*, 227, 76-83.
- Heinze, T. (1999), «Musaios», *Neue Pauly*, 7, 501-503.
- Henige, D. (1982), *Oral historiography*, London, Longman.
- Heninger, A. (2004), «Analyzing Unicode text with regular expressions», en *Proceedings of the 26th International and Unicode Conference. Unicode and Internationalization: best practices in the networked world*, San Jose, California, Unicode Consortium.
- Herman, D. (ed.) (2003), *Narrative theory and the cognitive sciences*, Stanford, CSLI Publications.

- Hermann, P. (1981), «Teos und Abdera in 5. Jahrhundert v. Chr.», *Chiron*, 11, 1-29.
- Heslin, P. (2000), «Parsing ancient Greek and other obsolete data formats with Perl», en *Yet Another Perl Conference (YAPC). The Art of Perl. London, september 22-24, 2000*, London, YAPC, Europe.
- Heslin, P. (2001), «TLG, Thesaurus Linguae Graecae, CD-ROM Disk E. Distributed by the Thesaurus Linguae Graecae for the Regents of the University of California, February 2000. ISBN 0-9675843-0-2», *BMCW*, 2001.09.23.
- Heubeck, A. (1979), *Schrift*, Göttingen, Vandenhoeck & Ruprecht.
- Heurley, L. (2010), «A revisão de texto: abordagem da psicologia cognitiva», *Scripta*, 14 26, 121-138.
- Higbee, K. L. (2001), *Your memory: how it works and how to improve it*, New York, Marlowe & Company.
- Higbie, C. (2010), «Divide and edit: a brief history of book divisions», *HSPH*, 105, 1-31.
- Hilbert, M. *et al.* (2004), «Making CONCUR work», *Proceedings of Extreme Markup Languages*.
- Hliaoutakis, A. *et al.* (2006), «Information retrieval by semantic similarity», *International Journal on Semantic Web and Information Systems*, 2 3.
- Hockey, S. (1994), «Electronic text in the humanities: a coming of age», en B. Sutton (ed.), *Literary texts in an electronic age: scholarly implications and library services*, Urbana, University of Illinois, 21-34.
- Hockey, S. (2004), «The history of humanities computing», en S. Schreibman *et al.* (eds.), *A companion to Digital Humanities*, Oxford: Blackwell, 3-20.
- Hoffman, T. (1999), «Probabilistic latent semantic analysis», en *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, New York, ACM, 50-52.
- Hopcroft, J. E. *et al.* (2001²), *Introduction to automata theory, languages, and computation*, Reading, Addison-Wesley.

- Hopkins, R. L. (1995), «Countering information overload: the role of the librarian», *The Reference Librarian*, 23, 305-333.
- Hornblower, S. (1987), *Thucydides*, Baltimore, Johns Hopkins University Press.
- Howald, E. (1944), *Vom Geist antiker Geschichtsschreibung*, Munich, Oldenbourg.
- Hubbard, T. (2008), «Getting the last word: publication of political oratory as an instrument of historical revisionism», en E. A. Mackay (eds.), *Orality, literacy, memory in the ancient Greek and Roman world*, Leiden, Brill, 185-202.
- Hudak, P. (1989), «Conception, evolution, and application of functional programming languages», *ACM Computing Surveys*, 21 3, 359-411.
- Hude, C. (ed.) (1927³), *Herodoti Historiae, recognovit brevique adnotatione critica instruxit Carolus Hude* (2 vols.), Oxonii, Clarendon.
- Huitfeldt, C. & Sperberg-McQueen, C. M. (2001), *TexMECS: an experimental markup meta-language for complex documents*, [página web], consultado el 23-9-2016, <<http://mlcd.blackmesatech.com/mlcd/2003/Papers/texmecs.html>>.
- Hummel, P. (2000), *Histoire de l'histoire de la philologie: étude d'un genre épistémologique et bibliographique*, Genève, Droz.
- Hunt, L., Lundberg, M. & Zuckerman, B. (2010), «Concrete abstractions: ancient texts as artifacts and the future of their documentation and distribution in the digital age», en W. Van Peursen, E. D. Thoutenhoofd & A. Van der Weel (eds.), *Text comparison and digital creativity: the production of presence and meaning in digital text scholarship*, Leiden, Brill, 149-172.
- Hunter, V. (1973), *Thucydides the artful reporter*, Toronto, Hakkert.
- Hunter, V. (1982), *Past and process in Herodotus and Thucydides*, Princeton, Princeton University Press.
- Immerwahr, H. R. (1964), «Book rolls on attic vases», en C. Henderson (ed.), *Classical, medieval, and renaissance studies in honor of B. L. Ullman*, Roma, Edizioni di Storie e Letteratura, 17-48.

- Immerwahr, H. R. (1966), *Form and thought in Herodotus*, Cleveland, Press of Western Reserve University.
- Immerwahr, H. R. (1973), «More book rolls on attic vases», *AK*, 16, 143-147
- Irigoin, J. (1997), *Tradition et critique des textes grecs*, Paris, Les Belles Lettres.
- Jacoby, F. (1913), «Herodotus», *RE Suppl.*, II , 205-520.
- Jacoby, F. (1949), *Atthis: the local chronicles of ancient Athens*, Oxford, Clarendon.
- Jagadish, H. V. *et al.* (2004), «Colorful XML: one hierarchy isn't enough», *Proceedings of the 2004 ACM SIGMOD International conference on Management of data*, 251-262.
- Jeffery, L. H. (1967), «*Archaia grammata: some ancient Greek views*», en W. C. Brice (ed.), *Europa: Studien zur Geschichte und Epigraphik frühen Aegaeis. Festschrift Ernst Grumach*, Berlin, De Gruyter, 152-166.
- Jeffery, L. H. (1976), *Archaic Greece: the city states c. 700-500 BC*, London, St. Martin's Press.
- Jeffery, L. H. (1990), *The local scripts of archaic Greece*, Oxford, Oxford University Press.
- Jensen, M. S. (1980), *The Homeric question and the oral-formulaic theory*, Copenhagen, Museum Tusulanum Press.
- Jockers, M. (2010), *Who's your DH blog mate: match-making the day of DH bloggers with topic modeling*, [artículo de blog], consultado el 15-11-2016, <<http://www.matthewjockers.net/2010/03/19/>>.
- Jockers, M. L. (2014), *Text analysis with R for students of literature*, Heidelberg, Springer.
- Johnson, W. A. (1994), «Towards an electronic Greek historical lexicon», *Emerita*, 62 (2), 253-261.
- Johnson, W. L. *et al.* (1968), «Automatic generation of efficient lexical processors using finite state techniques», *Communications of the ACM*, 11 (12), 805-813.
- Kauffmann-Samaras, A. (1992), «Mousaios», *LIMC*, 7 (1), 685-687.
- Kenyon, F. G. (1951²), *Books and readers in ancient Greece and Rome*, Oxford, Clarendon.

- Kleene, S. C. (1956), «Representation of events in nerve nets and finite automata», en C. E. Shannon *et al.* (eds.), *Automata studies*, Princeton, Princeton University Press, 3-41.
- Kleingünther, A. (1933), *Πρώτος ε ύρετής: Untersuchungen zur Geschichte einer Fragestellung*, Leipzig, Dieterich.
- Knox, D. (1958), «Silent reading in Antiquity», *GRBS*, 9, 421-435.
- Koenig, J. P. *et al.* (1992), «Evidence for different types of lexical representations in the cerebral hemispheres», *Cognitive Neuropsychology*, 9 1, 33-45.
- Köntges, T. (2016b), (*Meletē*)ToPān v.0.3, [programa], consultado el 1/05/2017, <<https://github.com/ThomasK81/ToPan>>.
- Köntges, T. (2016), «Researchers to your driving seats: building a graphical user interface for multilingual topic-modelling in R with Shiny», en *Digital Humanities 2016: conference abstracts*, Kraków, Jagiellonian University & Pedagogical University, 605-607.
- Kukkonen, K. (2013), «Form as a pattern of thinking: cognitive poetics and new formalism», en V. Theile & L. Tredennick (eds.), *New formalisms and literary theory*, New York, Palgrave Macmillan, 159-179.
- Kurke, L. (1999), *Coins, bodies, games and gold: the politics of meaning in Archaic Greece*, Princeton, Princeton University Press.
- Lambridunakis, W. & Wörrle, M. (1983), «Ein hellenistisches Reformgesetz über das öffentliche Urkundenwesen von Paros», *Chiron*, 13, 283-368.
- Lancashire, I. (2004), «Cognitive stylistics and the literary imagination», en S. Schreibman *et al.* (eds.), *A companion to Digital Humanities*, Oxford: Blackwell, 397-414.
- Lancaster, F. W. (1998), *Indexing and abstraction in theory and practice*, London, Library Association.
- Lang, M. L. (1984), *Herodotean narrative and discourse*, Cambridge, Harvard University Press.
- Langacker, R. W. (1987), *Foundations of cognitive grammar*, vol. I: *Theoretical prerequisites*, Standford, Standford University Press.

- Langacker, R. W. (2008), *Cognitive grammar: a basic introduction*, Oxford, Oxford University Press.
- Lara, D. (1997), *Iniciación a la lexicografía griega*, Madrid, Ediciones Clásicas.
- Latte, K. (1958), «Die Anfänge der griechischen Geschichtsschreibung», en *Histoire et historiens dans l'Antiquité*, Vandoeuvres-Genève, Fondation Hardt, 3-37.
- Laue, A. (2004), «How the computer works», en S. Schreibman *et al.* (eds.), *A companion to Digital Humanities*, Oxford: Blackwell, 145-160
- Lavency, M. (1964), *Aspects de la logographie judiciaire attique*, Louvain, Publications Universitaires de Louvain.
- Lazarini, M. L. (1998), «Questioni relative all'origine dell'alfabeto greco», en G. Bagnasco & F. Cordano (eds.), *Scritture mediterranee fra IX e VIII sec. a.C. Atti del Seminario: Università degli Studi di Milano, Istituto di Storia Antica, 23-24 febbraio 1998*, Milano, Edizioni Et, 53-61.
- Lebert, M. (2008), *Project Gutenberg (1971-2008)*, NEF, University of Toronto & Project Gutenberg.
- Legrand, Ph.-E. (1932), *Hérodote. Introduction. Notice préliminaire sur la vie et la personnalité d'Hérodote et sur la présente édition*, Paris, Les Belles Lettres.
- Legrand, Ph.-E. (ed. y trad.) (1932-1954), *Hérodote, Histoires, text établi et traduit par Ph.-E. Legrand*, Paris, Les Belles Lettres.
- Leguizamo León, A. V. & García Carrasco, J. (2011), «Semántica de las búsquedas de información en entornos virtuales de formación», *Education in the Knowledge Society*, 12 (1), 80-97.
- Lendle, O. (1992), *Einführung in die griechische Geschichtsschreibung: von Hekataios bis Zosimos*, Darmstadt, Wissenschaftliche Buchgesellschaft.
- Lens Tuero, J. (1995), «El tipo psicológico del político en las *Historias* de Heródoto, el *Filoctetes* de Sófocles y los *Relatos persas* de Ctesias», *Synthesis* 2, 3-22.
- Lewis, N. (1974), *Papyrus in classical Antiquity*, Oxford, Clarendon.

- Lida, M^a. R. (1949), *Heródoto. Los nueve libros de la Historia*, Buenos Aires, W. M. Jackson.
- Linforth, I. M. (1931), «Two notes on the legend of Orpheus», *TAPhA*, 62, 5-17
- Linforth, I. M. (1941), *The arts of Orpheus*, Berkeley, University of California Press.
- Lledó, E. (1992), *El silencio de la escritura*, Madrid, Centro de Estudios Constitucionales.
- López Facal, F. (1977), «Los diccionarios de autor. Tipos, metodología y estado actual», en F. Rodríguez Adrados *et al.* (eds.), *Introducción a la lexicografía griega*, Madrid, Instituto Antonio de Nebrija, 151-159.
- López Férez, J. A. (ed.) (1988), *Historia de la literatura griega*, Madrid, Cátedra.
- López Jimeno, A. (2000), «Nuevos sistemas de almacenamiento y difusión de la información científica: internet y el mundo griego», *Minerva*, 14, 239-263.
- López Muñoz, M. (2002), «Delenda est machina? Informática y Filología Latina», *RELat* 2, 235-250.
- López Yepes, J. (ed.) (1996), *Manual de información y documentación*, Madrid, Pirámide.
- Lord, A. (1960), *The singer of tales*, Cambridge, Harvard University Press.
- Luján Atienza, A. L. (2006), «El estudio de la poesía desde una perspectiva cognitiva: panorama y propuesta», *Revista de Literatura*, 68, nº 135, 11-39.
- Luque Moreno, J. (ed.) (1987), *Scriptores Latini de re metrica. Concordantiae-indices*, Granada, Universidad de Granada.
- Luque Moreno, J. (2001), «Palabras en verso», *RELat*, 1, 13-43.
- Luque Moreno, J. (2001b), «Numerus: la articulación rítmica del lenguaje», en M. J. Barrios Castro & E. Crespo (coord.), *Actas del X Congreso Español de Estudios Clásicos. Vol. II.*, Madrid, SEEC, 493-527.
- Luque Moreno, J. (2006), *Puntos y comas: la grafía de la articulación del habla*, Granada, Universidad de Granada.
- Lutz, M. & Ascher, D. (2004²), *Learning Python*, Sebastopol, O'Reilly.

