From dictionaries to cross-lingual lexical resources

Guadalupe Aguado-de-Cea, Elena Montiel-Ponsoda, Ilan Kernerman and Noam Ordan

1 Introduction

While the number of general resources that are connected as part of the linked open data paradigm increases, the need to relate and link linguistic data in multiple languages as a result of this trend has rocketed as well. The vision of a universe that allows linguistic information from different resources to be interlinked has attracted many scholars in search of "the magic wand" for solving the everlasting problem of the Tower of Babel, which now includes languages for machines in addition to human users. Currently, most linguistic resources are still in proprietary formats, making it difficult to be linked and interoperate on the Web. To achieve that envisioned linked cloud of linguistic resources, several issues have to be addressed, from representation models to linking processes, from querying interfaces to dataset maintenance solutions.

Great advances in methodologies and techniques for the publication of linked data are laying solid foundations for turning independent databases into a boundless cloud where users can make queries in an integrated environment using dedicated, standardized querying languages, thus catering for interoperability as well as fostering univocity of the elements described. Linked data relies on the Resource Description Framework (RDF) data model[1] as the main mechanism applied to describe data. These data in turn are linked to other similarly modelled data, and ultimately retrieved and manipulated by using Web standards such as the SPARQL[2] query language.

Many language resources have seen the advantages of complying to this new paradigm, and are currently available as part of the Linguistic Linked Open Data (LLOD) Cloud[3], a sub-cloud of the linked open data cloud that brings together linguistic resources formalized in RDF (from lexicons, dictionaries, and terminologies to metadata repositories and corpora). However, as in the case of the traditional Web, the LLOD is mainly English-oriented, though more non-English data sources are increasingly being published. As stated by Gracia et al. (2011), the new challenge is to overcome language barriers if we aim to attain a truly multilingual Semantic Web.

WordNet[4] (Fellbaum 1988), for example, which is the most widely used lexico-semantic resource in English with more than 117,000 synsets (sets of synonyms that account for a concept), has recently undertaken a new role in constructing the Semantic Web (Berners Lee et al. 2001). The W3C draft RDF/OWL Representation of WordNet[5] has defined URIs for the synsets covered by the WordNet lexical database. Many other efforts have been devoted to link WordNet to other resources. McCrae et al. (2012) used WordNet together with Wiktionary as a case study of the possible transformation of lexical resources into linked data compatible formats. In McCrae et al. (2014), the authors provide RDF-compliant Wordnet with links to other lexical resources, such as VerbNet[6], Lexvo[7] or lemonUby.[8]

As for multilingual linguistic resources which are part of the current LLOD cloud, it is worth mentioning IATE RDF[9] (Cimiano et al. 2015), AGROVOC[10] and EUROVOC in SKOS[11], or the APERTIUM[12] series of bilingual dictionaries (all of which are navigable and searchable from Datahub[13]). Several chapters of DBpedia[14] are now available in different languages, as well as some language versions of EuroWordNet (the Basque[15] and Catalan[16] versions present a case in point). However, what still remains

1 https://www.w3.org/standards/techs/rdf#w3c_all/
2 http://www.linkeddatatools.com/querying-semantic-data/
3 http://linguistic-lod.org/llod-cloud/
4 https://wordnet.princeton.edu/
5 https://www.w3.org/TR/wordnet-rdf/
6 http://verbs.colorado.edu/~mpalmer/projects/verbnet.html/
7 http://www.lexvo.org/
8 http://lemon-model.net/lexica/uby/
9 https://datahub.io/es/dataset/iate-rdf/
10 https://datahub.io/es/dataset/agrovoc-skos/
11 https://datahub.io/es/dataset/eurovoc-in-skos/
12 https://datahub.io/es/dataset/apertium-rdf/
13 https://datahub.io/
14 http://linghub.lider-project.eu/datahub/dbpedia/
15 http://linghub.lider-project.eu/datahub/basque-eurowordnet-lemon-lexicon-3-0/
16 http://linghub.lider-project.eu/datahub/catalan-eurowordnet-lemon-lexicon-3-0/

