Linked data in lexicography

Julia Bosque-Gil, Jorge Gracia and Asunción Gómez-Pérez

1. Introduction

The notions of *linked data* (LD) and *Web of Data* are increasingly gaining ground in digital humanities, linguistics, biomedicine, e-science, data journalism, etc. and lexicography is not staying behind. The LD paradigm meets the need to link isolated pieces of information which were in their own proprietary formats and were previously hard to discover and integrate. The term actually refers to a "set of best practices for exposing, sharing, and connecting data on the Web" (Bizer et al. 2009). In order to create LD there is a set of requirements to fulfill, among them, the use of Unique Resource Identifiers (URIs) and the establishment of links to other resources. The Resource Description Framework (RDF)[1] is the formal backbone giving support to this network of interlinked resources and allowing for the definition of *triplets* or statements of the form *subject-predicate-object*, where *subject* and *object* are resources and the *predicate* is the edge or *property* connecting the nodes. The result is a vast graph whose nodes can be practically anything, including lexical units, and this is where lexicography comes into play.

The work in models for the representation of linguistic information as LD (McCrae et al. 2012), as well as in best practices and guidelines for the conversion of mono- and multilingual language resources[2] has been continuous in recent years. The benefits that LD brings to lexicography have been already pointed out in recent works related to the conversion of bilingual and multilingual dictionaries as LD (e.g. Gracia 2015, Klimek and Brümmer 2015, Bosque-Gil et al. 2016) and etymological and dialectal dictionaries (Declerck et al. 2015, among others), as well as in recent initiatives and international projects that have embraced the use of semantic technologies[3] and in current e-lexicography work (McCracken 2015). The main advantages are the semantic and syntactic interoperability provided by RDF and linguistic vocabularies (LexInfo[4] or GOLD[5]), which enables the integration, exchange, and enrichment of lexicographic data among different resources, the reusability of the whole resource, which in turn prevents lexicographers from "re-inventing the wheel" in potential future projects, improved data visualization and querying, resource sustainability (Wandl-Vogt 2015), and easy discovery thanks to metadata repositories.[6]

In this context, this paper seeks to present, on the basis of our experience in the conversion of lexicographic data to LD[7], our reflections on the implications of converting lexical data to LD, drawing special attention to the advantages it offers from the eyes of a lexicographer or a linguist outside the realm of the Semantic Web, but as part of a discipline which can be already considered part of information science (Fuertes-Olivera and Bergenholtz 2011). Our goal is therefore twofold: to place LD in the context of lexicographic work in lexical networks, and to bring its benefits closer to the lexicographer so she can consider it a basis for future endeavours. To this end, we will first provide a brief overview of the work on the representation of lexical information as graphs outside the context of the Semantic Web with focus on WordNet (Miller 1995, Fellbaum 1998) and Polguère's lexical systems (Polguère 2012, 2014) implemented in the French Lexical Network (Gader et al. 2012). Then, we will dwell on the practical advantages of LD for representing both the macro- and the microstructure of a lexicon.

2. Lexical data as a graph

Modeling lexical information as a graph is not a novel notion coming from LD.

1 https://www.w3.org/TR/rdf11-primer/
2 https://www.w3.org/community/bpmlod/
3 Such as the ENel cost action (http://www.cost.eu/COST_Actions/isch/IS1305/) or the LIDER (http://lider-project.eu/) and LDL4HELTA (http://www.eurekanetwork.org/project/id/9898/) projects
4 http://www.lexinfo.net/ontology/2.0/lexinfo/
5 http://www.linguistics-ontology.org/
6 linghub.org, http://metashare.elda.org/
7 From October 2015 to February 2016, the Ontology Engineering Group at UPM worked on the development of a linguistic linked data prototype for K Dictionaries and Semantic Web Company as part of their LDL4HELTA project, and, more specifically, on the transformation to RDF of the Spanish dataset of K Dictionaries.