- Maas, P. (1957²), *Textkritik*, Leipzig, Teubner.
- Mackay, E. A. (ed.) (2008), *Orality, literacy, memory in the ancient Greek and Roman world*, Leiden, Brill.
- Maddy, P. (1988), «Believing the axioms, I», *Journal of Symbolic Logic*, 53 (2), 481-511.
- Magini, D. (1998), «Parole e *grammata* nella tragedia greca: Eschilo e Sofocle», *AAP*, 47, 21-47.
- Mandell, L. (2013), *Introduction to digital textual editing: An unofficial guide to the value of TEI*, [diapositivas], consultado el 4-4-2017, <<http://www.slideshare.net/mandellc/tei-and-xslt-23711832>>.
- Martínez Alonso, H. (2013), *Annotation of regular polisemy: an empirical assessment of the underspecified sense*, Barcelona, Universitat Pompeu Fabra / Københavns Universitet.
- Martínez Comeche, J. A. (1996), «El proceso informativo documental», en López Yepes, J. (ed.), *Manual de información y documentación*, Madrid, Pirámide, 29-48.
- Martínez Hernández, M. et al. (ed.) (2000), *Cien años de investigación semántica: de Michel Bréal a la actualidad. Actas del Congreso Internacional de Semántica*, vol. II, Madrid, Ediciones Clásicas.
- Martínez, M. (2003), «Definiciones del concepto “campo” en semántica: antes y después de la Lexemática de E. Coseriu», *Odisea*, 3, 101-130.
- Maurer, K. (2012), *Commonest abbreviations, signs, etc. used in the apparatus to a classical text*, [página web], consultado el 1/05/2017, <<http://udallasclassics.org>>.
- McCloskey, M., Wible, C. G. & Cohen, N. J. (1988), «Is there a special flashbulb-memory mechanism?», *Journal of Experimental Psychology: General*, 117 (2), 171-181.
- McRae, K. & Jones, M. (2013), «Semantic memory», en Reisberg, D. (eds.), *The Oxford handbook of cognitive psychology*, Oxford: Oxford University Press, 206-216.
- Meadow, C. T. et al. (2007³), *Text information retrieval systems*, Amsterdam, Elsevier

- Meeks, E. (2011), *Comprehending the Digital Humanities*, [artículo de blog], consultado el 23-11-2016, <<https://dhs.stanford.edu/comprehending-the-digital-humanities/>>.
- Meiggs, R. & Lewis, D. (ed.) (1969), *A selection of Greek historical inscriptions to the end of the fifth century B.C.*, Oxford, Clarendon.
- Meister, K. (1990), *Die griechische Geschichtsschreibung: von den Anfängen bis zum Ende des Hellenismus*, Stuttgart, Kohlhammer.
- Meister, K. et al. (2006), «Herodotus», *Neue Pauly*, 6, 265-271.
- Mejer, J. (1978), *Diogenes Laertius and his Hellenistic background*, Stuttgart, Steiner.
- Melena, J. L. (2001), *Textos griegos micénicos comentados*, Vitoria-Gastéiz, Eusko Legebiltzarra.
- Melgoza, P. et al. (2002), «Information overload», *Collection Building*, 21, 32-43.
- Miccoli, P. (1989), «*Mneme, anamnesis, mnemosyne*: sull'identità dell'uomo storico», *GM*, 11 3, 465-476.
- Mikheev, A. (2003), «Text segmentation» en R. Mitkov (ed.), *The Oxford handbook of computational linguistics*, Oxford, Oxford University Press, 201-218.
- Mimno, D. (2012), «Computational historiography: data mining in a century of classics journal», *Journal of Computing and Cultural Heritage*, 5 (1), 1-19.
- Minchin, E. (2002), «Speech acts in the everyday world and in Homer: the rebuke as a case study», en Worthington, I. & Foley, J. M. (eds.), *Epea and grammata: oral and written communication in ancient Greece*, Leiden, Brill, 71-97.
- Momigliano, A. (ed.) (1966), *Studies in historiography*, London, Weidenfeld & Nicholson.
- Momigliano, A. (1972), «Tradition and the classical historian», *H&T*, 11, 279-293.
- Montana, F. (2011), «The making of Greek scholiastic corpora», en F. Montanari & L. Pagani (eds.), *From scholars to scholia: chapters in the history of ancient Greek scholarship*, Berlin, De Gruyter, 105-163.
- Montanari, F. (ed.) (1988), *I frammenti dei grammatici Agathokles, Hellanikos, Ptolemaios Epithetes*, Berlin, De Gruyter.

- Montanari, F. & Pagani, L. (ed.) (2011), *From scholars to scholia: chapters in the history of ancient Greek scholarship*, Berlin, De Gruyter.
- Montejo-Ráez, A. *et al.* (2005), «Text categorization using bibliographic records: beyond document content», *Procesamiento del Lenguaje Natural*, 35, 119-126.
- Mooers, C. (1950), «EDITOR'S CORNER: Coding, Information Retrieval and the Rapid Selector», *Journal of the American Society for Information Science*, 1 (4), 225.
- Mooers, C. (1951), «Zatocoding applied to mechanical organization of knowledge», *Journal of the Association for Information Science and Technology*, 2 (1), 20-32.
- Moreiro González, J. A. *et al.* (2003), «Panorámica y tendencias en *topic maps*», *Hipertext.net*, 1.
- Moretti, F. (2005), *Graphs, maps and trees: abstract models for literary history*, London, Verso.
- Moya Anegrón, F. de (1996), «Sistemas de RI documental», en J. López Yepes (ed.), *Manual de información y documentación*, Madrid, Pirámide, 429-445.
- Mulvany, N. C. (2009), *Indexing books*, Chicago, Chicago University Press.
- Münzer, F. (1897), *Beiträge zur Quellenkritik der Naturgeschichte des Plinius*, Berlin, Weidmann.
- Murdock, J. & Allen, C. (2015), «Visualization techniques for topic model checking», *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, vol. VI, Austin, AAAI Press, 4284-4285.
- Murray, D. J. (1968), «Articulation and acoustic confusability in short term memory», *Journal of Experimental Psychology*, 78, 679-684.
- Murray, O. (1980), *Early Greece*, Brighton, Harvester Press.
- Mylonas Shear, I. (1998), «Bellerophon tablets from the Mycenaean world? A tale of seven bronze hinges», *JHS*, 118, 187-189.
- Nagler, M. N. (1974), *Spontaneity and tradition: a study in the oral art of Homer*, Berkeley, University of California.

- Nagy, G. (1983), «Sêma and noesis: some illustrations», *Arethusa*, 16, 35-55.
- Nagy, G. (1987), «Herodotus the logios», *Arethusa*, 20 (1), 175-184.
- Neches, R. *et al.* (1991), «Enabling technology for knowledge sharing», *AI Magazine*, 12 (3), 35-56.
- Neisser, U. (1967), *Cognitive psychology*, Englewood Cliffs, Prentice Hall.
- Nenci, G. (ed.) (1988), *Hérodote et les peuples non grecs*, Vandoeuvre-Genève, Fondation Hardt.
- Newell, A. & Simon, H. A. (1956), «The logic theory machine: a complex information processing system», *IRE Transactions on Information Theory*, 2 (3), 61-79.
- Newell, A. & Simon, H. A. (1959), «Report on a general problem solving program», *IFIP Congress*, 256, 64.
- Nieddu, G. F. (2004), *La scrittura «madre dell Muse»: agli esordi di un nuovo modello di comunicazione culturale*, Amsterdam, Hakkert.
- Nilsson, M. P. (1943), «Die Quellen der Lethe und der Mnemosyne», *Eranos*, 41, 1-7.
- Norman, D. A. (1993), *Things that make us smart: defending human attributes in the age of the machine*, Boston, Addison-Wesley Longman.
- Notopoulos, J. A. (1938), «Mnemosyne in oral literature», *TAPhA*, 69, 465-493.
- Novak, J. D. & Gowin D. B. (1984), *Learning how to learn*, Cambridge, Cambridge University Press.
- Oakley, S. P. (2016), «The “proto-history” of the text of Livy», en J. Velaza (ed.), *From the protohistory to the history of the text*, Frankfurt am Main, Peter Lang, 165-186.
- Ochoa, J. A. (1992), «Observaciones sobre la aplicación de la informática a la lexicografía general grecolatina», *EHum*, 14, 129-142.
- Olivier, J.-P. (2012), «Las escrituras egeas: “jeroglífica” cretense, lineal A, lineal B, chipriotas y escrituras silábicas chipriotas del I milenio antes de nuestra era», *Faventia, Supplementa 1: Actas del Simposio Internacional: 55 años de Micenología (1952-2007)*, 15-35.

- Olson, D. R. (1993), «How writing represents speech», *Language & Communication*, 12, 1-17.
- Olson, D. R. (1994), *The world on paper: the conceptual and cognitive implications of writing and reading*, Cambridge, Cambridge University Press.
- Orasan, C. (2000), «Hybrid method for clause splitting in unrestricted English texts», *Proceedings of ACIDCA'2000, Corpora and Natural Language Processing (Monastir)*, 129-134.
- Ostwald, M. (1991), «Herodotus and Athens», *ICS*, 16, 137-148.
- Osuna García, F. (2006), «De nuevo sobre el concepto de oración», *Journal of Theoretical and Experimental Linguistics*, 8, 5-33.
- Oswald, R. (1995), «Gedankliche und thematische Linien in Herodots Werk», *Grazer Beiträge*, 21, 47-59.
- Otlet, P. (1934), *Traité de documentation: le livre sur le livre: theorie et pratique*, Bruxelles, Editions Mundaneu.
- Ott, W. (1979), «A text processing system for the preparation of critical editions», *Computers and the Humanities*, 13, 29-35.
- Pack, R. A. (1965²), *The Greek and Latin literary texts from Greco-Roman Egypt. Second revised and enlarged edition*, Ann Arbor, University of Michigan Press.
- Palaima, T. (2004), «“Archives” and “scribes” and information hierarchy in Mycenaean Greek Linear B records» en M. Brosius (eds.), *Ancient archives and archival traditions: concepts of record-keeping in the ancient world*, Oxford, Oxford University Press, 153-194.
- Palmer, C. L. (2004), «Thematic research collections» en S. Schreibman *et al.* (eds.), *A companion to Digital Humanities*, Oxford: Blackwell, 348-366.
- Palomares Perraut, R. (1997), *Análisis de fuentes de información de estudios de traducción: creación de una base de datos (tesis doctoral dirigida por M. Pinto Molina)*, Málaga, Universidad de Málaga.

- Papp, A. H. R. E. (1948), *De Herodoti reliquiis in papyris et membranis Aegyptiis servatis*, Lugduni Batavorum, Brill.
- Parry, M. (1930), «Studies in the epic technique of oral verse-making I. Homer and Homeric style», *Harvard Studies in Classical Philology*, 41, 73-147.
- Pasquali, G. (1998), *Filologia e storia*, Firenze, Le Monnier.
- Peirce, C. S. (1988), *Escritos lógicos*, Madrid, Alianza.
- Pérez Álvarez-Ossorio, J. R. (1990), *Introducción a la Información y Documentación Científica*, Madrid, Alhambra.
- Pfeiffer, R. (1968), *History of classical scholarship: from the beginning to the end of the Hellenistic age*, Oxford, Clarendon.
- Piccaluga, G. (1988), «Potmos impar a scrivere», *Scrittura e Civiltà*, 12, 5-46.
- Pinto Molina, M. (1993²), *Análisis documental: fundamentos y procedimientos*, Madrid, Eudema.
- Pinto Molina, M. (1996), «Análisis documental de contenido», en López Yepes, J. (ed.), *Manual de información y documentación*, Madrid, Pirámide, 300-323.
- Plamboeck, G. (1959), *Erfassen, Gegenwärtigen, Innesein: Aspekte Homerischer Psychologie*, Kiel.
- Pohlenz, M. (1961²), *Herodot der erste Geschichtschreiber des Abendlandes*, Leipzig, Teubner.
- Pope, M. (1985), «A nonce-word in the *Iliad*», *CQ*, 35, 1-8.
- Posner, M. I. & Petersen, S. E. (1990), «The attention system of the human brain», *Annual Review of Neuroscience*, 13, 25-42.
- Powell, B. B. (1989), «Why was the Greek alphabet invented?», *Classical Antiquity*, 8, 321-350.
- Powell, J. E. (1938), *A lexicon to Herodotus*, Cambridge, Cambridge University Press.
- Powell, J. E. (1939), *The History of Herodotus*, Cambridge, Cambridge University Press.

- Pressley, G. M., Levin, J. R. & Delaney, H. D. (1982), «The mnemonic keyword method», *Review of Educational Research*, 52, 61-91.
- Priestley, J. & V. Zali (eds.) (2016), *Brill's companion to the reception of Herodotus in Antiquity and Beyond*, Leiden, Brill.
- Pritchett, W. K. (1993), *The liar school of Herodotus*, Amsterdam, J. C. Gieben.
- Pucci, P. (1988), «Inscriptions archaïques sur les statues des dieux», en M. Detienne (ed.), *Les savoirs de l'écriture*, Villeneuve d'Ascq, Presses Universitaires de Lille, 480-497.
- Putnam, A. L. (2015), «Mnemonics in education: current research and applications», *Translational Issues in Psychological Science*, 1 (2), 130-139.
- Raymond, D. R. et al. (1992), «Markup reconsidered», en *First International Workshop on Principles of Document Processing*, Washington D.C.
- Reichenberger, G. (1965), «Herodotus in Spain», *Romance Philology*, 19, 235-249.
- Reinhardt, A. (1994), «Managing the new document», *Byte*, 19 (8), 91-104.
- Reisberg, D. (ed.) (2013), *The Oxford handbook of cognitive psychology*, Oxford, Oxford University Press.
- Renehan, R. (ed.) (1991), reseña de Rosén, H. B. (1987), *Herodoti Historiae*, vol. I: *Libros I-IV continens*, edidit Haiim B. Rosén, Leipzig, Teubner.», *GGA*, 243, 23-44.
- Reydams-Schils, G. J. (ed.) (2011), *Thinking through excerpts: studies on Stobaeus*, Turnhout, Brepols.
- Reynolds, L. D. & Wilson, N. G. (1991³), *Scribes and scholars: a guide to the transmission of Greek and Latin literature*, Oxford, Clarendon.
- Richardson, N. J. (1987), «The individuality of Homer's language», en J. M. Bremer, I. J. F. De Jong & J. Kalff (eds.), *Homer: beyond oral poetry: recent trends in Homeric interpretation*, Amsterdam: Grüner, 165-184.
- Riemann, K.-A. (1967), *Das Herodoteische Geschichtswerk in der Antike*, München, Ludwig-Maximilians-Universität.

- Robb, K. (1978), «The poetic sources of the Greek alphabet: rhythm and abecedarium from Phoenician to Greek», en E. Havelock & J. P. Hershbell (eds.), *Communication arts in the ancient world*, New York: Hastings, 23-36.
- Rodríguez Adrados, F. (1977), «Ordenadores y lexicografía griega. El banco de datos», en F. Rodríguez Adrados *et al.* (eds.), *Introducción a la lexicografía griega*, Madrid, Instituto Antonio de Nebrija, 209-215.
- Rodríguez Adrados, F. (2000), «La semántica en el *Diccionario Griego-Español*», en M. Martínez Hernández *et al.* (eds.), *Cien años de investigación semántica: de Michel Bréal a la actualidad. Actas del Congreso Internacional de Semántica. II.*, Madrid, Ediciones Clásicas, 99-110.
- Rodríguez Adrados, F. (2003), «Diccionario Griego-Español, vol. VI», *ICS*, 27-28, 115-130.
- Rodríguez Adrados, F. (ed.) (2011), *Heródoto. Historias. Libro III*, Madrid, CSIC.
- Rodríguez Adrados, F. & Rodríguez Somolinos, J. (1994), «El banco de datos del *TLG*, el *DGE* y la lexicografía griega», *Emerita*, 62, 193-203.
- Rodríguez Adrados, F. *et al.* (ed.) (1977), *Introducción a la lexicografía griega*, Madrid, Instituto Antonio de Nebrija.
- Rodríguez Adrados, F. *et al.* (ed.) (2008), *Veinte años de filología griega*, Madrid, CSIC.
- Rodríguez Somolinos, J. (2008), «Lexicografía» en F. Rodríguez Adrados *et al.* (eds.), *Veinte años de filología griega*, Madrid, CSIC, 413-447.
- Rodríguez Somolinos, J. & Álvarez, I. (1991), «Informática y lexicografía: la experiencia del Diccionario-Griego-Español», *Emerita*, 59 (1), 81-99.
- Roemer, C. (2007), «The papyrus roll in Egypt, Greece and Rome», en S. Eliot & J. Rose (eds.), *A companion to the history of the book*, Malden / Oxford / Victoria, Blackwell, 84-95.
- Rollinger, R. (1993), *Herodots babylonischer Logos: eine kritische Untersuchung der Glaubwürdigkeitsdiskussion*, Innsbruck, Verlag des Instituts für Sprachwissenschaft der Universität Innsbruck.

