Towards a module for lexicography in OntoLex

Julia Bosque-Gil, Jorge Gracia and Elena Montiel-Ponsoda

1. Introduction

Over the past few years, more and more efforts are being devoted towards the conversion of dictionaries into Linguistic Linked Data (LLD), based on Lemon (McCrae et al., 2012) and its more recent version OntoLex-Lemon[1], a de facto standard to represent ontology-lexica on the Web. These works aim both to enrich the so-called Linguistic Linked (Open) Data cloud[2] with lexical information to be consumed by natural language processing (NLP) tools, and to build bridges between the lexicography and semantic web communities. Recent projects such as LIDER[3], or on-going ones such as ENeL[4], LDH4HELTA[5] and LiODi[6], promote the adoption of linked data technologies in the work with lexicographic resources focusing on language technologies, e-lexicography and linguistic research, respectively.

Nonetheless, the conversion of a lexicographic resources to OntoLex is not always straightforward. Lemon was initially developed to enrich a given ontology with a lexical layer, and not with the idea of rendering any already existent dictionary to LLD. A majority of scholars working on this field, however, are turning to Lemon or OntoLex in pursuit of the latter objective. The more numerous and resource-specific the annotations in a dictionary are, the more complex the modeling solutions are, especially if until then the dictionary was targeted at human users. We are aware that some solutions exceed the needs of lexical information that some NLP tools require. However, if we are also aiming to bring linked data to lexicography, all dictionary content must be taken into account and must be retrievable once converted to LLD, i.e, migrating to LLD should imply no information loss. This means that structural aspects of the dictionary, as for instance senses and homographs order, along with the sub-sense hierarchy some dictionaries display, should be kept in mind when offering modeling solutions. There is a range of dictionary annotations (domain of usage, region, frequent use tags, restrictions on number and gender depending on a sense, etc.) that affect word meaning and language usage and are not structural in nature. Collocations, idioms, context indicators, semantic selection, etc. are presented differently in dictionaries and modeling them is not trivial.

The natural doubt that would be entertained by many experts is whether OntoLex is supposed to provide the means to model all aspects of a dictionary or whether this is outside its scope of ontology lexicalization, and therefore should be tackled by another initiative. In this paper we motivate our insights on OntoLex to enable dictionary representation as LLD in all its granularity, and advocate for the creation of a lexicography-specific module that would gather elements concerning dictionary structure and annotations. The module could also link to other modules that might be proposed, such as an etymology-oriented one to support etymological dictionaries.

The rest of the paper is organized as follows: Section 2 goes through state-of-the-art of LLD and lexicography and some of the problems encountered during the representation of dictionaries as LLD. Our motivation for OntoLex to be able to tackle those and the issues presented throughout the paper are stated in that section as well. Section 3 describes five of a series of issues we identified in our work modeling and analyzing dictionary entries, and which we argue serve as input for discussion on the need for a module for lexicography. Our initial approaches towards such a module and a description of how it would solve the described issues are outlined in Section 4, while Section 5 offers some concluding remarks.

2. Background and motivation

There have been several reports in the literature on the conversion of dictionaries to LLD, most of them relying on Lemon or OntoLex. However, proprietary formats, such as that of K Dictionaries (KD)[7], often have XML tags used in their annotation schemes that refer to linguistic categories or features which are not present in available repositories of linguistic categories or which lack a compatible definition that prevents us from reusing the ontology entity at hand. Ad

1 http://www.w3.org/2016/05/ontolex/
2 http://linguistic-lod.org/llod-cloud
3 http://lider-project.eu/
4 http://elexicography.eu/
5 http://ldl4.com/
6 http://acoli.informatik.uni-frankfurt.de/liodi/

7 http://kdictionaries.com/

Julia Bosque Gil is a PhD student at the Ontology Engineering Group (OEG) since October 2014. 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, 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 and lexicography. jbosque@fi.upm.es