Guadalupe Aguado-de-Cea is Professor at Universidad Politécnica de Madrid (UPM). She received both her MSc in Translation and PhD in English Philology from Universidad Complutense de Madrid, and has been a member of the Ontology Engineering Group at UPM since 1996. Her current research activities include, among others: terminology and ontologies, the representation of lexical knowledge in ontologies, multilinguality in linked data, specialized languages as well as the linking between the ontological field and the natural language field, especially in its application to the Semantic Web. She has participated in several standardization projects, such as the Ontology Lexica Community Group (Ontolex) in the W3C, in particular regarding the representation of translation relations among languages with a view on the multilingual Web. She is the President of the Spanish Association for Terminology, and convenor of the AENOR CTN_191 Terminology Committee, the corresponding Spanish Committee of ISO TC 37.
lupe@fi.upm.es

Elena Montiel-Ponsoda is Associate Professor at the Applied Linguistics Department at Universidad Politécnica de Madrid (UPM) since 2012, and member of the Ontology Engineering Group since 2006. She got her PhD on Applied Linguistics from UPM in 2011. Her research interests are at the intersection between translation (and terminology) and knowledge representation, including among others: ontology localization and lexicalization, lexico-syntactic patterns for ontology development, functional models for deep semantics analysis, sentiment analysis, and linguistic linked data for content analytics. She is currently working on the representation of lexical resources according to the linked data paradigm, specifically, on how translation relations can help in the construction of the multilingual Web of Data.
emontiel@fi.upm.es

a challenging issue is the flawless linking of complementary resources in different natural languages. By complementary resources, we refer to resources that deal with the same (or closely related) parcels of knowledge, be it general or domain-specific knowledge, whose metadata descriptions as well as actual data are in different natural languages. In this sense, we argue that Semantic Web approaches and technologies are ripe enough to offer viable solutions to the linking issue in a principled manner.

Our objective in this contribution is to report on our experience in modelling the linked data version of the Spanish set of the K Dictionaries (KD) multi-language Global Series that will serve to transform a multilingual dictionary into a cross-lingual lexical resource. We would like this to set ground for discussion to define open issues for the linkage of lexical data in multiple languages, and some solutions are suggested on the base of *de-facto* standard *lemon-ontolex* model[17], initially designed to serve as an interface between an ontology and the natural language descriptions that lexicalize the knowledge represented in it, and currently widely adopted for exposing linguistic resources as linked data. Specifically, we describe how multilingual information in the RDF version of KD's dataset has been represented according to the *vartrans* module, a *lemon-ontolex* module for representing translations and term variants, and how this could contribute to enhance interoperability among the different language versions of the Global Series.

The paper is further structured as follows. In the next section we refer to the background and motivation, i.e. approaches to linking multilingual lexical and/or conceptual resources. Section 3 introduces the KD approach and Section 4 presents the formal solution we have adopted for its Spanish dataset in the linked data model, specifically, the *lemon-ontolex vartrans* module. The actual modeling of the Spanish dataset from the XML proprietary format of the dictionaries is spelled out in Section 5. In Section 6 we list some advantages of complying to this or similar formalisms in the context of the linked data paradigm, and our conclusions are presented in Section 7.

2 Background and motivation

When approaching this issue in the Semantic Web field, it is inevitable to refer to a former, much older discussion on how to bring together lexicons in different languages and how to solve discrepancies resulting from the idiosyncratic categorization of each language system/culture, some of which are reflected in the way different linguistic features, such as gender, pronouns or classifiers, are encoded (cf. Fellbaum and Vossen 2007). One of the resources that better materializes (and tries to solve) this problem is EuroWordNet (Vossen 1998), and subsequently derived projects such as MultiWordNet (Pianta et al. 2002). Broadly speaking, such databases connect wordnets or lexicons in different languages via a set core of categories, the so-called Interlingual Index, based on Princeton WordNet (Miller 1995). In the case of EuroWordNet there is an implicit bias towards English synonym sets which allegedly stand for concepts realized lexically by lexical items in different languages, and, in the case of MultiWordNet, the bias is more explicit, because the English WordNet is literally translated into the various languages, and gaps are declared by free translations that stand for those concepts, allowing linked concepts/synsets to percolate through the gaps.

