## DESARROLLO DE APLICACIONES PARA LA GENERACIÓN AUTOMÁTICA DEL LENGUAJE: LOS RECURSOS DEL PORTAL LEXICOGRÁFICO PORTLEX

Tal y como indica el título, el presente número monográfico presenta diferentes recursos alojados en el portal lexicográfico *PORTLEX*. A lo largo del volumen se pueden encontrar referencias al diccionario multilingüe de la valencia del nombre *Portlex*, el cual sirvió de catalizador para el desarrollo de nuevos proyectos. Los trabajos aquí compilados, no obstante, ponen el foco en la descripción de los fundamentos teóricos y metodológicos, así como en las técnicas y herramientas aplicadas para el desarrollo de recursos plurilingües de análisis y generación automática del lenguaje natural con aplicación lexicográfica. Se trata, en concreto, de los prototipos *Xera*, *XeraWord*, *Combinatoria* y *CombiContext*, los cuales generan automáticamente ejemplos a partir de parámetros de consulta sintácticos-semánticos. Por tanto, estos simuladores ofrecen información sobre el potencial combinatorio de sustantivos valenciales creando automáticamente ejemplos dinámicos con diferentes finalidades, entre ellas la ejemplificación lexicográfica. Estos ejemplos generados por los propios usuarios muestran el vocabulario que puede ocupar determinadas casillas funcionales. De este modo, dichos recursos aportan combinaciones tanto en el eje sintagmático como paradigmático. El acceso en la interfaz de usuario es, dependiendo del recurso, semasiológico u onomasiológico.

El público encontrará además detalladas descripciones de herramientas esenciales para el análisis y la extracción de datos de WordNet, el método seguido para la anotación semántico-ontológica de los datos recogidos, así como la propia ontología diseñada para la finalidad de los proyectos en los que se enmarcan los recursos. Destacan aquí, junto con los generadores, herramientas como *Combina*, *Lematiza*, *Ontología léxica*, *TraduWord* y el etiquetador en desarrollo ESMAS-ES$^{+}$, todos ellos detalladamente explicados en la monografía. En el volumen también se presta

atención a resultados ligados a la aplicación de redes neuronales y métodos predictivos para el análisis de la similitud semántica y las coocurrencias.

Junto con los temas señalados, la monografía aporta aproximaciones teóricas al estudio valencial de la frase nominal, discute dificultades en el análisis lingüístico y computacional para desarrollar herramientas de procesamiento del lenguaje natural y propone el uso de los generadores del lenguaje en la enseñanza de lenguas junto con modelos concretos para su aplicación.

El volumen se articula en los siguientes artículos y temas:

El estudio *LA AVENTURA DE LOS GENERADORES AUTOMÁTICOS DEL LENGUAJE NATURAL: DEL ANÁLISIS LINGÜÍSTICO AL PROCESAMIENTO AUTOMÁTICO DE DATOS* de María José Domínguez Vázquez dota al volumen de una descripción general de los generadores desarrollados en el portal *PORTLEX* y sus fundamentos. De este modo, se describe el marco teórico, la gramática y lexicografía de valencias, así como las características generales de los cuatro generadores automáticos del lenguaje y la tipología de datos que ofrecen.

Natalia Catalá Torres aporta en *SOBRE LA ESTRUCTURA DE LOS SINTAGMAS NOMINALES* una aproximación teórica amplia sobre el sintagma nominal y la estructura argumental, abordando los tipos y propiedades de nominalizaciones, los sustantivos no derivados y los sintagmas nominales en los proyectos *MultiGenera* y *MultiComb*.

Una visión de conjunto de diferentes recursos y herramientas manejadas en diferentes estadios de desarrollo de los simuladores de generación, así como las técnicas y estrategias aplicadas se encuentra en *GUÍA DE TÉCNICAS, ESTRATEGIAS Y HERRAMIENTAS EN EL DISEÑO Y DESARROLLO DE GENERADORES AUTOMÁTICOS DEL LENGUAJE* de Daniel Bardanca Outeiriño y María José Domínguez Vázquez. El estudio también refleja diferentes fases metodológicas hasta alcanzar el objetivo final, la generación de ejemplos dinámicos y esquemas argumentales del nombre anotados semánticamente.