Jorge Gracia is a postdoctoral researcher at Ontology Engineering Group, Universidad Politécnica de Madrid. 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. His current research focus is on 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/

hoc vocabularies were defined to migrate content from the German monolingual dictionary of KD's Global Series (Klimek and Brümmer, 2015) and its Spanish multilingual set (Bosque-Gil et al., 2016). These works approached issues which affect, for example, the relation between a lexical sense and the lexicalized phrases and idioms in which it occurs, regional restrictions, lexical and semantic selection (in general) of lexical entries, groups of homographs, tone and register indications, inflection groups, context of use, frequency modifiers to register, etc. Multilingual dictionaries pose further problems due to the modeling of examples and translations of examples, as well as alternative forms of those translations (e.g. an example in English translated to Japanese in kanji and hiragana, and that translation in turn with a transliteration in rōmaji). The set of thirteen dictionaries (dialectal, bilingual, monolingual, historical, etc.) converted as part of the ENel Action (Declerck and Wandl-Vogt, 2015) required the definition of new properties to encode different types of temporal information and etymological aspects.

Structures typically found in dictionaries, such as the sense and sub-sense hierarchy in an entry, are not trivial to model either. polyLemon (Khan et al., 2016), developed as part of the conversion of the Liddell-Scott Greek-English Lexicon to Lemon, was suggested in order to capture the sense and sub-sense structure in dictionaries using properties such as `senseChild` and `senseSibling` to relate senses and their parent senses in the dictionary entry.

The accurate representation of etymological information as LLD is key in the conversion of historical and etymological dictionaries. An extension to Lemon, Lemonet, to represent etymological information of lexical entries was proposed (Chiarcos and Sukhareva, 2014) and, more recently, a revisited version builds upon the properties suggested for the modeling of the etymological WordNet[8] to undertake the conversion of the Tower of Babel (Starling) in the LiODi project (Abromeit and Fäth, 2016). Some recent work on the conversion of the classical Arabic dictionary *Al-Qamus* to Lemon and LMF has been undertaken (Khalfi et al., 2016), but no pointers or traceback to the original structure are given in the work.

Alternatives to the use of OntoLex are available as well. The Oxford Global Languages Ontology (OGL) (Parvizi et al. 2016) has been developed to model

---

[8] http://www1.icsi.berkeley.edu/~demelo/etymwn/

and integrate multilingual linguistic data from Oxford Dictionaries and emerges as an ontology exclusively created to meet dictionary representation requirements. It accounts for a range of information found in dictionaries, from inflected forms to semantic relations, pragmatic features and etymological data. The focus is laid on the representation of grammatical information with cross-linguistic validity and the respect towards grammar traditions. However, some modeling decisions and class definitions differ from those suggested in OntoLex (e.g. `Form` in OntoLex vs. a `Form` in OGL) and the emphasis is not set on the reuse of available ontology entities.

In this position paper we do not focus on a particular kind of lexical information present in dictionaries (e.g., etymology or morphology) but we aim to highlight some difficulties in the modeling of dictionary entries without information loss. Thus, we will not target the representation of resource-specific features of particular dictionaries. Taken into account the problems reported in the literature, and after analyzing dictionary entries in e-dictionaries of English (*Oxford Living Dictionaries Online*; *Merriam Webster Dictionary Online*; *American Heritage Dictionary Online*; *COBUILD Advanced English Dictionary* and *Collins English Dictionary*), German (*Duden Online Wörterbuch*; *PONS Deutsch als Fremdsprache Online Wörterbuch*), and Spanish (*Diccionario de la Lengua Española*; *Clave*: *diccionario de uso del español actual*), we report on some of the issues we gathered which may pose problems for the modeling with OntoLex and which we believe call for the definition of a new module to account for them. Future steps include the analysis of dictionaries in languages that are underrepresented in the LLOD cloud (e.g. Japanese) to identify further representation challenges.

