Toward Linked Data-native Dictionaries

Jorge Gracia[1], Ilan Kernerman[2], Julia Bosque-Gil[1]

[1]Universidad Politécnica de Madrid, Campus de Montegancedo sn, Boadilla del Monte 28660 Madrid, Spain
[2] K Dictionaries Ltd, 8 Nahum Hanavi Street, 6350310 Tel Aviv, Israel
E-mail: jgracia@fi.upm.es, ilan@kdictionaries.com, jbosque@fi.upm.es

Abstract

The ways in which dictionaries are compiled and used have evolved dramatically in recent years owing to the processes of digitization. This evolution has found in the Web an optimal means to empower the visibility and usability of dictionaries. In this context, we witness nowadays increasing interest in the interoperability of linked data (LD) technologies for the development and representation of lexicographic data on the Web.
In this paper we propose the notion of *LD-native dictionaries* as a natural next step in the evolution of lexicography. These future dictionaries could be LD-native and, as such, graph-based. Their nodes are not dependent on any internal hierarchy and are uniquely identified at a Web scale. We analyze the advantages of such an approach and identify its possible impact on the dictionary representation, compilation, and usage processes. Some challenges related to interoperability and data aggregation issues are also discussed.

Keywords: linguistic linked data; linked data-native dictionaries; e-lexicography

1. Introduction

The dictionary concept has been evolving over the last generation alongside the advent of technology and digitization of modern life, both as regards the lexicographic compilation process and the dictionary's media, dissemination and forms of usage. In the first wave of the electronic era (1990's), dictionaries usually remained little more than that same old 'book of words' in new e-dress(es), but gradually more e-features were introduced, such as advanced search modes, dictionary as corpus, morphological connections, integrating with other language software, embedding audio and images, and so on.

This logical evolution has found in the Web an optimal means to empower the visibility and usability of dictionaries. In particular, we witness nowadays increasing interest in the interoperability with linked data (LD) technologies to develop and represent lexicographic data on the Web. LD refers to a set of best practices for exposing, sharing and connecting data on the Web (Bizer et al., 2009). In short, the LD paradigm requires that *resources* be represented on the Web via URIs (Unique Resource Identifiers) and that, once a resource is accessed via its URI, useful information can be obtained, along with links to other resources. The basic mechanism that enables this is the Resource Description Framework (RDF),[1] which follows the *subject-object-predicate* pattern. The result is a vast graph of linked resources on the Web, whose nodes can be practically anything, including documents, people, physical objects and abstract concepts (such as lexical entries or any other entity that lexicography needs to model).

Some of the advantages of using LD to represent lexicographic content have already been reported in the literature (e.g., Klimek & Brümmer, 2015; Declerck et al., 2015; Bosque-Gil et al., 2016a) and the number of initiatives applied toward the conversion of proprietary dictionary formats to LD continues to grow (e.g., Bosque-Gil et al., 2016b; Parvizi et al., 2016). Also the community of ontology lexica has shown interest in LD for lexicography and started discussing best practices and modelling issues on this topic (Bosque-Gil et al., 2017) in the context of the W3C Ontolex community group.[2]

[1] http://w3.org/TR/rdf11-primer/
[2] https://www.w3.org/community/ontolex/

As a natural next step, we envisage dictionaries that are born and evolve dynamically on the Web. These will not be (only) the result of transforming lexicographic data from previous electronic formats into LD, but will ensue from compiling dictionaries as LD from scratch. Thus, such future dictionaries are LD-native and, as such, graph-based. Their nodes are not dependent on any internal hierarchy and are uniquely identified at a Web scale. This will enable the enhancement of a vast network of interconnected linguistic elements through semantically well-defined lexical, syntactic, pragmatic, etc. relations, through which lexicographers and users navigate to edit, query, or aggregate data. Links to other lexical resources, including other dictionaries, would thus be quickly and naturally established.

