

Involving Lexicographers in the LLOD Cloud with LexO, an Easy-to-use Editor of Lemon Lexical Resources

Andrea Bellandi, Emiliano Giovannetti

Institute of Computational Linguistics
Via G. Moruzzi, 1 - Pisa, Italy
{name.surname}@ilc.cnr.it

Abstract

In this contribution we show LexO, a user-friendly web collaborative editor of lexical resources based on the *lemon* model. LexO has been developed in the context of Digital Humanities projects, in which a key point in the design of an editor was the ease of use by lexicographers with no skill in Linked Data or Semantic Web technologies. Though the tool already allows to create a *lemon* lexicon from scratch and lets a team of users work on it collaboratively, many developments are possible. The involvement of the LLOD community appears now crucial both to find new users and application fields where to test it, and, even more importantly, to understand in which way it should evolve.

Keywords: Semantic Web, tools for E-Lexicography, lemon model

1. Introduction

The increasing growth of the Linguistic Linked Open Data Cloud¹ (LLOD) witnesses the liveliness of activities carried out by the relative community (in which the Open Linguistics Working Group² stands out) in the last few years. In this context, the availability of models and tools can be crucial to attract people (such as linguists, lexicographers, terminologists) willing to be involved but who feel worried by the lack of standards and, at the same time, by the technological skills required to face the construction of a resource as a linked data.

A solution to the first issue has come from the definition of *lemon*, a model developed to create resources formally describing lexicalizations of ontological concepts (McCrae et al., 2012b). The *lemon* model has been designed to compensate for the way in which the Web Ontology Language (OWL) (McGuinness and Harmelen, 2004) allows to denote a concept, i.e. by simply labelling it with a string and not, as it can be desired, with a complex lexical unit. Today, *lemon* can be considered a *de facto* standard for the representation of computational lexicons in the Semantic Web (SW). The number of users potentially interested in editing or consuming *lemon* data is large (McCrae et al., 2017). However, the construction of a lexicon based on *lemon* can be a tedious process, due to the complexity of the model (constituted by a number of different modules) and on the need of relying on complex patterns, such as reification, to represent certain phenomena.

In order to face the second issue, relative to the aforementioned “technical bottleneck”, we have developed LexO³ (Bellandi et al., 2019; Bellandi and Khan, 2019), a collaborative, easy to use, web editor of lexical resources based on the *lemon* model. In the context of Semantic Web (SW), very few editors of lexicons exist, and just two of them (at

least to the best of our knowledge) handle the *lemon* model. The first one is “*lemon source*”, a Wiki-like web interface to manipulate and publish *lemon* data in a collaborative way (McCrae et al., 2012a). *lemon source* allows a user to upload a lexicon and share it with others. This tool was developed as an open-source project, based on the *lemon* API, and it is freely available for use online. However, it can only manage older versions of the *lemon* model, and it doesn’t seem to be updated anymore.

The tool that appears to be the most similar to LexO is VocBench (Stellato et al., 2017) a web-based, multilingual, collaborative development platform born to manage OWL ontologies and SKOS(/XL) thesauri and more recently updated to manage also Ontolex-Lemon lexicons and generic RDF datasets. VocBench is a very well engineered tool, conceived for users with experience in SW technologies and vocabularies. Indeed, the main difference between VocBench and LexO lies in target users and the required expertise to use the tool. As a matter of fact, LexO is aimed at lexicographers who have no (or very limited) skills in representing lexica using the OntoLex-Lemon model and, more in general, in using SW and LD related technologies. However, the ease with which a user can create a *lemon* lexical resource in LexO is counterbalanced by the little flexibility it currently offers, for example to extend the underlying lexical model with custom classes or relations, as VocBench allows to do.

2. An Overview of LexO

LexO⁴ is a collaborative web editor for easily building and managing *lemon* resources. The features of LexO were defined on the basis of our experience gained in the creation of lexical and terminological resources in the framework of several projects in the field of Digital Humanities. The context in which the first version of LexO was devel-

¹<https://linguistic-lod.org/>

²<http://linguistics.okfn.org>

³It is fully compatible with the latest version of *lemon*, developed by the Ontology Lexicon (Ontolex) community group. Please, see <https://www.w3.org/2016/05/ontolex/>

⁴The source code is available at <https://github.com/andreabellandi/LexO-lite>. You can find a demo at <http://klab.ilc.cnr.it/talmud/LexO>

oped is DiTMAO⁵, a project aimed at producing a digital-born multilingual medico-botanical terminology pivoting on Old Occitan and developed by philologists (Bellandi et al., 2018b). Then, a team of lexicographers applied LexO to create FdS⁶, a multilingual diachronic lexicon of Saussurean terminology in the framework of an Italian PRIN project (Bellandi et al., 2018a). Finally, a bilingual Chinese-Italian resource was created with LexO on the basis of Matteo Ricci’s Atlas in the context of Totus Mundus⁷ project (Piccini et al., 2018).

