

Título: Poesía, interoperabilidad y estándares para el tratamiento de datos poéticos. Una aproximación desde el proyecto POSTDATA.

Autores: Elena González-Blanco, Helena Bermúdez, Mariana Malta, Gimena del Rio, Diego Ferreyra, Clara Martínez, Pablo Ruiz y el resto del equipo de POSTDATA.
Modera el panel Pablo Ruiz Fabo.

Este panel integra y agrupa varias aproximaciones tecnológicas al tratamiento computacional de la poesía, que se suman y se trabajan en paralelo dentro del proyecto POSTDATA (www.postdata.linhd.es), un proyecto financiado por la Comisión Europea a través de su programa ERC Starting Grant, que busca la interoperabilidad de datos poéticos procedentes de diferentes fuentes, tradiciones y corpora mediante la utilización de estándares tecnológicos: los datos enlazados y la creación de un perfil de aplicación para el análisis poético; la utilización de estándares existentes en el ámbito de las Humanidades Digitales, como el TEI y la automatización de los procesos de análisis de los textos y específicamente de la poesía, gracias a la utilización combinada de las tecnologías del lenguaje, el aprendizaje automático y el análisis estadístico de los textos.

El panel consiste en una presentación de los resultados obtenidos hasta el momento mediante 1) la presentación del proceso de construcción del modelo de datos poéticos obtenido tras realizar el análisis de veinte (20) bases de datos de proyectos dedicados a poesía, y que servirá como base para la creación de un perfil de aplicación de metadatos; 2) la recolección y organización de los vocabularios controlados que cada una de las diferentes tradiciones poéticas, lingüísticas y culturales utiliza para nombrar y clasificar su información, gracias al uso de una herramienta común de gestión de estos vocabularios (TemaTres); 3) el análisis de los problemas específicos del modelado de la rima, y 4) los nuevos retos del proyecto y las líneas que están en marcha, como la cuestión de la transformación de los datos en XML/TEI para datos en RDF teniendo como base el modelo semántico del perfil de aplicación de metadatos, o la definición de los algoritmos que configurarán el "PoetryLAB", espacio de análisis y experimentación con PLN y tecnologías del lenguaje que buscan automatizar el proceso de análisis poético.

1) Título: *Validando un modelo de dominio para la poesía europea*

Autoría: Mariana Curado Malta (2,1), Helena Bermúdez Sabel (1), Elena González-Blanco García (1)

Filiación: (1) Laboratorio de Innovación en Humanidades Digitales (LINHD) - Universidad de Educación a Distancia, (2) CEOS.PP - Politecnico do Porto, Portugal

Contacto: mariana@iscap.ipp.pt, helena.bermudez@linhd.uned.es, egonzalezblanco@flog.uned.es

Si intentáramos hacer un estudio completo de cómo se construyen las tradiciones literarias europeas, obtendríamos una compleja red de relaciones entre las diferentes comunidades lingüísticas. Estas interacciones son tan importantes a la hora de analizar cualquier producto cultural que las comparaciones entre diferentes tradiciones literarias no están limitadas a la disciplina de la Literatura Comparada.

Los avances tecnológicos han facilitado el acceso al patrimonio cultural mundial. Hoy en día es posible consultar con la máxima facilidad la producción poética Húngara más antigua (Horváth et al. 1992-) o las canciones profanas latinas más tardías (Stella et al, 2011). Lo que (aún) no es posible es contrastar en una única búsqueda, por ejemplo, qué *leitmotifs* latinos se siguen desarrollando en la tradición húngara. Este es uno de los retos que el proyecto Poetry Standardization and Linked Open Data (POSTDATA)¹ quiere superar.

Uno de los objetivos del proyecto es el de proporcionar a la comunidad de práctica de la poesía europea los recursos necesarios para que puedan publicar sus datos como Datos Abiertos Enlazados (LOD por sus iniciales en inglés) con el fin de incrementar la interoperabilidad dentro de esta comunidad. Para poder publicar datos interoperables como LOD, es necesario desarrollar un Perfil de Aplicación de Metadatos (PAM) (Nilsson, Baker, y Johnston 2009), proceso en el cual se encuentra POSTDATA en la actualidad.

