The OntoLex-Lemon Model: development and applications

John P. McCrae[1], **Julia Bosque-Gil**[2], **Jorge Gracia**[2], **Paul Buitelaar**[1], **Philipp Cimiano**[3]

[1]Insight Centre for Data Analytics, National University of Ireland Galway
[2]Ontology Engineering Group, Universidad Politécnica de Madrid, Spain
[3]Cognitive Interaction Technology Excellence Cluster, Bielefeld University
E-mail: john@mccr.ae, jbosque@fi.upm.es, jgracia@fi.upm.es, paul.buitelaar@insight-centre.org, cimiano@cit-ec.uni-bielefeld.de

Abstract

The *lemon* model has become the primary mechanism for the representation of lexical data on the Semantic Web. The *lemon* model has been further developed in the context of the W3C OntoLex community group, resulting in the new OntoLex-Lemon model, recently published as a W3C report. In this paper, we describe the development and future outlooks for this model as well as briefly review some of its current applications. The recent evolution of *lemon* into OntoLex-Lemon, in the context of the community group, led to improvements on the model that further extends its application domain from formal applications such as question answering and semantic parsing to the representation of general machine-readable dictionaries, including WordNet and digitized versions of existing dictionaries.
We look at two use cases of the OntoLex-Lemon model in representing dictionaries and in the WordNet Collaborative Interlingual Index. Finally, we consider the future of the OntoLex-Lemon model, which we intend to continue to develop and have recently identified areas that increase the applicability and value of the model to more users.
Keywords: linked data; lexicography; ontologies; Semantic Web; ontology-lexicon interface

1. Introduction

Ontologies have become an increasingly important method for modelling domains and representing data in a variety of forms, most notably the Semantic Web. However the existing standards for ontologies, in particular the Web Ontology Language (McGuinness and Harmelen, 2004, OWL), provide little support for the representing information about how a word is expressed in language beyond a simple string. In order to close this gap, the *lemon* model (McCrae et al., 2012) was proposed, which created a separate lexicon that could describe how an ontological concept was lexicalized in more details. This builds on the paradigm of the ontology-lexicon interface, as well as existing models for lexicography including LMF Francopoulo et al., 2006 and the EAGLES[1] and ISLES projects[2], whereby how a concept is expressed in natural language and the formal description of the concept in the ontology is kept separated. This has several advantages, most notably in that by separating the ontological and the lexical layer we can easily switch an ontology from one language to another by changing its lexicon.

The *lemon* model was adopted by a number of projects (Ehrmann et al., 2014; Navigli and Ponzetto, 2012; Sérasset, 2015; Eckle-Kohler, McCrae, and Chiarcos, 2015) and several authors have proposed modifications, improvements or changes (Khan, Boschetti, and Frontini, 2014; Chavula and Keet, 2014; Bosque-Gil, Gracia, and Gómez-Pérez, 2016;

[1] `http://www.ilc.cnr.it/EAGLES/home.html`
[2] `http://www.ilc.cnr.it/EAGLES96/isle/ISLE_Home_Page.htm`

Gracia, Montiel-Ponsoda, et al., 2014) to the model. In order to accommodate these changes, it was decided that the model should be further developed under an open forum and some of the authors of this paper founded the OntoLex Community Group[3]. This group was part of the World Wide Web Consortium's Business and Community group program, a new initiative to support the development of emerging standards on the Web. The results of this group's work was the publishing of an updated version of the model in May 2016, namely the OntoLex-Lemon model.

The new OntoLex-Lemon model has already started to be applied in a number of cases. In this paper we will examine some of these use cases, in particular looking at the expanded use case of the model for representing existing dictionaries and look at the conversion of several existing commercial and free dictionaries. Secondly, we will look at the use of the OntoLex model in the recently proposed Global WordNet Interlingual Index (Vossen, Bond, and McCrae, 2016; Bond, Vossen, et al., 2016), whereby the model is used as a foundation for creating a truly interlingual concept index.