Currently LexO is being used in the context of the Babylonian Talmud Translation Project in Italian⁸, to create a multilingual Aramaic/Hebrew/Italian terminological resources of the talmudic domain. The examples that will follow are based on this work. During the development of LexO we have also been influenced by some of the latest works taking place in the European H2020 project ELEXIS (in which the Institute of Computational Linguistics of Pisa is involved as a partner). In particular, our work has been closely informed by the survey of lexicographic users’ needs conducted as part of the project and recently published as a deliverable⁹. The main aim of LexO is to encourage the involvement of lexicographers in the LLOD community by i) spreading the use of the *lemon* model also in communities of lexicographers new to the modelling of resources as linked data ii) make the model available to lexicographers already involved in LD and SW but who do not have the skills to handle the required technologies.

In this sense, LexO acts as a filter between the formal representation language encoding the lexicon and the user, allowing the latter to focus on the lexicographic work without having to take care of the underlying modeling technology. To bring an example, to establish a translation relation between two lexical senses the user does not have to know which properties are to be involved and how to instantiate the reification pattern required by the “variation and translation” (vartrans) module of *lemon*. Another key feature of LexO is its collaborative nature. As a web application, it allows a team of users with different roles (administrators, editors, viewers, etc.) to work on the same resource at the same time. The issue of concurrent access has been faced by implementing locking mechanisms. As soon as a user starts editing a certain lexical entry, the system locks it and all the related entities, avoiding the concurrent editing of other users.

Finally, LexO provides a set of services implemented with RESTful protocols that give software agents access to the resources managed by the tool. The REST interface implemented in the system allows to query individual resources and get information about lexical entries in JSON format.

⁵<https://www.uni-goettingen.de/en/487498.html>

⁶Demo available at <http://ditmao-dev.ilc.cnr.it:8082/saussure>

⁷Demo available at <http://lexo-dev.ilc.cnr.it:8080/TMLexicon>

⁸www.talmud.it

⁹See deliverable D1.1 “Lexicographic practices in Europe: a survey of user needs” at <https://elex.is/deliverables/>

In the current version, a first set of basic REST services is available, though which it is possible to list all the lexicon languages, the lexicon lemmas, the lemma metadata, the lemmas involved in a specific lexico-semantic relation with a given lemma, and some basic statistics.

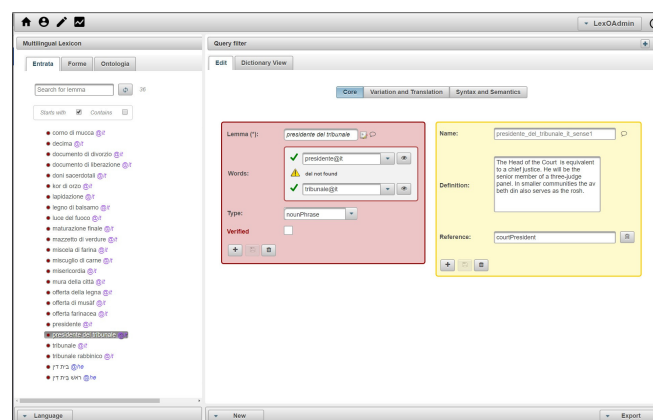


Figure 1: The main interface of LexO showing the core module.

The system interface has been conceptually designed on the basis of the *lemon* model¹⁰. As shown in figure 1, the linguistic information is organized in 3 modules¹¹:

- the core module, to define the structure of each entry, e.g., lemma, other forms, senses;
- the variation and translation module, to represent the variation of relations across entries in the same or different languages (e.g., dialectal, register, and translation relations, or morphological and orthographic ones);
- the syntactic and semantics module, to describe the syntactic behavior of an entry, its valence (the syntactic arguments involved by the situation the word refers to) and the link to the ontological structures representing the meaning of the entry.