Julia Bosque-Gil is a PhD student at the Ontology Engineering Group, Universidad Politécnica de Madrid. She holds a B.A. in German Linguistics and English from the Humboldt University of Berlin and an M.A. in Computational Linguistics from Brandeis University, Waltham. Her interests include the lexicon-ontology and the syntax-semantics interfaces, the relations between the lexicon and syntax, semantic annotation and the representation of (multilingual) language resources as linked data. She has been working in the conversion of multilingual terminologies to RDF as part of the LIDER project and in the modeling of lexicographic data as linked data in collaboration with K Dictionaries and Semantic Web Company. For her PhD thesis she is investigating the use of linguistic linked data for research in linguistics.
jbosque@fi.upm.es

Jorge Gracia is a post-doctoral researcher at the Ontology Engineering Group, Universidad Politécnica de Madrid, Spain. He got his PhD in Computer Science at University of Zaragoza in 2009, with a thesis about heterogeneity issues on the Semantic Web. His current research interests include multilingualism and linked data, linguistic linked data, and cross-lingual matching and information access on the Semantic Web. Currently he is exploring how to move language resources (lexica, dictionaries, corpora, etc) from their data silos into the multilingual Web of Data and make them interoperable, in order to support a future generation of linked data-aware NLP tools.
http://jogracia.url.ph/web/

WordNet already set a precedent (Miller 1995, Fellbaum 1998) as a graph-based lexico-semantic database where nodes represent the concepts (synsets or sets of cognitive synonyms) and hyponymy, meronymy and antonymy constitute the relations that link them together. Furthermore, there are other efforts in lexicography that emerge from a conception of the natural language lexicon as a network of entries rather than a list, which is what the organization of conventional dictionaries looks like. The entries are then viewed as part of a language system of related lexical elements. Polguère's notion of a lexical system, implemented in the framework of the French Lexical Network project, falls into this category. However, in contrast to the projects developed in lexical semantics, linguistic linked data (LLD) and the models proposed for converting resources into them (*lemon*[8], SKOS-XL[9], LIR (Montiel-Ponsoda et al. 2008)), do not arise as initiatives to model the (mental) natural language lexicon, nor make such claim, even though they entail the use of classes and properties such as *lexical entry*, *sense*, *lexical concept*, *syntactic frame*, *lexical form* or *definition*. LD emerges as a technological means to better represent, share, integrate and discover linguistic knowledge scattered over the Web and its underlying RDF formalism is not conceived from a theoretical perspective as an alternative to structure mental lexical information. Nonetheless, knowing the direction into which lexical semantics and lexicography move, as well as the similarities between the representations suggested there and those proposed from an LD perspective, will help us in building bridges for collaboration between experts from both sides. LD, as best practices for data representation, should be compatible with the representation of any lexical network, even though this implies the extension of vocabularies currently available on the so-called linguistic linked open data (LLOD) cloud[10] or models to encode all the data that the theory on which the resource is based addresses.

An analysis of what modeling the lexicon as a graph in WordNet entails and which needs are met is given in Polguère (2014) and McCracken (2015): lexical entries were previously analyzed and presented independently one from another and a novel approach reflecting what the structure of the mental lexicon might resemble was called for.[11] WordNet falls under the category of ontology-based lexical network (Polguère 2014: 3), i.e. a network of lexical units with an ontology as backbone, including word senses arranged in a hierarchy and related by synonymy, hyponymy and meronymy relations. It is worth mentioning that LLD relies on linguistic ontologies or vocabularies, but the creation of an ontology of word senses or concepts is actually optional and it is not a required step in order to publish LD. Accordingly, we can state that the entry *enthusiasm* in an English lexicon has as the part-of-speech `lexinfo:noun`, which is defined along with, for instance, `lexinfo:reflexivePersonalPronoun`, as an individual of type `lexinfo:PartOfSpeech` in the linguistic ontology LexInfo.[12] We are thus linking two resources without establishing the concept denoted by *enthusiasm* in any hierarchy (e.g. as a child of *feeling*). LD resources such as BabelNet[13], DBpedia[14], and WordNet RDF (McCrae et al. 2014) have an underlying ontology, but this is not implied in the conversion of every resource to LLD. In relation to this, LLD builds upon the notion of *semantics by reference* (McCrae et al. 2012): the meaning of a word and the word itself (the *signifier -- signified* opposition) are separated in two different layers, with `ontolex:LexicalEntry` and `skos:Concept` respectively, and the relation between the two is "reified" in a class that aims at encoding *a sense* (`ontolex:LexicalSense`). All the linguistic information pertaining to the word itself or to the use of that word with that specific meaning is separated from the actual meaning, which, ideally, is language independent. Hierarchic conceptual relations would be established, if they are, at the level of the concept.

