ELEXIS - European Lexicographic Infrastructure: Contributions to and from the Linguistic Linked Open Data

Thierry Declerck[1,2], John McCrae[3], Roberto Navigli[4], Ksenia Zaytseva[1], Tanja Wissik[1]
[1]Austrian Centre for Digital Humanities at the Austrian Academy of Sciences
[2]DFKI GmbH, Multilingual Technologies Lab
[3]Insight Centre for Data Analytics at the National University of Ireland Galway, Ireland
[4]Sapienza University of Rome
[1]Sonnenfelsgasse 19,1010 Vienna, Austria
[2]Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
[3]IDA Business Park, Lower Dangan Galway, Ireland
[4]Via Regina Elena, 295 - 00161 Roma, Italy
[2]declerck@dfki.de, [1]{Ksenia.Zaytseva,Tanja.Wissik}@oeaw.ac.at,
[3]john.mccrae@insight-centre.org, [4]navigli@di.uniroma1.it

Abstract

In this paper we outline the interoperability aspects of the recently started European project ELEXIS (European Lexicographic Infrastructure). ELEXIS aims to integrate, extend and harmonise national and regional efforts in the field of lexicography, both modern and historical, with the goal of creating a sustainable infrastructure which will enable efficient access to high quality lexical data in the digital age, and bridge the gap between more advanced and lesser-supported lexicographic resources. For this, ELEXIS will make use of or establish common standards and solutions for the development of lexicographic resources and develop strategies and tools for extracting, structuring and linking lexicographic resources.

Keywords: eLexicography, Linguistic Linked Data cloud, BabelNet, OntolexLemon

1. Introduction

The field of lexicography has a long tradition of proposing as accurate as possible descriptions of languages. As stated in (Køhler Simonsen, 2017): "Lexicography is a four thousand year old discipline and dictionaries have been an integral part of commerce and human cultural history for centuries".

Since the 1980s, lexicographers have started to utilize computers and to apply computational methods. Online dictionaries are no longer only a reference work, but are also seen as platforms for supporting advanced search facilities. This emerging field of e-lexicography, nevertheless, is still not clearly shaped, and methods and workflows not yet fully agreed on. We see for example in a recent article (Rundell, 2015), in which the author describes the current situation of e-lexicography as being in a transitional phase, a quotation of Robert Lew stating that "It seems that the web community, while enthusiastically embracing the novelty of online collaboration, propagates the traditional model of lexicographic description"[1]. This transitional status is even more patent, when we consider the relations between the fields of lexicography and Natural Language Processing (NLP)[2], although both sides could greatly benefit from each other, as this was already pointed out in (Kilgarriff, 2000). Lexicographic work is also under-represented in the Linked Data (LD) cloud and in Semantic Web technologies.

In recent years, however, new developments have emerged in the field of e-lexicography, like the eLex conference series[3], which started in 2009, the Globalex initiative[4], which was established at eLex 2015 and which organized two workshops at LREC[5], thus directly addressing the Language Technology community, or the recently ended ENeL COST action[6], which is described below in more details.

In 2013, the European lexicographic community was brought together for the first time in the European Network of e-Lexicography (ENeL) COST action. This initiative was set up to improve the access for the general public to scholarly dictionaries and make them more widely known to a larger audience. In the context of this network, a clear need emerged for a broader and more systematic exchange of expertise, for the establishment of common standards and solutions for the development and integration of lexicographical resources, and for broadening the scope of application of these high quality resources to a larger community, including the Semantic Web, artificial intelligence, NLP and Digital Humanities. For describing such an integrative approach, the term "virtuous circle" re-emerged, as it characterizes very well the intended spiralling development of lexicographic data on the basis of a cross-disciplinary exchange of knowledge and the incremental contributions of the different methods and technologies to be involved.

We write "re-emerge", as the term was already coined in (Kilgarriff, 2000): "In the best of all possible worlds, computational enhancement and lexicographical upgrading would build upon each other in a virtuous circle that knew

[1]The quotation was taken from (Lew, 2014).