A particular consideration has to be made regarding the core module (OntoLex). As already said, LexO is based on the *lemon* model, which was designed to lexicalize ontologies in an onomasiological perspective (i.e. from the concept to the word). LexO, on the contrary, aims at supporting in the construction of lexical resources by starting from text, in a more semasiological perspective (i.e. from the word to the concept), without requiring that an ontology must be available from the beginning. As a matter of fact, the system does not force the user to link each lexical sense to a concept. However, to stay compliant to the *lemon* model, it links an *OWL:Nothing* class to each new lexical sense, though in a transparent way to the user. If a user does want to specify the conceptual reference of a lexical sense, LexO provides a way to import an OWL ontology and link

¹⁰For describing the properties of linguistic objects, LexO relies on the LexInfo vocabulary.

¹¹In LexO, the value of linguistic Catalog is automatically set to LexInfo, an ontology that provides data categories for the OntoLex-Lemon model.

a sense to any of its ontological entities (class, instance or property).

2.1. Working Examples

The main interface of LexO, as shown in Figure 1, is designed for the editing of a multilingual lexicon. It is mainly composed of 2 parts. The leftmost column allows a user to browse lemmas, forms and senses, according to the OntoLex-Lemon core model. If the resource is multilingual, a lexicographer can filter lemmas, forms and senses by language. Information related to the selected entry is shown in the central panel where the system shows the lexical entry of reference, alongside the lemma (red box), its forms (blue boxes) and the relative lexical senses (yellow boxes). By means of the plus button it is possible to add other forms and other senses. Figure 1 shows the entry “presidente del tribunale” (court president) modeled with the decomp¹² module. LexO automatically tries to link each multiword component to the correspondent lexical entry, if already available in the lexicon.

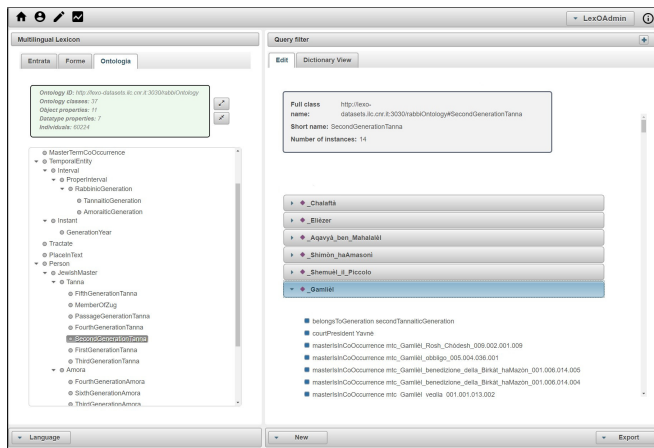


Figure 2: The management of the imported ontology.

As figure 2 depicts, LexO gives users the possibility to import an existing domain ontology, visualize its hierarchical structure, its properties and instances, and allows the association of an ontological entity to a lexical sense. In our example of figure 2, the Talmud domain ontology, which is being developed within the project, is shown. The sense of the term “presidente del tribunale” is linked to the object property *talmud:courtPresident* binding Jewish Rabbis to geographical places.

Through the select button at the top center of the interface of figure 1, a lexicographer can switch to the “variation and translation” or the “syntax and semantics” part. The former (figure 3) consists of two kinds of relations: i) the semantic relations holding between senses and including terminological relations (dialectal, register, chronological, discursive, and dimensional variation) and the translation relation; ii) the relations linking lexical entries and/or forms, which describe, for example, the morphological and orthographic variations of the word. Figure 3 shows an example of relation between lexical entries, i.e. “presidente” is the head of

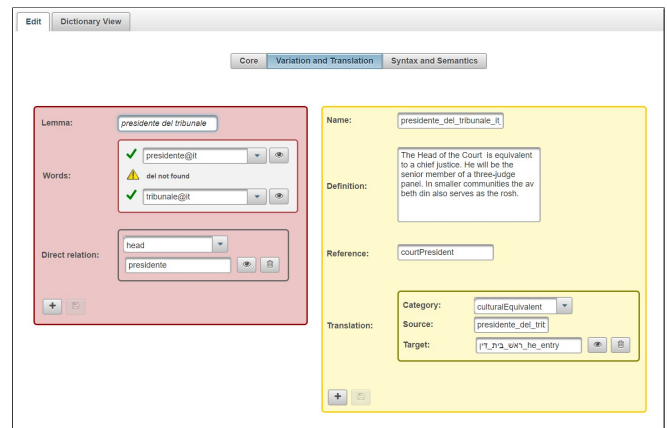


Figure 3: The interface to the variation and translation module.