Coyle y Baker (2009) definen un PAM como un constructo genérico para diseñar registros de metadatos. Para el desarrollo de un PAM para la poesía europea se están siguiendo sistemáticamente las actividades definidas por Me4MAP (Curado Malta y Baptista n.d.). Una de estas actividades se marca como objetivo la definición de un modelo de dominio, un modelo conceptual común que debe representar las necesidades informativas de la comunidad de práctica de la poesía europea. Para extraer estas necesidades, se tiene que definir los requisitos funcionales de la comunidad, así como analizar los modelos de datos de los repositorios digitales ya existentes. Además, se

¹ <http://postdata.linhd.es/>

evalúan los intereses de los usuarios de esas bases de datos cuya opinión queda recogida a través de una encuesta².

Para poder tener la certeza de que las necesidades de la comunidad están siendo consideradas, todo el proceso de análisis de los modelos de datos de los repertorios digitales es realizado con la ayuda de las personas que están al cargo de dichos repertorios. Además, es necesario establecer vías de contacto que permitan una comunicación fluida entre los agentes que desarrollan el PAM y la comunidad a la que pretenden representar.

La versión 0.1 del modelo de dominio fue validada en Madrid un taller al que asistieron representantes de las diferentes bases de datos utilizadas para construir el modelo de dominio³, todos ellos filólogos y ninguno con experiencia previa en modelación semántica. Esta circunstancia hizo que, en cierta manera, se ralentizara el proceso de validación. Para superar esta dificultad, el equipo de POSTDATA preparó una comunicación de dos horas que incluyó una presentación de posibles líneas de investigación que la comunidad podría desarrollar una vez que los datos poéticos se publiquen en LOD. Acto seguido, se expusieron los fundamentos de LOD y, por último, el método utilizado para representar el modelo de dominio, esto es, un modelo de datos como un diagrama de clases UML (Fowler 2003). Esta sesión formativa fue de vital importancia para que los participantes pudieran conocer el contexto del proyecto y sus ventajas. Si la comunidad de práctica no se involucra en las fases de desarrollo del modelo de dominio, se dificulta el proceso de aceptación del mismo (véase la idea del *Technological Acceptance Model* desarrollada por Davis 1985).

A partir de los resultados de dicho taller, el equipo de POSTDATA realizó una versión 0.2 del modelo de dominio. Con base en esa versión, se implementó una nueva ronda de análisis de bases de datos y de interfaces Web de proyectos de poesía. Este análisis está actualmente en su etapa final y dará lugar a una V0.3 del modelo de dominio. Este modelo será validado por la comunidad de poesía, a través de un proceso de validación cuya definición está siendo definida.

En esta comunicación se presentará este proceso de desarrollo.

² <http://postdata.linhd.es/limesurvey/index.php/113575>

³ Véase <http://postdata.linhd.es/workshop> para consultar la agenda académica de dicho evento.

REFERENCIAS BIBLIOGRÁFICAS

- Coyle, Karen, y Thomas Baker. 2009. «Guidelines for Dublin Core Application Profiles». <http://dublincore.org/documents/profile-guidelines/>.
- Curado Malta, Mariana, y Ana Alice Baptista. n.d. Me4MAP V1.0: A method for the development of metadata application profiles. Submitted for publication.
- Davis, F. D. (1985). A technology acceptance model for empirically testing new end-user information systems: Theory and results. Massachusetts Institute of Technology.
- Fowler, Martin. 2003. *UML Distilled: A Brief Guide to the Standard Object Modeling Language*. 3 edition. Boston: Addison-Wesley Professional.
- Horváth, I et al. 1992-: Répertoire de la poésie hongroise ancienne, <http://rpha.elte.hu>
- Stella, F. et al. 2011: Corpus Rythmorum Musicum (saec. IV-IX), <http://www.corimu.unisi.it>
- Nilsson, Mikael, Thomas Baker, y Pete Johnston. 2009. «Interoperability Levels for Dublin Core Metadata». <http://dublincore.org/documents/interoperability-levels/>.