[2]In this paper, we will use the terms NLP or Language Technology (LT) interchangeably.

[3]https://elex.link/

[4]https://globalex.link/

[5]http://ailab.ijs.si/globalex/

[6]http://www.elexicography.eu/.

no bounds". The implementation of such a virtuous cycle for the generation of high-quality e-lexicographic resources is a central objective of the recently started ELEXIS project, described in section 2.

2. ELEXIS

ELEXIS (European Lexicographic Infrastructure) is fostering cooperation and information exchange among lexicographical research communities. The infrastructure is a newly granted project under the H2020-INFRAIA-2016-2017 call, with the topic "Integrating Activities for Starting Communities", and started in February 2018[7].

ELEXIS is building on infrastructures defined in other projects and initiatives, especially CLARIN[8] and DARIAH[9], which allow language or Digital Humanities resources (both tools and data) to be shared. In this, the partners of ELEXIS will get support for easily sharing their lexicographic resources, yet this does not necessarily lead to any interoperability of such resources. In order to support interoperability, ELEXIS enables stakeholders to encode their lexicographic data with common concepts and entities from models such as BabelNet[10], DBpedia[11] or Wikidata[12], which are accessible as nodes in the Linked Data cloud[13].

Moreover, to ensure that there is integration of lexical resources at even the most basic level, ELEXIS will define a minimal common data model capturing the core concepts of a lexicographic resource such as entries (single-word, multi-word), senses, syntactic and semantic frames, etymologies etc. and linguistic relationships such as synonymy/antonymy, translation, domain/region/register classification, relatedness, etc. that will be compatible with existing models used in the community, including TEI[14], Wikidata[15], LMF[16] and OntoLex-Lemon[17]. The data converted to this model will be available in RDF[18], facilitating linking and publishing on the Web as linked data.

A key goal of the ELEXIS project is thus to enable stakeholders to link their existing lexicographic resources, either as dictionaries or as standalone lexical descriptions encoded, and so to create a huge multilingual registry, a kind of "Matrix Dictionary" (see Section 3.) that connects lexicographic resources across common concepts. A possible infrastructure for hosting this registry is the Linguistic Linked (Open) Data cloud, which is described in section 4.. In this scenario, ELEXIS would also follow the W3C recommendations for "accessing, updating, creating and deleting resources from servers that expose their resources as Linked Data", as those are stated by the Linked Data Platform (LDP)[19].

3. A Matrix Dictionary for ELEXIS

A key goal of ELEXIS is the creation of a "Matrix Dictionary", that will be formed of links created between lexicographic resources in different languages, domains and forms. With this, ELEXIS will create a universal repository of linked senses, meaning descriptions, etymological data, collocations, phraseology, translation equivalents, examples of usage and all other types of lexical information found in all types of existing lexicographic resources, monolingual, multilingual, modern, historical, etc. In order to reach this goal, ELEXIS will develop strategies, tools and standards for extracting, structuring and linking the high quality semantic data from lexicographic resources and make them available to the Linked (Open) Data family. Those processes are necessary, as current lexicographic resources, both modern and historical, have different levels of structure and are not equally suitable for applications in advanced NLP technologies, for which they should be disclosed to or from which they could benefit.

The project will also work on interlinking lexical content with other structured or unstructured data – corpora, multimodal resources, etc. – on any level of lexicographic description: semantic, syntactic, collocational, phraseological, etymological, translation equivalents, examples of usage, etc. By creating an integrated, linked and interlinked resource, a huge amount of high quality lexical data will not only become available to the linguistic, NLP and Semantic Web communities, it will also facilitate cutting-edge research in Digital Humanities.

