

Eveling Garzón Fontalvo (coord. ed.), *Verba notata. Retos de anotación en las lenguas clásicas*, Madrid, CSIC, 2023, 158 pp. ISBN: 978-84-00-11211-0, e-ISBN: 978-84-00-11212-7

Lucía Madrigal Acero

Universidad Complutense de Madrid ✉

<https://dx.doi.org/10.5209/cfcg.98091>

La anotación lingüística de *corpora* extensos de tal manera que permita la recuperación de información sintáctica y semántica con precisión es uno de los grandes retos tanto de la lingüística de corpus como de la lingüística computacional. Las numerosas herramientas que hoy existen se han elaborado con el asiduo trabajo de un extenso número de investigadores, puesto que uno de los principales obstáculos que plantea este tipo de trabajo es la imposibilidad de mecanizar el proceso en su totalidad. A esto se suman las dificultades que se derivan de las diferencias entre las lenguas humanas y los lenguajes informáticos. Es de sobra conocido que una de las principales limitaciones de muchas bases de datos y *treebanks* —que recogen información morfológica, sintáctica y semántica en forma de análisis arbóreo— es su imposibilidad de representar la ambigüedad lingüística, ya que la inclusión de varios análisis diferentes para un mismo segmento de texto requeriría duplicar datos y, por lo tanto, falsear las estadísticas y complicar las búsquedas. Además, cada nuevo proyecto se asienta sobre unos objetivos y fundamentos teóricos concretos que condicionan los análisis que se ofrecen, lo que a su vez da lugar a incompatibilidades entre distintas bases de datos.

El monográfico coordinado y editado por Eveling Garzón Fontalvo, con prólogo de Esperanza Torrego Salcedo, recoge las soluciones propuestas para diversos problemas que surgieron durante el diseño de una nueva herramienta, COMREGLA, destinada al análisis lingüístico de textos en griego y en latín. Se articula en seis capítulos, de los cuales el primero, escrito por Cristina Tur y Berta González Saavedra, sirve como introducción, estado de la cuestión y presentación del proyecto COMREGLA.

En una primera parte de este capítulo, se presenta de manera narrativa un recorrido por los principales recursos digitales de los que se ha servido la lingüística latina y griega, desde el *Index Thomisticus* hasta el *Universal Dependencies Treebank*, con un comentario crítico de sus propiedades y sus limitaciones. En una segunda parte, se presenta el origen de COMREGLA, una base de datos lingüísticos en latín y griego concebida como una versión más amplia y homogénea a nivel teórico de la base de datos de REGLA (*Rección y Complementación del Griego y el Latín*) en XML, con modificaciones tales respecto de la versión anterior que facilitan su compatibilización con otros recursos digitales como los presentados en la sección inmediatamente anterior. La principal diferencia entre REGLA y COMREGLA es que la primera divide el análisis lingüístico en dos niveles, palabras (*words*) y oraciones (*sentences*), mientras que en COMREGLA se añade un tercero, la predicción (*clause*). Esta división se amolda mejor al marco teórico de la Gramática Funcional, que distingue tres niveles de complementos verbales según su nivel de integración en la oración, a saber, argumentos, adjuntos y disjuntos. Esta solución abre la puerta a la anotación de información pragmática, además de la información sintáctica y semántica que ya incluía la primera versión de REGLA.

El segundo capítulo, “Entre la oración y la palabra la predicación como unidad de análisis”, por Cristina Tur, profundiza en la justificación teórica y práctica para la adición de este tercer nivel de análisis en COMREGLA. Los *corpora* se pueden dividir de manera automatizada en palabras, esto es, las unidades de texto comprendidas entre espacios, y oraciones, las secciones de texto comprendidas entre puntos. Esta división se llama *tokenización* y se puede automatizar. El nivel de la predicación se encuentra entre estos dos niveles y, al carecer de una delimitación ortográfica clara, resulta mucho más difícil de automatizar. En un esquema arbóreo, el análisis sintáctico de una oración se traslada gráficamente a un esquema con divisiones bimembres, lo que hasta la fecha se ha considerado la forma más eficiente de almacenar y recuperar la información dadas las características de los lenguajes informáticos. No obstante, este compromiso práctico entra en conflicto con los conceptos teóricos más aceptados en la actualidad, principalmente con la anotación de disjuntos y la representación de la subordinación.

En los principales *treebanks* se anotan en el mismo nivel los argumentos y los adjuntos que los disjuntos, que se encuentran fuera de la predicación y por lo tanto afectan al conjunto de la misma y no tan solo al verbo. Por lo tanto, la decisión de incluir un tercer nivel de anotación permite jerarquizar los componentes de la oración en COMREGLA y distinguir niveles de integración sintáctica en los complementos del verbo. La creación del nivel de la predicación permite además atribuirle características semánticas que no pertenecen ni al nivel superior ni al inferior, a saber, control, aspecto léxico, dominio léxico, fuerza ilocutiva y polaridad, es decir, permite recuperar información adicional.

Eveling Garzón Fontalvo analiza las “Estructuras comparativas en griego antiguo y latín”. En las primeras páginas del capítulo se presentan los componentes de las estructuras comparativas y su clasificación de acuerdo con los elementos que se pueden elidir en ellas. Es precisamente la cuestión de la elisión la que plantea mayores retos para la anotación digital y no hay homogeneidad entre las soluciones propuestas por distintos recursos. En COMREGLA, para salvar este escollo, se proponen tres etiquetas semánticas —grado, base de comparación y estándar de comparación— para marcar los elementos que forman parte de una estructura comparativa. Cuando el segundo término de la comparación es una predicación, se crea un “nodo vacío” (pp. 47, 72) y se establece un análisis independiente para esta, mientras que en los casos de elipsis se hace necesario reconstruir las palabras elididas y etiquetarlas como “elípticas” (p. 73). Respecto al grado, cuando este se marca de manera sintética, se anota en el nivel del análisis morfológico de la palabra.