Finally, in this paper we will also provide an outlook of the next steps that we aim to achieve for the model, in particular in terms of the new modules that we aim to create in order to address concerns raised in the community. Thus, we briefly sketch four modules on morphology, lexicography, etymology (and diachronicity) and lexical categories.

2. The OntoLex Community Group

The OntoLex Community Group[4] was founded in December 2011 to support the development of a model for the representation of lexical information relative to ontologies. The group provided a number of tools for collaboration on this task including a wiki and a public mailing list for discussion of topics. Moreover, the group chaired by the authors of this paper organized public telephone conference calls, of which over 70 have taken place between 2012 and 2016. The group developed the model firstly by collecting relevant use cases[5], and then distilling this into a set of essential requirements[6] for the model. Then, the development of the model took place in two stages: firstly the *core* model was defined, which incorporates the basic elements that it was assumed that all applications of the model would use and then in the second stage, four extra modules were defined: Syntax and Semantics, Decomposition, Variation and Translation, and Metadata (Lime). Finally, these models were combined and documented in a final report that was published by the W3C (Cimiano, McCrae, and Buitelaar, 2016) along with technical model files in OWL.

One significant difference in the creation of this standard, in contrast to the processes of other standard organizations, was the degree of openness in the development of this model. The community group has over one hundred members from a very diverse number of institutes and this is due to the fact that admission to the group was conditioned only on assenting to a short agreement that any contributions would be open. Moreover, many issues of the model were decided by open conversation or votes and all of these contributions are available publicly in the form of wiki contributions and mailing list posts, all of which are archived on the Web and accessible to anyone[7].

[3] `http://www.w3.org/community/ontolex`
[4] `https://www.w3.org/community/ontolex/`
[5] `https://www.w3.org/community/ontolex/wiki/Specification_of_Use_Cases`
[6] `https://www.w3.org/community/ontolex/wiki/Specification_of_Requirements`
[7] `https://lists.w3.org/Archives/Public/public-ontolex/`