Polguère places lexical systems on the other side of the balance: they are lexical networks that are not ontology-based. Lexical systems are conceived with the relations among the lexical elements as focus and relegate to the background the classification of units or property inheritance (Polguère 2014: 3). A key aspect of lexical systems is that relations are not limited to synonymy, hyponymy, etc. but they include

8 http://www.lemon-model.net/lemon/
9 https://www.w3.org/TR/skos-reference/skos-xl.html/
10 http:/linguistic-lod.org/llod-cloud/

11 However, most lexical semantics research addresses different aspects with different levels of granularity, but it does not analyze all word types and the semantic structure of the lexicon as a whole (Swanepoel 1994)
12 http://www.lexinfo.net/ontology/2.0/lexinfo/
13 http://babelnet.org/
14 http://wiki.dbpedia.org/

paradigmatic and syntagmatic relations drawn from the set of lexical functions of the Meaning Text Theory (Mel'čuk 1996). The result is a multi-dimensional graph with a wide range of relations linking the nodes (lexical elements), which instantly brings RDF to mind. There are two important points to bear in mind when comparing lexical systems with resources migrated to LLD: first, the nodes in lexical systems are already "disambiguated", each node represents one specific meaning of the lexical unit at hand (Polguère 2014: 5). The closest counterpart we have in RDF is `ontolex:LexicalSense`, which is a unique relation between a word and a meaning. Secondly, the nodes in a lexical system are not atomic and each one records the information we would find in a lexicographic article. Grammatical information, semantic label, syntactic government pattern (collocations are implemented by edges), etc. are stored inside the node. In LLD, some of these data would be linked to the entry at hand or to one of its `ontolex:LexicalSense`(s) by means of specific properties (edges) and elements available in linguistic vocabularies, identifiable with their own URIs: lexical entries, word forms, senses, part of speech tags, gender, number, subcategorization, etc. To see which entries are related, the SPARQL query language[15] allows to perform queries on the graph and trace the connections between ontolex lexical senses or ontolex lexical entries. LexInfo, *lemon-ontolex*, SKOS, GOLD, etc. already provide a high number of relations, which can be extended with new ones or new vocabularies can be created as needed.

In sum, the idea of representing lexical information as a graph is not new, and LD are not presented as a novelty in this regard. However, they allow for the implementation of networks or the integration of already available ones on the basis of a homogenous format. Thus LD meet the need for linking lexical elements that were previously isolated by using sets of relations and elements that are defined externally and can be extended as required, relying or not on an underlying ontology of word senses. This does not mean that LD is equivalent to any of the efforts mentioned above or forms a better option to the structures in which they are implemented, and, as said, it does not make claims on the structure of our mental lexicon. RDF is, however, a model to represent data worth taking into consideration for lexicographic projects aiming at the creation of lexical networks because it provides a basis for their implementation while retaining all the benefits related to interoperability, visibility and NLP-services compliance.

All in all, the LD paradigm is agnostic with respect to the different theories in modern lexicography, and it poses a number of tangible benefits that we enumerate in the following sections.

3. Benefits of a lexicon in linked data: macro-structure

Having placed LD in context, what are the actual benefits of creating or converting a lexicon to LD? The most evident advantage is that LD enable the integration with other external resources thanks to the semantic and syntactic interoperability achieved by the use of RDF and linguistic ontologies. Besides this fact and focusing on the lexicon itself, we have identified the following benefits in the course of our work towards the migration of language resources to LLD, some of them also highlighted in the literature.

Firstly, the entries of a dictionary become internally reusable (Klimek and Brümmer 2015) thanks to their URIs. This does not seem novel given that entries might already have numeric identifiers to point to each other, but the choice of transparent URIs, i.e, human-readable, which reflect the semantic content, and a suitable URI naming strategy play a crucial role (Bosque-Gil et al. 2016): the editor of a dictionary entry will be able to refer to another entry without the need to know its identifier in advance. Following this, the entry `:lexiconEN/risk-n` can be linked to `:lexiconEN/risky-adj` through a relation of morphologic derivation without the need of an identifier. If, later on, the noun *risk* occurs as an entry in another dictionary of the same or a different family of dictionaries, the information can be integrated in a straightforward manner without relying on dictionary-dependant numeric IDs.