In this regard, we would argue that different language-culture couplings (we see this as a binomial) can exhibit different levels of granularity when representing and categorizing knowledge. Even among culturally-related languages, such as Italian and English, it has been shown that a medium-sized dictionary of English to Italian contains around 7.8% lexical gaps, where there is no equivalence and a free translation is needed to fill the gap (Bentivogli and Pianta 2000). Therefore, and in order to address these issues, the Global WordNet Grid (Fellbaum and Vossen 2007; Vossen et al. 2016) initiative aims at providing a platform for centralising all wordnets and their linkage, and coordinating the inclusion of new concepts for multiple languages. As such, this latter approach represents an important step towards a more principled solution to the multilingual (still unresolved) issue.[18]

Another approach that also builds on WordNet, but which has been born in the Semantic Web era, is BabelNet (Navigli and Ponzeto 2012). This is a semantic network and ontology that aims at bringing together words and terms in different languages, from various resources, which refer to the same concept, with the objective of serving as valuable sources of translation or equivalent relations. According to Moro and Navigli (2015), in BabelNet it is possible "to find the

17 https://www.w3.org/community/ontolex/wiki/Final_Model_Specification/

18 cf. http://compling.hss.ntu.edu.sg/omw/

concept *medicine* (bn:00054128n), which is represented by both the second word sense of *medicine* in WordNet and the Wikipedia page *Pharmaceutical drug*, among others, together with synonyms such as *drug* and *medication* in English and lexicalizations in other languages, such as *farmaco* in Italian and *medicamento* in Spanish". In this way, BabelNet combines the general-specific approach taken from WordNet with the specific knowledge extracted from Wikipedia (and other resources, e.g. OmegaWiki). As for the English-language bias issue, it is probably propagated to this resource, since WordNet is taken as a starting point. However, it can also be reduced, because of the use of Wikipedia entry pages for categories not initially included in the original WordNet.

Apart from acknowledging the great value of such a resource, we have also spotted some flaws that will undoubtedly be solved in the future, and which are probably due to automating the linking process. For instance, some synsets contain words that belong to different categories. An example is the synset for *paella* (typical Spanish rice dish), which also includes the pan used to cook it. As for the translations in BabelNet, when different options are offered, we would suggest that additional information is required, such as confidence scores associated to the proposed translation, pragmatic restrictions (for instance, the frequency with which a word in language A is translated with the proposed equivalent in language B), or directionality of the translations. Means such as these would positively contribute to enhance this resource's functionality.

All in all, and although many advances have been made in the alignment and linking of resources in different languages, it is still necessary to cater for certain aspects in order to make the most of the multilingual information contained in such resources.

3 The K Dictionaries approach

The dictionary data used as input in this research belong to the Global Series of K Dictionaries (KD)[19]. KD is a technology-oriented-content creator that specializes in developing pedagogical and multilingual lexicographic data. In 2005 it launched the Global Series, which today includes lexical resources for 24 languages. The approach followed in this series is to compile for each language a core vocabulary as a standalone project, and have it translated to other languages in more projects. In other words, there is no bias towards any language, each is represented on its own terms, and only at a later phase it is translated to another, creating a pair-specific, and thus pair-sensitive, interlingual representation.

The outset of each language dataset in this series concerns mapping its components to identify, categorize and interlink them, including semantic and grammatical information. Each language core then serves as a base for adding translation equivalents in other languages and developing bilingual and multilingual versions. All the different language datasets share the same common methodological framework and technical infrastructure. The entries in the different languages also have the same microstructure, which still enables each one to convey its peculiarities. The data is structured in XML format and is currently being modeled in RDF. The French dataset, for instance, has the most extensive multilingual reach so far with 18 language pairs, the German lexical dataset groups 8 more languages, Spanish has 7, Japanese – 7, English – 6, Norwegian – 6, etc. Now that several language sets have become so lexically rich, they are ripe to start networking with each other, such as by connecting L2 translations to their corresponding entries in the L1 lexical dataset and from there on to translations in other languages, and so on.

As explained in the introduction, we reflect here on some interesting issues spotted when transforming the Spanish lexical core of the Global dataset, focusing on multilingual ones. We leave aside the methodology followed in the modeling part, which has been described in greater detail in Bosque-Gil et al. (2016a and 2016b), and move on to the resulting representation of translations in the proposed model.