3. The OntoLex-Lemon Model

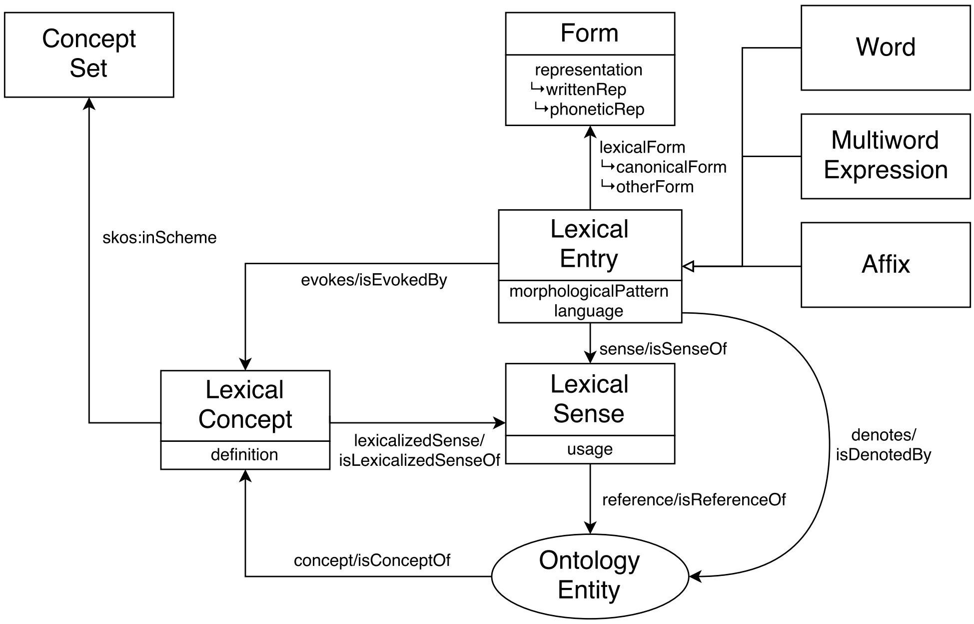

Fig. 1: The Core OntoLex-Lemon Model

Here we provide a brief summary of the OntoLex-Lemon model, for a more complete description please see Cimiano, McCrae, and Buitelaar, 2016. The OntoLex-Lemon model is based around the core module, as depicted in Fig. 1. The primary element of this is the *lexical entry* which represents a single word and thus collects together all morphological expressions of that word, which correspond to *forms* in the model, and all possible concepts in the ontology it can refer to, which correspond to *lexical senses* in the model. It is important to note that the actual meaning of a word is given by reference to an ontological concept and *lexical senses* represent only the mapping from a word to a concept. In contrast to the previous *lemon* model, a third semantic element called the *lexical concept* was introduced that allows for a meaning to be defined independently of an ontology. For example, the verb 'to die' may refer to different ontological properties such as `deathDate` and `deathPlace` while still referring to a single concept of *Dying*. The model also supports some other features including marking the canonical form (lemma), whether an expression is a multiword expression and giving a *usage* condition describing when a particular word expresses a given concept (for example the register), which is annotated on the lexical sense showing its role in giving a mapping between concepts.

In addition to the core, there are four modules defined by the specification:

Syntax and Semantics The syntax and semantics module describes how particular lexical constructs, e.g., verb frames, can be mapped to constructs in the ontology. As a simple example, this concerns how a transitive verb frame such as 'X knows Y' can be mapped to the subject and object arguments of a property such as `A foaf:knows B`. In this case there are only two options based on whether the grammatical subject (X) refers to the property's subject (A) or object (B), however more complex multi-argument structures are also covered.

Decomposition The decomposition module allows for multiword lexical entries to be decomposed into individual words, which are also represented as *components*. Components are allowed to be marked with their own grammatical properties and are said to *correspond to* either a lexical entry (i.e., for the word), an argument in a frame or another frame (to model phrasal arguments).

Variation and Translation The variation and translation model represents relationships between words at three levels: (purely) lexical, sense (lexico-semantic) and conceptual. These correspond to the levels of the model, with lexical relations being between lexical entries and as such not considering the meaning of a word and only its syntactic properties. Similarly, a conceptual relationship is between concepts and does not consider the lexical form and hence language of a relation. Sense relations require knowledge of both the word form and the meaning and translation is thus considered a special case of a sense relation. The module also allows technical modelling of a relation either as a single triple or as a dereferenceable entity in itself, which allows for further annotation of metadata about the link. This module integrates previously proposed extensions to *lemon* such as the translation module (Gracia, Montiel-Ponsoda, et al., 2014).

Metadata (Lime) The Linguistic Metadata (Lime) module (Fiorelli, Pazienza, and Stellato, 2013) adds modelling for grouping sets of lexical entries together into a lexicon and providing simple metadata such as the number of entries, senses, etc. Note that as linked data is intended to be published on the Web together the necessity to have all words grouped into a lexicon is not considered a core but still a useful feature.

4. Use cases

4.1 Representing dictionaries with OntoLex

In the past few years, the linguistic linked data community has showed a growing interest in the publication of dictionaries as linked data. The benefits of representing lexicographic content as linked data (LD) are twofold: on the one hand, LD resources are easily reused, gain in visibility and accessibility at a Web scale, their content can be seamlessly aggregated with content from external lexical resources (not necessarily dictionaries), as well as integrated and exploited by LD-aware Natural Language Processing (NLP) tools (Klimek and Brümmer, 2015; Gracia, Villegas, et al., 2016). On the other hand, LD brings about several advantages to the modeling of the macro and micro-structure of a dictionary (Bosque-Gil, Gracia, and Gómez-Pérez, 2016): moving beyond traditional cross-references, dictionary entries and each of their components are uniquely identified at a Web scale and become internally reusable thanks to URIs; hierarchical ordering of information is replaced by graph structures, where each node becomes a potential entry point to traverse the whole graph, and any relation between two elements is typed and defined in a vocabulary over which previously consensus was reached. The dictionary allows thus for an interpretation as a vast interoperable typed network of lexical elements, as opposed to the more traditional list-inspired view of it.