This in turn relates to a second advantage: we no longer depend on the order of appearance of lexical entries or senses in cross-references, which is usually indicated by a superscript in numeric form in printed or electronic format, e.g: *bow*[2], meaning, for instance, the second homograph of the word *bow*. There are ways of keeping track of the order and the lexical entry to which that position refers, but a change in the original order of entries or the integration with other dictionaries in which the order differs would then require the update of all cross-references to any of the ordered entries. Since entries and senses are now identifiable throughout the data and graphs are not actually ordered, cross-references can be direct pointers to the entry or sense to which they refer.

The third advantage is intrinsically

15 https://www.w3.org/TR/rdf-sparql-query/

Asunción Gómez-Pérez is Vice-Rector for Research, Innovation and Doctoral Studies and Full Professor at Universidad Politécnica de Madrid. She is Head of the Department of Artificial Intelligence since 2008, Director of the Ontology Engineering Group since 1995, Academic Director of the Master's Degree in Artificial Intelligence since 2009, and Coordinator of the PhD Programme in Artificial Intelligence since 2009. Some of her main research areas are Ontological Engineering, Semantic Web, Linked Data, Multilingualism in Information and Management of Knowledge. She has been the coodinator of the OntoGrid, SemSorGrid4Env, SEALS, Interactivex and LIDER research projects, and she is currently taking part in three European H2020 research projects. She has also participated in numerous research projects of the Spanish National Plan of Basic Research, Networks, Special Actions and Technological Transference (ZENITH, Hundred, Advances, Profit, etc) and directed multiple projects of national and international enterprises. She is the head of the first node of the Open Data Institute in Spain and the main researcher of the first research project that uses IBM-Watson in a Spanish university (2015).
asun@fi.upm.es

The Ontology Engineering Group (OEG), led by Prof. Asunción Gómez-Pérez, is based at the Computer Science School at Polytechnic University of Madrid (UPM). It ranks eighth among the two hundred research groups of UPM and is widely recognized in Europe in the areas of Ontology Engineering, Semantic Infrastructure, Linked Data, and Data Integration. Its main research areas are Ontological Engineering, Open Science, Data-driven Language Technologies, Data on the Web, and Data Science. The OEG was the coordinator of LIDER, a European project that promoted the creation of a linked data-based ecosystem of interlinked multilingual language resources to support content analytics tasks. http://oeg-upm.net/

related with the first one, too: we can represent an "abstract" lexicon that gathers all the entries in a specific language. In other words, have a "pool" of lexical entries extracted from different dictionaries of the same or different type, monolingual or multilingual, without losing provenance information about which data comes from which dictionary. Thanks to an appropriate URI naming strategy, this pool of entries will grow dynamically (Gracia 2015) with each dictionary converted into RDF that has any information about an entry in that specific language.

If the approach mentioned above is applied in the conversion of multilingual dictionaries, for instance, Spanish-French and French-English, linking the French entries from the ES-FR dictionary with their corresponding entries in the FR-EN dictionary will bring us a fourth advantage: translation relations can be established through a language acting as a pivot (Villegas et al. 2016).

The fifth benefit concerns the onomasiological view that LD enables. The source dictionary has probably been compiled from a semasiological perspective, by putting the word as the center of attention and listing its different senses. Given the semantics by reference in LLD mentioned above, the synonym of a word and the word itself will point to the same concept, which is modeled as a node in the graph and has therefore a URI. Accessing that node will allow us to see which words lexicalize it, i.e. putting the concept as our focus and traversing the graph from it to the lexical elements related to it. This way of thinking is well illustrated in the case of multilingual dictionaries in LLD, where we can see how a concept is verbalized in different languages. The potential is however no less interesting in monolingual dictionaries. In the authors' work on the migration of language resources to LLD, definitions have been encoded at the level of the concept. Even though definitions can be fine-grained and are not presented in the form of keywords, SPARQL queries over them are feasible. For instance, we can search for concepts in whose definition the word *sunrise* occurs, which will yield the series of concepts that words like *dawn, morning, daylight*, etc. denote and which are semantically related, although these relations are not implemented. Through these concepts we could not only access *dawn, morning,* etc. but also their antonyms *dusk, twilight*, etc. Thus, by taking the concept as entry point we can get a set of concepts that are related but are not necessarily equivalent, which is not a trivial task when searching in a conventional online dictionary.