- Rommel, T. (2004), «Literary studies», en S. Schreibman *et al.* (eds.), *A companion to Digital Humanities*, Oxford: Blackwell, 88-97.
- Rosén, H. B. (ed.) (1987), *Herodoti Historiae*, vol. I: *Libros I-IV continens*, edidit Haiim B. Rosén, Berolini, De Gruyter.
- Rosén, H. B. (ed.) (2008), *Herodoti Historiae*, vol. II: *Libros V-IX continens, indicibus criticis adiectis*, edidit Haiim B. Rosén, Berolini, De Gruyter.
- Roueché, C. & Flanders, J. (2014), *Introduction for Epigraphers. EpiDoc guidelines: ancient documents in TEI XML*, [página web], consultado el 4-4-2017, <<http://www.stoa.org/epidoc/gl/dev/intro-eps.html>>.
- Rudhardt, J. (1988), «Mnémosyne et les Muses», en P. Borgeaud (ed.), *Le mémoire des religions*, Genève, Labor et Fides, 37-63.
- Ruipérez, M. S. & Melena, J. L. (1990), *Los griegos micénicos*, Madrid, Historia 16.
- Russell, D. B. (1967), *COCOA: A word count and concordance generator for Atlas*, Chilton, Atlas Computer Laboratory.
- Russo, J. & Simon, B. (1971), «Psicología omerica e tradizione epica orale», *QUCC*, 12, 40-61.
- Ryan, M.-L. (2003), «Cognitive maps and the construction of narrative space» en D. Herman (ed.), *Narrative theory and the cognitive sciences*, Stanford: CSLI Publications, 214-242.
- Saadani, L. & Bertrand-Gastaldy, S. (2000), «Cartes conceptuelles et thésaurus: essai de comparaison entre deux modèles de représentation issus de différentes traditions disciplinaires», en *Proceedings of the Annual Conference of CAIS / Actes du congrès annuel de l'ACSI*, Edmonton, University of Alberta.
- Saenger, P. (1997), *Space between words: the origins of silent reading*, Stanford, Stanford University Press.
- Sagredo Fernández, F. & Espinosa, B. N. (1996), «Tecnologías de la información y documentación: recursos y aplicaciones», en López Yepes, J. (ed.), *Manual de información y documentación*, Madrid, Pirámide, 157-183.

- Säid, S. (1985), *Sophiste et tyran, ou: le problème enchaîné*, Paris, Klincksieck.
- Sánchez Lasso de la Vega, J. (1963), «Hombres y dioses en los poemas homéricos», en Gil, L. (ed.), *Introducción a Homero*, Madrid, Guadarrama, 237-318.
- Sánchez Ruenes, D. (2007), *Domain ontology learning from the web*, Tarragona, Universitat Politècnica de Catalunya.
- Sandborg-Petersen, U. & Øhrstrøm, P. (2010), «Towards an implementation of Jacob Lohard's ontology as a digital resource for historical and conceptual research in early seventeenth century thought», en W. Van Peursen, E. D. Thoutenhoofd & A. Van der Weel (eds.), *Text comparison and digital creativity: the production of presence and meaning in digital text scholarship*, Leiden, Brill, 57-77.
- Sang, E. E. et al. (2001), «Introduction to the CONLL-2001 shared task: clause identification», en W. Daelemans & R. Zajac (eds.), *Proceedings of the Computational Natural Language Learning Workshop (CoNLL-2001) (Toulouse)*, New Brunswick, Association for Computational Linguistics, 53-57.
- Sarcisi-Weendenburg, H. W. A. & Kuhrt, A. (ed.) (1987), *Achaemenid history*, vol. II: *The Greek sources*, Leiden, Nederlands Instituut voor het Nabije Oosten.
- Schacter, D. L. et al. (2011), *Psychology*, New York, Worth.
- Schadewalt, W. (1982), *Die Anfänge der Geschichtsschreibung bei den Griechen: Herodot, Thukydides*, Frankfurt am Main, Suhrkamp.
- Schepens, G. (1975), «Some aspects of source theory in Greek historiography», *AncSoc*, 6, 257-274.
- Schick, A. G. et al. (1990), «Information overload: a temporal approach», *Accounting Organizations and Society*, 15, 199-220.
- Schinke, R. et al. (1998), «Retrieval of morphological variants in searches of Latin text databases», *Computers and the Humanities*, 31, 409-432.
- Schroni, F. (ed.) (2004), *I frammenti di Aristarco di Samotracia negli etimologici bizantini. Etymologicum genuinum, Magnum, Symeonis, Μεγ ἀλη γραμματικὴ , Zonarae lexicon. Introduzione, edizione critica e commento*, Göttingen, Vandenhoeck & Ruprecht.

- Schmidt, D. (2010), «The inadequacy of embedded markup for cultural heritage texts», *Digital Scholarship in the Humanities*, 25 (3), 337-356.
- Schrader, C. (1988), «Historiografía», en López Férez, J. A. (ed.), *Historia de la literatura griega*, Madrid, Cátedra, 503-536.
- Schrader, C. (trad.) (1992), *Heródoto, Historia. Libros I-II*, introd. de Francisco R. Adrados; trad. y notas de C. Schrader, Madrid, Gredos.
- Schreibman, S. et al. (ed.) (2004), *A companion to Digital Humanities*, Oxford, Blackwell.
- Scodel, R. (2002), «Homeric signs and flashbulb memory», en I. Worthington & J. M. Foley (eds.), *Epea and grammata: oral and written communication in ancient Greece*, Leiden, Brill, 99-116.
- Scodel, R. (2008), «Social memory in Aeschylus' *Oresteia*», en E. A. Mackay (eds.), *Orality, literacy, memory in the ancient Greek and Roman world*, Leiden, Brill, 115-143.
- Scott, M. & Tribble, C. (2006), *Textual patterns: key words and corpus analysis in language education*, Amsterdam/Philadelphia, John Benjamins Publishing.
- Searle, J. R. (1969), *Speech acts*, Cambridge, Cambridge University Press.
- Sebesta, R. W. (2012), *Concepts of programming languages*, Boston, Pearson.
- Shadbolt, N. et al. (2006), «The semantic web revisited», *IEEE Intelligent Systems*, 21 (3), 96-101.
- Shannon, C. E. & Weaver, W. (1949), *The mathematical theory of communication*, Urbana, University of Illinois Press.
- Shenk, D. (1997), *Data smog: surviving the Information Glut*, New York, HarperCollins.
- Shive, D. (1987), *Naming Achilles*, Oxford, Oxford University Press.
- Signes Codoñer, J. (2004), *Escritura y literatura en la Grecia arcaica*, Madrid, Akal.
- Signes Codoñer, J. (2009), «La escritura en la Grecia arcaica: un debate metodológico», *Cultura Escrita y Sociedad: Revista Internacional de Historia Social de la Cultura Escrita*, 9, 14-64.

- Signes Codoñer, J. (2010), «Nuevas perspectivas en el estudio de los orígenes de la escritura en la Grecia arcaica (siglos IX-VIII a.C.)», en G. Carrasco Serrano & J. C. Oliva Mompeán (eds.), *El mediterráneo antiguo: lenguas y escrituras*, Madrid, Centro de Estudios del Próximo Oriente, 247-314.
- Silk, M. (1987), *Homer, The Iliad*, Cambridge, Cambridge University Press.
- Sipser, M. (2006²), *Introduction to the theory of computation*, Boston, Thomson.
- Skupin, A. (1998), *Organizing and visualizing hypermedia informations spaces*, Buffalo, State University of New York.
- Skupin, A. & Battenfield, B. P. (1997), «Spatial metaphors for visualizing information spaces», en *Annual convention & exposition technical papers. ACSM 57th Annual Convention. ASPRS 63rd Annual Convention*, vol. V: *Auto-Carto 13*, Bethesda, American Society for Photogrammetry and Remote Sensing & American Congress on Surveying and Mapping, 116-125.
- Skydsgaard, J. E. (1968), *Varro the scholar*, Copenhagen, Munksgaard.
- Slakta, D. (1985), «Grammaire de texte: synonymie et paraphrase», en C. Fuchs, (eds.), *Aspects de l'ambigüité et de la paraphrase dans les langues naturelles*, Berna, Peter Lang, 123-140.
- Small, J. P. (1997), *Wax tablets of the mind*, London, Routledge.
- Smith, D. et al. (2000), «The Perseus Project: a digital library for the Humanities», *Literary and Linguistic Computing*, 15, 15-25.
- Sommerstein, A. H. (2000), «The prologue of Aeschylus' *Palamedes*», *RhM*, 143, 118-127.
- Sowa, C. A. & Sowa, J. F. (1974), «Thought clusters in early Greek oral poetry», *CHum*, 8 (3), 131-146.
- Sowa, J. F. (2000), *Knowledge representations: logical, philosophical and computational foundations*, Pacific Grove, Brooks-Cole.
- Spence, P. (2014), «Centros y fronteras: el panorama internacional de las humanidades digitales», en S. López Poza & N. Pena Sueiro (eds.), *Humanidades digitales: desafíos, logros y perspectivas de futuro*, A Coruña, Universidade da Coruña, SIELAE, 37-63

- Sperberg-McQueen, C. M. (1991), «Text in the electronic age: textual study and text encoding, with examples from medieval texts», *Literary and Linguistic Computing*, 61 (1), 34-47.
- Sperberg-McQueen, C. M. (2004), «Classification and its structures» en S. Schreibman *et al.* (ed.), *A companion to Digital Humanities*, Oxford: Blackwell, 161-177.
- Spiro, L. (2010), *Digital Humanities education*, [grupo de Zotero], consultado el 4-4-2017, <https://www.zotero.org/groups/digital_humanities_education/>.
- Squire, L. R. (2004), «Memory systems of the brain: a brief history and current perspective», *Neurobiology of Learning and Memory*, 82, 171-177.
- Stein, H. (ed.) (1893-1908⁶), *Herodoti Historiae, ad recensionem suam recognovit Henricus Stein*, 2 vols., Berolini, Weidmann.
- Steiner, D. (1994), *The tyrant's writ. Myths and images of writing in ancient Greece*, Princeton, Princeton University Press.
- Strasburger, H. (1982), «Herodot und das perikleische Athen», en W. Marg (ed.), *Herodot, eine Auswahl aus der neueren Forschung*, Munich, Wissenschaftliche Buchgesellschaft, 574-608.
- Strasburger, H. (1982b), «Herodots Zeitrechnung» en W. Marg (ed.), *Herodot, eine Auswahl aus der neueren Forschung*, Munich, Wissenschaftliche Buchgesellschaft, 688-736.
- Striegnitz, K. *et al.* (2003), *Algorithms for computational linguistics*, Saarbrücken, Universität des Saarlandes.
- Studer, R. *et al.* (1998), «Knowledge engineering: principles and methods», *IEEE Transactions on Knowledge and Data Engineering*, 25 (1), 161-197.
- Svenbro, J. (1988), «J'écris, donc je m'efface: l'énonciation dans les premières inscriptions grecques», en M. Detienne, (ed.), *Les savoirs de l'écriture*, Villeneuve d'Ascq, Presses Universitaires de Lille, 459-479.
- Svenbro, J. (1988), *Phrasikleia: an anthropology of reading in ancient Greece*, Paris, Presses Universitaires de Lille.

- Svenbro, J. (2004), «La Grecia arcaica y clásica: la invención de la lectura silenciosa», en Cavallo, G. & Chartier, R. (eds.), *Historia de la lectura en el mundo occidental*, Madrid, Santillana, 67-109.
- Talmy, L. (2000), *Toward a cognitive semantics*, vol. I: *Concept structuring systems*, Cambridge, MIT Press.
- Tardieu, M. (1988), «Théorie de la mémoire dans les traditions mystiques de l'Inde» en P. Borgeaud (ed.), *Le mémoire des religions*, Genève, Labor et Fides, 105-115.
- Téllez Muñoz, J. A. (2005), *La comprensión de los textos escritos y la psicología cognitiva: más allá del procesamiento de la información*, Madrid, Dykinson.
- Templeton, C. (2011), *Topic modeling in the Humanities: an overview*, [artículo de blog], consultado el 24-11-2016, <<http://mith.umd.edu/>>.
- Tennison, J. & Piez, W. (2002), «The layered markup and annotation language», *Proceedings of Extreme Markup Languages Conference*.
- TEI Consortium (2016), *TEI P5: Guidelines for electronic text encoding and interchange. Version 3.0.0. Last updated on 29th March 2016, revision 89ba24e*, TEI Consortium.
- Thomas, R. (1989), *Oral traditions and written records in classical Athens*, Cambridge, Cambridge University Press.
- Thomas, R. (1992), *Literacy and orality in ancient Greece*, Cambridge, Cambridge University Press.
- Thompson, K. (1968), «Programming techniques: regular expression search algorithm», *Communications of the ACM*, 11 (6), 419-422.
- Toffler, A. (1970), *Future shock*, Toronto, Bantam.
- Toffler, A. (1973), *El «shock» del futuro*, trad. J. Ferrer Aleu, Barcelona, Plaza & Janes.
- Torres Guerra, J. B. (2012), «Las bases del canon: literatura griega», *Minerva*, 25, 21-48.
- Trier, J. (1931), *Der deutsche Wortschatz im Sinnbezirk des Verstandes: Die Geschichte eines sprachlichen Feldes*, Heidelberg, Winter.

- Tufte, E. (1997), *Visual explanations: images and quantities, evidence and narrative*, Cheshire, Graphics Press.
- Tulving, E. (1984), «Precis of elements of episodic memory», *Behavioural and Brain Sciences*, 7 (2), 223-268.
- Turner, E. G. (1952), *Athenian books in the fifth and fourth centuries BC*, London, H. K. Lewis & Co.
- Urbano, C. (2010), «Algunas reflexiones a propósito de la "infoxicación"», *Anuario ThinkEPI*, 4, 304-308.
- Van den Branden, R. *et al.* (2010), *TEI by example*, [página web], consultado el 3-6-2017, <<http://www.teibyexample.org>>.
- Van den Hoek, A. (1996), «Techniques of quotation in Clement of Alexandria: a view of ancient literary working methods», *VChr*, 50, 223-243.
- Van Dijk, T. A. (1977), *Text and Context: explorations in the semantics and pragmatics of discourse*, London / New York, Longman.
- Kintsch, W. & Van Dijk, T. A. (1978), «Toward a model of text comprehension and production», *Psychological Review*, 85 (5), 363-394.
- Van Dijk, T. A. (1978), *La ciencia del texto: un enfoque interdisciplinario*, Barcelona, Paidós.
- Van Dijk, T. A. (1980), *Texto y contexto: semántica y pragmática del discurso*, Madrid, Cátedra.
- Vandendorpe, C. (1999), *Du papyrus à l'hypertexte: essai sur les mutations du texte et de la lecture*, Paris, La Découverte.
- Vansina, J. M. (1973), *Oral tradition: a study in historical methodology*, Harmondsworth, Penguin Books.
- Vansina, J. M. (1985), *Oral tradition as History*, Wisconsin, University of Wisconsin.
- Vergés, F. (2008), «La memoria como resistencia», *Revista Internacional de Filosofía Política*, 31, 49-64.