2) La formalización del conocimiento en las Humanidades Digitales. Experiencias de integración de vocabularios controlados

Autoría: Gimena del Rio Riande (CONICET, Argentina), Diego Ferreyra (CONICET, Argentina), Elena González-Blanco García (LINHD-UNED)
Presenta: Gimena del Rio Riande

Los vocabularios controlados permiten la formalización de dominios conceptuales y constituyen una herramienta indispensable en el contexto de la comunicación científica. En líneas generales, estas taxonomías construyen infraestructuras socioculturales dedicadas a resolver cuestiones tales como la visibilidad de elementos categorizables (o no categorizables), la comparabilidad semántica, y el alcance y el significado de términos. Así, los vocabularios controlados aportan facilidades para la búsqueda, recuperación, visualización, análisis y representación comparada en cada campo y dominio científico, permitiendo a la comunidad científica valerse de herramientas para optimizar y mejorar las condiciones de representación y descripción de recursos, como para expresar sus líneas de investigación. Este tipo de infraestructuras digitales de información sirven de soporte para actividades de investigación, apelando a estándares y formas persistentes de identificación (servicios web, formatos de metadatos estandarizados estrategias de reutilización de servicios web, enriquecimiento de contenido semántico). Así y todo, en el ámbito de las

Humanidades Digitales, los vocabularios controlados no han sido demasiado explotados.

El presente trabajo trae el ejemplo de tres instancias de explotación e integración de vocabularios controlados en el marco de la colaboración entre el LINHD de la UNED y la línea de Humanidades Digitales del CAICYT (CONICET).

El primer caso que se estudiará es el de la traducción y mejora de la Taxonomía sobre Actividades de investigación digital (TaDIRAH) del consorcio europeo DARIAH y de su integración de en objetos digitales a través de *Popterms* y alimentación de forma distribuida (<http://www.vocabularyserver.com/tadirah/es/index.php>). Estas actividades generaron un insumo útil y polivalente que permite diseminar y definir herramientas y las categorías de análisis y descripción relevantes en el campo científico de las Humanidades Digitales y disponer de una fuente formal de referencia para la formalización de un campo disciplinar, siendo además buen ejemplo de utilización empoderada de medios digitales para favorecer la adopción de marcos teóricos y metodológicos específicos y formalizar un dominio y un marco de análisis teórico determinado.

El segundo caso se circunscribe a la integración del vocabulario controlado para la Poesía Medieval Castellana (<http://www.caicyt-conicet.gov.ar/vocabularios/pmc/>) y el de Historia de Occidente (<http://www.caicyt-conicet.gov.ar/vocabularios/historiaoccidente/>), en el marcado en XML-TEI de los textos editados, e interfaz de distintos proyectos de investigación en Humanidades Digitales: Diálogo Medieval (IIBICRIT-CONICET, Argentina, <http://dialogo.linhd.es/>), Poetriae (UNED, España, <http://poetriae.linhd.es/>) y Obras Completas de José Luis Romero (UDESA; Argentina, <http://jlromero.com.ar/>).

El último caso es el desarrollado para el proyecto europeo POSTDATA (ERC STG-679528, LINHD-UNED, <http://postdata.linhd.es/>). POSTDATA busca comparar las distintas tradiciones poéticas europeas -principalmente, las medievales- a partir del análisis y estudio de objetos digitales de la más diversa naturaleza, utilizando las tecnologías de la web semántica. Desde esta aproximación hace dialogar a objetos digitales (sus objetos de estudio) y los resignifica desde la interoperabilidad semántica, para dar cuenta de los modos en los que la tecnología nos permite detectar patrones métricos, rimáticos y textuales, y arrojar luz sobre las formas de la poesía medieval en un marco pan-europeo, a través de la combinación de vocabularios controlados procedentes de 15 bases de datos en diferentes lenguas y tradiciones poéticas. En este proyecto la utilización de vocabularios controlados apunta a identificar, analizar, recoger y establecer los vocabularios controlados para las diferentes tradiciones poéticas que son objeto de estudio del proyecto, con el fin de lograr una descripción coherente y cohesiva del campo de las tradiciones poéticas, proporcionar un dispositivo capaz de describir y operacionalizar fenómenos en el campo de la poesía. Como resultado final, ofrecerá una taxonomía cohesiva, inclusiva y expresiva que dará cuenta de las principales características de la poesía europea (<http://www.vocabularyserver.com/lab/postdata/>).