Uno de los recursos centrales, diseñado *ad hoc* para el desarrollo de los simuladores de generación automática del lenguaje, es la ontología léxica *bottom up*. Rosa María Martín Gascueña ofrece en su investigación *DISEÑO DE UNA ONTOLOGÍA DE SEMÁNTICA LÉXICA EN LOS PROYECTOS MULTIGENERA Y MULTICOMB* un estudio detallado de las ontologías de WordNet, elemento clave en el desarrollo de los prototipos de generación. La autora presenta las diferentes fases en el desarrollo de la ontología léxica propia aplicada en todos los simuladores para el etiquetado semántico.

El estudio de Carlos Valcárcel Riveiro y Laura Pino *HERRAMIENTAS Y DIFICULTADES EN EL ANÁLISIS DEL GRUPO NOMINAL EN FRANCÉS PARA SU PROCESAMIENTO COMPUTACIONAL* revisa el trabajo desarrollado para la lengua francesa en los diferentes proyectos del portal *PORTLEX*, como ejemplo extrapolable a las otras lenguas incluidas en los recursos. Se presentan los resultados obtenidos y se hace un análisis detallado de las herramientas utilizadas por los equipos de trabajo para el análisis, la extracción y el procesamiento de datos en francés atendiendo tanto a sus funcionalidades como a sus limitaciones.

Cierra el volumen el trabajo de Nerea López Iglesias *MUCHO MÁS QUE EJEMPLOS: APLICACIONES DIDÁCTICAS DE LOS GENERADORES AUTOMÁTICOS*, el cual incide en la importancia del contexto en el desarrollo de la competencia léxica. La autora describe cómo los ejemplos generados automáticamente por los generadores pueden ser de aplicación directa en el aula de lenguas extranjeras, pero también propone actividades concretas en línea diseñadas a partir de los datos que ofrecen los prototipos de generación automática.

Un hilo conductor de todos los trabajos es la importancia concedida a la interoperabilidad y sostenibilidad. De este modo, los prototipos de generación automática del lenguaje se retroalimentan entre sí y sus datos son integrables en otros recursos. Otro ejemplo de ello es el desarrollo del nuevo recurso compilado en el portal lexicográfico, el etiquetador plurilingüe semántico y

automático ESMAS-ES+, actualmente en elaboración. Este bebe de las fuentes de los datos lingüísticos anotados semánticamente, de las herramientas diseñadas para la generación automática del lenguaje y de la traducción automática del caudal léxico. Algunas de las optimizaciones de los recursos de generación automática presentados resultan de la investigación al abrigo de ESMAS-ES+.

Las diferentes herramientas, recursos y simuladores se han desarrollado u optimizado al abrigo de diferentes proyectos competitivos:

- *MultiGenera. Generación multilingüe de estructuras argumentales del sustantivo y automatización de extracción de datos sintáctico-semánticos*. Fundación BBVA. Ayudas Fundación BBVA a Equipos de Investigación Científica - Humanidades Digitales. 2017-2020. http://portlex.usc.gal/multigenera/
- *MultiComb. Generador multilingüe de estructuras argumentales del sustantivo con aplicación en la producción en lenguas extranjeras*. FI2017-82454-P: Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia, Generación de Conocimiento. MCIN/AEI/ FEDER "Una manera de hacer Europa" (EXCELENCIA 2017, 2017-PN091). 2018-2021. http://portlex.usc.gal/multicomb/
- *Ferramentas TraduWord e XeraWord: tradución de caudal léxico e xeración automática da linguaxe natural en galego e portugués*. 2020-PU004. Convocatoria proyectos de colaboración. Universidade de Santiago de Compostela. https://ilg.usc.gal/xeraword/
- *Etiquetador semántico multilingüe automático y sostenible*. ESMAS-ES+. PID2022-137170OB-I00: Programa Estatal para Impulsar la Investigación Científico-Técnica y su Transferencia, del Plan Estatal de Investigación Científica, Técnica y de Innovación 2021-2023. Generación de Conocimiento. MCIN/AEI//FEDER "Una manera de hacer Europa". 2023-2027.

A su vez, los resultados de investigación han sido propiciados, son objeto de estudio o son aplicados por el grupo de investigación Humboldt (Grupo GI 1920, Universidad de Santiago de Compostela), el grupo de innovación docente MeReLing (Universidad de Vigo) y el Instituto da Lingua Galega (Universidad de Santiago de Compostela), entre otros.

Nuestro más sincero agradecimiento a las instituciones que apoyan nuestro trabajo, a los evaluadores y las evaluadoras por sus contribuciones, así como a la revista por permitirnos presentar los resultados de nuestra investigación en este foro.

Los editores