We ground our proposal for a lexicography module on the following four points: (1) the use of OntoLex by the majority of the community to convert linguistic resources to LLD instead of to lexicalize ontologies, (2) the nature of Lemon being descriptive but not prescriptive and the respect towards different lexicographic views, (3) the coming together of the lexicography and the Semantic Web communities and potential benefits that LLD may bring about to lexicography, assuming it involves no information loss, and (4) the reuse of already available mechanisms in OntoLex.

3. Issues

In the following we report on some of the issues we have come across after our experiences in converting dictionaries to

LLD and our analysis of dictionary entries in English, German and Spanish. Here we restrain ourselves to issues that reveal current limitations of the OntoLex model, i.e, cases in which applying the Lemon core implies a different view on the data than the one provided in the original resource and, therefore, an information loss (type 1, hence T1), and missing entities, e.g. a property or a class, to account for information mostly found in dictionaries (type 2, hence T2). We have already raised some of these issues as input for discussion to the OntoLex community.[9]

Issue 1 (T1). Headwords that can take different parts-of-speech
Both Lemon and OntoLex specify a lexical entry as a word, a multiword expression or an affix with a single part-of-speech, morphological pattern, etymology and set of senses.[10] However, a headword in a dictionary may occur with different parts-of speech depending on context and its senses are nonetheless defined in the same dictionary entry, all of them derived from the same etymology (no homonymy involved). Applying the OntoLex model would imply the generation of several `ontolex:LexicalEntr[ies]`, one per each part-of-speech the headword can take. Splitting the dictionary entry into several lexical entries would cause loss of information (shared etymology, pronunciation, senses implicitly related) and does not keep track of the dictionary representation. Examples: *poison*, *bread*, *water* (noun and verb), and Spanish *lento* 'slow, slowly' (adjective and adverb) and *alto* 'tall, loudly, height' (adjective, adverb and noun).

Issue 2 (T1). Lexical sense requiring a particular form
Some senses of a dictionary headword require a particular form, e.g. in English a plural form or in Spanish a masculine or feminine one. Since the meaning in these cases is associated with the form and it may differ significantly from other senses that do share gender or number features, splitting the dictionary entry into different lexical entries would be an option (see Issue 1). An alternative is the linking of that sense to elements in a catalog of grammatical categories which encode those grammatical restrictions, but we would need an exhaustive list of them for this option to be applicable. Examples: *refreshment(s)*,

Spanish *cometa* (m.) 'comet', (f.) 'kite'. In these cases, the dictionary entry can be a single one (e.g. *refreshment* in English or *cometa* in Spanish) but one of its senses indicates a preferred form. In the case of *refreshment*, the plural form is used if the intended meaning is snacks and beverages; with *cometa*, the feminine form is applicable when referring to a kite, the masculine when denoting a comet. Further examples are *good(s)*, *manner(s)*; and Spanish *frente* (m.) 'front', (f.) 'forehead'.

Issue 3 (T2). Usage examples and their translations
Usage examples of a word or multiword expression are often provided in the definition of each of a dictionary entry's senses. Lexinfo[11] includes a property `lexinfo:senseExample` to describe an example of a sense (as a subproperty of `Lemon:definition`) and which is linked to the example data category in ISOCat.[12] Nonetheless, due to it being a datatype property, it does not enable including further information on the example or to establish translation relations among examples, which is common practice in bilingual and multilingual dictionaries. The Lemon model included a `Lemon:UsageExample` class and a property `Lemon:example` to link to it, but OntoLex does not cover this aspect yet. Examples: Spanish *preocuparse* 'worry'; *no hay por qué preocuparse* 'there is nothing to worry about' (*Collins English-Spanish Dictionary*).

Issue 4 (T2). Sense and homograph order
The order of senses may be based on frequency of use, date of origin, concreteness (from the most concrete to most abstract sense, etc.). Homographs are also given according to some ordering criteria that may vary from dictionary to dictionary. Their order should be searchable and retrievable as to recover the information provided in the original resource. Examples: *Boa*: noun. (1) any of a family (Boidae) of large snakes that kill by constriction and that includes the boa constrictor, anaconda, and python (2) a long fluffy scarf (*Merriam Webster Dictionary*)[13]; *bat1*: n. 1. A stout wooden stick; a cudgel [. . .]; *bat2*: n. Any of various nocturnal flying mammals of the order Chiroptera [. . .] (*American Heritage Dictionary*).