“presidente del tribunale”, and a relation between senses, i.e. the cultural Italian translation¹³ with the original sense associated with the Hebrew term. The encoding generated by LexO is the following:

```
:presidente_del_tribunale_it_entry a
  ontolex:LexicalEntry,
  ontolex:MultiwordExpression ;
lexinfo:partOfSpeech lexinfo:NounPhrase ;
decomp:constituent :presidente_it_comp0,
  :del_it_comp1, :tribunale_it_comp2 ;
ontolex:canonicalForm
  :presidente_del_tribunale_it_lemma ;
ontolex:sense
  :presidente_del_tribunale_it_sense1 .

:presidente_del_tribunale_it_lemma a ontolex:Form ;
ontolex:writtenRep "presidente del tribunale"@it .

:presidente_it_entry lexinfo:head
  :presidente_del_tribunale_it_entry .

:presidente_del_tribunale_it_sense1 a
  ontolex:LexicalSense ;
skos:definition "The Head of the Court ..."@en ;
ontolex:reference talmud:courtPresident .

:rosh_beit_din_heb_sense1 a
  ontolex:LexicalSense ;

:trans a vartrans:Translation ;
vartrans:source
  :presidente_del_tribunale_it_sense1 ;
vartrans:target :rosh_beit_din_heb_sense1 ;
vartrans:category transcat:culturalEquivalent .
```

Figure 4 shows the syntactic behavior of a word and its government pattern, namely the actants introduced by the word, their syntactic functions and their morpho-syntactic realization. These syntactic frames need also to be bound to the ontological structures representing their meaning. As a consequence, LexO makes it possible to map the argument of a predicate defined in an ontology and the syntactic argument introduced in a given syntactic frame.

Figure 4 depicts an example of noun frame of the term having a subject and an (optional) object that has the preposition “di” (of) as marker. An example of frame instance in the Italian translation of the Babylonian Talmud is: “Ma quando giunsero presso il tribunale a Yavnè, rabbàn Gamlièl, che era il presidente del tribunale, accettò la loro

¹²<https://www.w3.org/2016/05/ontolex/#decomposition-decomp>

¹³LexO also uses the vocabulary *transcat* defined at <http://purl.org/net/translation-categories>

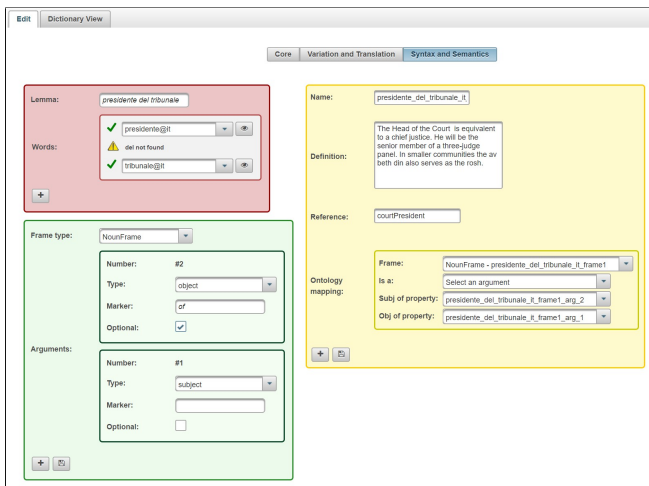


Figure 4: The interface to the syntax and semantics module.

testimonianza” (“But when they arrived at the court in Yavnè, rabbàn Gamlièl, who was the president of the court, accepted their testimony”). Basically, the term verbalizes the property *talmud:courtPresident* where the subject corresponds to the first argument of the property (i.e., the property domain class), and the object corresponds to the second argument of the property (i.e., the property range class). The encoding generated by LexO is the following:

```

:presidente_del_tribunale_it_entry a
  ontolex:LexicalEntry,
  ontolex:MultiwordExpression ;
  synsem:synBehavior
  :presidente_del_tribunale_it_frame1 .

:presidente_del_tribunale_it_frame1 a
  synsem:SyntacticFrame, lexinfo:NounFrame ;
  lexinfo:subject
    :presidente_del_tribunale_it_frame1_arg1 ;
  lexinfo:object
    :presidente_del_tribunale_it_frame1_arg2 .

:presidente_del_tribunale_it_frame1_arg1 a
  synsem:SyntacticArgument .

:presidente_del_tribunale_it_frame1_arg2 a
  synsem:SyntacticArgument ;
  synsem:marker "oE" ;
  synsem:optional "true"^^xsd:boolean .

:presidente_del_tribunale_it_sense1 a
  synsem:OntoMap, ontolex:LexicalSense ;
  synsem:ontoMapping
    :presidente_del_tribunale_it_sense1 ;
  ontolex:reference talmud:courtPresident ;
  synsem:subjOfProp
    :presidente_del_tribunale_it_frame1_arg2 ;
  synsem:objOfProp
    :presidente_del_tribunale_it_frame1_arg1 .