In this paper we analyze this vision and its advantages as compared to a more traditional tree-based view of lexicographic data. We also explore its impact on the editorial process, both on the content itself and on the way lexicographers work. Some challenges about interlinking and data aggregation are discussed as well.

The rest of the paper is organized as follows. In Section 2 the vision of LD-native dictionaries is presented. Then in Section 3 the impact of the LD-native dictionaries notion on the editorial process is discussed. Some challenges related to data integration are presented in Section 4. Finally, our conclusions are presented in Section 5.

2. The vision of LD-native dictionaries

Several experiences have been reported in the literature related to the conversion of different types of dictionaries as LD (e.g., Klimek & Brümmer, 2015; Declerck et al., 2015; Bosque-Gil et al., 2016b; Gracia et al., 2016), which illustrate the growing interest for LD in lexicography. Nevertheless, the idea of developing dictionaries as LD in a native way, rather than converting already existent ones from their proprietary formats into LD, has received little attention so far.

2.1 LD in lexicography

There are, of course, a number of advantages in using LD in lexicography (Bosque-Gil et al., 2016b) that do not depend on whether the dictionary data have been converted from previous formats or have been built as LD from scratch. For instance, the main models developed for representing linguistic information as LD (e.g., OntoLex-lemon[3]) do not make claims on the structure of our mental lexicon, being agnostic of the particular linguistic theory underlying the lexicographic data. Thus, LD constitutes an ideal common representation framework for dictionaries that have been built by following different practical and theoretical perspectives, while retaining all the benefits related to interoperability, visibility and NLP-services compliance. Another evident advantage is the fact that LD enables a seamless integration with other internal and external resources (via links among entities, expressed for example in RDF), allowing for a natural graph-based representation of dictionary data on the basis of Web standards.

These and other benefits have been reported as a result of the initial experiences of converting already existent dictionaries to LD format. We envision, however, a situation

[3] https://www.w3.org/2016/05/ontolex/

in the near future when dictionaries will be developed natively as LD, that is, by compiling them from scratch in an RDF-based environment and directly following the LD principles. This will have an impact on the process of dictionary compilation, representation, and interoperation with other resources.

2.2 Issues of tree-like dictionary structures

In modern electronic dictionaries, entries are typically represented as a tree (usually encoded in XML), following a hierarchical data structure where every element has at most one parent. As discussed by Měchura (2016), this choice of data structure makes some aspects of the lexicographer's work unnecessarily difficult, from deciding where to place multiword items to reversing an entire bilingual dictionary. This is a consequence of the fact that dictionary writing, although assisted by computing methods, still tends to replicate what lexicographers would be doing on paper or with a word processor. This raises a number of issues. Although we are not exhaustive in describing them (see, e.g., Bosque-Gil et al., 2016a, for a more detailed analysis) we illustrate them through a couple of examples. First we can mention the problem of headword selection for multiword phrasemes (Měchura, 2016), e.g., under which entry to place *bow and scrape* (meaning *to be overly polite*), *bow* or *scrape*? Ideally, it should be placed under both entries. However, in a tree-like representation, this obliges the lexicographer to copy the same information in both places, which makes the data more difficult to be maintained or updated (changes in one place need to be propagated into other places). Of course, clever search mechanisms can be built to work around this problem, as modern digital dictionaries do, in which a lemma is provided just once and the system is able to search it wherever it appears. However, that does not solve the problem at source, and the search mechanism is not able to infer the particular sense or homograph of the parent entries that should be associated to the phraseme. For instance, our previous example *bow and scrape* would be associated to the sense of *bow* that corresponds to the action of *inclining to show respect*.

Another example of an issue caused by tree-based view of the dictionary information is that cross-references typically depend on the order of appearance of lexical entries or senses, being 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 entry *bow*. The problem of this approach is that the introduction of new elements in the middle of the sequence obliges to review and redefine all the involved cross-references across the dictionary, making this modelling technique very sensitive to any change in the ordering criteria. Techniques such as the latter are prone to errors and might result in the collision of identifiers. Again, mechanisms have been implemented that reduce such a problem, although they do not solve it at source.