The sixth advantage is related to cross-references in the sense of any reference to another entry that might occur inside the lexicographic article: orthographical variants, synonyms, antonyms, genus terms, semantic types, etc. Not only are the entries reusable throughout the data (first advantage), but the pointers to them are now *typed* (Klimek and Brümmer 2015, McCracken 2015). This might not seem like an evident benefit to the user of online dictionaries, for whom the label *antonym* or a typographical mark may suffice, but typed properties allow users to perform queries not dependant on the (proprietary) format of the data and LD-aware systems to find any needed information. At the same time, by virtue of being defined in a public external vocabulary, e.g. LexInfo, the same properties can be reused in the conversion of other lexica of the same series into LLD, thus gaining interoperability. This responds to the need of standardization among the high number of heterogeneous annotation schemas, tagsets, and proprietary DTDs that are being used to create language resources.

Furthermore, given that these vocabularies are extensible, new properties and individuals or classes can be added. If the hierarchy defined in a linguistic ontology is not compatible with the view other domain experts might have, new vocabularies can be created and aligned to the ones already available. As we could experience during our work on the conversion of dictionaries to LD, a detailed comparison of the elements (and their classification) present in external vocabularies with the proprietary data model of a company specialized in lexicography is actually a significant step towards the improvement, refinement or even reconsideration of the elements that configure that data model.

As the last paragraphs suggest, the concept of *reusability* lies at the heart of LD. If the enterprise of compiling a dictionary is seen through the looking glass of LD from the very beginning, it will affect the whole process. Decisions such as, for example, keeping independent lexical entries for an entry and its homographs will have to be considered from the point of view of lexicography (two words that share form but are not related etymologically could thus be regarded as independent entries) and LD. At the same time, how do we model homographs in such a way that enables us to identify each entry but also to integrate content from another source that we do not know to which of the homograph entries it pertains? It will not be a matter of converting lexical data to LD, but of creating them from scratch in a reusable, interoperable and linguistically accurate way.

4. Benefits of a lexicon in linked data: micro-structure

The previous section dwelled on the benefits of representing a lexicon as LD but it did not deepen into the modeling of information present in a single lexicographic article. As opposed to lexical systems (Section 2), this information (definitions, grammatical data, syntactic frames, etc.) is also modeled as a graph.

Ideally, everything in the lexicographic entry can be modeled as a node (McCracken 2015) but, in general, and on the basis of *lemon-ontolex*, the representation revolves around lexical entries (`ontolex:LexicalEntry`), concepts, the relation between entries and concepts reified as lexical senses (`ontolex:LexicalSense`), word forms (`ontolex:Form`), definitions, phonetic representations, register, syntactic frames, etc. Relations between nodes have a well-defined domain and range, and, with actual data, every node will be an instance of a class defined in an ontology. Following the *lemon-ontolex* model, the English entry *cloud*, with the sample URI *lexiconEN/cloud-n* will have `rdf:type ontolex:LexicalEntry`, will denote as many `skos:Concepts` as senses or meanings it has, and the relation from the word to the concept will be encoded as `ontolex:LexicalSense`. Word forms (*cloud, clouds*) will be recorded at the `ontolex:Form` level, together with grammatical number information and phonetic transcription. Definitions, usage examples, etc. are likewise linked to the entry through edges and intermediate nodes.

On the one hand, one of the consequences of this configuration is that elements previously embedded in the lexicographic article become entry points in the graph and are no longer subsumed under any entry, since the hierarchy is lost. This implies that an idiom or collocation, for instance, will not be encapsulated under the container of the entry in which it was originally defined, but will be related to it with the suitable property. Since the idiom now becomes a node, we are able to link it to any other node from any other entry in the lexicon: *like a cat on a hot tin roof* could then be linked, for example, to the appropriate sense of *cat,* of *hot* and of *roof*, if desired, which will allow to access the idiom from any of those entries. Also, in the case of idioms and frequent collocations, we are creating new lexical entries that were not originally conceived as such in the lexicon. As lexical entries, they will be also linked to their corresponding `skos:Concept`(s), which brings us back to the possibility of an onomasiological perspective on the data.