REFERENCIAS BIBLIOGRÁFICAS

Beall, J. (2006) Metadata Schemes Points of Comparison. Recuperado Agosto 26, 2007, de : <http://eprints.rclis.org/archive/00005544/01/comparingschemes.pdf>

European Committee for standardization (2005). CWA 15453:2005. Harmonisation of vocabularies for eLearning. Recuperado octubre 20, 2009, de : <http://fire.eun.org/cwa15453-00-2005-Nov.pdf>. Fecha de consulta: 10/10/2008

González-Blanco, E. Martínez Cantón, C., Martos Pérez, M.D. y Del Rio Riande, G. (2014). Una propuesta de integración del sistema de formularios de bases de datos MYSQL con etiquetado TEI: ReMetCa, Repertorio digital de la métrica medieval castellana, Humanidades Digitales: desafíos, logros y perspectivas de futuro. Sagrario López Poza y Nieves Pena Sueiro (editoras), Janus [en línea], Anexo 1 (2014), 209-219, publicado el 11/04/2014, consultado el 20/08/2014. URL: <http://www.janusdigital.es/anexos/contribucion.htm?id=19>

Heery, Rachel; Patel, Manjula. (2000). Application profiles: Mixing and matching metadata schemas, En: Ariadne, (25). Recuperado Agosto 26, 2007, de : <http://www.ariadne.ac.uk/issue25/app-profiles/intro.html>.

International Federation of Library Associations and Institutions. (2009). Guidelines for multilingual thesauri. The Hague: International

Johnson, S. (2003). Sistemas emergentes: o qué tienen en común hormigas, neuronas, ciudades y software. México DF: Ediciones Turner.

Lagoze, C., Lynch, C. A., and Daniel, R. (1996). The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata. Cornell Computer Science Technical Report TR96-1593.

Malta, M., González-Blanco, E., Martínez, C. and del Río, G. (2016). Digital repertoires of poetry metrics: towards a Linked Open Data ecosystem, Proceedings of the First Workshop on Digital Humanities and Digital Curation co-located with the 10th Conference on Metadata and Semantics Research (MTSR 2016) Goettingen, Germany, November 22, 2016 edited by Ernesto William De Luca and Paolo Bianchini en <http://ceur-ws.org/Vol-1764/>.

Méndez Rodríguez, E. M. (2007). Singapur, metadatos y cepillos de dientes. Recuperado octubre 30, 2007, de : <http://www.thinkepi.net/repositorio/singapur-metadatos-y-cepillos-de-dientes/>

Millerand, F., Bowker, G. C. (2007). Metadata Standards: Trajectories and Enactment in the Life of an Ontology. Recuperado Agosto 26, 2007, de :
http://grm.uqam.ca/static/textes/MillerandBowker07Metadata_0.pdf

National Information Standards Organization (U.S.). (2005). Guidelines for the construction, format, and management of monolingual controlled vocabulary. National information standards series. Bethesda, Md: NISO Press. Recuperado octubre 20, 2009, de :
http://www.niso.org/kst/reports/standards/kfile_download?&pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVa86hjP37r_hONsJghRDv2N-zj4TZCh8Dp01rZbmK3O-8vcVjh4hezP

Pérez Agüera, José Ramón Automatización de tesauros y su utilización en la web semántica. En: BiD : textos universitaris de biblioteconomia i documentació (13), 2004. Disponible en <http://eprints.rclis.org/archive/00004176/>. Fecha de consulta: 10/10/2008

Tudhope, Douglas; Alani, Harith; Jones, Christopher. "Augmenting thesaurus relationships: Possibilities for retrieval", Journal of digital information, 2001, n. 3, pp. 123-148.

3) Título: *Acuerdos y desacuerdos. Cómo hacer interoperables conceptos métricos complejos. El caso de la rima*

Autoría: Clara Isabel Martínez Cantón, Helena Bermúdez Sabel, Elena González-Blanco

La búsqueda de una interoperabilidad que dé lugar a un verdadero enriquecimiento del conocimiento es una cuestión ineludible, aunque espinosa, para los estudios literarios y métricos: las relaciones de equivalencia o semejanza entre conceptos de distintas tradiciones poéticas.

Para establecer estas relaciones entre conceptos, en POSTDATA se han analizado, en una primera fase, once bases de datos distintas que modelizan la poesía y sus características formales, pertenecientes a nueve lenguas distintas. En este análisis han ido surgiendo circunstancias diversas que han supuesto un reto a la hora de modelizar y hacer interoperables los campos de las bases de datos. Entre los más relevantes destacamos la diversidad con que cada tradición ha conceptualizado los fenómenos

poéticos (González-Blanco García, Martínez Cantón, Martos Pérez, & del Río Riande, 2014) y la disparidad existente a la hora de modelar conceptos equiparables.