- Villén Rueda, L. (2006), *La indización y el acceso por materias en los catálogos de bibliotecas: el desafío de la recuperación de información*, Guadalajara, Universidad de Guadalajara.
- Vizcaíno Ortega, F. (1997), «Una aproximación interdisciplinar a los procesos de comprensión de un texto: lingüística y psicología cognitiva», *LFE: Revista de Lenguas para fines específicos*, 4, 225-274.
- Wachter, R. (1996), «Alphabet», *Neue Pauly*, 1, 536-547.
- Waters, K. H. (1985), *Herodotus the historian: his problems, methods and originality*, London, Croom Helm.
- Watt, A. (2005), *Beginning regular expressions*, Indianapolis, Wrox.
- Waugh, N. C. (1961), «Free versus serial recall», *Journal of Experimental Psychology*, 62 (5), 496-502.
- Weinrich, H. (1977³), *Tempus: Besprochene und erzählte Welt*, Stuttgart, Kohlhammer.
- Wellington, J. S. (2003), *Dictionary of bibliographic abbreviations found in the scholarship of classical studies and related disciplines*, Connecticut/London, Praeger.
- Wells, G. (1987), «Apprenticeship in literacy», *Interchange*, 18 (1-2), 109-123.
- Wendel, C. (1949), *Die griechisch-römische Buchbeschreibung verglichen mit der des vorderen Orients*, Halle, Niemeyer.
- West, S. R. (1985), «Herodotus epigraphical interests», *CQ*, 35, 278-305.
- West, S. R. (2011), «The papyri of Herodotus» en D. Obbink & R. Rutherford (eds.), *Culture in pieces. Essays on ancient texts in honour of Peter Parsons*, Oxford, Oxford University Press, 50-68.
- Wians, W. (ed.) (2009), *Logos and muthos: philosophical essays in Greek literature*, Albany, N.Y., State University of New York Press.
- Wilson, J.-P. (2009), «Literacy», en A. Raaflaub & H. van Wees (eds.), *A companion to archaic Greece*, Oxford, Oxford University Press, 542-563.
- Wilson, N. G. (2011), «Maasiana on Herodotus», *ZPE*, 179, 57-70.

- Wilson, N. G. (2015), *Herodotea: studies on the text of Herodotus*, Oxford, Oxford University Press.
- Wilson, N. G. (ed.) (2015b), *Herodoti Historiae*, vol. I: *Libri I-IV, recognovit brevique adnotatione critica instruxit N.G. Wilson*, Oxford, Oxford University Press.
- Wilson, N. G. (ed.) (2015c), *Herodoti Historiae*, vol. II: *Libri V-IX, recognovit brevique adnotatione critica instruxit N.G. Wilson*, Oxford, Oxford University Press.
- Wilson, P. (1996), «Interdisciplinary research and information overload», *Library Trends*, 44, 192-204.
- Wirbelauer, E. (2004), «Eine Frage von Telekommunikation? Die Griechen und ihre Schrift im 9.-7. Jahrhundert v. Chr.», en R. Rollinger & C. Ulf (eds.), *Griechische Archaik: interne Entwicklungen, externe Impulse*, Berlin, Akademie Verlag, 187-206.
- Wolff, P. (1963), *Breakthroughs in mathematics*, New York, New American Library.
- Wood, H. (1972), *The Histories of Herodotus: an analysis of the formal structure*, La Hague, Mouton.
- Woolridge, R. (2004), «Lexicography», en S. Schreibman *et al.* (eds.), *A companion to Digital Humanities*, Oxford, Blackwell, 69-79.
- Worthen, J. B. & Hunt, R. R. (2011), *Mnemonology: mnemonics for the 21st century*, New York, Taylor & Francis.
- Worthington, I. (1991), «Greek oratory, revisions of speeches and the problem of historical reliability», *C&M*, 42, 55-74.
- Worthington, I. & Foley, J. M. (ed.) (2002), *Epea and grammata: oral and written communication in ancient Greece*, Leiden, Brill.
- Yang, C. C. *et al.* (2003), «Visualization of large category map for internet browsing», *Decision Support Systems*, 35 (1), 89-102.

V. ANEXOS

V.1. TOPIC MODELS

V.1.1. LISTA DE STOPWORDS

1. Artículos: ό, ή, τό (23341)
2. Partículas: άν [ήν] (501), άρα (2), άρα (68), άτάρ (14), γάρ (1479), γε (237), δέ (8075) [δ' (322; aunque aparece ante vocales 3278 veces)], δή (1390), δήθεν (14), δήτα (5), ή (14), καίπερ (17), καίτοι (19), κε (5), κ ή (2), κ η (27), κόθεν (4), κοθεν, κοτε (60) [= ποκα, 4, 159, 3; = ποτε (5)], κου (47) [= που (3)], κω (51), κως (44), μέν (2567), μέντοι (116), νυν (329), περ (4), τε (2735), τοι* (22), τοιγάρ (2), τοίνυν (22), ώ* (180),
3. Conjunciones: αλλά (515), †διό (2), †διότι (14), ει (441) [αι : 4, 157, 2], είτε (68), είπερ (13), έπεάν (158), έπί (143), έπειδή (49), έπίτε (229), έπίν (4), έστε (11), εΰτε (5), έως (15), ή (430), ήπερ (17), ήτοι (7), ήν (209 = ει άν), ήνπερ (5), ίνα (127), και (7551), κατάπερ (69), μηδέ (55)[μεδ' (8)], μήτε (70), όκως (148), όμως (25), ότι (258), ούδέ (240), οΰτε (447), ώς (1461), ώστε (154).
4. Adverbios: άγχοϋ (22), †άγχόθεν, άγαν (2), άγχιστα (10), άγχοτάτω (14), †άδεως, αιεί[= αεί : MSS., 1, 97, 1; 2, 79, 2. αιέν : 7, 106, 2] (110), αι [1, 27, 3, «as utinam in Latin»], άλλη (20), άλλως (34), άμα (175), άμφίς, †άναμίξ (3), †άναξίως, άνέκαθεν, άλλοθεν, άλλοθι (2), άνω (53), άνωθεν (7), †άνωμοτί, άπαξ (9), άπαρτί (2), †άπειρος, †άπεναντίον, άπονητί (4), άπονητότατα, άπονητάτω, άπόνως, †άρεστώς, άρτι (2), άρχήθεν (8), άσσον (4), άτε (64), άτρεκέως (42), α υ (11), α ύτίκα (116), α ύτις (61) [La variante α ύθις no está recogida en el léxico de Powell], α ύτόθεν (4), α ύτόθι (34), †αΰτομάτως, †αΰτόσε, α ύτοϋ (59), α ύτως (6) [= ώσαύτως], †άφύκτως, γενναίως, δειν ώς (14), δεϋρο [2, 116, 5, tomado de Homero], διακριδόν, δίς (14), δίχα (6), †δίχοϋ, έγγύς (3), έθειλοντί (4), ε ίνεκα (84), ε ίνεκεν (46), έκάς (17), έκάστοτε (15), †έκατέρωθεν, έκατέροθι (2), εκει* (4), εκειθεν*, εκειθι (2), εκεισε (3), έκποδών (3), εκτός (10), έμπαλιν (7), έμπεδον, έμποδών (9), έμπροσθε (19), έναλλάξ (3), ένδοθεν, ένδον, ένερθε (11), ένθα (50), ένθαϋτα (182), ένθεν (16), ένθένδε, ένθεϋτεν (111), †ένιαχή (2), έντός (27), έξαιφνης, έξαπίνης (2), έξής, έξω (47), έξωθεν (13), †έπάνω (3), έπειτα (84), έπειτε (5), έπέκεινα, †επεξής

(10), ἐπιπᾶν (8), ἐπίτηδες (4), ἔσωθεν (8), †έσωτέρω, ἐτέρωθεν, ἐτέρωθι (7), ἔτι (225), εὖ (125), εὐθύς [= ἰθύς], ἤδη (135), ἥμισυ (21), κάρτα (93), †κατιθύ, κατύπερθε (50), κάτω (28), κάτωθεν (2), †κατωτάτω, †κατωτέρω, κότε (2), κοῦ (4), κῶς (16), λάθρη (6), λίην (3), μάλα (15), μάλιστα (219), μεγάλως (53), μεγαλωστί (3), μεταξύ (36), μετα ὕτις, μεταυτικά (2), μετέπειτα (15), μέχρη (118), μή (503), μηδαμά* (4), μηδαμῆς (4), μήκοτε (7), μόγις (10), νῦν (250), ὄθεν (17), ὄθι [4, 29, tomado de Homero], οἰκότως (3), ὄκη (7), ὀκόθεν (11), ὀκότε (4), ὄκου (25), ὀμοῦ (18), †ὄνομαστί (3), ὄπισθε (29), ὀπίσω (126), ὄταν (12), ὄτε (44), ὄτεπερ (2), οὐ/κ/χ (1479), οὐδαμά (26), οὐδαμῆ (8), †οὐδαμόθι (6), οὐδαμοῦ (4), οὐδαμῶς (34), οὐδέκοτε (2), οὐδέκω, οὐκέτι (42), οὐκί (12), οὔτω(ς) (516), πάλαι (25), πάλιν (8), πάμπαν (3), πανδημεῖ (9), πάντως (33), †πανοικίη (3), †πανστρατιῆ (5), †παντάπασιν, πανταχῆ (6), †πανταχόθεν (2), †πανταχοῦ, παντελέως (6), πάντῃ (12), πάντοθεν (7), †παντοίως, παράπαν (19), παραυτικά (16), πάρος, †πέρηθε, πέρην (12), πέριξ (16), Περσιστί (2), πλησίον (3), πολλάκις (22), πολλαχῆ (4), †πολλαχοῦ, πρὶν (86), †πρόκατε (4), †προσέτι, πρόσθε (14), πρόσω (27), πρώην, πρωί, σαφηνέως (3), σαφέως (6) [los términos derivados de la raíz σαφ- son utilizados por Heródoto para manifestar su opinión con respecto a sus fuentes; así lo refleja la traducción «reliable», «on good authority» que da Powell (s. v. σαφής)], σχεδόν (17), τάχα (4), τετράκις (2), τότε (207), τρίς (10), ὑψοῦ (3), χωρίς (37), ὧδε (128), ὥς (20), ὡσαύτως (19), ὥσπερ (34).

5. Preposiciones: ἀμφί (35), ἀνά (66), ἄνευ (27), ἀντί (71), ἀπό (642; ἀπ' [139]; τῷπὸ [2] : 1, 199, 4; 3, 40, 4*), ἄχρι (2), διά (335), ἐγγύς (3), ἐκ (1158), ἐν (1794), ἐπί (1336), ἐς (2748), κατά (662) [Powell añade aparte †κατά = κατ' ἅ (9)], †καταντίον (4), μετά (403), παρὰ (453), περὶ (538), πρό (71), πρὸς [= ποτί] (822), σύν (72), ὑπέρ (71), ὑπό (554).

Quedan pues, por analizar, verbos, sustantivos y adjetivos. Dentro de estas tres categorías se ha realizado también una pequeña criba. En primer problema que se plantea es, evidentemente, el sustantivo λόγος, con 447 usos: 1. «palabra» (30); 2. «cosa dicha» (82); 3. en pl. «talk» [ej. 2, 32, 1] (38); 4. «relato» (144); 5. «estimation» (59); 6. «condition» (10); 7. en perífrasis del tipo λόγον ἐωυτῷ διδόναι (ὅτι, ὡς...), «deliberar» (21); 8. λόγον (περὶ) τινος ποιέεσθαι, «hablar de» (8); 9. ἀληθεί λόγῳ, «expresión verdadera» (15); 10. frases hechas (27) [Sin embargo, sí que incluimos ῥῆμα, por su sentido más específico de «dicho» (6), y, por idéntico motivo, ῥήσις (3)]. Otro caso semejante es οὔνομα (277). El sustantivo πλῆθος no se ha tenido en cuenta, puesto que aunque puede indicar «multitud»

por sí solo, los usos con este sentido son relativamente pocos (17 de 92). Otro sustantivo con una pobre carga semántica es *πρῆγμα* (210) [1. «cosa, asunto» abstracto (88); 2. «affairs, business» (23); 3. «la situación» de forma vaga (27); 4. en plural con un adjetivo posesivo o un genitivo, indica «la causa o el origen» de algo, *τὰ τῶν Περσέων πρήγματα αὐξανόμενα* (1, 46, 1) (48); 5. el plural, «el gobierno, el Estado» (7); 6. *πρήγματα νεώτερα* «revuelta» (8); 7. en singular formando parte de expresiones hechas (9)]; igual de *ἔργον*; sin embargo, sí que puede tenerse en cuenta *συμφορῆ* (46), puesto que especifica un tipo de circunstancias, normalmente negativas (cf. Powell 1939, s. v. *συμφορῆ*); el sustantivo *χρόνος* (295), *ὥρη* (30), *ἡμέρη* (280) como expresiones de tiempo. Algunos sustantivos usados la mayoría de veces de forma apositiva y que en realidad funcionan como adjetivos: *ἀνήρ*, *βασιλεύς*, *πόλις*. El caso de *χώρα* y *ποταμός* es similar al de *πόλις*; además, en estos dos casos se trata de términos recogidos en el *index nominum* de Wilson, como modificadores de las palabras clave utilizadas, y, por lo tanto, son objeto de análisis por otro lado. No queremos que rompan la estadística.

1. Verbos:

1. Por su uso como parte de fórmulas: *λέγω* (1096) [de las que 1038 son «decir»], *μανθάνω* (149), *πάρειμι* (en expresiones del tipo «no sé quién estaba presente») o perífrasis verbales con *μέλλω* (116):
2. El caso de *ἔχω* (1142) es delicado. He aquí un análisis más en detalle de sus acepciones y de la frecuencia en cada caso:
 1. (A) usado como transitivo (886):
 1. «tener», y sentidos derivados, «tomar», «dominar», «controlar» (555);
 2. «sostener» (213);
 3. «impedir», «contener», en aoristo y futuro (10);
 4. otros significados (108).
 2. (B) empleado intransitivamente (162):
 1. como verbo auxiliar con aoristo y participio (8);
 2. sinónimo de «ser, estar» (83);
 3. «extender» (37);
 4. como impersonal (34); «extender» (37).

3. (C) en voz media con genitivo objetivo (73):

1. «estar junto a» (46);
2. «ser de tal o cual clase» (8);
3. «desistir de» (3);
4. «apegarse a» (4).

4. (D) en voz pasiva (21):

1. «ser tomado cautivo» (5);
2. «ser atrapado» (15).

5. ἐπι ξυρο ὁ ἀκμῆς ἔχεται τὰ πρήγματα (6, 11, 2*). [ἄλις ἔχειν (ἔστω): 1, 119, 5; IX, 27, 5*, «enough of a thing»], [†ἀνακῶς ἔχειν τινός: 1, 24, 7; VIII, 109, 4, «give heed to»] [†ἀπείρωσ ἔχειν τινός: 2, 45, 2, «be ignorant of»] [ἔγκοτον (5) ἔχειν τινί, «bear one a grudge»].

6. Compuestos de ἔχω: ἐνέχω (9).

3. El verbo εἰμί (3729) [Powell da una entrada aparte para el lema ἐστί y εἰσι, como enclíticos (631 y 206); también para ἔστί = ἔξεστί, «es posible»; la forma de infinitivo εἶναι es muy utilizada (618), y más aún la de participio, ἐὼν, ἐούσα, ἐὸν (923); en el imperfecto, la forma ἦν (476) [La posible confusión se limita a 2, 34, 1, donde equivale a ἐξῆν]. Sus compuestos: ἔνειμι (44), ἔξεστί (31), ἔπειμι (30) [Este sí debería ser analizado, por su sentido de «estar al mando de»], πάρεμι (167), πρόσειμι (6), σύνειμι (2). Powell da una entrada aparte para el participio de presente ὄν (455).
4. Sinónimos o similares verbos con poca carga semántica: γίνομαι (1359) [= γίγνομαι; más de la mitad de sus usos son «pobres» semánticamente: 402 usos como «happen, take place»; 340 como «become». No obstante, sí que deben analizarse compuestos como †προσγίνομαι, «unirse, apoyar» (10)], καλέω (352), de los que 294 corresponden al uso predicativo «llamar» o «nombrar» a algo o alguien; κείμαι (98) [pero sí que podemos incluir πρόκειμαι, ya que de sus 38 apariciones, sólo 11 tienen un sentido local, «estar en»]; παραγίνομαι (26) «estar presente»; φέρω (219) [1. transitivo (133); 2. intransitivo (69); 3. en voz pasiva (16)]; sólo unos pocos usos de φύω (8 de 36) equivalen más o menos a «ser», por lo que, en este caso, no lo incluimos en la lista de palabras no analizadas.