This will be achieved by creating an infrastructure dedicated to automatic segmentation and structuring of content for dictionaries that are currently produced in digital environments but are typically encoded in their own custom data format. ELEXIS conversion and alignment tools will provide users of the infrastructure with the possibility to harmonise and convert their lexicographic resources to a uniform data format that allows their seamless integration in Linked Open Data.

This infrastructure is responding to one of the missions of ELEXIS consisting in enabling the integration of (big) data in different modalities into the lexicographic process, pre-

[7] See http://www.elex.is/.

[8] See https://www.clarin.eu/.

[9] See https://www.dariah.eu/.

[10] See http://babelnet.org/ and (Navigli and Ponzetto, 2012).

[11] See Seehttp://wiki.dbpedia.org/. See also (Unger et al., 2013) for a first study on how to publish a DBpedia based ontology lexicon as linked data.

[12] See Seehttps://www.wikidata.org/wiki/Wikidata:Main_Page.

[13] See http://linkeddata.org/ for more details.

[14] TEI stands for "Text Encoding Initiative". See http://www.tei-c.org/index.xml.

[15] See https://www.wikidata.org/wiki/Wikidata:Main_Page.

[16] LMF stands for "Lexical Markup Framework", an ISO standard. See http://www.lexicalmarkupframework.org/.

[17] OntoLex-Lemon is the result of a W3C Community Group, building on and extending LMF and an earlier version of *lemon* (lexicon model for ontologies, (McCrae et al., 2012)). See https://www.w3.org/2016/05/ontolex/ for the final W3C Community report and (McCrae et al., 2017) for the current status of OntoLEx-Lemon.

[18] RDF stands for "Resource Description Framework", a W3C standard model for interchanging data on the Web. It is a building stone for the realisation of the Linked Data cloud (see for this point http://linkeddata.org/faq).

[19] The source we quoted from LDP: https://www.w3.org/TR/ldp/.

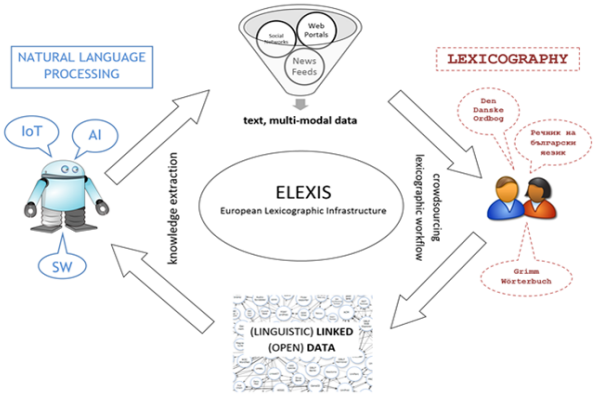

Figure 1: The virtuous cycle of e-lexicography

pared and visualised for human end users. Figure 1 is displaying this development, which is of cyclic nature.
The existence of common data models and standards that are produced bottom-up from within the lexicographic community fostered by ELEXIS is a necessary condition for the successful development of the whole platform. Standards will be developed and tested during the project on the data provided by the lexicographic partners and implemented in the newly-developed service.

4. Linguistic Linked Open Data

The Linguistic Linked Open Data (LLOD)[20] is an initiative started by the Open Linguistics Working Group (OLWG)[21] aims at breaking the data silos of linguistic data and thus encourage NLP applications that can use data from multiple languages, modalities (e.g., lexicon, corpora, etc.) and develop novel algorithms. Figure 2 gives a partial view of the current state of the LLOD cloud.