---

9 http://w3.org/community/ontolex/wiki/Lexicography
10 http://w3.org/2016/05/ontolex/#lexical-entries

11 http://lexinfo.net/ontology/2.0/lexinfo.owl
12 http://isocat.org/
13 Example of logical order of senses inspired by *Diccionario de la Lengua Española, Guía de Consulta*, http://dle.rae.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

OntoLex 2017
1st workshop on the
OntoLex model

The W3C OntoLex Community
Group was launched in 2011,
with Paul Buitelaar (INSIGHT.
National University of
Ireland, Galway) and Philipp
Cimiano (CITEC, Bielefeld
University, Germany) as
chairs, with the goal to define
a model for allowing to
represent lexical knowledge
in connection to ontologies
[1]. The rationale behind the
model is that semantics are
captured by ontologies, and the
role of the lexicon-interface
is to link lexical entries to
ontological entities expressing
their denotational meaning,
following a principle called
Semantics by Reference.
Based on five years of intensive
discussions and work, the new
OntoLex-Lemon model was
launched in May 2016 [2]
and is becoming the primary
method for representing linked
lexical resources on the Web
of Data, not only for capturing
the lexicon-ontology interface
but for the representation of
lexicographic resources as well.
The model facilitates bridging
the gap between the NLP and
data science communities by
making available and linking
large amounts of quality lexical
information to the knowledge
represented on the semantic
web, for example in graphs
such as DBpedia, applications

Issue 5 (T2). Semantic selection
Some dictionaries indicate the semantic
features of the lexical items that an entry (in
one of its senses) selects or even the exact
lexical items with which it collocates. This
is usually indicated either with a specific
tag (e.g. KD's Range Of Application), or
in-between parentheses at the beginning
of a definition. Examples are, for instance,
the dictionary entry for the German verb
dämmen, which in its sense 'to insulate,
absorb, mute' selects arguments that denote
warmth or sound (German *Wärme*, *Schall*,
etc.) (KD), the adjective *cozy*, meaning
beneficial to all those involved and possibly
somewhat corrupt if predicated from a
transaction or an arrangement (*Google
Dictionary*); or the collocational measure
words of *luck*: stroke, piece of (*Oxford
Collocations Dictionary*). The OntoLex
Syntax-Semantics Module (*synsem*) class
`synsem:OntoMap` allows to map a
syntactic frame to an ontology entity, so that
the frame and its arguments are linked to
the ontology elements that they lexicalize.
Even though dictionaries commonly include
information on subcategorization (transitive/
intransitive/reflexive etc. annotations for
verbs, for instance), details on the syntactic
frame are not always provided beyond those
annotations. Since in dictionary conversion
we often lack a given ontology and detailed
syntactic information is not provided, the
mapping between syntactic arguments and
ontology entities seems difficult to establish
automatically via `synsem:OntoMap`:
how do we automatically represent that the
adjective *cozy* has a meaning only applied to
transaction or agreement or that the measure
words that collocate with *luck* are *stroke* or
piece if the morphosyntactic information
provided in the dictionary is just that *cozy* is
an adjective and *luck* a noun? Furthermore,
`synsem:condition` (in its turn
subsuming `synsem:propertyRange`
and `synsem:propertyDomain`)
enables us to state constraints on the
arguments of a predicate in a given
ontology.14 The possibility of reusing it to
state the constraints on syntactic arguments
even in cases in which we lack a given
ontology and therefore are not mapping to
given ontology properties has to be further
analyzed. In addition, the potential links
between the modeled entries (e.g. *piece*
and *luck*), i.e. the links at the lexical level,
are also to be considered, for instance, by
taking into account recent proposals on the
representation of lexical functions as LLD
(Fonseca et al. 2016).