5. En general, verbos con una carga semántica pobre: ποιέω (1215) [(A) 831 usos con el sentido de «hacer algo»: I. «make» (195): 1. «crear, manufacturar» (70); 2. «componer» (dicho de poetas etc.) (16); 3. «render» con objeto predicativo (88); 4. «bring about» (14); 5. «put», τοῦτο ἐπὶ νόον τινί (1, 27, 3*) (3); 6. «make a fuss», δεινὰ π. (4); II. «do» (636): 1. uso cognado con acusativo (549); 2. con doble acusativo, «hacer algo a alguien» (27); 3. con adverbios (38); 4. uso final o prospectivo, ποιέειν ὅπως con fut. ind. (8); 5. absoluto, «actuar» (14); (B) en voz media (384): I. «make» (167); II. «considerar, pensar» (93); III. formando perífrases verbales, del tipo ἀγῶνα π. = ἀγωνίζομαι (122); IV. δι' ἀγγέλων, διὰ χρηστηρίων, ποιούμενος, «deal through» (2)], y su sinónimo, πρήσσω (70) [sin embargo, sí que analizamos aquellos compuestos con un sentido pleno, como προσποιέομαι (8)]; el caso de συμφέρω es difícil [de 49 usos, en voz activa: 1. «bring together» (2); 2. «be profitable» (4); 3. impers. «ocurrió» (17); en voz pasiva: 4. «collide with», sólo una vez en 6, 50, 3* (μεγάλῳ κακῷ); 5. dentro de un relato, «estar de acuerdo con» (11); 6. «ocurrir a (alguien o algo)» (14)]; τίθημι (99), τρέπω (104); τυγχάνω (110) y compuestos: ἐντυγχάνω (10), παρατυγχάνω (5); χρεόν (ἔστι) (31+3), χροή (38), χροῶ (330) [sólo 59 con el sentido de «dar un oráculo» y 2 significado «prestar», frente a los 268 casos en los que significa simplemente «usar»]; δίδωμι (302).
6. Verbos de opinión y conocimiento: δόκω (386) y compuestos con sentido similar: προσδοκῶ (4); νομίζω (125), οἶμαι, πιστεύω (12); πυνθάνομαι (277) [sólo 19 usos en el sentido de «preguntar», frente a los 258 en el sentido de «aprender»; este verbo forma parte del léxico utilizado para expresar la procedencia de la información; este conjunto de términos «meta-literarios» o «metasemánticos» está presente en casi todas las informaciones, de modo que hay que eliminarlo del análisis automático para que no altere los resultados, del mismo modo que eliminamos todos aquellos elementos que tiendan a ser ubicuos o cuyo sentido depende del contexto]; οἶδα (152), σύνοιδα (7), φρονέω (33), φροντίζω (19).
7. Verbos y sustantivos de lengua, especialmente λόγος, por su participación en fórmulas introductorias: καλέω, φημί (354), φράζω (83), λόγος, εἶπον etc.

8. Ciertos verbos de percepción, por su uso en las fórmulas explicativas de Heródoto, del tipo «he visto esto en...»: ὄρω (356).

9. De movimiento: ἀφικνέομαι, ἔρχομαι,

2. . Adjetivos:

1. . Numerales: δέκα (66), δέκατος (22), δευτερος (122), δισμύριοι (8), δισχίλιοι (12), †διχοί (2), δύο (164), δυοκαίδεκα (3), δώδεκα (39) [pero δώδεκα, 6, 108, 4], δυωδέκατος (2), εἴκοσι (62), εἰκοστός (4), †εἰνακισχίλιοι (2), εἷς (216), ἑκατόν (63), ἑκτός (11), ἕνδεκα (17), ἑνδέκατος (6), ἑνενήκοντα (5), ἑννέα (6; más dos como parte del nombre de lugar Ἐννέα Ὀδοί), ἕξ (40), ἑξακόσιοι (18), ἑξακισμύριοι, †ἑξακισχίλιοι (7), †ἑξακόσιοι (18), ἑξήκοντα (31), †ἑξηκοστός (2), ἑπτά (73), †ἑπτακαίδεκα (6), †ἑπτακισμύριοι, †ἑπτακισχίλιοι (2), †ἑπτακόσιοι (8), πρῶτος (πρότερος), πέμπτος (17), †πεντακισμύριοι (2), †πεντακισχίλιοι (9), πεντακόσιοι (23), πέντε (64), πεντεκαίδεκα (13), †πεντακαιδέκατος, πεντέκοντα (38), τέσσαρες, ὀκτώ, †ὀκτακισχίλιοι (4), †ὀκτακόσιοι (10), ὀκτωκαίδεκα (6), πρότερος (57), πρῶτος (211), τεσσεράκοντα (35), τέσσαρες (61), †τεσσερεσκαίδεκα (4), †τεσσερεσκαιδέκατος (2), τέταρτος (22), †τετρακισχίλιοι (12), †τετρακόσιοι (24), τέως (17), †τηνικαῦτα (19), τρεῖς (104), τριήκοντα (49), †τριηκόντερος (3), †τριηκόσιοι (62), †τριηκοστός (2), τρίξοι (4), τρισμύριοι (6), τρισχίλιοι (16), τρίτος (61), †τριφάσιοι (3), χίλιας (19), χίλιοι (57).

2. Cantidad y cualidad (relativa): ἄλλος (989), ἀλλοῖος (3), ἄλλοτε* (2), ἀλλότριος (3), ἀλλήλους (71), ἐκῶν (25), ἐλάσσων (59), ἐλάχιστος (23), ἕτερος (173), ἥμισυς (32), ἥσων (33), ἴσος (45), κοῖος (17), κόσος (6), κότερος (13), μακρός (56), μέγας (368), μέσος (108), μηδεῖς (84), μόνος (211), οὐδεῖς (564), οἶκα (63; es un participio pero por su función lo incluimos entre los adjetivos, además de por el tipo de carga semántica que comporta), οἶος (173), †οἶόσπερ (9), ὀκοῖος (15), ὀκόσος (17), ὀκότερος (21), ὄλος (7), ὄμοιος (60), ὄσος (250), ὄσοσπερ (16), πολύς, πᾶς (1134), ἅπας (118), οὐδέτερος (17), οὕτος (4592), παραπλήσιος (24), πιστός (28), πλείστος (64), πλέων (156) [alternancia en la raíz entre -εο y -ευ.], ποδαπός (3), πολλός (670) [πολύς, ύ, ύν MSS.], σμικρός (40), συνάπας

(12), τοῖος (3), τοιόσδε (212), τοιοῦτος (188), τόσος (3), τοσόσδε (24), τοσοῦτος (126).

3. Posesivos, por las mismas razones que eliminamos los pronombres personales: ἐμός (88), σός* (70), ἡμέτερος (40), σφέτερος (64).
4. De lugar: ἀγχότατος (3), †ἀγχιστέες, †ἀγχότερος, ἐναντίος (26), πλησίος, ὑπεναντίος (4).
5. De tiempo: ἐνιαυτός (25), ἐπέτειος (9), πρώιος, τριταῖος (4), ὕστατος (15), ὕστερος (91).
6. .Distributivos: ἕκαστος (260), ἐκάτερος (54).

3. Los pronombres han sido igualmente excluidos del análisis, en tanto que su carga semántica depende de un antecedente que el algoritmo no puede detectar: αὐτός (1101) (y su variante τούτο), αὐτόν/ήν/ό (1412) [pronombre anafórico la mayoría de ocasiones (1289), pero αὐτόν, reflexivo, en el epigrama citado en 4, 88, 2.], οὔτος σφεῖς, ἐγώ/ἡμεῖς (238; incluye las instancias en plural), ἔγωγε (13), ὄδε (ἦδε, τόδε) (709), ἐαυτός, τις (τίς), σύ* (142), ὑμεῖς, μιν (329), οἱ (dat. sg., 566), ὅς (ἦ, τό) (2039), ὅσπερ (102), ὅστις (215), †ὅστισπερ, †οὐδαμοί (29), ἐκεῖνος (244), ἐνιοι (6), ἐωυτοῦ (528), σεωυτοῦ* (37), σφε (la suma de todas sus formas: 4+5+22+64+228+16+61+9+80+6+628 = 1123), τις (656), ὑμέτερος* (17).

V.1.2. RESUMEN POR TEMA

Tema 1

Términos incluidos en el tema, en orden descendente de frecuencia:

Σπάρτη, μάχομαι, φεύγω, Λακεδαιμονίων, ἀποθνήσκω, πόλεμος, ἀθηναῖοι, ἀθηναίων, Πausanίας, σπαρτιητέων, σύμμαχος, μάχη, Λακεδαιμόνιοι, ἄριστος, πέρθω, ἄγγελος, Λακεδαιμονίοισι, ἀπαλλαξείω, κλέω, στρατόπεδον, νικάω, κήρυξ, αἰρέω, γελάω, τάξις.

Observaciones a la lista. Errores de lematización: Errores claros son la separación en dos o más entradas de lemas que deberían ser únicos:

1. Ἀθηναῖοι: el programa separa ἀθηναῖοι y ἀθηναίων.

2. Λακεδαιμόνιοι : además de la forma correcta, el programa distingue entre Λακεδαιμονίων y Λακειμονίοισι.

Los términos más característicos de este tema, esto es, aquellos cuya frecuencia total y frecuencia local (en este tema en concreto) es más similar, son: Σπάρτη , Λακεδαιμονίων (y -οισι), Παισάνιας , Σπαρτιήτων , κλέω y γελάω . Esto no quiere decir que una descripción de lo que representa este tema deba basarse exclusivamente en este tipo de términos. Siguiendo la lógica subyacente al proceso de detección o generación de estos *topic models*, el objetivo de este proceso es resaltar la especial relación que se establece entre palabras que aparecen a lo largo de todo el texto; por lo tanto, un análisis de las palabras contenidas en un tema debe explicar activamente estas relaciones.

Tema 2

πέρθω, κυρέω , μάγος , κακός, ἄρχω, γυνή , πατήρ , γνώμη, ἀρχή, Κύρος, δαρείου (Δαρείος), ἀδελφός, στρατεύω, ἀγαθός, Ξέρξης, ἀμείβω, θυγάτηρ, μήδων, δαρείος (Δαρείος), ἐφίστημι, Καμβύσης, δεσπότης, ἀπόλλυμι, σαυτοῦ, τελευτάω.

Observaciones a la lista: errores de lematización:

1. El nombre propio de Darío (Δαρείος) ha dado algún quebradero de cabeza al programa, que ha separado el genitivo del lema al que debería pertenecer.

2. Algo semejante ocurre con Ciro, cuyo genitivo (Κυρέω) es analizado por separado.

Este tema posee pocos términos claramente específicos del mismo, si atendemos exclusivamente a la frecuencia. Sin mencionar σαυτοῦ, que no debería aparecer¹, sólo podemos apreciar μάγος como término característico, seguido con cierta distancia de Ciro

¹ Debe de tratarse de un error de lematización, puesto que el pronombre αὐτός está en la lista de *stopwords* excluidos del análisis.

(más en su forma de genitivo; aquí el error ha podido revelar accidentalmente un uso diferenciado), μεδών, Καμβύσης, δεσπότης y ἀμείβω.

Tema 3

ῥέω, ὄρος, λίμνη, ὁδός, ἥλιος, ἄνεμος, ὕδωρ, γαῖα, βορέας, ἐσπέρα, χῶρος, κόλπος, καθύπερθε, ἐκδίδωμι, νέμω, ἔθνος, ἠώς, λιβύης, χειμών, σκυθικός, κείμαι, πλέω, ὀρύσσω, νῆσος, πόντος.

Observaciones a la lista: errores de lematización: no hay.

Los términos más característicos de este tema son: ῥέω, κόλπος, Βόρεας, ἐσπέρα, seguidos de Λιβύης, χειμών y πόντος. La inclusión del adverbio homérico καθύπερθε parece un error por nuestra parte, al no estar incluido en la lista de stopwords.

Tema 4

πέρθω, στρατός, στρατιά, Ξέρξης, ὁδός, ἄσιος, περεύω, διαβαίνω, τάσσω, παίονες, φωκεύς, Ἑλλάδα (ἤ Ἑλλάς?), ἔθνος, ἐλαύνω, γαῖα, σκύθας, πεζός, στρατηγός, εισβάλλω, βουλεύω, πειράω, συμβάλλω, Ἑλλησπόντον, σκύθαι, γνώμη.

Observaciones a la lista: errores de lematización: Ἑλλάδα en lugar del lema Ἑλλάς; entrada separada para el acusativo plural de Σκύθαι.

Los términos más característicos de este tema son: παίονες, φωκεύς, seguidos de Ἑλλησπόντον, Σκύθας y πορεύω.

Tema 5

ναύς, πλέω, ναυμαχία, ἔλληνες, στρατός, πεζός, ναυτικός, στρατηγός, διαφθείρω, τάσσω, ἀθηναῖοι, ξέρξω (Ξέρξης), ἀριθμός, βουλεύω, νικάω, ἄριστος, γνώμη, Θεμιστοκλῆς, ναυμαχέω, ἰσθμός, φοῖνιξ, πλῆθος, Σαλαμίνα (Σάλαμις), πλοῖον, ἀθηναίων.

Observaciones a la lista: en este tema tenemos tres casos de lematización por separado de casos pertenecientes al mismo lema: el genitivo de Ξέρξης y ἀθηναῖοι, y el acusativo de Σάλαμις.

Los términos más característicos de este tema son: να ὑς², Θεμιστοκλής, ναυμαχέω, Σαλαμίνα y ναυμαχία.

Tema 6

ιερός, χρύσεος, θύω, βο ὑς, καρπός, θεός, ο ἴνος, γα ἰα, χρυσός, κρέας, πρόβατον, κε ἴμαι, κτ ἦνος, π ὑρ, χρ ἦμα, θυσία, ἄγαλμα, νεκρός, ξύλον, ναός, χάλκεος, νύξ, κεφαλή, δένδρον, φύω.

No parece haber errores de lematización en este tema. Los términos más característicos de este tema son οἶνος, κρέας, κτήνος, πρόβατον, καρπός y βούς.

Tema 7

Σάμος, να ὑς, στρατιά, χρ ἦμα, τύραννος, α ἰρέω, ἰώνων, πλέω, στρατηγός, ἄρχω, κακός, πέρθω, Σάρδις, στρατεύω, Μίλητον, ἀθηναῖοι, λαμβάνω, Μιλήτου, ἰστιαῖος, πολιορκέω, ἀφεστήξω, Λακεδαιμόνιοι, Δαρείος, ἰωνίην.

Observaciones a la lista: errores de lematización: el genitivo Μιλήτου ha sido separado de Μίλητον; por lo demás, no hay errores.

