

Manuel Márquez

UN MODELO ESTRUCTURAL DE DATOS LEXICOGRÁFICOS PARA LA CODIFICACIÓN EN XML DE UN DICCIONARIO DE APRENDIZAJE DE LATÍN: LA DTD DE LOS LEMAS VERBALES

Keywords Computational lexicography; DTD; XML; learner's dictionaries

La evolución tecnológica ha contribuido notablemente a la mejora de la praxis lexicográfica, especialmente – pero no solo – durante las dos últimas décadas. La versatilidad que proporciona, por ejemplo, la codificación de los datos lexicográficos en un formato XML permite tanto su libre transferencia como su análisis y gestión por diferentes aplicaciones sin necesidad de modificar dichos datos. Asimismo, el desarrollo de los Dictionary Writing Systems (DWS) facilita el proceso de confección de un diccionario, toda vez que proporcionan un entorno de trabajo adecuado y productivo con vistas al almacenamiento, la gestión y la explotación de los datos lexicográficos (Rubio López et al. 2021; Woldrich et al. 2020; Simões et al. 2019; Abel 2012). No obstante, la tecnología no es el único pilar que sustenta la confección de una obra lexicográfica. Es necesario que unos principios teóricos sólidos fundamenten la estructura de los datos con los que se trabaja.

Teniendo en cuenta lo expuesto en las líneas precedentes y la experiencia adquirida durante cinco proyectos de innovación educativa dedicados a la confección y al testado de dos diccionarios de aprendizaje de lenguas declinables, nos planteamos la posibilidad de mejorar la explotación de este tipo de diccionarios nivelando la información que proporcionan.

Afrontamos el reto de diseñar una DTD que define una estructura de datos lexicográficos codificados en un formato XML que permite una codificación de los contenidos fundamentales de un diccionario de aprendizaje de lenguas (Heuberger, 2018) al tiempo que una explotación variada de esos datos en atención al nivel cognitivo del usuario. El prototipo diseñado reutiliza una selección de los datos básicos de un diccionario electrónico de latín-español de iniciación al estudio de dicha lengua ya validado, reorganizando y añadiendo nueva información codificada en XML. La DTD diseñada permite validar la estructura y la sintaxis del documento para que tanto las búsquedas como la información devuelta sean correctas.

Como logros del diseño y de su posterior experimentación se citan los siguientes:

- 1) Crear un modelo estándar de codificación de datos lexicográficos que permita una explotación nivelada, complementando la DTD de diccionarios de TEI.
- 2) Explotar los datos XML mediante el interfaz de un único diccionario electrónico que resulte efectivo para responder a las diferentes necesidades cognitivas del alumnado de latín (atención a la diversidad) en un mismo contexto educativo.
- 3) Facilitar diferentes ritmos de aprendizaje mediante el uso de los datos lexicográficos.
- 4) Diseñar un sistema de estructura de datos que, una vez testado, pueda ser aplicado a otras lenguas, declinables y no declinables, como el griego (clásico y moderno), el alemán y el inglés.

Presentamos en este trabajo un fragmento de una DTD, concretamente aquel que corresponde al tratamiento de los verbos, unidad léxica principal de la estructura lexicográfica. Dicha DTD viene a complementar y mejorar la propuesta de TEI (<https://tei-c.org/release/doc/tei-p5-doc/es/html/DI.html>), por cuanto potencia la medioestructura del diccionario resultante (estableciendo vínculos entre verbos que pertenecen a una misma clase) y nivela la información relativa a cada verbo: clase de verbo (se sigue, a grandes rasgos, la clasificación proporcionada por ADESSE: <http://adesse.uvigo.es/data/clases.php>), significados, complementación verbal (descripción de su valencia cuantitativa y cualitativa, esta última en atención a la descripción morfológica de los argumentos verbales y a su ontología), colocaciones, etimología y ejemplos. La cantidad de información y su grado de complejidad se incrementa según el nivel de conocimientos lingüísticos del estudiante: n1 (básico), n2 (medio), n3 (avanzado)

El diseño de la DTD se fundamenta en los siguientes principios lingüísticos: la Gramática Dependencial de Tesnière (1959) en lo que se refiere a la descripción de la valencia verbal; la Lingüística Funcional de Dik (1997) en cuanto a los roles semánticos que desempeñan los argumentos verbales (agente, procesado, fuerza, meta...); por último, una ontología adaptada de las entidades de primer y segundo orden de Lyons (1980) utilizada para la descripción de la caracterización léxica de los argumentos verbales (+/- animado, +/- humano, +/- concreto, lugar). En cuanto a los principios lexicográficos, entendemos que, en el ámbito del aprendizaje de lenguas, la Teoría Funcional de la Lexicografía proporciona un marco teórico idóneo por cuanto establece unas bases teórico-prácticas adecuadas para determinar las fases de confección del diccionario y la estructura de la información codificada en atención a los diferentes tipos de necesidades de los potenciales usuarios de los datos lexicográficos (Tarp 2013). Por último, confirmamos que está previsto volcar la estructura XML en *Lexonomy*, un DWS que permite la confección y publicación de diccionarios electrónicos reduciendo a mínimos las habilidades técnicas requeridas en sus usuarios, con el objetivo de facilitar en un futuro la ampliación de los datos por parte del equipo de lexicógrafos y la apertura del sistema en forma de diccionario electrónico para su libre consulta.

Bibliografía

- Abel, A. (2012): Dictionary Writing Systems and beyond. In: Granger, S./Paquot, M. (eds.): *Electronic lexicography*. Oxford, p. 83–106.
- Dik, S. C. (1997): *The theory of functional grammar. Part 1: The structure of the clause*. Berlin/New York.
- Heuberger, R. (2018): Dictionaries to assist teaching and learning. In: Fuertes Olivera, P. A. (ed.): *The Routledge handbook of lexicography*. New York, p. 300–316.
- Lyons, J. (1980): *Semántica*. Barcelona.
- Měchura, M. (2017): *Introducing lexonomy: an open-source dictionary writing and publishing system*. <https://www.lexonomy.eu/docs/elex2017.pdf> (last access: 22-03-2022).
- Rubio López, R. J./Bonilla Huérfano, J. E./Bernal Chávez, J. A. (2021): Dictionary Writing Systems y otras herramientas informáticas para la elaboración, administración y publicación de diccionarios. In: *Lingüística y Literatura* 80, p. 340–360.
- Simões, A./Salgado, A./Costa, R./Almeida, J. J. (2019): LeXmart: a smart tool for lexicographers. In: Kosem, I. et al. (eds.): *Electronic Lexicography in the 21st Century: Proceedings of the eLex 2019 Conference: Smart Lexicography*. Sintra, p. 453–466.

Tarp, S. (2013): Necesidad de una teoría independiente de la lexicografía: el complejo camino de la lingüística teórica a la lexicografía práctica. In: *Círculo de Lingüística Aplicada a la Comunicación* 56, p. 110–154.

Tésnière, L., (1959): *Éléments de syntaxe structurale*. Paris.

Woldrich, A./Goli, T./Kosem, I./Matuška, O./Wissik, T. (2020): ELEXIS: Technical and social infrastructure for lexicography. In: *K Lexical News* 28, p. 45–52.

Información del contacto

Manuel Márquez

Universidad Complutense de Madrid

manmarqu@ucm.es

Agradecimientos

Agradecemos al grupo de investigación ILSA de la Universidad Complutense de Madrid (<http://ilsa.fdi.ucm.es/>) su inestimable labor en el campo de la investigación sobre gramáticas formales como herramienta de desarrollo de software, así como sobre los repositorios y colecciones de objetos digitales.