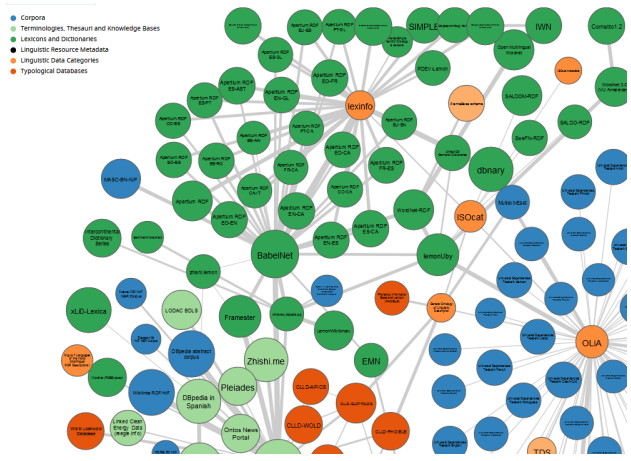

Figure 2: A (partial) view on the Linguistic Linked Open Data cloud, July 2017.

The rapid development of the LLOD cloud[22] was also supported by the European LIDER ("Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe") project[23]. LIDER has set up some basis for the further development of the Linguistic Linked Open Data and published a series of guidelines on how to publish linguistic data in the Linked Data framework. Those guidelines are used in relation to the task of making the LLOD actionable for language intensive use cases, with a focus on multilingual application. Those guidelines will be used and extended in the context of ELEXIS.
A cooperation established between LIDER and the aforementioned ENeL Cost Action, also in the form of short term exchanges of junior researchers and of the participation of ENeL members to a datathon organized by LIDER[24] has been in fact instrumental in the formulation of some of the central objectives of the ELEXIS project, which will also stress the need of community integration besides the technological one, whose description is the focus of this paper.
The successful development of the LLOD is also based and linked to the development of the Lexicon Model for Ontologies (*lemon*)[25] and its successor the OntoLex-Lemon model[26]. And although *lemon*, which stands for "LExicon Model for ONtologies", was originally developed in order to model language data used in ontologies, experience has shown that *lemon* or OntoLex-Lemon can indeed be used for modelling lexicographic data[27] or some specific lexical phenomena[28].

5. Further Developments of the LLOD and OntoLex-Lemon within ELEXIS

While looking in details at the current state of the Linguistic Linked Open Data (LLOD)[29], one can see that the data sets published in this cloud are classified along the lines of six categories:

- Corpora
- Terminologies, Thesauri and Knowledge Bases
- Lexicons and Dictionaries
- Linguistic Resource Metadata
- Linguistic Data Categories
- Typological Databases

[20] See `http://linguistic-lod.org/`.
[21] See (Chiarcos et al., 2012) and (McCrae et al., 2016).
[22] The full LLOD cloud can be accessed at `http://linguistic-lod.org/llod-cloud`. There one can click on the various nodes and get more details about the data sets represented by the "bubbles".
[23] LIDER was an FP7 Coordination and Support Action from 2013-11-01 to 2015-12-31. See also `http://lider-project.eu/`.
[24] See `http://datathon.lider-project.eu/`.
[25] See (McCrae et al., 2012)
[26] `https://www.w3.org/2016/05/ontolex/`. See also for a kind of historical view on the development of *lemon* towards OntoLex-Lemon (McCrae et al., 2017).
[27] See (Declerck et al., 2017) or (Tiberius and Declerck, 2017).
[28] See (Declerck and Lendvai, 2016).
[29] See again `http://linguistic-lod.org/llod-cloud`.

However, as of today, the LLOD is not populated by many lexicographic resources, due to the lack of a dedicated infrastructure for resource interlinking and of effective ontology alignment algorithms, which depend on multilingual semantic similarity, entity linking and word sense disambiguation.

One goal of ELEXIS can be to have a specific lexicographic category containing its specific data sets and linking those to both data sets included into the other LLOD categories and to data sets included in the global Linked Data cloud[30]. This step is responding for example to insights described in (Gracia et al., 2017), in the abstract of which we can read: "[...] future dictionaries could be LD-native and, as such, graph-based. Their nodes are not dependent on any internal hierarchy and are uniquely identified at a Web scale". ELEXIS will address this view on the generation of linked data-native dictionaries and facilitate their publication in the LLOD cloud as lexicographic data sets. (Declerck, 2018) proposes a similar approach, but considering all types of lexical data, not only those included in a dictionary.