---

14 http://w3.org/2016/05/
ontolex/#conditions

4. A module for lexicography
The previous section dealt with some of
the issues we encountered in our work
with dictionaries and the potential ones that
may rise with other lexicographic works
that have not been migrated to LLD yet. In
the following we draft a potential solution
which can serve as a basis for a new module
in OntoLex specifically developed for the
representation of dictionaries after thorough
revision and improvement according to the
community's feedback.

In order to keep track of the dictionary
representation and prevent any loss of
information mentioned in Issue 1, related
to the splitting of dictionary entries in
several lexical entries, we propose a new
class `DictionaryEntry`. This new
class would both enable to group together
lexical entries as well as to associate any
information shared by all of them. In
our view, we distinguish lexical entries
and lexicons (as containers of lexical
entries), from the original dictionary entry
(the new class `DictionaryEntry`)
and the original dictionary resource
(`Dictionary`), which would serve in turn
to record the provenance of each dictionary
entry. Mirroring the `lime:Lexicon-
ontolex:LexicalEntry` relation
we suggest a `Dictionary-
DictionaryEntry` one. Any lexical
entry created during the conversion to LLD
but not originally provided in the resource
would then belong to a `lime:Lexicon`,
but not to the instance of `Dictionary`
representing that resource. A
`lime:Lexicon` in English, for example,
could aggregate lexical entries created on
the fly by the LLD expert or original ones
coming from as many English dictionaries
as desired. These dictionaries can in turn
differ in their modeling and their views on
the data, their criteria of sense ordering or
their structure.

Regarding Issue 2, the
`DictionaryEntry` class would allow
to divide a single lexical entry into several
ones if desired, each with a different
preferred form, while maintaining the
original dictionary representation. If the
dictionary entry is not split, the option of
linking a sense to a grammatical restriction
on gender or number from an external
catalog would solve the issue, although the
implications of this solution (its benefits and
drawbacks) will need further analysis.

In order to represent usage examples
and their translations (Issue 3) we propose
to go back to `lemon:UsageExample`
and link it to a `LexicalSense`. A new
class `ExampleCluster` would link to
`UsageExamples` that are translations
from each other. The use of the vartrans

module to model translations among senses would imply the creation of lexical senses for each example, and therefore treating the example as a lexical entry, which we deem is beyond the definition of lexical entries.

Issue 4 was concerned with the order of senses in a dictionary entry and the order of homographs in the macrostructure of the dictionary. There are different possible approaches to resolve this: reusing already available RDF mechanisms, reifying the sense order in a new class `SenseOrder`, or defining a new property `senseOrder` attached to the lexical sense. The first option involves the reuse of `rdfs:Container[s]` to declare with e.g. `rdf:_1, rdf:_2` that a particular sense is the first or the second one. However, cases in which a set of senses allows for various orderings, depending on the ordering criterion, or in which some senses come from different dictionaries (each with its order), should also be accounted for. The second option suggests that the sense order is reified in a class `SenseOrder` linked to the lexical sense. This class would enable us to record the position of that sense, its provenance (presumably an instance of the class `Dictionary`), and, if desired, the ordering criterion. If repeated senses were identified (e.g. senses that share a definition in both dictionaries), `SenseOrder` would allow us to have one single lexical sense with two different positions according to the two different orderings and dictionaries, in a similar fashion as two containers with two different sequences of senses. Alternatively, if we assume that a lexical sense always comes from just one dictionary source, a property `senseOrder` would suffice.