Initiatives such as the European Network for e-Lexicography (ENeL)[8], Linked data lexicography for high-end language technology application (LD4HELTA)[9] or the Linked Open Dictionaries (LiODi) project[10] foster the conversion of dictionaries to linked data as part

[8] `http://www.elexicography.eu/`
[9] `http://www.eurekanetwork.org/project/id/9898`
[10] `http://acoli.cs.uni-frankfurt.de/liodi/home.html`

of the adoption of the new technological advances in the Semantic Web by digital humanities.

As *lemon* and OntoLex gradually become widespread models for the conversion of lexical resources to linked data, dictionaries represented with them are easily integrable with other resources previously converted to RDF without any remodelling efforts. This means, in turn, that in many cases dictionaries go from a proprietary data model to one widely accepted by the community. In fact, dictionary conversion to linked data was already receiving much attention prior to OntoLex, and several contributions put forward LD-versions of dictionaries based on *lemon*. Examples of these are the family of bilingual dictionaries Apertium RDF (Gracia, Villegas, et al., 2016), the Germ monolingual dictionary in K Dictionary's Series (Klimek and Brümmer, 2015), sentiment lexica (Vulcu et al., 2014), the Parole-Simple lexica (Villegas and Bel, 2015), the Pattern Dictionary of English Verbs (El Maarouf, Bradbury, and Hanks, 2014), the classical Al-Qamus dictionary (Khalfi, Nahli, and Zarghili, 2016) and DBpedia lexicalizations such as DBlexipedia (Walter, Unger, and Cimiano, 2015), just to mention a few. Some of these efforts, e.g. Dbnary (Sérasset, 2015), called for the definition of new properties that at that time were not covered by *lemon* (e.g. `dbnary:isTranslationOf`) and that nowadays have a counterpart in OntoLex or by extension vocabularies such as LexInfo Cimiano, Buitelaar, et al., 2011. Recently, the interest is being directed towards the transformation of dictionaries which contain a variety of rich annotations and which are developed both for NLP purposes and human users. These include multilingual (Bosque-Gil, Gracia, Montiel-Ponsoda, and Aguado-de-Cea, 2016), dialectal (Declerck and Mörth, 2016), etymological (Abromeit et al., 2016), and ancient Greek (Khan, Díaz-Vera, and Monachini, 2016) dictionaries, among others (Declerck, Wand-Vogt, and Mörth, 2015). The work on these resources and the dictionaries mentioned above has lead to the proposal of extensions and modifications to OntoLex to account for specific informations ranging from etymological annotations, translations of examples, groups of inflections and temporal information to the sense-subsense hierarchy in a dictionary entry.

4.2 The Collaborative Interlingual Index

Princeton WordNet (PWN, Fellbaum, 2010) is the most widely used lexicographic resource for natural language processing, but yet is only available for English. There have been many versions of wordnets for other languages and these have been collected together in the Open Multilingual WordNet (Bond and Foster, 2013), however they have primarily been created by the *extend approach*, where existing synsets from PWN have been translated and then new synsets are added for words which do not exist in English. Unfortunately, this has led to a degree of fragmentation, where certain concepts may be independently defined by different wordnets. In order to address this issue, it has been proposed that all wordnets contribute to a single index of concepts (Pease, Fellbaum, and Vossen, 2008). This has recently been realized by the Collaborative Interlingual Index (CILI, Bond, Vossen, et al., 2016), in which all wordnets are converted to a common format and linking is made between the synsets. In order to do this, it is assumed that each concept must have both an English definition and a link to a synset already defined in the CILI.

```
{
  "@context": "http://globalwordnet.github.io/schemas/wn-json-context-1.0.json",
    "@graph": [{
      "@context": { "@language": "en" },
      "@id": "example-en",
      "@type": "ontolex:Lexicon",
      "label": "Example wordnet (English)",
      "language": "en",
      "email": "john@mccr.ae",
      "rights": "https://creativecommons.org/publicdomain/zero/1.0/",
      "version": "1.0",
      "entry": [{
          "@id": "w1",
          "lemma": { "writtenForm": "grandfather" },
          "partOfSpeech": "noun",
          "sense": [{
              "@id": "example-en-10161911-n-1",
              "synset": "example-en-10161911-n"
          }]
      }],
      "synset": [{
          "@id": "example-en-10161911-n",
          "ili": "i90287",
          "partOfSpeech": "noun",
          "definition": [{
              "gloss": "the father of your father or mother"
          }],
          "relations": [{
              "relType": "hypernym",
              "target": "example-en-10162692-n"
          }]
      }]
  }]
}
```