En esta comunicación presentaremos la problemática que se plantea a la hora de crear categorías en un modelo que aspira a representar las necesidades informativas de toda la comunidad que trabaja (y trabajará) con poesía europea. Para ello, nos centraremos en el caso de la rima. Hemos escogido la rima como caso particular porque en métrica siempre ha sido considerada como un fenómeno de gran complejidad que reúne varias funciones. Señalan, por ejemplo, Wellek y Warren:

La rima tiene su mera función eufónica como repetición (o cuasi repetición) de sonidos. [...] Pero, aunque esta faceta pueda ser fundamental, evidentemente sólo constituye un aspecto de la rima. Estéticamente es mucho más importante su función métrica de señalar la terminación de un verso, o bien de organizadora –a veces única- de las formas estróficas. (2002:190)

La tercera función de la rima y, para las autoras, de máxima importancia, es la referente a su significado, es decir, su función semántica.

Esta multifuncionalidad de la rima, señalada por muchos tratadistas hace que haya sido reflejada en las distintas bases de datos con mayor o menor nivel de detalle y según las convenciones propias de cada lengua. A partir del análisis de las once bases de datos se han identificado 28 conceptos relacionados con la rima. En la Figura 1 recogemos únicamente una pequeña muestra en la que se presenta el término escogido para el modelo semántico siguiendo los criterios de: *lingua franca*, neutralidad y eficiencia semántica (Bermúdez Sabel et al. 2017), su definición, las bases de datos en las que está recogida y con qué nombre, y un ejemplo de los valores que en estas bases de datos toma.

Term in model	Definition	BBDD where it is contained	Term in BBDD	Value example
rhymeScheme	A rhyme scheme is the pattern of rhymes at the end of each line of a poem or song. It is usually referred to by using letters to indicate which lines rhyme; lines designated with the same letter all rhyme with each other	Metrique en Ligne (WorkPattern)	SCHEMA	abba
		Metrique en Ligne (StanzaPattern)	SCHEMA	abba
		Versologie (WorkPattern)	rhymeScheme	abab', 'abab', 'aaX', 'aaX'
		MedDB (StanzaPattern)	ESQRIM_TIPO	abbaCC
		CSM (WorkPattern)	rhymecode	AA bbba
		Occitane (StanzaPattern)	SchemaRimique	aaabcccbdeD
		Remetca (StanzaPattern)	formulaRimatica	aabb
clausula	The rhythmic ending of a line determined by the number of unstressed syllables following the last stressed one.	RPHA (StanzaPattern)	AAA	
		CSM (StanzaPattern)	rhyme_type	aguda/grave/esdrúxula
		Metrique en Ligne (LinePattern)	GENRE	m/f
		Metrique en Ligne (Rhyme)	GENRE	m/f
clausulaScheme	Pattern of clausulae	Versologie (LinePattern)	clause	male/female/acatalectic
rhymingLines	Lines that rhyme	Occitane (StanzaPattern)	SchemaGonique	-----f-f
isAssonant	Whether the rhyme of the complete work is assonant (only stressed vowels are repeated)	Corimu (WorkPattern)	RIMA	3,2-3 Baptizatus in Iordani
isUnissonant	Whether all stanzas have the same rhyme	Remetca (WorkPattern)	asonancia	boolean
hasEcho	Some lines repeat the word rhyme. In case this happens the boolean is 1, otherwise, 0.	Remetca (WorkPattern)	unisonancia	boolean
ending	Contents from the last stressed vowel of the rhyme to its end	RPHA (WorkPattern)		boolean
		CSM (Rhyme)	rhymea, rhymeb, rhymec	ia eu ou
		Occitane (Rhyme)	Rimes	ets,e,il,aire,e,ir,an,i,,a,
rhymePhoneme	Phonetic transcription of rhyme	MedDB (Rhyme)	Ri (rima)	ia
rhymeGrapheme	Grapheme of the syllabic nucleus of rhyme	Metrique en Ligne (Rhyme)	PHONEME	a
rhymePrefinalConsonant	Onset of rhyme syllable	Metrique en Ligne (Rhyme)	GRAPHEME	a
rhymeFinalConsonant	Coda of rhyme syllable	Metrique en Ligne (Rhyme)	CONS_SUB	mm
		Metrique en Ligne (Rhyme)	CONS_FIN	a

Figura 1.