Issue 5, dealing with semantic selection, has been brought up for further discussion in this paper to see whether it could be covered by synsem module mechanisms or whether it would require new entities in the context of the lexicography module. As part of the conversion of the KD's *Global Spanish Multilingual Dictionary* (Bosque-Gil et al. 2016), the semantic selection information provided by KD's tag RangeOfApplication was captured by the use of `synsem:condition`. In that approach, `synsem:condition` would link a lexical sense to a blank node[15] with an `rdf:value` recording the strings given as arguments in the original data. This modeling allowed us to deal with the lack of a given ontology and detailed information on the syntactic frames of lexical entries for

---

15   `synsem:condition` has `rdfs:Resource` defined as its range.

each of their senses. Thus, the focus was set on representing the data just as it was in its original format while being compliant with the OntoLex formal specification and reusing its elements as much as possible. We argue that the lexicography module should aim to set the basis to exploit at the dictionary's macro-structure level the potential benefits of establishing semantic relations among lexical senses based on lexical selection or among syntactic frames and arguments and the ontology entities that they denote. To this aim, overcoming the lack of detailed syntactic information in the dictionary as well as the lack of a given ontology to lexicalize becomes essential.

5. Conclusion

OntoLex is increasingly being used to convert linguistic resources to LLD outside the scope of ontology lexicalization. In this position statement we have drawn attention to a series of issues raised in the literature on LLD related to the conversion of dictionaries to LD and to five of the ones we came across in the same line of work and after a later analysis of several additional dictionaries. We argue that the OntoLex model should enable the preservation of the content and the structure of the original resource, even if the LLD expert opts for a different representation that is better suited to the data exploitation by external applications or is more in line with his or her view on the lexicon-ontology interface. We have outlined some of our insights on how to address these issues in a new module for lexicography. It would be compatible with the mechanisms suggested in the state-of-the-art on dictionaries represented as LLD, as of the moment of writing, and also with other potential modules for the encoding of specific lexical aspects (e.g. etymology). The final module is intended to be dictionary-agnostic in the sense that it should be applicable (and combined with other modules if necessary) to different kinds of dictionaries (e.g., general, collocations, learner's, etymological, historical, etc.). This would bring linked data (LD) closer to lexicography not only with the aim of leveraging already available dictionaries in LD for NLP tasks, but also for introducing LD in the work carried out in that discipline.

Acknowledgments

as part of the Global WordNet Association Collaborative Interlingual Index, or existing resources such as BabelNet, DBnary and commercial dictionaries.

OntoLex 2017 was co-located with the Language, Data and Knowledge conference [3, and see p.13] and presented the first opportunity for practitioners to meet to discuss the model, its applications and future development [4]. The participants have shown interest in continuing the development of OntoLex-Lemon, particularly with regard to lexicographic resources. In consequence, the group is starting to work on a new best practice document that will provide modeling examples and guidelines for how to use OntoLex specifically to represent lexicographic resources such as dictionaries. Then it will be decided whether to stipulate the proposed vocabulary elements or modeling solutions into the status of a new module or to keep the best practices proposal as an informal document.

Philipp Cimiano
Universität Bielefeld

[1] https://w3.org/community/ontolex/
[2] https://w3.org/2016/05/ontolex/
[3] http://ldk2017.org/
[4] http://ontolex2017.linguistic-lod.org/

KDICTIONARIES

KD API

K Dictionaries is completing the development of an online API which will provide programmatic access to its rich cross-lingual lexicographic resources for 50 languages.

In addition to the well-formatted XML data, the API outputs developer-friendly JSON-LD, which complies with the familiar RESTful API standard. The JSON-LD encodes RDF linked data, making it highly compatible with complementary open linked linguistic data sets. Users will be authenticated and given access according to their account type.

There is also an editing pipeline in development in which editors, translators and crowdsourced suggestions will use API calls to consolidate suggested changes to K Dictionaries' data and direct them through quality assurance.

The API will be launched in September 2017. Registration is open at api@kdictionaries.com.

Morris Alper

Ministry of Education, Culture and Sports through the Formación del Profesorado Universitario (FPU) program.

This contribution was presented at the 1st Workshop on the OntoLex Model, co-located with the Language, Data and Knowledge Conference (LDK 2017) in Galway, Ireland, on June 18.