2.3 Building a graph-based reusable structure

A key aspect of an LD-based dictionary is that every lexical element (headword, sense, written form, grammatical attribute, etc.) is treated as a first-class citizen, being identified by its own URI at a Web scale, and being attached to its own descriptive information and linked to other relevant elements through RDF statements. That allows for a graph-based view of the lexicographic information where the above referred issues can be easily avoided.

Continuing with the example cited above, in an LD-native dictionary the *bow and scrape* multiword expression will be a headword on its own with its own URI, and links will be drawn to relate it to the two parent entries *bow* and *scrape*, directly pointing to their suitable senses or homographs whenever appropriate. In that way, changes will be done in a single place, avoiding the need for copying information and reducing the risk of bad maintenance. 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 dictionary.

Similarly, LD solves the issue of maintaining cross-references. Since entries and senses are now uniquely identifiable throughout the dictionary data and graphs are not actually ordered, cross-references can be direct pointers to the entry or sense to which they are referring. Cross-references will not (only) be manual annotations for human consumption but real links between nodes in the dictionary graph.

Differently from other graph-based approaches for representing lexicographical information (Miller, 1995; Polguère, 2014), LD is based on Web standards, has interoperability as its main focus, and is agnostic of the particular lexicographic theory underlying the dictionary data.

Of course, the conversion of already existent XML-based dictionary data into LD might solve the aforementioned issues, and other similar ones, at the modelling level, but still not at the source. We argue that, by solving such issues at source, LD-native dictionaries will make lexicographers' work more efficient and will make the *consistency of lexicographic data* easier to maintain, given that redundancies are more easily avoided.

LD-native dictionaries will maximise *re-usability of lexical knowledge* during the lexicographic compilation process. For instance, a lexical entry can be characterised by synonyms. In a hierarchical arrangement, such synonyms are nested under their associated entry and there is no guarantee of their existence as lexical entries for themselves. In an LD set-up, each synonym is designed as a new node in the graph and then linked to the initial lexical entry through a synonymy relation. Such a new lexical entry only needs to be defined once, no matter the number of times it appears in the dictionary. External re-use of lexical knowledge is also enhanced via link declarations (in RDF) to other LD sources. That enables, for instance, the re-use of grammatical categories already defined in external catalogues (e.g., LexiInfo[4]), the import of additional semantic descriptions from encyclopaedic resources such as BabelNet[5] or DBpedia (Auer et al., 2007), or the connection of different LD-based dictionaries.

Conceiving a dictionary as LD from scratch has also another advantage. In previous XML to LD conversion experiences, it was necessary to preserve as much information content as possible in order to keep the process reversible. This has led to the propagation of superfluous information into RDF, such as internal dictionary identifiers of the lexical elements or information related to how lexicographic data are displayed in a user interface. In the latter case, we argue that such information should be maintained apart from the purely lexicographic graph. In the former case, the definition of URIs for every lexical element makes the internal identifiers redundant. Further, well designed URIs will avoid

[4] http://www.lexinfo.net/ontology/2.0/lexinfo.owl
[5] http://babelnet.org/

collision of identifiers when integrating several dictionaries, which might be a risk if only internal dictionary identifiers were used.

3. Impact on the editorial process

The lexicographic compilation process generally attempts to represent language in a faithful and authoritative manner, whether inscriptively or descriptively, author-created or corpus-based, for reception or production purposes, and to present the results of the lexicographer's investigation and analysis in one dictionary format or another, as considered to be the most suitable for that editorial concept and most beneficial to the user. The entry microstructure is determined accordingly, to best reflect the items of linguistic knowledge selected by the lexicographer, and is arranged in some hierarchical system, whether historically or by order of frequency, with or without definitions, descriptions or translation equivalents, and examples of usage or citations, accompanied or not by relevant attributes such as synonyms and antonyms, register and geographical or dialect information, grammatical, usage or etymological notes, etc. The dictionary can thus resemble a closed world, with each element minutely selected and designed by the creator, and the end result expressing that mastermind and vision.