As the development of the LLOD cloud is closely related to OntoLex-Lemon and related vocabularies, a working group was built in order to study the representation of lexicographic data (sets) and to propose a lexicographic module to be added to Ontolex-Lemon[31], so that their linking to all types of lexical data covered by the Ontolex-Lemon Model is guaranteed. In this, ELEXIS partners are contributing to standardisation of the formalisation of lexicographic data.

6. Interoperability and Quality

To provide conceptual interoperability, services enabling linking of ELEXIS lexicographic resources will be developed and made available in the ELEXIS linking tools segment of the platform (see Figure1). This will provide the possibility to link lexical entries, senses and fundamental concepts in different lexical resources, using a semi-automatic approach. BabelNet[32], as an existing multilingual resource to provide cross-lingual linking, will be exploited for this purpose. Extensive linking of existing lexicographic resources by pivoting through BabelNet will enable the creation of what we call the ELEXIS matrix dictionary[33]. Data from this new resource will be available through ELEXIS matrix dictionary RESTful Web service as part of the platform.

This work will be achieved through four principle steps:

- **Common access and models**: We will define a set of common protocols, in the form of REST API calls that can allow dictionaries involved in the project to be accessed through a single interface. This model will be based on existing web standards and models including RDF, SPARQL and OntoLex-Lemon. Furthermore, the task will define common metadata and concept properties for use within the project. The outputs of this task will be technical documentation describing the formats and tools to allow resources in OntoLEx-Lemon RDF or TEI to be compliant with this protocol.
- **Semi-automatic dictionary linking**: Linking lexical resources is a challenge that requires impractical amounts of human efforts, but is still not easy to solve automatically. We will develop a semi-automatic system that will make the linking problem viable for large resources, by using state-of-the-art semantic and natural language processing techniques, especially deep learning methods such as LSTMs (Tai et al., 2015), with a human in the loop. Furthermore, we will apply constraint-based optimisation of the linking, which can quickly find the correct mapping in an active learning setting with only a small amount of human input. As such, we will develop a single tool where a user can upload two lexical resources and interactively link them. We will then evaluate this tool by developing gold standard mappings in the context of a shared task.
- **Cross-lingual linking through BabelNet**: In order to link lexical resources across languages, we will use one highly multilingual lexicon, BabelNet, as the basis for a cross-lingual mapping system. As such, we will extend our linking tools to cross language boundaries by pivoting through BabelNet. We will further allow for resources linked through BabelNet to be used to be submitted to BabelNet, so they can extend the resource in future releases.
- **Validation and quality assurance**: We will develop tools to automatically verify the quality of lexical resources at three levels: Firstly, the *technical quality*, which means ensuring that the resource maintains the validity of its output and does not make errors in encoding, this will be achieved by Web services that validate TEI and RDF data as provided by producers. Secondly, *operational quality* ensures that the lexical resources remain available and responsive as they are deployed on the Web, in particular, a service will measure uptime of each resource. Finally, *scientific quality* ensures that the results of the service are correct in the task they try to perform and will work by creating benchmarks for tasks in NLP, with Web services to automatically check resource performance against existing gold standards.

7. Lexicography for Natural Language Processing

To show the effectiveness of the interlinking across lexical resources, ELEXIS will study the impact of an enriched LLOD on several NLP tasks:

- **Multilingual Word Sense Disambiguation:** a long-standing issue of supervised Word Sense Disambigua-

[30] `http://lod-cloud.net/`. There the reader can observe that "Linguistics" is listed as a domain-specific sub-set of the cloud.

[31] See (Bosque-Gil et al., 2017) and for the current state of the discussions on the lexicographic module `https://www.w3.org/community/ontolex/wiki/Lexicography`.

[32] See again `http://babelnet.org/` and (Navigli and Ponzetto, 2012).