Este tema tiene pocos términos exclusivos con respecto al resto de la obra, apenas tres: Μίλητον (idéntica proporción de aparición tanto en genitivo como en nominativo), Ἴστιαῖος y Ἴωνίην.

Tema 8

χρηστήριος, ἀθηναίων, γνώμη, θεός, Κροῖσος, κελεύω, χρ ἦμα, ἐπερωτάω, χρησημός, ἀναιρέω, πυθίη, μαντεῖον, Δελφούς, ἄρχω, ἀθηναῖοι, λοιπός, ἐπέρομαι, ἐφίστημι, Μιλτιάδης, ἀμείνων, τίθημι, ἄριστος, Ἴππίας, δῆμος, συμβουλεύω.

Observaciones a la lista: el lema ἀθηναῖοι solo aparece recogido en este tema en genitivo; algo parecido ocurre con Δέλφοι, recogido en acusativo.

² Aunque el término no es exclusivo de este tema, su frecuencia relativa con respecto al resto de la obra y dentro de su propio tema es muy superior al resto.

Los términos más característicos de este tema son *χρησμός*, *Ἰππίας* y *Μιλτιάδης*.

Tema 9

γυνή, *κελεύω*, *κεφαλή*, *λαμβάνω*, *Ἄρπαγος*, *κακός*, *νεκρός*, *θάπτω*, *ἐντέλλω*, *χρῆμα*, *ἀποτέμνω*, *ἀφίημι*, *ἐφίστημι*, *ἀποθνήσκω*, *ὁδός*, *μηχανάομαι*, *δε ἵπνον*, *θυγάτηρ*, *φυλακός*, *τέκνον*, *ἐργάζομαι*, *ἀποπέμπω*, *παιδίον*, *μήτηρ*, *οἰκία*.

No se aprecian errores de lematización. Los términos más característicos de este tema son: *φυλακός*, *παιδίον*, *δείπνον* y *Ἄρπαγος*.

Tema 10

ιερός, *θεός*, *Αἰγύπτιοι*, *γυνή*, *Ἕλληνες*, *ἄγαλμα*, *θύω*, *γλῶσσα*, *Αἰγυπτίων*, *ἰδρύω*, *ἐρῶ*, *Αἰγύπτου*, *Ἑλένη*,

ιερεὺς, *Διός*, *Ἕλλησι*, *Δῆλος*, *ὀνομάζω*, *φοῖνιξ*, *τέμενος*, *ἔπος*, *χρηστήριος*, *τίθημι*, *Αἰγυπτίοισι*, *λαμβάνω*. Como hemos visto ya en varios temas, la mayoría de errores de lematización se centran en nombres propios, creando entradas separadas para instancias del mismo nombre en casos diferentes. Aquí se aprecia en *Ἕλληνες* (entrada separada para el dativo plural) y *Αἰγυπτίοι* (genitivo y dativo analizados por separado). El genitivo suele separarse con frecuencia, sustituyendo por completo a la forma base, como ocurre aquí con *Διός* y *Αἰγύπτου*.

En cuanto a los términos más característicos de este tema sólo podemos hablar con seguridad de *Ἑλένη* y *Διός*, seguidos a cierta distancia de términos sin una relación muy clara entre ellos: *Αἰγυπτίοισι*, *τέμενος*, *Δῆλος*, *ὀνομάζω*.

Tema 11

τάλαντον, *λίθος*, *νομός*, *βασιλεύω*, *πεντήκοντα*, *δεκάω*, *ἴστημι*, *ιερός*, *μέγαθος*, *τριάκοντα*, *φόρος*, *λίμνη*, *στάδιος*, *μῆκος*, *στάδιον*, *Αἰγύπτου*, *διακόσιοι*, *ἕξ*, *ὀκτώ*, *Αἰγυπτίων*, *σταθμός*, *τεσσαράκοντα*, *ἐλαύνω*, *πυραμῖς*, *πήχυς*.

Detectamos errores de lematización con στάδιον (στάδιος), Αἰγύπτου (sólo en genitivo), Αἰγυπτίοι (sólo en genitivo plural); por otra parte, se han «colado» dos numerales, πεντήκοντα y τεσσαράκοντα.

Los términos más característicos de este tema, excluyendo numerales como πεντήκοντα, τριάκοντα y τεσσαράκοντα, son τάλαντον, σταθμός, πυραμῖς, φόρος y νομός.

Tema 12

πέρθω, τεῖχος, ἴωνες, πολιορκέω, ἱερός, Σάρδις, μάχομαι, ἄστυ, στρατόπεδον, στρατός, ἀκρόπολις, χῶρος, ἀλίσκομαι, ἰώνων, ἀφεστήχω, στρατεύω, ἐμπύμπρημι, ἐσσομαι, ὀρμάω, κήρυξ, θεός, λυδῶν, εὐρίσκω, πόλεμος, ἐλαύνω.

El único error de lematización que se aprecia es el genitivo plural Λυδῶν en lugar del esperado Λύδοι. Por otra parte, es difícil establecer cuáles son los términos más característicos de este tema, puesto que ninguno es exclusivo. Por lo tanto, debemos situar en primer lugar aquellos con una frecuencia de aparición en este tema mayor que en el resto de la obra: Λυδῶν, ἐμπύμπρημι, ἀκρόπολις, πολιορκέω y Ἴωνες.

Tema 13

πέρθω, κεφαλή, ἐσθής, ἄρχω, φορέω, ἀσπίς, ἔθνος, αἰχμή, τόξον, πίπτω, χεῖρ, λαμβάνω, κάμηλος, σκευάζω, μάχη, ἐνδύω, ὀφθαλμός, ὀπισθεν, βάλλω, τέλος, ἔλκω, μικρός, βραχύς, ἐπιχώριος, ἔπειμι.

No se aprecian errores de lematización. Los términos más característicos de este tema son αἰχμή, ἀσπίς, ἐνδύω, μικρός, βραχύς, κάμηλος, ὀφθαλμός y φορέω.

Tema 14

γυνή, θυγάτηρ, νόμος, τίκτω, πατήρ, γαμέω, ἀδελφός, Περίανδρος, βασιλείος, τέκνον, ἄρσην, καλός, νέος, ἀποκτείνω, Ἄδραστος, συνοικέω, ἱερός, γάμος, παρθένος, Κλείσθηνς, δέχομαι, βουλεύω, λαμβάνω, μίγνυμι, τελευτάω.

No se aprecian errores de lematización. Los términos más característicos de este tema son Περίανδρος, γαμέω, γάμος, ἄρσην, Ἄδραστος y Κλεισθένης.

Tema 15

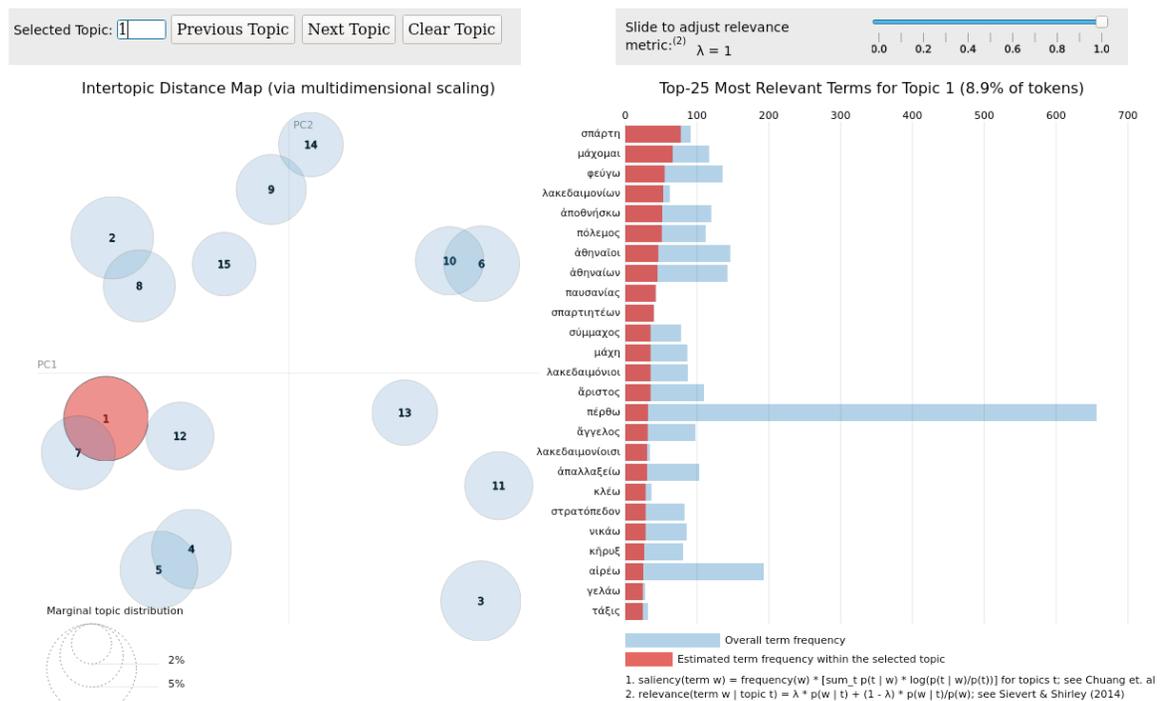
πέρθω, τε ἶχος, α ἰρέω, α ἴτιος, δίκαιος, Α ἰγυπίων, βας ἰλειος, στρατιά, Αἴγυπτον, Ἄμασις, δόκιμος, δίκη, ἄστν, κελεύω, Α ἰγύπτου, πολυκράτεος, πύργος, ἡμαρ, κῦρος, Αἰγύπτιοι, ἡμίονος, πύλη, κτείνω, Δαρειός, ἀρχή.

Aparte de una lematización errónea de Α ἰγύπτος y Α ἰγύπτιοι (en genitivo y acusativo), no se aprecian más errores.

Según la gráfica, este tema se encuentra aislado de los demás. Los términos más característicos son Πολύκρατης, πύργος, ἡμαρ, ἡμίονος y Ἄμασις.

V.1.3. REPRESENTACIONES GRÁFICAS

Tema 1

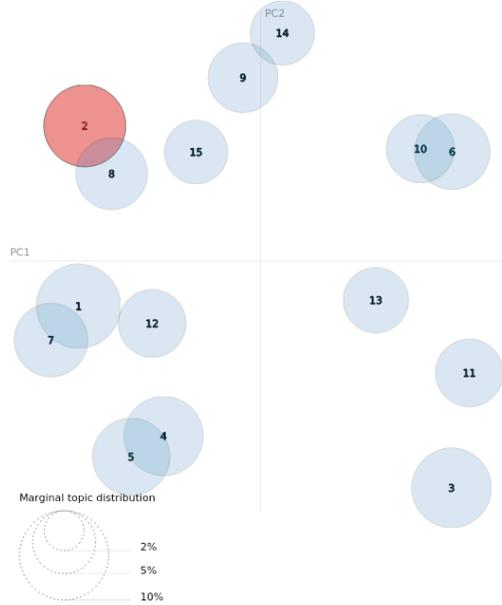


Tema 2

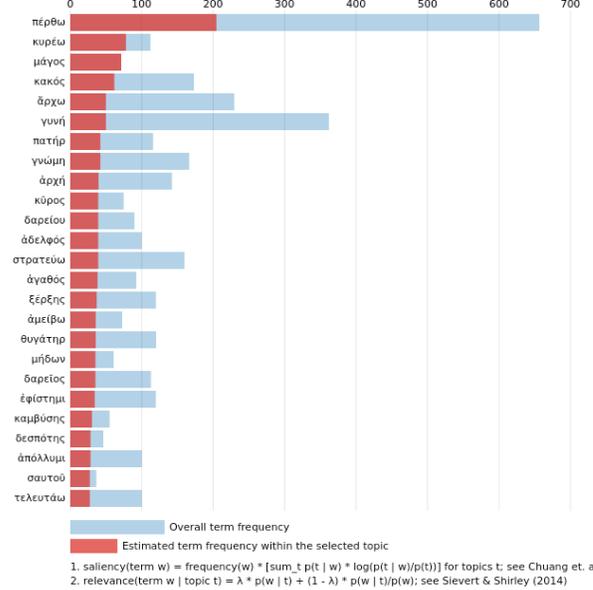
Selected Topic: Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)



Top-25 Most Relevant Terms for Topic 2 (8.4% of tokens)

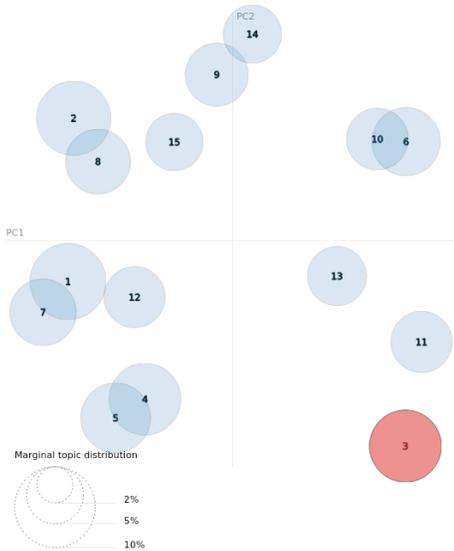


Tema 3

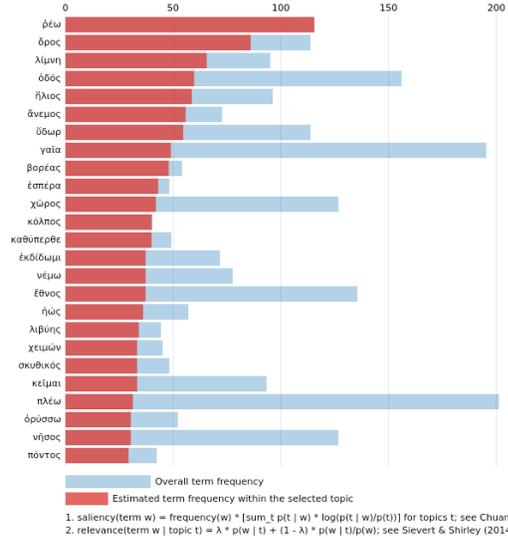
Selected Topic: Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)



Top-25 Most Relevant Terms for Topic 3 (8% of tokens)



Tema 4

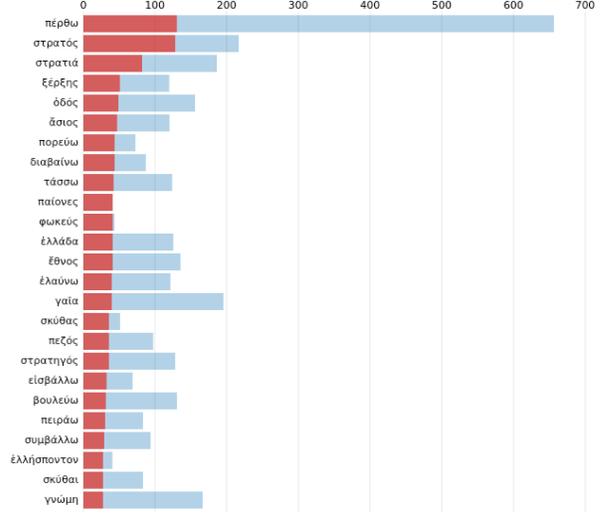
Selected Topic: Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)



Top-25 Most Relevant Terms for Topic 4 (7.9% of tokens)



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))]