Fig. 2: An example WordNet in the Global WordNet JSON-LD format

In order to implement this, it has been necessary to define a common format for the definition of wordnets[11]. This format allows for three forms: XML, JSON and RDF, all of which can be converted without any loss of information. The XML format is based on the existing Lexical Markup Framework (Francopoulo et al., 2006) and in particular on the WordNet-LMF variant (Soria, Monachini, and Vossen, 2009). Both the JSON and RDF formats are based on this OntoLex model described in this paper, and the RDF version of this format is considered a limited *profile* of the OntoLex model, suited particularly for the case of representing wordnets. The JSON version more precisely defines its semantics by means of a JSON-LD context (Sporny et al., 2014). An example of this is given in Fig. 2, in which the term "grandfather" is defined. In this example, a number of required standard metadata properties are defined using widely-used vocabularies, namely Dublin Core (Weibel et al., 1998) and Schema.org[12]. Then the file contains two sections `entry` and `synset`, which define the *lexical entries* and *lexical concepts* in this lexicon. They both have a part-of-speech property, with specific values defined in a custom WordNet ontology[13]. The senses of the model correspond to the *lexical senses* of the OntoLex model. For synset and sense relations the variation modules is used that enables relationships between senses to be further described with metadata.

The use of linked data to represent the interlingual index has a number of advantages, most specifically that each ILI identifier is associated with a unique URL, where further information about the term can be found. For example, for `i1234` information about this resource can be obtained at `http://ili.globalwornet.org/ili/i1234`, including the definition of the concept in English as well as links to the PWN and other wordnets who have contributed their links to the ILI. The URL thus allows for a stable identifier that can be referred to unambiguously as opposed to the current method of referring to offsets in release files.

5. Extensions and Future Plans

The OntoLex Community Group released its "final report" on the 10th May 2016, however the work of the group has not yet stopped and the group has an ambition to develop more modules in response to critical analysis and novel uses case (such as Chavula and Keet, 2014). In particular, the group has recently aimed to develop four new modules in order to further extend the applicability of the model:

Morphology The first *lemon* proposal contained a module for "inflectional and agglutinative morphology", which primarily defined morphological processes by means of regular expressions. This methodology was very simple to implement in any programming language that support Perl-like regular expressions, however does not very accurately represent the phonological process that occur in word morphology. As such, under this model certain regular cases like the plural of 'leaf' to 'leaves' would be modelled as distinct morphological paradigms even though it is generally considered part of the normal paradigm of pluralization in English. Thus the original model was not included in the OntoLex model and has been made available as a standalone ontology called LIAM (Lemon Inflectional Agglutinative Morphology)[14]. There have since

[11] `https://globalwordnet.github.io/schemas`
[12] `https://webmasters.googleblog.com/2011/06/introducing-schemaorg-search-engines.html`
[13] `http://globalwordnet.github.io/schemas/wn`
[14] `http://lemon-model.net/liam`

been a number of new proposals for morphology and in particular the group is discussing the adoption of the MMoOn Ontology of (Klimek, Arndt, et al., 2016; Klimek, 2017), which enables the documentation of the morphological data of any inflectional language in RDF

Lexicography Previous experiences in the representation of dictionaries using OntoLex-Lemon, as the ones described in Section 4.1, have led to a number of issues[15] Bosque-Gil, Gracia, and Montiel-Ponsoda, 2017. In particular, these issues include associating senses with forms and syntactic information such as grammatical gender, adding examples, geographic information and ordering senses in terms of importance, along other aspects of dictionary information that are not always explicitly covered in the core Ontolex-Lemon model. As such, the OntoLex community has perceived the necessity of adding extra modelling to cover such issues. To this end, a new OntoLex Lexicography module will be built targeted at the representation of dictionaries and which will address structures and annotations commonly found in lexicography. Such a module is intended to be compatible with other modules in OntoLex (e.g., Etymology and Diachronicity) and should constitute a viable mechanism for lexicographers in the development of dictionaries as linked data.