[33] The motivation behind the ELEXIS matrix dictionary has been described in 3.

tion (WSD) – the task of automatically determining the meaning of words occurring in context – is that huge amounts of sense-annotated sentences need to be manually created. This endeavour, which as of today, is incomplete even for English, needs to be repeated for each new domain and language, something that makes the task arduous to replicate in most European languages. The ELEXIS lexicographic resources will be utilized to bootstrap large training datasets for WSD in dozens of languages.

- **Multilingual Semantic Parsing:** semantic parsing aims to map sentences to formal representations of their meaning. It has deeper relationships to syntactic parsing than WSD. However, most semantic parsing approaches in the literature either work in a supervised fashion with even higher annotation costs than those of WSD or require knowledge resources such as DBpedia or Wikidata which seem to work only in domain-restricted specific tasks such as question answering. In ELEXIS we will develop innovative algorithms that exploit the huge multilingual network of interlinked lexical knowledge to perform multilingual semantic parsing.

- **Word sense clustering:** where development of semi-automatic procedures to bring together subtle sense distinctions in clusters of meanings will be shown to improve the performance of tasks such as Word Sense Disambiguation;

- **Domain labelling of text:** where the aggregated information obtained from the lexicographic network of resources will be shown to improve automatic tagging of text with domain labels in arbitrary languages thanks to developing innovative neural techniques.

- **Study of the diachronic distribution of senses:** the use of the most frequent sense in NLP is a solid baseline used in WSD and other tasks. However, it is not systematic and it is useful only for the English language. We will develop novel techniques for aggregating the predominance information of senses a) from the multitude of resources b) considering evolution over time, so as to have important impact on disambiguation and corpus analysis.

8. The User Perspective

While the description of the foreseen ELEXIS platform can at first sight look like an academic exercise, it should be stressed that the project responds to the needs formulated by publishers and other professionals in the e-lexicographic field. Some of those needs were already articulated by industrial/commercial partners in the ENeL Cost action. The changing technological context calls not only for adaptations of the lexicographic workflows but also for the establishment of new business models, as this is for example expressed in (Køhler Simonsen, 2017). We quote from this eLex 2017 paper: "[...] the biggest problem of lexicography is that lexicographic products are no longer perceived as relevant for the vast majority of people. Most people in fact do not use dictionaries, and if they need to find help when communicating or when looking for data, they simply use the Internet instead. So dictionaries are in fact not being used as much as we want them to be. The most important question is: why do not people use online or mobile dictionaries? Obviously, there are a number of reasons, but I would argue that the most important reason is that most lexicographic resources are not tool-integrated and not specifically related to the user's job tasks".

We can see that ELEXIS is (at least partly) responding to this situation if we compare some of the technological goals of ELEXIS with the six theses that are formulated in (Køhler Simonsen, 2017) and which describe the components of what could and should be the ingredients of a viable business model for the modern e-elexicography. In those 6 theses, (Køhler Simonsen, 2017) requires among others that lexicographic products are moving to lexicographic services, the integration of lexicographic data in lexicographic platform and distribution, and to take increasingly into account the lexicographic users and their needs. Topics that are at the core of ELEXIS, as well as the move "from dictionary to lexicographic data in software [and] artificial intelligence".

9. Conclusion

We described in this paper the main technological challenges that the ELEXIS project will try to solve, based on existing initiatives, projects, infrastructures and standards. A focus will be on the interrelation between e-lexicography and the technologies used in the context of the Linguistic Linked Data, in order to generate high-quality lexicographic data that can be then immediately re-used in NLP and Semantic Web applications, which are themselves based on the LLOD. ELEXIS implements thus a "virtuous cycle" scenario to make sure that lexicographic resources and expertises are played a central role in high-quality language applications, also beyond the era of dictionary-based lexicographic products.

Acknowledgements

This work was supported in part by the H2020 project "ELEXIS" with Grant Agreement number 731015 and by Science Foundation Ireland under Grant Number SFI/12/RC/2289 (Insight).