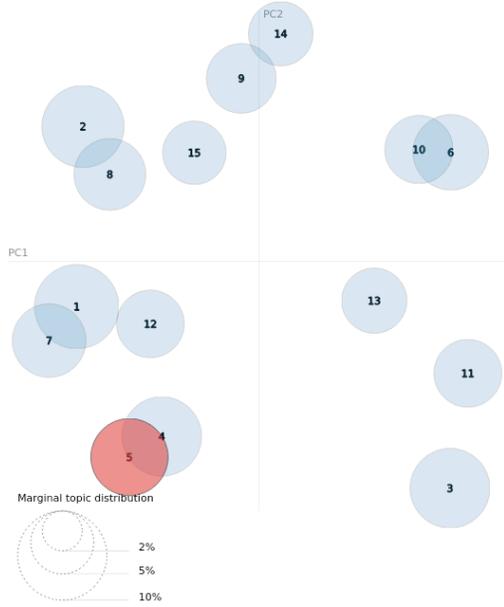
2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

Tema 5

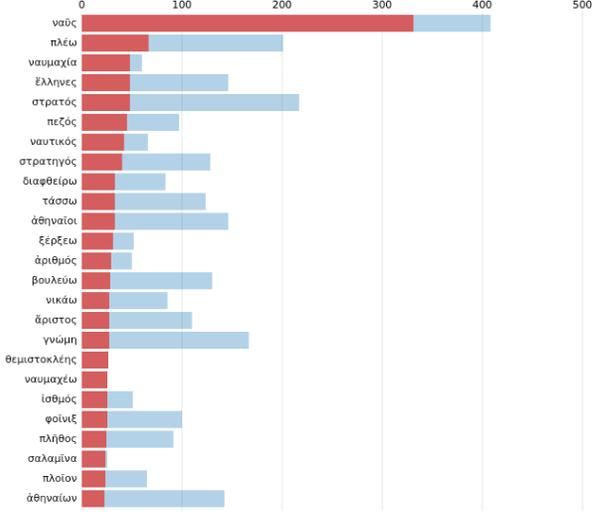
Selected Topic: Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)



Top-25 Most Relevant Terms for Topic 5 (7.5% of tokens)



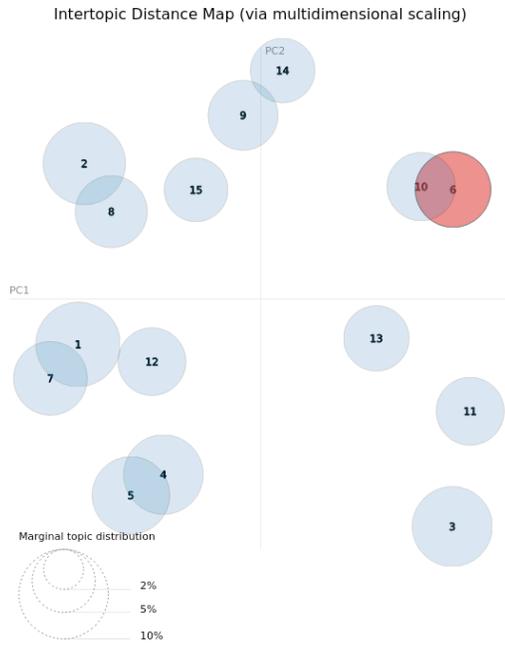
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))]

2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

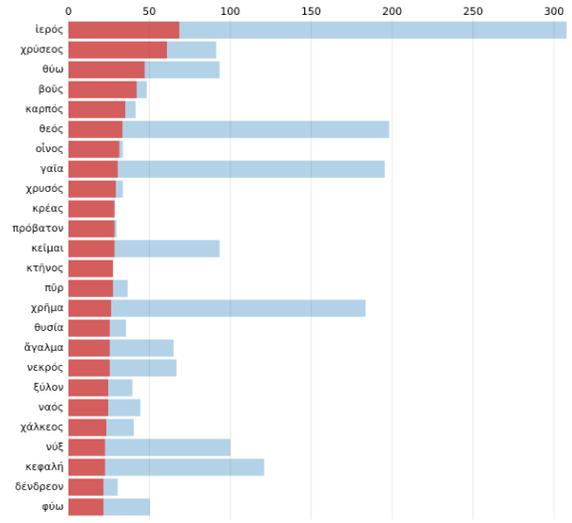
Tema 6

Selected Topic:

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$



Top-25 Most Relevant Terms for Topic 6 (7.1% of tokens)

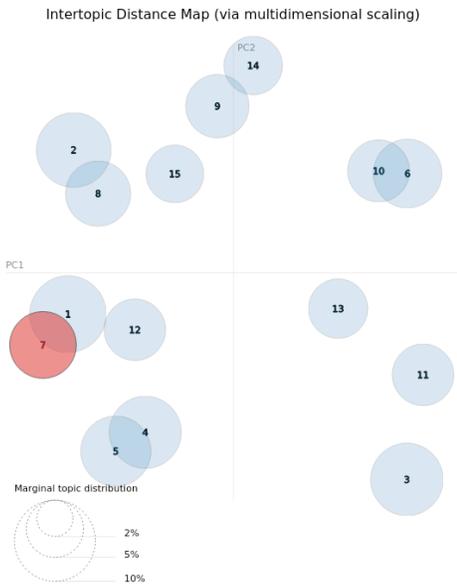


1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))]; see Chuang et. al
2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

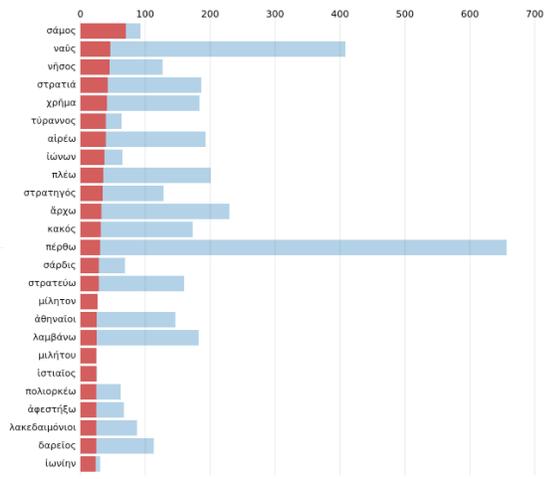
Tema 7

Selected Topic:

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$



Top-25 Most Relevant Terms for Topic 7 (6.8% of tokens)



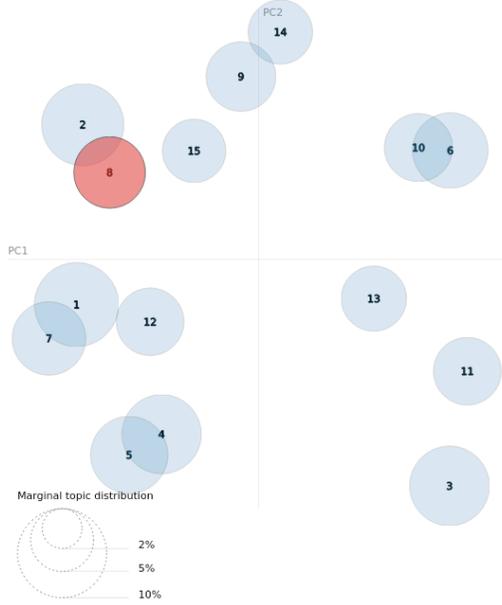
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))]; see Chuang et. al
2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

Tema 8

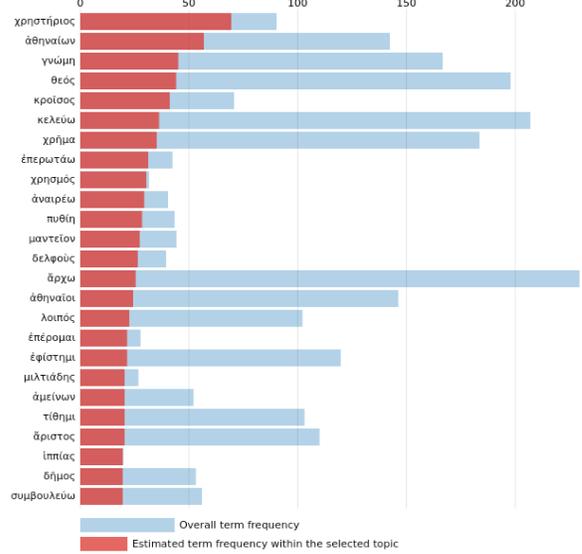
Selected Topic: 8 Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$ 0.0 0.2 0.4 0.6 0.8 1.0

Intertopic Distance Map (via multidimensional scaling)



Top-25 Most Relevant Terms for Topic 8 (6.4% of tokens)



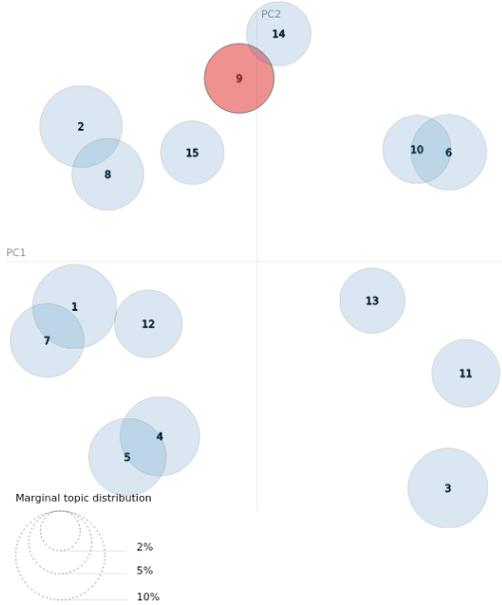
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al
 2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

Tema 9

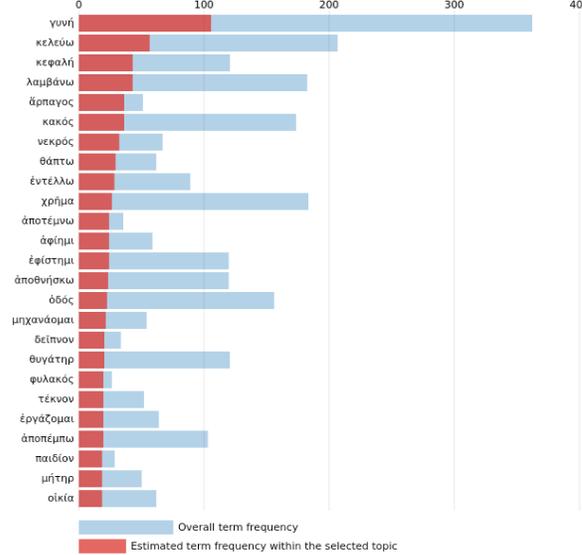
Selected Topic: 9 Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$ 0.0 0.2 0.4 0.6 0.8 1.0

Intertopic Distance Map (via multidimensional scaling)



Top-25 Most Relevant Terms for Topic 9 (6.1% of tokens)



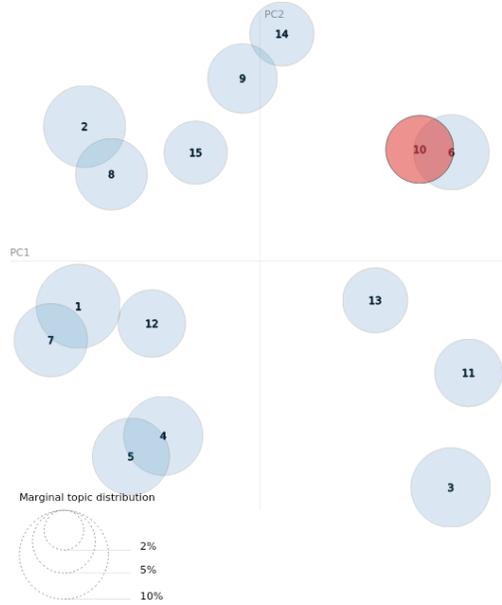
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al
 2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

Tema 10

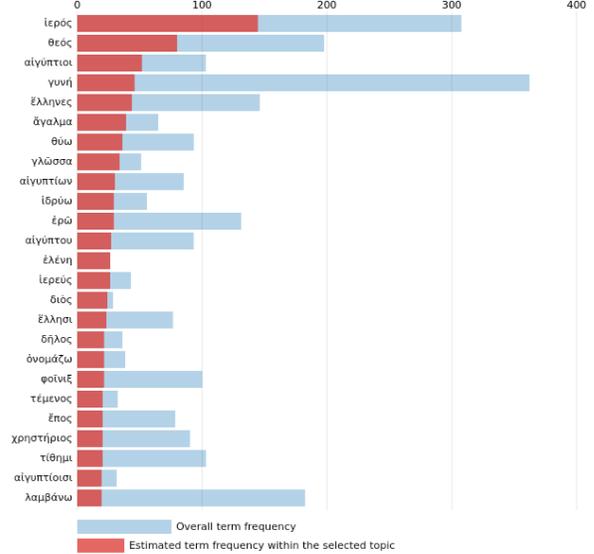
Selected Topic: Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)



Top-25 Most Relevant Terms for Topic 10 (5.8% of tokens)



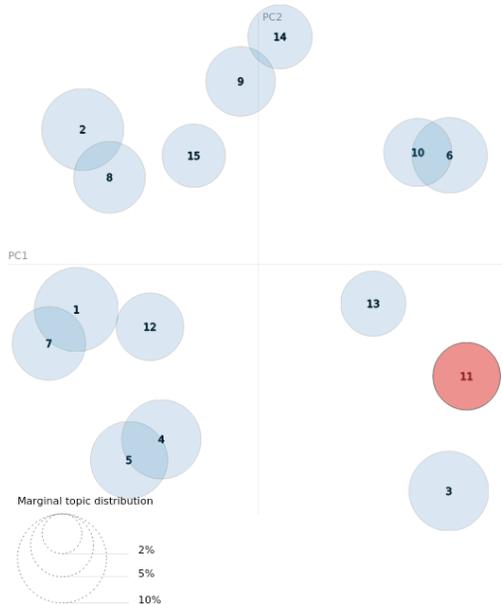
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al
2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

Tema 11

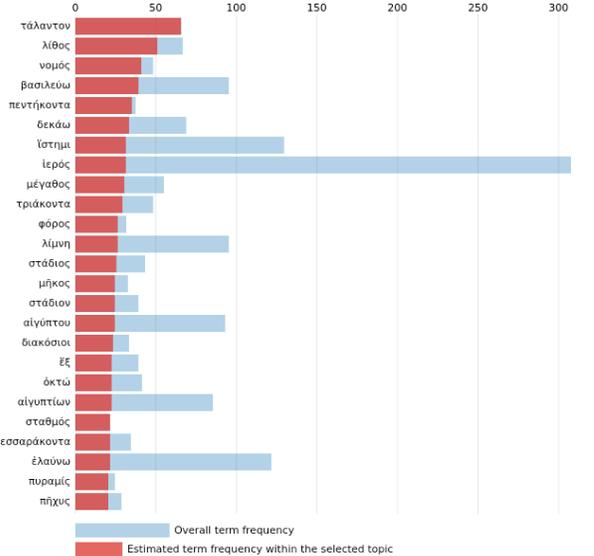
Selected Topic: Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)



Top-25 Most Relevant Terms for Topic 11 (5.8% of tokens)