Overall, this approach is still valid today for lexicography at the wake of the LD era, even though the resources in service of the lexicographer are tremendously multiplied. At this stage, we are only starting to reveal and get acquainted with the new possibilities and horizons offered by LD lexicography, alongside its related requirements and priorities. In this section we analyze the impact LD-native dictionaries will have on the work of lexicographers in several aspects and their related challenges.

Modelling. For a dictionary to be created in LD, we first have to select the kind of information its entries will cover, and make sure this information is indeed representable as LD by available mechanisms. Once the information that an entry will capture is decided upon (syntatic, semantic, pragmatic, phonetic, etc.), the selection of available vocabularies, and the models to represent them, will proceed in order to create the model that will be the backbone of the editing tool that the lexicographer will be using to generate the data. Modelling challenges include the representation of the sense hierarchy, translations, examples, inflections, homographs or multimedia content in a way that stays true to the lexicographer's view and maximizes re-usability according to the LD principles. However, as mentioned above, the major shift that lexicography would experience involves a transition from a hierarchical ordering of the information recorded in a dictionary entry into a graph structure with its nodes uniquely identified by URIs, whose form should be also determined by the editor. The lexicographer will be required to identify the precise nature of the relation between two pieces of information by using ontological properties rather than unbounded textual descriptions. This echoes the difference between compiling dictionaries with only the human as target or creating them for both humans and computers.

Basic knowledge on LD. Even though expert knowledge of RDF and SPARQL[6] should ideally not be required on the lexicographer's part, he or she would need to assimilate the principles of LD lying at the heart of lexicographic compilation. By doing so, the editor will be able to unlock the potential of both using different URI naming strategies and linking to diverse external or internal resources to enrich his or her own data, for example.

[6] https://www.w3.org/TR/rdf-sparql-query/

Technical needs. Even though developing dictionaries natively in LD would allow them to be integrated into bigger knowledge systems and consumed by LD-aware NLP applications from the very beginning, the daily tasks that human users perform with the help of dictionary data should not be relegated to the background. In this respect, a clear challenge that we must face as we envision this ecosystem is the lack of a well-established and solid mechanism for everyday dictionary users to query LD resources without the need to rely on Semantic Web and LD knowledge. In order to build LD-native dictionaries, tools for graph editing and visualization would be called for to enable the compilation without expert knowledge of Semantic Web formalisms. Natural language and guided interfaces on top of SPARQL would evolve into essential tools for the editor to query the different LD versions created during the editorial process and thus control the project's progress. A paradigm-shift in lexicography would involve reconsidering the skills that are required both from lexicographers and editors as well as from the potential users of such linked dictionaries. Just as new natural language or guided interfaces will be called for in order for non-experts to query the datasets, their maintenance in terms of modification, enrichment and quality control on part of the editor will require new mechanisms as well.

Quality control. As reported in recent surveys on LD quality (Zaveri et al., 2016), there are aspects concerning data quality that are original to LD and therefore will need to be taken into account in LD-driven lexicography. Quality can be assessed through different dimensions, ranging from availability, licensing and security (accessibility dimensions) to data accuracy, consistency, etc. (intrinsic dimensions) and reputation and verifiability, among others (trust dimensions) (see Zaveri et al. (2016) for a state-of-the-art account on LD quality). Although each dictionary data provider may define its own criteria, they all share a common goal with respect to the intrinsic data, namely to provide lexical information that is semantically and syntactically correct, compact (i.e, without redundant data), complete (gathering all available data concerning an entry), and logically consistent (without contradictions or conflicting values). Processes aimed at evaluating the quality of the ontology in support of the dictionary editing phase, as well as for assessing the quality of the generated instances would need to take place as part of the regular lexicographic workflow.