4 *lemon-ontolex* at a glance: The *vartrans* module

In order to link and represent the linguistic data included in KD's Global Spanish dataset we relied on the *lemon-ontolex vartrans* module. It presents wide possibilities to link lexical senses and variants in different languages from the same or different data sets. As shown in Figure 1, the lexico-semantic generic class addresses the relation between two lexical entries or two lexical senses. This relation is established by means of two properties: `lexicalRel` and `senseRel`. Thus, `lexicalRel` relates two lexical entries that are grammatically or stylistically connected, such as acronyms, derivatives and other forms.

The second class, `senseRel`, represents the relation between two senses whose meanings are related. Not only can

19 http://kdictionaries.com/

Ilan Kernerman is CEO of K Dictionaries, leading its lexicographic development and international cooperation. He edits and publishes *Kernerman Dictionary News*, co-edited and published two collections of conference papers (1998, with Tom McArthur, and 2010, with Paul Bogaards), and is associate editor of *Lexicography – Journal of Asialex* and guest co-editor of the special *IJL* issue on bilingual learners' dictionaries (2016, with Arleta Adamska-Sałaciak). His interests include multilingual and pedagogical lexicography, and interoperability with NLP and knowledge systems. Currently he is president of Asialex (2015-2017) and on the preparatory board of Globalex.
ilan@kdictionaries.com

Noam Ordan studied translation, linguistically and computationally, and completed his PhD at Bar Ilan University under the supervision of the late Miriam Schlesinger. He has published extensively, in particular on automatically identifying translated texts and statistical machine translation, worked as researcher and teacher in universities in Israel and Germany, and took part in various projects in the industry. Currently he coordinates research innovation at K Dictionaries and designs algorithms for using human-crafted lexicographic data for computational tasks, such as cross-lingual information retrieval. Dr Ordan also serves as an adjunct teacher at the English Language Department at the Arabic Academic College in Haifa.
noam@kdictionaries.com

lexico-semantic relations, such as synonymy, antonymy or hypernymy-hyponymy be represented in this way, but also term variants and translations. The purpose of such a representation is to account for two lexical senses of terms (in the same or different language) that are semantically related in the sense that they can be exchanged in most contexts, but their surface forms are not directly related. Additionally, other types of semantic and pragmatic information, such as dialectal, registerial, chronological, discursive, and dimensional variation can also be captured by `senseRel`.

5 Modelling multilingual entries in the KD data with *vartrans*

The starting point in the transformation of the multilingual information (translations) contained in KD's Global Spanish dataset was a 'Translation cluster' that encompassed a set of translations for the original Spanish lexical entry, including syntactic-semantic and pragmatic information about the translations (e.g. grammatical gender), and usage examples of the headword (commonly a short phrase), as well as translations of those examples.

See Example 1 for the XML encoding of the headword *acalorado* (*heated*), which contains a synonym, namely, *agitado* (*lively* or *passionate*), a definition, *que es muy animado* (*of a discussion or debate, that is heate*d), and translations into Dutch (*verhit* and *vurig*) and Norwegian (*ivrig*, *oppsatt*, and *opphetet*). Moreover, this sense of *acalorado* is complemented with a usage example (*una sesión acalorada*), and its equivalents in Dutch (*vurige zitting*) and Norwegian (*et opphetet møte*), respectively, are all included in the ExampleCtn type and identified by means of a translation cluster identifier given in the XML, TC00001664.