Etymology and Diachronicity Some authors (Khan, Boschetti, and Frontini, 2014; Abromeit et al., 2016; Khan, Bellandi, et al., 2017) have proposed using the OntoLex model to represent dictionaries of historical languages, and moreover many dictionaries contain some etymological information. As such, the ability of a dictionary to represent the change of lexical items over time is important. Thus, it has been recognized that the development of a module to capture meaning of words over time is a key use case of the model.

Lexico-syntactic categories The OntoLex model follows a principle of avoiding prescriptive modelling, for example allowing individual applications to define their own categories. This is helpful as in the example of part-of-speech values in wordnets discussed above, where this approach allows the resource to define categories that may not be accepted by other lexicographers[16]. However, the definition of standard categories greatly helps interoperability between resources and the LexInfo ontology (Cimiano, Buitelaar, et al., 2011) has been used by a number of authors for this purpose (Buitelaar et al., 2013; Villegas and Bel, 2015; Sérasset, 2015). This resource, originally derived from the ISOcat (Kemps-Snijders et al., 2008) categories, is currently maintained as a single OWL file. As such, the group aims to re-evaluate this model and establish a procedure for adding new categories to a single ontology. This will still only be a suggestion for data categories and we expect particular communities to define their own ontologies.

6. Conclusion

The OntoLex model has been developed under an open process and as such represents on the most significant open models for the representation of electronic lexicographic resources. While the model as proposed has remained close to the proposal of McCrae et al., 2012, it has also significantly innovated in order to allow new use cases. In particular, the application of the model beyond the Semantic Web community has required new modelling, in particular the introduction of *lexical concepts* and dereferenceable relations.

[15] For more details see: `https://www.w3.org/community/ontolex/wiki/Lexicography`
[16] In particular, PWN defines ‘adjective satellite’ as a distinct category to ‘adjective’

These developments have seen the model adapted in a wider community and as such have consequently lead to requests for new features. The group remains committed to developing the model and new use cases in morphology and diachronic lexicography will further show the flexibility of this linked data based model.

Acknowledgement

We would like to thank all members of the OntoLex group for their contributions and discussion of the model. This work was supported in part by the Science Foundation Ireland under Grant Number SFI/12/RC/2289 (Insight), by 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 (UPM), and the CITEC excellence initiative funded by the DFG (Deutsche Forschungsgemeinschaft).

References

Book Sections

Fellbaum, Christiane (2010). “WordNet”. In: *Theory and applications of ontology: computer applications*. Springer, pp. 231–243.

Paper in conference proceedings

Abromeit, Frank et al. (2016). “Linking the Tower of Babel: Modelling a Massive Set of Etymological Dictionaries as RDF”. In: *5th Workshop on Linked Data in Linguistics: Managing, Building and Using Linked Language Resources*, p. 11.

Bond, Francis and Ryan Foster (2013). “Linking and Extending an Open Multilingual Wordnet.” In: *Proceedings of the 2013 Annual Meeting of the Association for Computational Linguistics (ACL)*. Vol. 1, pp. 1352–1362.

Bond, Francis, Piek Vossen, et al. (2016). “CILI: the Collaborative Interlingual Index”. In: *Proceedings of the Global WordNet Conference 2016*.

Bosque-Gil, Julia, Jorge Gracia, and Elena Montiel-Ponsoda (2017). “Towards a Module for Lexicography in OntoLex”. In: *Proceedings of the LDK workshops: OntoLex, TIAD and Challenges for Wordnets*.