Dictionaries cited

American Heritage Dictionary Online. Retrieved 25/05/2017 from http://ahdictionary.com/

Clave: diccionario de uso del español actual (online). Retrieved 25/05/2017 from http://clave.smdiccionarios.com/app.php

COBUILD Advanced English Dictionary. Retrieved 25/05/2017 from http://collinsdictionary.com/

Collins English Dictionary. Retrieved 25/05/2017 from http://collinsdictionary.com/

Duden Online Wörterbuch. Retrieved 25/05/2017 from http://duden.de/woerterbuch

Merriam Webster Dictionary Online. Retrieved 25/05/2017 from http://merriam-webster.com/

Oxford Living Dictionaries Online. Retrieved 25/05/2017 from https://en.oxforddictionaries.com/

PONS Deutsch als Fremdsprache Online Wörterbuch. Retrieved 25/05/2017 from http://de.pons.com/

Real Academia Española, Diccionario de la Lengua Española (DLE). Retrieved 25/05/2017 from http://dle.rae.es

References

Abromeit, F., and Fäth, C. 2016. Linking the Tower of Babel: Modelling a Massive Set of Etymological Dictionaries as RDF. In *LDL 2016 5th Workshop on Linked Data in Linguistics*, 11.

Bosque-Gil, J., Gracia, J., Montiel-Ponsoda, E., and Aguado-de-Cea, G. 2016. Modelling multilingual lexicographic resources for the Web of Data: The K Dictionaries case. In *GLOBALEX 2016 Workshop at LREC 2016*, 65.

Chiarcos, C., and Sukhareva, M. 2014. Linking Etymological Databases. A Case Study in Germanic. *In 3rd Workshop on Linked Data in Linguistics: Multilingual Knowledge Resources and Natural Language Processing*, 41.

Declerck, T., and Mörth, K. 2016. Towards a Sense-based Access to Related Online Lexical Resources. In *Proceedings of Euralex 2016*, 342-355.

Declerck, T., and Wandl-Vogt, E. 2014. Cross-linking Austrian dialectal Dictionaries through formalized Meanings. In *Proceedings of Euralex 2014*.

Declerck, T., and Wandl-Vogt, E. 2015. Towards a Pan European Lexicography by Means of Linked Open. Data. In *Proceedings of eLex 2015*, 342-355.

Declerck, T., and Wandl-Vogt, E. 2014. How to semantically relate dialectal Dictionaries in the Linked Data Framework. In *8th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities LaTeCH 2014*, 9-12.

Del Gratta, R., Frontini, F., Khan, F., and Monachini, M. 2015. Converting the Parole Simple Clips Lexicon into RDF with lemon. *Semantic Web 64*, 387-392.

El Maarouf, I., Bradbury, J., and Hanks, P. 2014. PDEV-lemon: a Linked Data implementation of the Pattern Dictionary of English Verbs based on the Lemon model. In *Proceedings of Language Resources and Evaluation LREC*.

Fonseca, A., Sadat, F., and Lareau, F. 2016. Lexfom: a lexical functions ontology model. In *COLING 2016*, 145.

Gracia, J., Villegas, M., Gómez-Pérez, A., and Bel, N. 2016. The Apertium Bilingual Dictionaries on the Web of Data. *Semantic Web Journal*.

Khalfi, M., Nahli, O., and Zarghili, A. 2016. Classical dictionary *Al-Qamus* in lemon. In *Information Science and Technology CiSt., 2016 4th IEEE International Colloquium*, 325–330.

Khan, F., Díaz-Vera, J. E., and Monachini, M. 2016. Representing Polysemy and Diachronic Lexico-semantic Data on the Semantic Web. In *Proceedings of the Second International Workshop on Semantic Web for Scientific Heritage co-located with 13th Extended Semantic Web Conference ESWC 2016.*, Heraklion, Greece, 37-46.

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

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

Parvizi, A., Kohl, M., González, M., and Saurí, R. 2016. Towards a Linguistic Ontology with an Emphasis on Reasoning and Knowledge Reuse. In *Proceedings of Language Resources and Evaluation LREC*.

Villegas, M., and Bel, N. 2013. PAROLE/SIMPLE "Lemon" ontology and lexicons. *Semantic Web Journal* II.