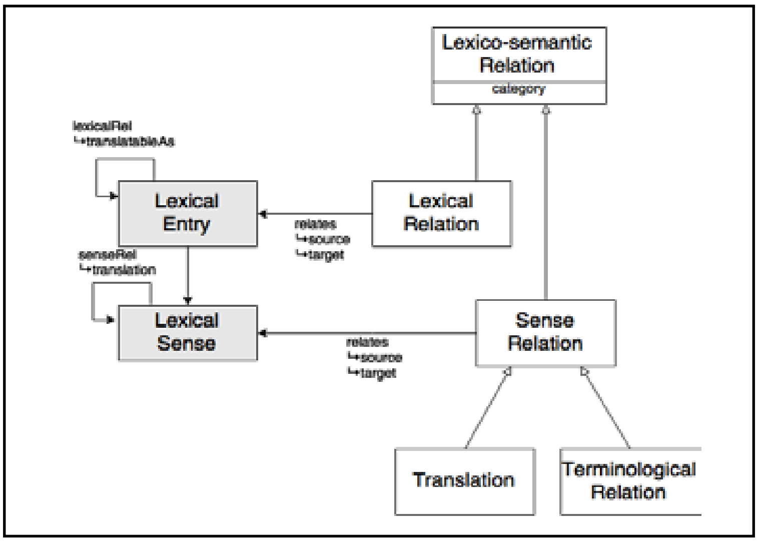

Figure 1. Classes and properties in the *vartrans* module

According to *lemon-ontolex*, a dictionary entry or headword in the KD set is modeled as an `ontolex:LexicalEntry` and its corresponding `ontolex:LexicalSense` and `skos:Concept`, as can be seen in Figure 2. Then, according to the *vartrans* module, synonym relations are modeled as relations between lexical senses that point to (`ontolex:reference`) the same concept (`skos:Concept`). Thus, for example, the lexical entry for the headword *acalorado* is linked to its corresponding sense and concept, and an artificial sense is created for the synonymous lexical entry *agitado*, so that a sense relation of the type synonymy can be established between them. Should *agitado* have also its own headword in the dictionary, a link could be established between the lexical senses later on, or lexical senses could be merged. Both lexical senses refer to the same `skos:Concept`, and a definition is also attached to the latter.

Similarly, translations are modeled as relations among lexical senses. Again, if we analyze Figure 2, the lexical sense for the entry in the source language (*acalorado*) is available, and the sense for the target language (*verhit*) has to be artificially created, since no pointer to that entry in other dictionaries is provided in the XML data (once the Dutch and Norwegian datasets are converted to RDF, these entities can support the automatic linking and growth of both datasets). The usage examples that accompany the senses are represented by means of the property `skos:example` and the class `kd:UsageExample`. Moreover, examples of usage are commonly translated into other languages and grouped by the `kd:TranslationExampleCluster`, a grouping made in the original datasets and maintained here.

The modeling solution proposed by the *vartrans* module for representing a translation relation by means of a reified class instead of a property or relation facilitates the further description of the translation object. In this sense, `translationSource` and `translationTarget` can be further specified, as done for the current version of the KD Spanish set. Also, other features that describe a certain translation relation could be added. For example, a confidence value can be assigned to the translation pair if available. A context could be determined to restrict the validity of the translation pair and differentiate it from other possible translations of the original entry into the target language. In fact, if we consider the usage examples available for *acalorado* in the XML dataset, *una sesión acalorada* (*a heated session*) has been translated into Dutch as *vurige zitting*, and not as *verhite zitting*, which was the synonym provided.

And the same happens with the Norwegian alternatives, the phrase is translated as *et opphetet møte*, and as learners of Norwegian we may wonder if the other two synonyms offered for *opphetet*, namely, *ivrig* and *oppsatt*, can be interchangeably used in that phrase.

Additionally, we may want to specify the type of translation relation that exists between a pair of translation equivalents. Gracia et al. (2014) propose a classification of translation equivalents into three types: direct equivalents (lexical entries in the translation pair that are semantically equivalent), cultural equivalents (lexical entries that are not semantically equivalent, but are pragmatically so), and lexical equivalents (the target lexical entry – or translation equivalent – verbalizes the original entry in the target language but is not a semantic or pragmatic equivalent). For more details we address the interested reader to the above-cited paper.

Therefore, apart from specifying the origin and target of the translation pair, the other descriptions that could further enrich the information related to were not available in the original source and have not been implemented in the current version. That does not mean such descriptions could not be added or imported from another resource that contains data to that respect. In fact, this is one of the main benefits of adopting the linked data paradigm, namely, being able to link to resources containing complementary information.

6 Advantages of cross-lingual lexical resources

Our reflections in this paper are made to point out some advantages of linking multilingual datasets in the aim of getting the most of the multilingual data value chains in the cloud of linked data. We argue that the linked data representation formalism offers an innovative way of bringing together resources in which either the vocabularies or models, or the data itself, are described in different natural languages, contributing to the construction of a truly multilingual Semantic Web. The challenge here is to account for as comprehensibe as possible specifics of each language taken *individually* while at the same time to represent links with meaningful labels across languages within a multilingual graph.