Bosque-Gil, Julia, Jorge Gracia, Elena Montiel-Ponsoda, and Guadalupe Aguado-de-Cea (2016). “Modelling multilingual lexicographic resources for the Web of Data: The K Dictionaries case”. In: *GLOBALEX 2016 Lexicographic Resources for Human Language Technology Workshop Programme*, pp. 65–72.

Buitelaar, Paul et al. (2013). “Linguistic linked data for sentiment analysis”. In: *Proceedings of the 2nd Workshop on Linked Data in Linguistics*, pp. 1–8.

Chavula, Catherine and C Maria Keet (2014). “Is lemon Sufficient for Building Multilingual Ontologies for Bantu Languages?” In: *OWLED*. Citeseer, pp. 61–72.

Declerck, Thierry and Karlheinz Mörth (2016). “Towards a Sense-based Access to Related Online Lexical Resources”. In: *Proceedings of the 17th EURALEX International Congress*. Ed. by George Meladze Tinatin Margalitadze. Tbilisi, Georgia: Ivane Javakhishvili Tbilisi University Press, pp. 660–667. ISBN: 978-9941-13-542-2.

Declerck, Thierry, Eveline Wand-Vogt, and Karlheinz Mörth (2015). “Towards a Pan European Lexicography by Means of Linked (Open) Data”. In: *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.* Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd. Ljubljana/Brighton.

Ehrmann, Maud et al. (2014). “Representing Multilingual Data as Linked Data: the Case of BabelNet 2.0”. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14).* Ed. by Nicoletta C. Chair et al. Reykjavik, Iceland: European Language Resources Association (ELRA).

El Maarouf, Ismaıl, Jane Bradbury, and Patrick Hanks (2014). “PDEV-lemon: a Linked Data implementation of the Pattern Dictionary of English Verbs based on the Lemon model”. In: *Proceedings of the 3rd Workshop on Linked Data in Linguistics (LDL): Multilingual Knowledge Resources and Natural Language Processing at the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland*, pp. 88–93.

Fiorelli, Manuel, Maria T. Pazienza, and Armando Stellato (2013). “LIME: Towards a Metadata Module for Ontolex”. In: *2nd Workshop on Linked Data in Linguistics: Representing and Linking lexicons, terminologies and other language data.* Pisa, Italy.

Francopoulo, Gil et al. (2006). “Lexical Markup Framework (LMF)”. In: *Proceedings of 5th International Conference on Language Resources and Evaluation (LREC)*, pp. 233–236.

Gracia, Jorge, Elena Montiel-Ponsoda, et al. (2014). “Enabling Language Resources to Expose Translations as Linked Data on the Web”. In: *Proc. of 9th Language Resources and Evaluation Conference (LREC'14), Reykjavik (Iceland).* European Language Resources Association (ELRA), pp. 409–413. ISBN: 978-2-9517408-8-4.

Kemps-Snijders, Marc et al. (2008). “ISOcat: Corralling Data Categories in the Wild.” In: *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)*, pp. 887–891.

Khalfi, Mustapha, Ouafae Nahli, and Arsalane Zarghili (2016). “Classical dictionary Al-Qamus in lemon”. In: *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt).* Ed. by Mohammed El Mohajir et al. IEEE, pp. 325–330.

Khan, Fahad, Andrea Bellandi, et al. (2017). “The Challenges of Converting Legacy Lexical Resources to Linked Open Data using OntoLex-Lemon:The Case of the Intermediate Liddell-Scott Lexicon”. In: *Proceedings of the LDK workshops: OntoLex, TIAD and Challenges for Wordnets.*

Khan, Fahad, Federico Boschetti, and Francesca Frontini (2014). “Using lemon to Model Lexical Semantic Shift in Diachronic Lexical Resources”. In: *3rd Workshop on Linked Data in Linguistics.*

Khan, Fahad, Javier E Díaz-Vera, and Monica Monachini (2016). “Representing polysemy and diachronic lexicosemantic 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.* Vol. 1595, pp. 37–46.