En la imagen podemos apreciar cómo hay ciertas categorías que podemos llamar equivalentes, como *rhymeScheme* y que se repiten en muchas bases de datos y cuya única gran diferencia consiste en los valores que este esquema puede tomar. El término *isAssonant* recoge la condición de rima consonante o asonante de un poema o estrofa, y se aplica solamente en Remetca, base de datos de poesía medieval castellana. Encontramos un concepto semejante, que hemos llamado *assonance* en la base de datos CORIMU, que recoge la poesía rítmica latina musicada entre los siglos IV y IX. El carácter sistemático que tiene la rima en castellano y totalmente asistemático y estilístico que presenta en CORIMU hace que no podamos considerarlos un mismo concepto, si bien es interesante relacionarlos. Los términos del modelo *rhymePrefinalConsonant* y *rhymeFinalConsonant*, solo relevantes para la poesía francesa recogida en Metrique en Ligne están ligadas a los grafemas del francés y sus convenciones métricas, pero no parecen ser útiles para ninguna otra base de datos. Además, la combinación de ciertas categorías hace que podamos inferir categorías no explicitadas: la combinación de *rhymeGrapheme* con *rhymePrefinalConsonant* y *rhymeFinalConsonant* constituirían el *ending* recogido en otras bases de datos.

Así, tanto en la rima como en otras categorías, un correcto análisis de datos y su modelado, propone interesantes cuestiones: 1) sobre la necesidad de exhaustividad (esto es, recoger todas las características poéticas existentes), para plasmar sin pérdidas la expresividad de todas las tradiciones; 2) sobre el deseo de simplicidad para incentivar el uso de un modelo común. Esta presentación irá plantear todos estas problemáticas.

Bibliografía

- Bermúdez Sabel, H., Curado Malta, M., & González-Blanco, E. (2017). Towards Interoperability in the European Poetry Community: the Standardization of Philological Concepts. En Gracia, J. et al. (eds.), *Language, Data, and Knowledge: First International Conference, LDK 2017, Galway, Ireland, June 19-20, 2017, Proceedings*. Springer - Lecture Notes in Computer Science (pp. 156-165) DOI: 10.1007/978-3-319-59888-8_14
- González-Blanco García, E., Martínez Cantón, C. I., Martos Pérez, M. D., & del Río Riande, M. G. (2014). Una propuesta de integración del sistema de formularios de bases de datos MYSQL con etiquetado TEI: ReMetCa, Repertorio digital de la métrica medieval castellana. En López Poza, Sagrario & Pena Sueiro, Nieves (Eds.), *Humanidades Digitales: desafíos, logros y perspectivas de futuro* (pp. 209-219). Recuperado a partir de <http://www.janusdigital.es/anexos/contribucion.htm?id=19>
- Wellek, R., & Warren, A. (2002). *Teoría literaria*. (J. M. Gimeno, Trad.). Madrid: Gredos.

4) Título: Retos, perspectivas y próximos pasos dentro de POSTDATA
Presentan: Helena Bermúdez Sabel (1) Pablo Ruiz Fabo (1)

Filiación: (1) Laboratorio de Innovación en Humanidades Digitales (LINHD) - Universidad de Educación a Distancia,
Uno de los objetivos del proyecto Poetry Standardization and Linked Open Data (POSTDATA) es el de proporcionar un modelo semántico de datos a la comunidad de práctica de la poesía europea. Con el fin de potenciar la interoperabilidad entre los datos de esta comunidad, se necesitan establecer los medios necesarios para que estos puedan ser publicados como Datos Abiertos Enlazados (LOD). Debido al alto uso de XML-TEI dentro de la comunidad filológica, desde POSTDATA se ha considerado necesario ofrecer un esquema TEI que represente los conceptos del modelo de dominio, es decir, el modelo conceptual común que recoge todas necesidades informativas de la comunidad de práctica de la poesía europea.

Desde el punto de vista metodológico, el paso de un modelo conceptual a un esquema XML-TEI presenta varios retos. La mayor dificultad no se encuentra en buscar dentro de un vocabulario controlado cuál es el término equivalente a cada uno de los conceptos que conforman el modelo de dominio, sino encontrar el lugar más adecuado dentro de la jerarquía de TEI.