In the specific case of the lexical resources under examination, we argue that by representing translations as links between lexical senses (and, in turn, lexical entries), whenever new datasets that contain information in the target languages are also represented according to this paradigm, links will be flawlessly established. As already mentioned in previous sections of this paper, once the different datasets of the Global Series are available in RDF, links will be established among the different entities, contributing to an automatic growth of the resources. If we take the example of KD's Global Spanish dataset, since it contains translations into Brazilian Portuguese, Dutch, English, Japanese, and Norwegian it is reasonable to assume that relying on those translations, links will be easily created among the different datasets.

Although this is still a visionary

```
<SenseGrp identifier=”SE00000730” version=”1”>
     <Synonym>agitado</Synonym>
     <Definition>que es muy animado</Definition>
     <TranslationCluster identifier=”TC00001663” text=”que es muy
animado” type=”def”>
       <Locale lang=”nl”>
         <TranslationBlock>
           <TranslationCtn>
             <Translation>verhit</Translation>
           </TranslationCtn>
           <TranslationCtn>
             <Translation>vurig</Translation>
           </TranslationCtn>
         </TranslationBlock>
       </Locale>
       <Locale lang=”no”>
         <TranslationBlock>
           <TranslationCtn>
             <Translation>ivrig, oppsatt, opphetet</Translation>
           </TranslationCtn>
         </TranslationBlock>
       [...]
     </TranslationCluster>
     <ExampleCtn type=”sid” version=”1”>
       <Example>sesión acalorada</Example>
         <TranslationCluster identifier=”TC00001664” text=”sesión
acalorada” type=”exmp”>
         <Locale lang=”nl”>
           <TranslationBlock>
             <TranslationCtn>
               <Translation>vurige zitting</Translation>
             </TranslationCtn>
           </TranslationBlock>
         </Locale>
         <Locale lang=”no”>
           <TranslationBlock>
             <TranslationCtn>
               <Translation>et opphetet møte</Translation>
             </TranslationCtn>
           </TranslationBlock>
         [...]
       </TranslationCluster>
```

Example 1: XML with the translations in Dutch and Norwegian of the Spanish headword *acalorado* sense of *heated*