Klimek, Bettina (2017). "Proposing an OntoLex - MMoOn Alignment: Towards an Interconnection of two Linguistic Domain Models". In: *Proceedings of the LDK workshops: OntoLex, TIAD and Challenges for Wordnets*.

Klimek, Bettina, Natanael Arndt, et al. (2016). "Creating Linked Data Morphological Language Resources with MMoOn - The Hebrew Morpheme Inventory". In: *Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC)*, pp. 892–899.

Pease, Adam, Christiane Fellbaum, and Piek Vossen (2008). "Building the global WordNet grid". In: *Proceedings of the 18th International Congress of Linguists (CIL18)*.

Soria, Claudia, Monica Monachini, and Piek Vossen (2009). "Wordnet-LMF: fleshing out a standardized format for wordnet interoperability". In: *Proceedings of the 2009 International Workshop on Intercultural Collaboration*. ACM, pp. 139–146.

Vossen, Piek, Francis Bond, and John P. McCrae (2016). "Toward a truly multilingual Global Wordnet Grid". In: *Proceedings of the Global WordNet Conference 2016*.

Vulcu, Gabriela et al. (2014). "Generating Linked-Data based Domain-Specific Sentiment Lexicons from Legacy Language and Semantic Resources". In: *5th International Workshop on Emotion, Social Signals, Sentiment and Linked Open Data, co-located with LREC 2014*.

Walter, Sebastian, Christina Unger, and Philipp Cimiano (2015). "DBlexipedia: A Nucleus for a Multilingual Lexical Semantic Web". In: *NLP-DBPEDIA@ISWC*. Ed. by Heiko Paulheim et al. Vol. 1581. CEUR Workshop Proceedings. CEUR-WS.org, pp. 87–92.

Journal Articles

Bosque-Gil, Julia, Jorge Gracia, and Asunción Gómez-Pérez (2016). "Linked data in lexicography". In: *Kernerman Dictionary News*, pp. 19–24.

Cimiano, Philipp, Paul Buitelaar, et al. (2011). "LexInfo: A declarative model for the lexicon-ontology interface". In: *Web Semantics: Science, Services and Agents on the World Wide Web* 9.1, pp. 29–51.

Eckle-Kohler, Judith, John P. McCrae, and Christian Chiarcos (2015). "lemonUby - a large, interlinked, syntactically-rich lexical resource for ontologies". In: *Semantic Web* 6.4, pp. 371–378.

Gracia, Jorge, Marta Villegas, et al. (2016). "The Apertium Bilingual Dictionaries on the Web of Data". In: *Semantic Web Journal*. ISSN: 1570-0844.

Klimek, Bettina and Martin Brümmer (2015). "Enhancing lexicography with semantic language databases". In: *Kernerman Dictionary News* 23, pp. 5–10.

McCrae, John P. et al. (2012). "Interchanging lexical resources on the Semantic Web". In: *Language Resources and Evaluation* 46.6, pp. 701–709.

Navigli, Roberto and Simone Paolo Ponzetto (2012). "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network". In: *Artificial Intelligence* 193, pp. 217–250.

Sérasset, Gilles (2015). "DBnary: Wiktionary as a lemon-based multilingual lexical resource in RDF". In: *Semantic Web* 6.4, pp. 355–361.

Villegas, Marta and Nuria Bel (2015). "PAROLE/SIMPLE 'lemon' ontology and lexicons". In: *Semantic Web* 6.4, pp. 363–369.

Technical Reports

Cimiano, Philipp, John P. McCrae, and Paul Buitelaar (2016). *Lexicon Model for Ontologies: Community Report.* W3C Community Group Final Report.
McGuinness, Deborah L. and Frank van Harmelen (2004). *OWL Web Ontology Language Overview.* W3C Recommendation.
Sporny, Manu et al. (2014). *JSON-LD 1.0.* W3C Recommendation.
Weibel, Stuart et al. (1998). *Dublin Core Metadata for Resource Discovery.* RFC 2413.