4. Making the graph grow

LD technologies enable the vision of an ecosystem of linked lexicographical resources in the form of a giant cloud of lexicographic data at a Web scale. This heterogeneous cloud could consist of several hubs of dictionaries, each containing data from the same dictionary family or type. This does not mean that all the information must be open and publicly accessible. Different licences and exploitation schemes could be supported,[7] including public and free dictionary data, data with conditional access (e.g., accessible under payment) or closed data internal to a company. Lexical resources are conceived from different theoretical backgrounds and with dissimilar goals and use cases, so that not all of them are equally integrable into a single dictionary. Such an ecosystem will be explorable along several dimensions (language, grammatical information, granularity level). In that sense the traditional notion of a dictionary is diluted because different views or aggregations of data are possible depending on the user's needs (Spohr, 2012).

[7] Declared by means of specialised vocabularies such as http://purl.oclc.org/NET/ldr/ns

4.1 Dictionary data as an asset for LD

As a result of our initial experiences in adapting existing, pre-LD, data into LD, the first thing we discover is that regardless of how fine and well-structured such data may be, and how successful its conversion from e.g. XML to RDF format is, there is a fundamental difference stemming from how such data were originally conceived. Basically, what we look for are the best points of automatic connection to other linguistic data sources and among any sets of data, which can be optimized by further annotation for its use in NLP applications such as word sense disambiguation and induction. Moreover, that, in turn, might lead us deeper into standardization, which facilitates such linking. Our principle observations from this experience so far can be summed as follows:

Metalanguage. The metalanguage that is part of the lexicographic editorial process (e.g., names of attributes, parts of speech, language tags, etc.) is an asset for LD'fying the content, as it helps to uniformize the names of the entry components and their various bits of information, and thus to enhance the communication with other datasets.

Free text. Some of the texts that are written freely by the lexicographer as additional semantic, syntactic or pragmatic information besides the predefined labels seem to be the least valuable for LD, as it is harder to relate them to specific and precise details in other sources. This does not concern definitions and examples of usage, which often contain semantic categories, semantic relations, collocates and so on, which may be useful for sense disambiguation and thus for LD.

Subject field. Tagging the 'domain' of each sense of an entry tends to generate the most accurate *sense-to-sense* linking to other data resources. Unless the specific sense is tagged appropriately, we perform general *word-to-word* linking and might obtain poor results for polysemous lemmas. Different resources do not necessarily use the same 'subject field' tag, for example the monetary aspect of *bank* can be labelled *finance* in one place and *economics* or *commerce* in others, but the relation between these domains is fairly simpler to establish. There is no standard list of domains that is applied universally, not even borrowed from the world of terminology. One of the most highly regarded domain lists is that of the Library of Congress,[8] which is more complex and detailed than lists often used in dictionaries, but its system of sub-classification (e.g. *Art* includes *Painting*, *Sculpture*, *Architecture*, etc.) makes it more precise and suitable for LD.

Attributes. Various types of attributes that can be very helpful for word sense disambiguation in lexicography play a minor role for LD. For example, the register and geographical groups are not relevant enough, nor is grammatical information and patterns in general, such as 'range of application' and inflected forms. Synonyms and antonyms form a group of their own, though failing to offer full one-to-one linking, they serve to expand the semantic field of a word or a sense and may be helpful for indirect linking (surprisingly enough, antonyms tend to be more precise than synonyms in carrying relevant information, and could therefore be more useful for word sense disambiguation and thus for LD). This perhaps accentuates the function of LD as a vehicle for Semantic Web technologies, which must nourish primarily on semantic information.

[8] http://loc.gov/catdir/cpso/lcc.html

4.2 Challenges of interconnecting LD dictionaries

In the rest of this section we discuss some challenges related to making LD-native dictionaries grow and interconnecting them. In particular, we discuss aspects related to interlinking and data integration.