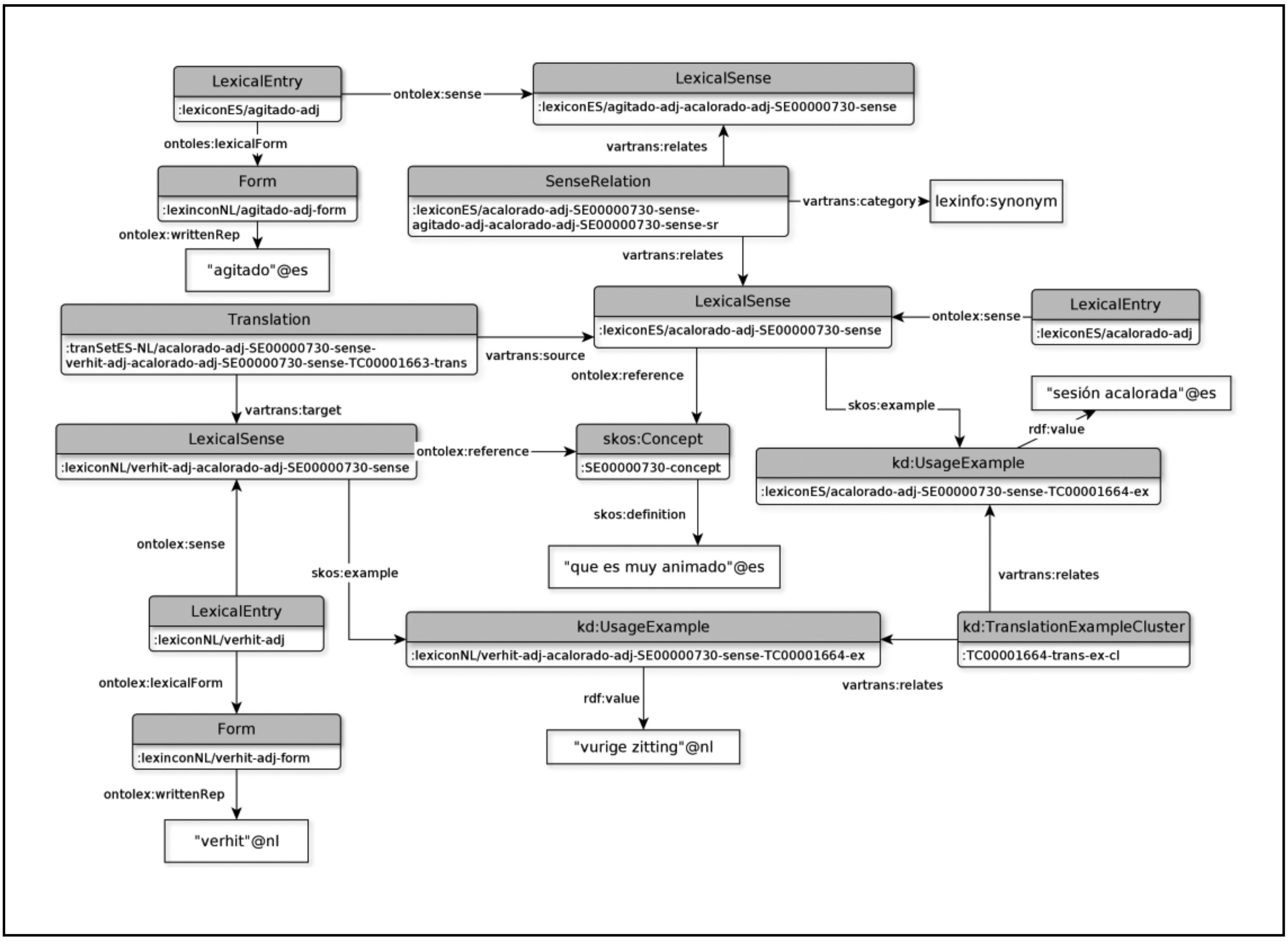

Figure 2. Modeling of a KD multilingual entry with *lemon-ontolex*

concept, representing lexical resources according to this approach will enable the emergence of a cross-lingual graph in a bottom-up fashion. This will maintain the distributed fashion of the linked data graph, and datasets will be easily connected, disconnected or contextualized for specific users and uses.

Contrary to the approaches described in state-of-the-art projects within the Global Grid initiative, we believe no common set of concepts or intermediary conceptualization would be needed to establish cross-lingual relations, but links would emerge among datasets at a different pace. Put differently, instead of relying on a common conceptualization to act as intermediary, the *burden of the cross-lingual connection* would be carried by the links.

At a monolingual level, since the relation between synonyms or terminological variants has been also reified in the `TerminologicalRelation` class, we could also determine precisely if a certain synonym or term is used in a specific context, or if all the synonyms related to the same concept can be interchangeably used. In the example of the BabelNet *medicine* concept mentioned in Section 2, we could identify accurately the specific uses of *medicine* versus *Pharmaceutical drug*, *drug* or *medication*. Are they used in the same contexts? Which is the most appropriate translation for *medicamento* in Spanish in an informal setting?

This is also specifically relevant in those cases in which complex linguistic descriptions are associated to conceptual structures. Let us consider the example of *biosanitary waste,* in general, and *hospital waste*, only for the waste produced in hospitals. If the difference between these concepts is established at the conceptual level, the two terms will most probably be associated to two different concepts. Conversely, if only one concept is represented in the ontology, we may still want to account for both terminological variants in the linguistic model, and explicitly state the motivation behind each denomination. In this way, we would also facilitate the linking of this data source to another data source contained in a different dataset and to which only the term *biosanitary waste* has been associated.

7 Conclusions

Following the experiences in this project we can claim that the publication of lexical and terminological resources as linked data will result in an enriched unified graph of lexical entries, senses and translations on the Web. Consequently, more information (additional notes, glosses, descriptions) will be retrieved by querying the linked data resources by means of SPARQL queries. Moreover, they could be enriched with pictures, audio, and the like, as has been successfully implemented in BabelNet, for example. However, having stated the benefits of linking linguistic resources, and more specifically the advantages of this initiative when applied to multilingual lexical resources, we are also aware of the challenges that still need to be tackled and that have been discussed in Section 6.