Un modelo de dominio representa todos los conceptos y las relaciones existentes en el contexto que se desea modelar. Por ejemplo, el modelo de dominio elaborado por POSTDATA para la poesía europea presenta un concepto "Person" que se vincula con otro concepto como "Bibliographic

Source” a través de relaciones “creates” e “edits”, entre otras. En TEI, los datos se describen a través de elementos, atributos y de la relación de parentesco existente entre los diferentes elementos. Por tanto, podríamos pensar que el equivalente en TEI del concepto del modelo dominio “Person” es el elemento `<person>`. Sin embargo, hay que tener presente todas las posibilidades de interacción del concepto “Person” para poder traducirlo como `<editor>`, `<author>` o, incluso, `<person role=“author”>`. La flexibilidad inherente a TEI se convierte en este tipo de transformaciones que apuestan por la estandarización en un hándicap.

Así pues, se analizarán algunos de los problemas y las opciones tomadas en el trabajo de transformar el modelo de dominio (esto es, un modelo conceptual) de POSTDATA en una jerarquía ordenada serán expuestos en esta comunicación.

Finalmente, se abordará una aproximación al concepto de PoetryLAB, un espacio que incluirá herramientas para automatizar análisis de poesía a varios niveles, que reposan sobre técnicas de procesamiento del lenguaje natural. Por un lado, análisis de fonética y métrica para español contemporáneo y variedades históricas. Se han comenzado también tareas de análisis automático del encabalgamiento (Ruiz Fabo et al, 2017). En cuanto a análisis de contenido, se ha implementado reconocimiento y clasificación de entidades nombradas en español medieval (Díez Platas et al, 2017). Se prevé también la utilización de técnicas de semántica distribucional para otros análisis semánticos.

Hay diferentes retos relacionados con estas tareas. La complejidad de los textos poéticos presenta dificultades particulares para el procesamiento del lenguaje natural. Otro desafío es la escasez de recursos lingüísticos disponibles para el análisis de variedades lingüísticas históricas. Finalmente, desde el punto de vista literario, inspirándose en la idea de análisis distante (Moretti, 2013), los análisis automáticos sobre un gran volumen de datos ofrecen una fuente de información que puede complementar análisis más detallados sobre volúmenes de texto más pequeños. Existen por otra parte retos para la evaluación e interpretación de los análisis automáticos.

Referencias

- Cummings, James. 2007. «The Text Encoding Initiative and the Study of Literature». En *A Companion to Digital Literary Studies*, editado por Raymond George Siemens y Susan Schreibman, 451-76. Malden, MA: Blackwell Pub.
- Díez Platas, M. L., Tobarra Abad, L., Ros Muñoz, S., González-Blanco, E., Robles Gómez, A., Caminero Herráez, A., Del Río Riande, G. (2017).

- «Hispanic Medieval Tagger (HisMeTag): una aplicación web para el etiquetado de entidades en textos medievales». En III Congreso Internacional de Humanidades Digitales Hispánicas. Málaga, Spain.
- Drucker, Johanna. 2012. «Humanistic Theory and Digital Scholarship». En *Debates in the Digital Humanities*, editado por Matthew K Gold, 85-95. Minneapolis: Univ Of Minnesota Press.
- Flanders, Julia. 2011. «Collaboration and Dissent: Challenges of Collaborative Standards for Digital Humanities». En *Collaborative Research in the Digital Humanities*, editado por Marilyn Deegan y Willard McCarty, 67-80. Farnham, Surrey, England; Burlington, VT: Ashgate Pub.
- Hockey, Susan. 2004. «The History of Humanities Computing». En *A Companion to Digital Humanities*, editado por Susan Schreibman, Raymond George Siemens, y John Unsworth, 3-19. Malden, Mass.: Blackwell Pub.
- Ide, Nancy M., y Christopher Michael Sperberg-McQueen. 1995. «The TEI: History, goals, and future». *Computers and the Humanities* 29 (1): 5–15.
- Moretti, F. (2013). *Distant Reading*. Verso Books, London & New York.
- Ruiz Fabo, Pablo, Clara Martínez Cantón, Thierry Poibeau and Elena González-Blanco. (2017). «Enjambment detection in a large diachronic corpus of Spanish sonnets.» En *LaTeCH-CLFL 2017, Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Vancouver, Canada. Association for Computational Linguistics.
- Sperberg-McQueen, C. M. 1997. «What is XML and Why Should Humanists Care?» <http://cmsmcq.com/1997/drh97.html>.