Interlinking. As a first step, lexical or general conceptual resources would need to be identified as suitable linking targets (Villazón-Terrazas & Corcho, 2011; Vila-Suero et al., 2014). Among the numerous datasets already available in the cloud of linguistic linked open data,[9] BabelNet[10] and DBpedia (Auer et al., 2007) emerge as the conceptual encyclopaedic resources with the highest in-degree of links, thus acting as pivotal elements among multiple language datasets. LD-based systems aimed to support the automatic discovery and validation of such relations among language resources would be required to assist the lexicographer at this stage.

Data integration. Services should be developed on top of the LD-based ecosystem that, given a query, aggregate data from the different entries and offer users a unified representation. In this way, the system acts as a 'single dictionary' that is actually the sum and combination of many of them, which are in turn managed separately and developed independently. The major challenge that we would meet here is the fact that information about the same dictionary entry would be sometimes repeated and scattered throughout the cloud of linked dictionaries. Each dictionary would be likely to show some differences in its underlying schema even though elements of the de facto standards had been re-used, especially if the editorial choice involved the use of a custom ontology. Some of the tasks that we would face in this stage have been already addressed in the literature in the LD integration context (Bleiholder & Naumann, 2009; Knap & Michelfeit, 2012), namely: schema matching, duplicate detection, and data fusion.

Schema matching refers to the detection of equivalent schema elements in the different sources (Bleiholder & Naumann, 2009). Proprietary schemas developed for the compilation of a dictionary often have equivalent counterparts in linguistic data category registries, such as LexInfo, but this is not always the case: mismatches between proprietary schema values for a specific DTD tag and individuals of an homologous class in an already available linguistic vocabulary can occur as well. Mappings between the dictionary editorial's custom ontology and other models thus become crucial for overcoming these difficulties.

Duplicate detection is the task of detecting equivalent resources to integrate data into one single and consistent representation (Bleiholder & Naumann, 2009). This means that information repeated across different linked dictionaries, e.g. the part of speech of a lexical item, should be presented only once in the answer to a query on the datasets. The problem arises when dissimilar values are extracted from different dictionaries and conflicts need to be resolved as part of the **data fusion** step. Compatible values which however are different in granularity (e.g. *noun* and *common noun*) would need to be distinguished from different and contradictory ones for the same dictionary entry (e.g. *common noun* and *proper noun*). As reported in the literature (Bleiholder & Naumann, 2009; Knap & Michelfeit, 2012), these conflicts would need to be either avoided (in our *proper noun* and *common noun* example, no information about the part of speech would be given), ignored

[9] http://linguistic-lod.org/llod-cloud
[10] http://babelnet.org/

(both values are presented as parts of speech), or resolved with a set of conflict handling strategies, which, for example, identify some sources as more trustworthy than others.

5. Conclusions

LD is generating a rising interest in the area of lexicography, and many dictionaries have been already converted into LD. In this paper, however, we have focused on what constitutes a step beyond by introducing the notion of LD-native lexicography. That is, dictionaries that will be compiled as LD from scratch. We have analyzed the main advantages of this networked approach in contrast with the more traditional tree-oriented view. We have also discussed its potential impact on the lexicographic data and on the work of lexicographers.

In this current intermediate phase between traditional and LD-driven lexicography, the observations described in this paper prompt us to revise existing lexicographic resources with LD in mind, and prioritize and emphasize certain ingredients, such as the subject field, and modify the entry structure. At the next stage, in the aim of being instantly understood by machines as part of *machine-to-machine* communication for the benefit of human beings, future LD-native lexicography will be considerate of these and other factors from its very conception and inception and throughout its compilation and usage. Although the LD format is displayed within an equalized non-hierarchical graph, its linking points are absolutely crucial. Metaphorically, while all the skin of our body is a living organ and sensitive to touch, it has a few points that serve most commonly for touching, like the fingertips. LD-native lexicography will attribute special attention to any such fingertips, as its precious communicative agents.

6. Acknowledgements

Supported by the LDL4HELTA EUREKA project,[11] the Spanish Ministry of Economy and Competitiveness through the ReTeLe Excellence Network (TIN2015-68955-REDT) and the Juan de la Cierva program, and by the Spanish Ministry of Education, Culture and Sports through the FPU program.