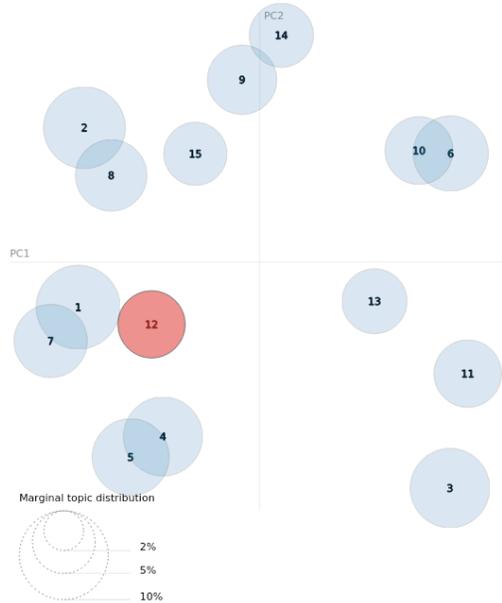
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al
2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

Tema 12

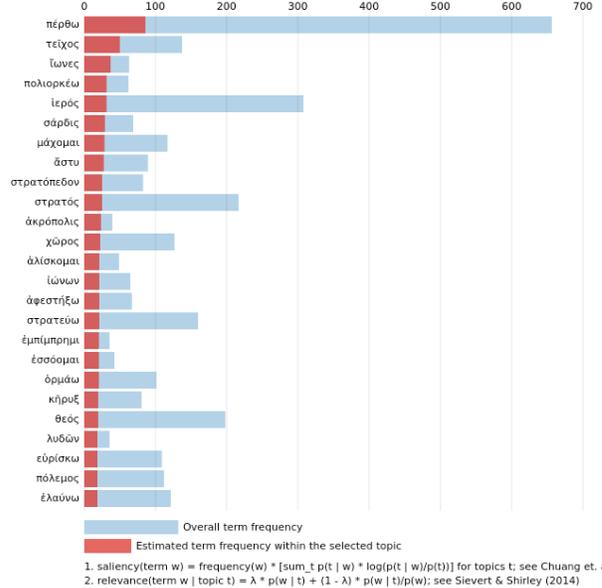
Selected Topic:

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)



Top-25 Most Relevant Terms for Topic 12 (5.7% of tokens)

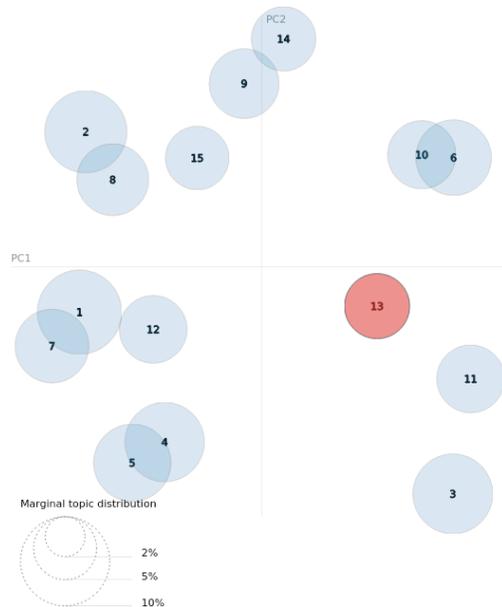


Tema 13

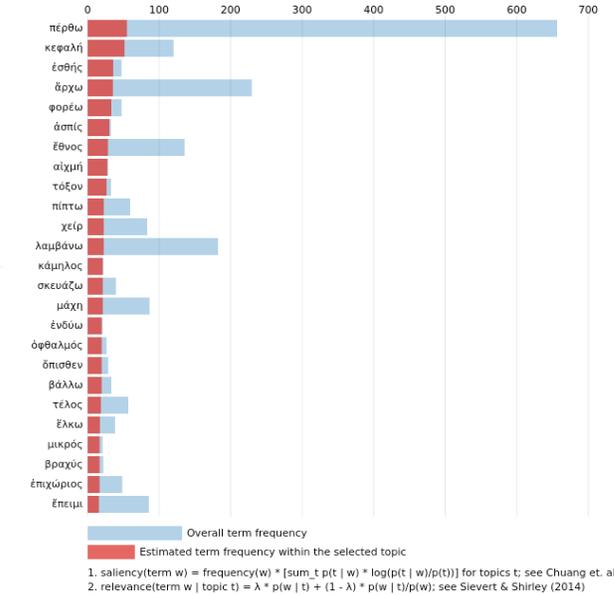
Selected Topic:

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)



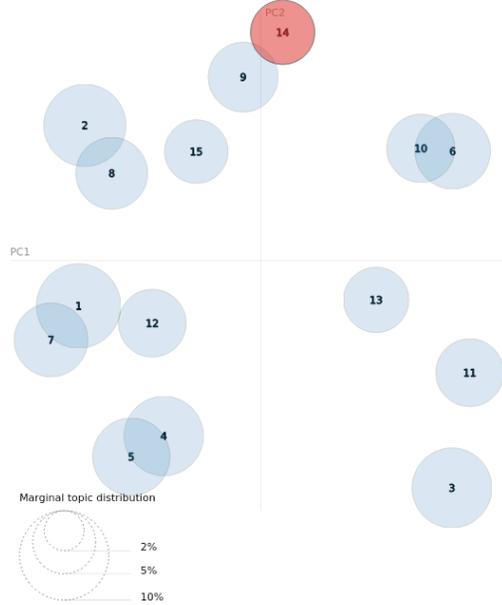
Top-25 Most Relevant Terms for Topic 13 (5.3% of tokens)



Tema 14

Selected Topic: 14 Previous Topic Next Topic Clear Topic

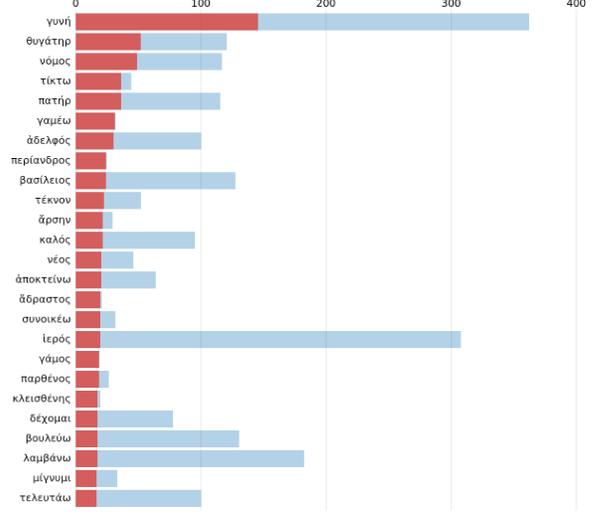
Intertopic Distance Map (via multidimensional scaling)



Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$



Top-25 Most Relevant Terms for Topic 14 (5.2% of tokens)



Overall term frequency

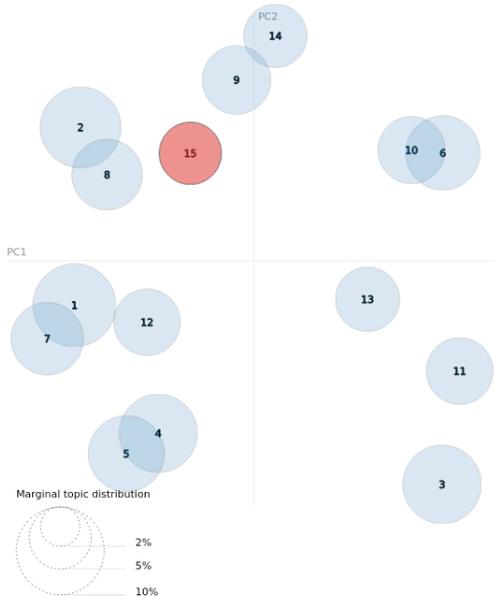
Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))]
2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

Tema 15

Selected Topic: 15 Previous Topic Next Topic Clear Topic

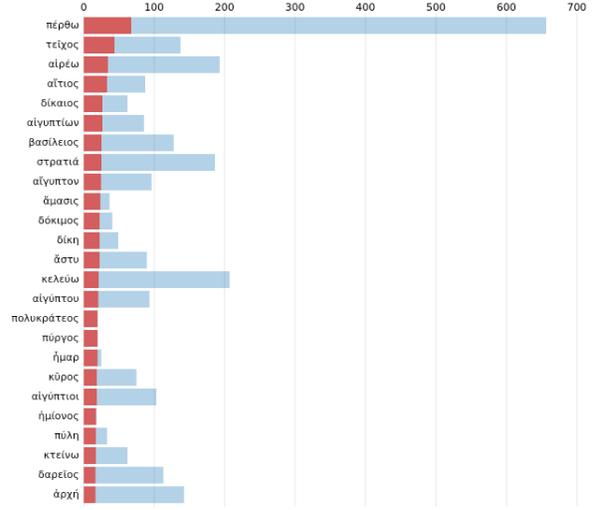
Intertopic Distance Map (via multidimensional scaling)



Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$



Top-25 Most Relevant Terms for Topic 15 (5% of tokens)



Overall term frequency

Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))]
2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

ÍNDICE GENERAL

I. INTRODUCCIÓN	1
I.1. BÚSQUEDA SEMÁNTICA Y SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN (SRI)	7
I.2. ESTRUCTURA DEL SIGNIFICADO: DATOS E INFORMACIÓN.....	10
I.3. MATERIAL DE ESTUDIO: EL TEXTO.....	12
<i>I.3.1. Definición de texto</i>	12
<i>I.3.2. El texto oral</i>	16
I.3.2.1. La oralidad y la memoria.....	17
I.3.2.1.a. El lenguaje formular	22
I.3.2.2. Funcionamiento de la memoria.....	28
I.3.2.2.a. La memoria como soporte de información	28
I.3.2.2.b. El estudio de la memoria	28
I.3.2.2.c. La atención y los tipos de memoria	30
I.3.2.2.d. La memoria a corto plazo	31
I.3.2.2.e. La memoria a largo plazo.....	32
I.3.2.2.f. Aportación del estudio de la memoria al análisis del texto	34
I.3.2.3. La memoria en la Antigüedad	35
I.3.2.3.a. La memoria como concepto	35
I.3.2.3.b. La μνήμη	37
I.3.2.3.c. Adaptación del recuerdo	39
I.3.2.3.d. Impacto de la escritura en el concepto antiguo de memoria.....	40
I.3.2.3.e. Μνήμη y mnemotecnia	41
I.3.2.3.f. Las huellas de la μνήμη en el texto	42
<i>I.3.3. El texto escrito</i>	46
I.3.3.1. La escritura en Grecia	48
I.3.3.1.a. Orígenes.....	49
I.3.3.1.b. Usos	54
I.3.3.2. La conciencia documental y el cambio de actitud hacia la escritura	60
I.3.3.2.a. Funciones de la escritura	60
I.3.3.2.b. El libro.....	63
I.3.3.2. Codificación visual	66

I.3.3.2.a. Filología y metatexto	66
I.3.3.2.b. Composición visual del texto	69
II. METODOLOGÍA	73
II.1. BÚSQUEDAS DE INFORMACIÓN Y BÚSQUEDAS SEMÁNTICAS	77
II.1.1. <i>Criterios de clasificación</i>	78
II.1.2. <i>Búsqueda directa</i>	79
II.1.2.1. La lectura	80
II.1.2.2. Búsqueda de coincidencias	82
II.1.3. <i>Búsqueda mediada</i>	84
II.1.3.1. Reducción	84
II.1.3.1.a. Reducción no invasiva	85
II.1.3.1.b. Reducción invasiva	91
II.1.3.2. Reconstrucción	93
II.1.3.2.a. Reconstrucción textual	94
II.1.3.2.b. Reconstrucción visual	96
II.1.3.2.c. Relación	101
II.2. PROCESAMIENTO DE INFORMACIÓN ASISTIDO POR ORDENADOR	107
II.2.1. <i>El ordenador</i>	107
II.2.1.1. Estructura física: <i>hardware</i>	108
II.2.1.2. Funcionamiento: <i>software</i>	109
II.2.2. <i>El texto digital</i>	111
II.2.2.1. Código Beta y Unicode	112
II.2.3. <i>Expresiones regulares</i>	113
II.2.4. <i>Lenguajes descriptivos</i>	117
II.2.4.1. Historia	118
II.2.4.2. Ventajas	119
II.2.4.3. <i>Extensible Markup Language</i>	119
II.2.4.4. Funcionamiento	122
II.2.4.5. El documento <i>XML</i>	123
II.2.4.6. El esquema <i>XML</i>	124
II.2.4.7. La Text Encoding Initiative	126
II.2.5. <i>Lenguajes procedimentales</i>	127
II.2.5.1. Historia y definición	127
II.2.5.3. <i>R</i>	130
II.3. MÉTODOS DE BÚSQUEDA SEMÁNTICA POR ORDENADOR	131
II.3.1. <i>Análisis semántico latente</i>	131
II.3.1.1. Topic modeling	132

II.3.2. <i>Ontologías</i>	133
II.4. CONCLUSIÓN. METODOLOGÍA ELEGIDA	136
III. ANÁLISIS DEL TEXTO	137
III.1. LAS <i>HISTORIAS</i> DE HERÓDOTO	139
III.1.1. <i>El autor</i>	140
III.1.2. <i>Heródoto el historiador</i>	142
III.1.3. <i>Importancia de su obra</i>	144
III.2. EL TEXTO DE LAS <i>HISTORIAS</i>	147
III.3. ETIQUETADO <i>XML</i>	150
III.3.1. <i>Digitalización de los textos clásicos</i>	150
III.3.1.1. Heródoto en el <i>TLG</i>	150
III.3.1.2. <i>Perseus</i> y el primer Heródoto en <i>XML</i>	151
III.3.2. <i>la normativa TEI aplicada a las Historias</i>	152
III.3.2.1. Objetivo y utilización de la normativa. Un marco abierto a la personalización.....	154
III.3.2.2. Nuestra propuesta: descripción de <i>estructuras</i> y <i>entidades</i>	156
III.3.3. <i>Descripción de estructuras</i>	157
III.3.3.1. Estructura y contenido de las <i>Historias</i>	157
III.3.3.3. Estructuras simples	163
III.3.3.3.a. El sistema CTS-URN.....	163
III.3.3.3.b. Recursos XML para describir estructuras simples	164
III.3.3.4.a Recursos XML para describir estructuras complejas	173
III.3.4. <i>Descripción de entidades</i>	189
III.3.4.1. <i>Personas</i>	190
III.3.4.1.a. Nombres propios	190
III.3.4.1.b. Nombres de «organizaciones»	191
III.3.4.1.c. Datos biográficos y prosopográficos.....	191
III.3.4.2. <i>Lugares</i>	192
III.3.4.3. <i>Relaciones entre entidades</i>	193
III.4. <i>TOPIC MODELS</i>	197
III.4.1. <i>Topic modeling en Humanidades</i>	197
III.4.2. <i>Topic modeling de Tucídides</i>	198
III.4.3. <i>Topic modeling de las Historias</i>	198
III.4.3.1. Preparación del texto.....	198
III.4.3.2. Resultados	200
IV. CONCLUSIONES	201
LISTA DE ABREVIATURAS	205

BIBLIOGRAFÍA.....	207
V. ANEXOS	247
V.1. TOPIC MODELS.....	249
V.1.1. <i>Lista de stopwords</i>	249
V.1.2. <i>Resumen por tema</i>	255
Tema 1	255
Tema 2	256
Tema 3	257
Tema 4	257
Tema 5	257
Tema 6	258
Tema 7	258
Tema 8	258
Tema 9	259
Tema 10	259
Tema 11	259
Tema 12	260
Tema 13	260
Tema 14	260
Tema 15	261
V.1.3. <i>Representaciones gráficas</i>	261