En el capítulo sobre “La numeración argumental y la diátesis”, Iván López Martín discute la decisión de algunas bases de datos de utilizar etiquetas numéricas para los argumentos de la predicación de acuerdo con la jerarquía de las funciones semánticas. En algunos casos, se combina una numeración incremental de los argumentos con la asociación de un valor numérico a una función semántica, lo que da lugar a discontinuidades que se pueden considerar incoherentes en la numeración, p. ej. asignar a un verbo un argumento 0 y 4 sin que medien los argumentos 1, 2 y 3. En otras ocasiones, los argumentos se numeran de acuerdo con su probabilidad de aparición, lo que dificulta la recuperación de información sintáctica o semántica. En REGLA se adoptó también un sistema numérico de anotación de los complementos, de donde surgieron dificultades para sistematizar los cambios diatéticos y otras construcciones como los acusativos internos. En COMREGLA, para solucionar estas ambigüedades y facilitar la recuperación de información, se hace la propuesta de anotar los complementos atendiendo a tres categorías: nivel —argumento, satélite, núcleo o colocativo—, función sintáctica y función semántica. Esta información se puede recuperar y filtrar por medio de búsquedas cruzadas.

Guillermo Salas Jiménez, en “La anotación del aspecto léxico en latín”, señala, al igual que los capítulos anteriores, los diferentes enfoques que han adoptado distintas bases de datos para la anotación del aspecto, bien gramatical o bien léxico, en el caso de que lo anoten. La solución de REGLA adoptaba categorías de la Gramática Funcional que omitían rasgos léxicos relevantes como la duración y la telicidad. En cambio, en COMREGLA se usan las categorías de Vendler de estado, actividad, logro y realización, que a juicio del autor evitan las ambigüedades que se dan en otros sistemas de anotación y facilitan la organización de la información. Además, el aspecto

léxico ya no se anota en el nivel de las palabras, sino en el nuevo nivel de la predicación, lo que es más coherente con el estado actual de la cuestión de acuerdo con el cual el aspecto léxico es una cualidad de la predicación y no de la palabra.

El siguiente capítulo ("La expresión morfológica de las funciones semánticas: otro enfoque sobre las relaciones", de Berta González Saavedra) presenta las nuevas etiquetas para la anotación de las funciones semánticas que incluye COMREGLA y aporta un estudio de caso sobre las funciones semánticas de las preposiciones *ab* en latín y *ἐκ* en griego y su expansión hacia otros campos semánticos, apoyándose en los datos recuperados de la base de datos. A continuación, traza los mapas semánticos para ambas preposiciones y comprueba sus similitudes y diferencias. Al final, concluye que la naturaleza semántica de los verbos a los que complementan las preposiciones y los sustantivos que rigen se correlacionan con los cambios semánticos que estas experimentan.

Por último, Alberto Pardal Padín ("Los sintagmas de *περί* + genitivo como expresión de la referencia") aborda los problemas de sistematización de estructuras lingüísticas ambiguas como los sintagmas de *περί* + genitivo, que pueden funcionar como argumentos, adjuntos o disjuntos en distintos contextos, incluso cuando van regidos por el mismo tipo de verbos. Al centrarse en el análisis de las predicaciones, COMREGLA permite una jerarquización de argumentos, adjuntos y disjuntos. Si esta estructura se replica a nivel textual, se podría introducir en la base de datos un análisis pragmático del texto.

Entre las muchas virtudes de este libro, se encuentra destacadamente su coherencia interna, puesto que, a pesar de la disparidad en los temas de análisis de cada capítulo, el hilo conductor es muy claro y destaca el argumento central de la monografía: los beneficios que tiene añadir un tercer nivel para la predicación en bases de datos digitales dedicadas a la recopilación de información sintáctica y semántica. Todo ello viene, además, ilustrado con abundancia de ejemplos traducidos y comentados, esquemas y tablas, que facilitan la visualización de las partes más abstractas de la argumentación.

La labor editorial en esta publicación resulta casi impecable, con un índice de materias y de recursos digitales muy necesario para quienes se inicien en las Humanidades Digitales. A modo de crítica constructiva, quizá falte una columna en las tablas de la p. 66 indicando el tipo de oración comparativa al que se refiere cada ejemplo, pero esto se debe probablemente a las limitaciones de espacio en la página. Igualmente, las erratas son casi inexistentes, con un «tal y como ha se ha descrito» (p. 25), «nivel tectogramatical» (p. 25) y la omisión del verbo en «en griego, este ES el caso de adjetivos y adverbios». Quizá la más llamativa, es «M.^a D Dolores Jiménez Martínez» (p. 12), a quien se atribuye la codirección de varios proyectos de investigación junto con José Miguel Baños, en realidad llamada M^a. Dolores Jiménez López.

Verba notata no solo es una lectura instructiva para quienes deseen iniciarse en la anotación de textos digitales, sino que también presenta una propuesta muy lúcida para resolver distintos problemas a los que ya se han enfrentado numerosos investigadores al centrar su análisis en un nuevo nivel, el de la predicación.