On the other hand, thinking in terms of LD forces us to constantly question what is the nature of the relation between two pieces of information. An LD-native dictionary will require a specification on the part of lexicographers of which kind of relations between which type of elements will be encountered when modeling lexicographic articles. This brings us to the difference between compiling dictionaries with only the human as target, and creating them for (both humans and) computers. The fact that an XML tag, for instance, can occur at different levels in the dictionary entry (e.g. a geographical usage indication attached to a pronunciation vs. a geographical usage indication attached to a sense) seems straightforward enough for a human, but an NLP application needs to be able to distinguish between a description of a string (e.g. [kɑː] is the transcription of the British pronunciation of *car*) and the restriction on the *usage* of a sense (e.g. the *floor* with the meaning *the floor above the ground level floor* is only used in the UK). Modeling data as LLD thus entails a reflection of which information affects which elements, and which properties are the most suitable ones to be used in which case, taking all nuances and human implicit knowledge into account.

5. Conclusion and future lines of work

LLD emerge as a promising option to represent and publish current lexicographic projects and to serve as a structural backbone for undertaking new ones. They allow for the creation of an interoperable lexical network that is endowed with all the benefits that LD offers: data aggregation, easy discovery, LD-aware services compliance, improved data querying, sustainability and reusability. In this paper we have offered a brief overview of LLD, placing them in the context of lexical networks, and analyzing some of the benefits of the conversion of lexical data into LD in terms of macro- and microstructure. The modeling of lexicographic data to LLD poses challenges for which bridging the gap between LD experts and lexicographers is crucial. Moreover, the relation of LD to functional lexicography has not been explored to its full potential and, although there has been some work on RDF and OWL as building blocks for an architecture of mono- and plurifunctional dictionaries (Spohr 2011, 2012), this remains a challenging line of work, partly due to the increasing need of natural languages interfaces for the Web of Data. However, current trends in LD-based NLP and in publishing language resources as LD, including lexical data, show that we will be getting there hopefully soon.

Acknowledgments

This work is supported by the Spanish Ministry of Economy and Competitiveness through the project 4V (TIN2013-46238-C4-2-R), the Excellence Network ReTeLe (TIN2015-68955-REDT), the Juan de la Cierva program, and the Spanish Ministry of Education, Culture and Sports through the Formación del Profesorado Universitario (FPU) program. This contribution is inspired by the work towards the development of a linked data prototype for the Spanish dataset of the Global Series of K Dictionaries, carried out by the authors as part of the Linked Data Lexicography for High-End Language Technology Application (LDL4HELTA) project of Semantic Web Company and K Dictionaries.

LDL4HELTA

Linked Data Lexicography for High-End Language Technology Application (LDL4HELTA) is a 24-month EUREKA project (July 2015 – June 2017) within the framework of the Austria-Israel Bilateral R&D Agreement, carried out by Semantic Web Company (SWC, http://semantic-web.at/) and K Dictionaries (KD, http://kdictionaries.com/), with funding from the Austrian Research Promotion Agency (FFG) and the Israeli Office of the Chief Scientist (OCS). The aim is to combine multi-language lexical resources with semantic technologies expertise and develop new products and services for the international language technology market, in reply to the needs for language-independent, specific-language and cross-language solutions, to enable cross-lingual search and data management approaches. The main tasks consist of converting KD lexicographic data from XML to RDF, developing an API for enhanced data streaming and dissemination, and incorporating it in SWC's PoolParty Semantic Suite (https://poolparty.biz/). The RDF modeling is designed by the Ontology Engineering Group of Universidad Politécnica de Madrid (UPM), which is involved also in the word sense disambiguation aspects. An advisory board consists of Christian Chiarcos (Goethe University, Frankfurt), Orri Erling (Google), Asunción Gómez-Pérez (UPM), Sebastian Hellmann (Leipzig University), Alon Itai (Technion, Haifa), and Eveline Wandl-Vogt (Austrian Academy of Sciences). http://ldl4.com/

References

Bizer, C., T. Heath, and T. Berners-Lee. 2009. Linked data – the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1-22.

Bosque-Gil, J., J. Gracia, E. Montiel-Ponsoda, and G. Aguado-de Cea. 2016. Modelling multilingual lexicographic resources for the Web of Data: The K Dictionaries case. In *Proceedings of GLOBALEX 2016 Workshop at the 10th Language Resources and Evaluation Conference (LREC 2016),* Portorož, (Slovenia).