10. Bibliographical References

Bosque-Gil, J., Gracia, J., and Montiel-Ponsoda, E. (2017). Towards a module for lexicography in ontolex. In *Proceedings of the LDK 2017 Workshops: 1st Workshop on the OntoLex Model (OntoLex-2017), Shared Task on Translation Inference Across Dictionaries & Challenges for Wordnets co-located with 1st Conference on Language, Data and Knowledge (LDK 2017), Galway, Ireland, June 18, 2017.*, pages 74–84.

Chiarcos, C., Hellmann, S., and Nordhoff, S., (2012). *Linking Linguistic Resources: Examples from the Open Linguistics Working Group*, pages 201–216. Springer Berlin Heidelberg, Berlin, Heidelberg.

Declerck, T. and Lendvai, P. (2016). Towards a formal representation of components of german compounds. In Micha Elsner et al., editors, *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*. Humboldt University, ACL, 8.

Declerck, T., Tiberius, C., and Wandl-Vogt, E. (2017). Encoding lexicographic data in lemon: Lessons learned. In John P. McCrae, et al., editors, *Proceedings of the LDK workshops: OntoLex, TIAD and Challenges for Wordnets*. CEURS, 8.

Declerck, T. (2018). Towards a linked lexical data cloud based on ontolex-lemon. In John P. McCrae, et al., editors, *Proceedings of the 6th Workshop on Linked Data in Linguistic (LDL-2018)*. ELRA, 5.

Gracia, J., Kernerman, I., and Bosque-Gil, J. (2017). Toward linked data-native dictionaries. In Iztok Kosem, et al., editors, *Proceedings of the eLex 2017 conference*, pages 550–559. INT, Trojína and Lexical Computing, Lexical Computing CZ s.r.o., 9.

Kilgarriff, A. (2000). Business models for dictionaries and nlp. *International Journal of Lexicography*, 13(2):107–118.

Køhler Simonsen, H. (2017). Lexicography: What is the business model? In Iztok Kosem, et al., editors, *Electronic Lexicography in the 21st Century*, pages 395–415. Lexical Computing CZ s.r.o.

Lew, R. (2014). User-generated content (ugc) in online english dictionaries. *OPAL - Online publizierte Arbeiten zur Linguistik*, 4:8–16.

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

McCrae, J. P., Chiarcos, C., Bond, F., Cimiano, P., Declerck, T., de Melo, G., Gracia, J., Hellmann, S., Klimek, B., Moran, S., Osenova, P., Pareja-Lora, A., and Pool, J. (2016). The Open Linguistics Working Group: Developing the Linguistic Linked Open Data Cloud. In *10th Language Resource and Evaluation Conference (LREC)*, pages 2435–2441.

McCrae, J. P., Buitelaar, P., and Cimiano, P. (2017). The OntoLex-Lemon Model: development and applications. In Iztok Kosem, et al., editors, *Proceedings of eLex 2017*, pages 587–597. INT, Trojína and Lexical Computing.

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.

Rundell, M. (2015). From print to digital: Implications for dictionary policy and lexicographic conventions. *Lexikos*, 25(1).

Tai, K. S., Socher, R., and Manning, C. D. (2015). Improved semantic representations from tree-structured long short-term memory networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 1556–1566.

Tiberius, C. and Declerck, T. (2017). A lemon model for the anw dictionary. In Iztok Kosem, et al., editors, *Proceedings of the eLex 2017 conference*, pages 237–251. INT, Trojína and Lexical Computing, Lexical Computing CZ s.r.o., 9.

Unger, C., McCrae, J. P., Walter, S., Winter, S., and Cimiano, P. (2013). A lemon lexicon for dbpedia. In *Proceedings of the NLP & DBpedia workshop co-located with the 12th International Semantic Web Conference (ISWC 2013), Sydney, Australia, October 22, 2013.*