Acknowledgements

This work is supported by the 4V Spanish National Project (TIN2013-46238-C4-2-R), the Spanish Excellence Network ReTeLe (TIN2015-68955-REDT), and the LDL4HELTA project under the EUREKA program.

References

Bentivogli, L., and Pianta, E. 2000. Looking for lexical gaps. In *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000*, pp. 663-669.

Berners-Lee, T., Hendler, J. and Lassila, O. 2001. The Semantic Web. *Scientific American*, May 2001. 29-37.

Bosque-Gil, J., Montiel-Ponsoda, E. Gracia, J. and Aguado-De-Cea, G. 2016a. Terminoteca RDF: a Gathering Point for Multilingual Terminologies in Spain. *12th International Conference on Terminology and Knowledge Engineering (TKE 2016*).

Bosque-Gil, J., Gracia, J., Montiel-Ponsoda, E. and Aguado-De-Cea, G. 2016b. Modelling multilingual lexicographic resources for the Web of Data: The K Dictionaries case. In *Globalex 2016 Lexicographic Resources for Human Language Technology*.

Cimiano, P., McCrae, J., Rodríguez-Doncel, V. Gornostay, T. Gómez-Pérez and B. Simoneit. 2015. Linked Terminology: Applying Linked Data Principles to Terminological Resources. In *Proceedings of the 4th Biennial Conference on Electronic Lexicography*.

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

Fellbaum C. and P. Vossen 2007. Connecting the Universal to the Specific: Towards the Global Grid. In *Proceedings of The First International Workshop on Intercultural Collaboration (IWIC 2007)*, Kyoto, Japan, January 25-26.

Gracia, J. Montiel-Ponsoda, E. Cimiano, P. Gómez-Pérez, A. Buitelaar, P. and J. McCrae. 2011. Challenges for the multilingual Web of Data. In *Web Semantics: Science, Services and Agents on the World Wide Web* 11: 63-71.

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

McCrae, J., Cimiano, P., and Montiel-Ponsoda, E. 2012. Integrating WordNet and Wiktionary with lemon. In C. Chiarcos, S. Nordhoff, S. Hellmann, (eds.), *Linked Data and Linguistics: Representing and Connecting Language data and Language Metadata. Heidelberg & New York: Springe*r, 25-34.

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

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

Montiel-Ponsoda, E., Bosque-Gil, J.,Gracia, J. Aguado-de-Cea, G. and Vila-Suero, D. 2015. Towards the integration of multilingual terminologies: an example of a linked data prototype. *TIA 2015, Granada, Spain*.

Moro A. and Navigli R. 2015. SemEval-2015 Task 13: Multilingual All-Words Sense Disambiguation and Entity Linking. In *Proceedings SemEval 2015*, Denver, Colorado, June 4-5, 2005, 288-297.

Navigli, R. and Ponzetto, S. P. 2012. BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. *Artificial Intelligence*, 193, 217-250

Pianta, E., Bentivogli, L., and Girardi, C. 2002. Developing an aligned multilingual database. In *Proceedings of the 1st International Conference on Global WordNet*.

Vossen, P. 1998. *A multilingual database with lexical semantic networks*. Dordrecht: Kluwer Academic Publishers.

Vossen, P., Bond, F. and McCrae, J. 2016. Toward a truly multilingual GlobalWordnet Grid. In *Proceedings of the 8th Global Wordnet Conference* (GWC 2016).

This paper was presented at META FORUM 2016 in Lisbon, Portugal on 5 July 2016.
http://meta-net.eu/events/meta-forum-2016/

Elena Montiel-Ponsoda at META

Interns @ KD 2015-2016

● Universitat Jaume I, Castelló
Lucia Belles Calvera.
Lidia Gallen Martinez.
Miriam Martinez Garcia.
Catalan-Spanish index and *English-Catalan dictionary*

● KU Leuven, Antwerp
Rhiannon Telery Hincks.
English-Welsh dictionary
Zrinka Knezovic.
Croatian-English index
Liubava Panchenko.
Ukrainian-English index

● Université de Lorraine, Nancy
Pauline Pierrot.
French dictionary