Declerck, T., E. Wandl-Vogt, and K. Mörth. 2015. Towards Pan European Lexicography by Means of Linked (Open) Data. In *Electronic lexicography in the 21st century: Linking lexical data in the digital age. Proceedings of the eLex 2015 conference*, Herstmonceux Castle, UK, 342–355.

Fellbaum, C. (ed.). 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Fuertes-Olivera, P.A., and H. Bergenholtz. 2011. Introduction: The construction of Internet dictionaries. In Fuertes-Olivera, P.A., and H. Bergenholtz (eds.), *eLexicography: The Internet, Digital Initiatives and Lexicography*. London & New York: Continuum, 1-16.

Gader, N., V. Lux-Pogodalla, and A. Polguère. 2012. Hand-Crafting a Lexical Network With a Knowledge-Based Graph Editor. In *Third Workshop on Cognitive Aspects of the Lexicon (CogALex III)*, Mumbai, India, 109-125.

Gracia, J. 2015. Multilingual dictionaries and the Web of Data. *Kernerman Dictionary News*, (23), 1–4.

Klimek, B., and M. Brümmer. 2015. Enhancing lexicography with semantic language databases. *Kernerman Dictionary News,* (23), 5-10.

McCracken, J. 2015. The Exploitation of Dictionary Data and Metadata. In P. Durkin (ed.), *The Oxford Handbook of Lexicography*. Oxford: Oxford University Press, 501-514.

McCrae, J., G. Aguado-de-Cea, P. Buitelaar, P. Cimiano, T. Declerck, A. Gómez-Pérez, J. Gracia, L. Hollink, E. Montiel-Ponsoda, D. Spohr, and T. Wunner. 2012. Interchanging lexical resources on the Semantic Web. *Language Resources and Evaluation*, *46*(4), 701-719.

McCrae, J., C. Fellbaum, and P. Cimiano. 2014. Publishing and Linking WordNet using lemon and RDF. In *Proceedings of the 3rd Workshop on Linked Data in Linguistics*.

Mel'cuk, I. 1996. Lexical functions: a tool for the description of lexical relations in a lexicon. *Lexical functions in lexicography and natural language processing*, (31), 37-102.

Miller, G.A. 1995. WordNet: A Lexical Database for English. *Communications of the ACM 38*(11), 39-41.

Montiel-Ponsoda, E., G. Aguado de Cea, A. Gómez-Pérez, and W. Peters. 2008. Modelling Multilinguality in Ontologies. In *The 22nd International Conference on Computational Linguistics (COLING 2008)*, August 18-22, Manchester, UK, 67-70.

Polguère, A. 2012. Like a lexicographer weaving her lexical network. In *Proceedings of CogALex-III Workshop of the 24th International Conference on Computational Linguistics (COLING 2012)*, December 8-15, Bombay, India. 1-4.

Polguère, A. 2014. From Writing Dictionaries to Weaving Lexical Networks. *International Journal of Lexicography*, *27*(4), 396-418.

Spohr, D. 2011. A Multi-layer Architecture for "Pluri-monofunctional" Dictionaries. In Fuertes-Olivera, P.A., and H. Bergenholtz (eds.), *eLexicography: The Internet, Digital Initiatives and Lexicography*. London & New York: Continuum, 103-120.

Spohr, D. 2012. *Towards a Multifunctional Lexical Resource: Design and Implementation of a Graph-based Lexicon Model*. Lexicographica Series Maior (141). Berlin: de Gruyter.

Swanepoel, P.H. 1994. Problems, theories and methodologies in current lexicographic semantic research. In W. Martin et al. (eds.), *Proceedings of the Sixth International Euralex Congress,* Amsterdam, 11-26.

Villegas, M., M. Melero, N. Bel., and J. Gracia. 2016. Leveraging RDF graphs for crossing multiple bilingual dictionaries. In *Proceedings of the 10th Language Resources and Evaluation Conference (LREC 2016),* Portorož, (Slovenia).

Wandl-Vogt, E. 2015. How to innovate lexicography by means of research infrastructures: The European examples of DARIAH, CLARIN and COST IS 1305 ENeL [slides]. http://www.slideshare.net/ewv/how-to-innovate-lexicography-by-means-of-research-infrastructures/ (June 5, 2016).