7. References

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R. & Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. *The semantic web*, pp. 722–735.

Bizer, C., Heath, T. & Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3), pp. 1–22.

Bleiholder, J. & Naumann, F. (2009). Data fusion. *ACM Computing Surveys (CSUR)*, 41(1), p. 1.

Bosque-Gil, J., Gracia, J. & Gómez-Pérez, A. (2016a). Linked data in lexicography. *Kernerman Dictionary News*, pp. 19–24. URL http://kdictionaries.com/kdn/kdn24.pdf#page=19.

Bosque-Gil, J., Gracia, J. & Montiel-Ponsoda, E. (2017). Towards a Module for Lexicography in OntoLex. In *Proc. of the LDK workshops: OntoLex, TIAD and Challenges for Wordnets at 1st Language Data and Knowledge conference (LDK 2017), Galway, Ireland.* CEUR-WS.

[11] https://ldl4.com/

Bosque-Gil, J., Gracia, J., Montiel-Ponsoda, E. & Aguado-de Cea, G. (2016b). Modelling Multilingual Lexicographic Resources for the Web of Data: the K Dictionaries case. In *Proc. of GLOBALEX'16 workshop at LREC'15, Portoroz, Slovenia.*

Declerck, T., Wandl-Vogt, E. & Mörth, K. (2015). Towards a Pan European Lexicography by Means of Linked (Open) Data. In I. Kosem, M. Jakubíček, J. Kallas & S. Krek (eds.) *Proceedings of eLex 2015. Biennial Conference on Electronic Lexicography (eLex-2015), electronic lexicography in the 21st century: Linking lexical data in the digital age, August 11-13, Herstmonceux, United Kingdom.* Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd.

Gracia, J., Villegas, M., Gómez-Pérez, A. & Bel, N. (2016). The Apertium Bilingual Dictionaries on the Web of Data. *Semantic Web Journal.* URL http://www.semantic-web-journal.net/system/files/swj1419.pdf.

Klimek, B. & Brümmer, M. (2015). Enhancing lexicography with semantic language databases. *Kernerman DICTIONARY News*, 23, pp. 5–10.

Knap, T. & Michelfeit, J. (2012). Linked Data Aggregation Algorithm: Increasing Completeness and Consistency of Data. *Provided by Charles University.*

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

Měchura, M. (2016). Data structures in lexicography: from trees to graphs. In *Proc. of 10th Workshop on Recent Advances in Slavonic Natural Language Processing (RASLAN 2016).* URL http://www.lexiconista.com/raslan2016.pdf.

Parvizi, A., Kohl, M., González, M. & Saur'ı, R. (2016). Towards a Linguistic Ontology with an Emphasis on Reasoning and Knowledge Reuse. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016.* URL http://www.lrec-conf.org/proceedings/lrec2016/summaries/523.html.

Polguère, A. (2014). From Writing Dictionaries to Weaving Lexical Networks. *International Journal of Lexicography*, 27(4), pp. 396–418. URL https://hal.archives-ouvertes.fr/hal-01097112/document.

Spohr, D. (2012). *Towards a multifunctional lexical resource: Design and implementation of a graph-based lexicon model*, volume 141. Walter de Gruyter.

Vila-Suero, D., Gómez-Pérez, A., Montiel-Ponsoda, E., Gracia, J. & Aguado-de Cea, G. (2014). Publishing linked data: the multilingual dimension. *Towards the Multilingual Semantic Web*, pp. 101–118.

Villazón-Terrazas, B. & Corcho, O. (2011). Methodological guidelines for publishing linked data. *Una Profesión, un futuro: actas de las XII Jornadas Españolas de Documentación: Málaga*, 25(26), p. 20.

Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J. & Auer, S. (2016). Quality assessment for linked data: A survey. *Semantic Web*, 7(1), pp. 63–93.