```

Finally, by selecting the “dictionary view” tab, a dictionary-like rendering of all the information related to the selected entry is shown in the central panel of LexO. The rendering layout can be modified by replacing the relative CSS files¹⁴.

3. Next Steps

In this work we have introduced LexO, a web collaborative tool for the construction of lexicographic resources based on the *lemon* model. LexO does not require any knowledge

¹⁴This operation requires the skills for editing the Java project, substituting the CSS files, and recompiling the LexO project.

on Linked Data and Semantic Web technologies to be used by a lexicographer. Though similar tools already exist (in primis the cited VocBench), we intend to specialize LexO in the ease with which it makes it possible to create lexical LLOD resources, though at the expense of flexibility and generality.

Regarding further developments of LexO, a number of upgrades have already been planned, among which we list the following: i) the integration of a module for the management of text, to allow a user to import a corpus of texts and use them as a source of lexicographic data; ii) the integration of tools for text processing to support in the construction of the lexicographic resource (for example, given a lemma, by automatically suggesting candidate forms), similarly to what Lexonomy (Měchura, 2017) and Sketch Engine (Kilgarriff et al., 2014) allow to do; iii) the possibility of easily extending the underlying lexical model by the user, through high level interfaces hiding properties and classes; iv) the integration of an importing module, to enable a user to upload an existing lexical resource encoded in other models, such as TEI-LexO (Bański et al., 2017), or LMF; v) the enhancement of LexO’s REST APIs for allowing users to get the lexical data in either JSON-LD, LMF, or TEI-LexO¹⁵; vi) the possibility to manage the module for lexicography (Bosque-Gil et al., 2017) targeted at the representation of dictionaries¹⁶.

To figure out where to start from and to what extent we should reassess our planned activities we would like to have feedback from the LLOD community and from lexicographers willing to create resources in this field. As a matter of fact, it is now crucial both to find new users and application fields where to test it, and, even more importantly, to understand in which way LexO should evolve.

Finally, we are planning to perform an evaluation of LexO focused on its usability by involving some potential users coming from the LLD community.

4. Acknowledgements

This work was supported by the TALMUD project and carried out within the scientific collaboration between S.c.a.r.l. PTTB and ILC-CNR (09/11/2017).

5. Bibliographical References

- Bański, P., Bowers, J., and Erjavec, T. (2017). Tei-lexO guidelines for the encoding of dictionary information on written and spoken forms. In *Proceedings of Electronic Lexicography in the 21st Century*, pages 485–494.
- Bellandi, A. and Khan, A. F. (2019). Lexicography and the semantic web: A demo with lexo. In *Proceedings of Electronic Lexicography in the 21st Century*, pages 53–56.
- Bellandi, A., Giovannetti, E., , and Piccini, S. (2018a). Collaborative editing of lexical and termino-ontological resources: a quick introduction to lexo. In *XVIII EU-RALEX International Congress Lexicography in Global*

¹⁵Considerations and results coming from ELEXIS project will take into account. See for example https://elex.is/wp-content/uploads/2019/02/ELEXIS_D2_1_Interface_for_Interoperable_Resources.pdf

¹⁶<https://www.w3.org/2019/09/lexicog/>

- Contexts – Book of Abstracts*, pages 23–27, Ljubljana, Slovenia.
- Bellandi, A., Giovannetti, E., , and Weingart, A. (2018b). Multilingual and multiword phenomena in a lemon old occitan medico-botanical lexicon. *Information*, 9(3) 52.
- Bellandi, A., Khan, A. F., and Monachini, M. (2019). Enhancing lexicography by means of the linked data paradigm: Lexo for clarin. In Kirl Simov et al., editors, *Proceedings of the CLARIN Annual Conference 2019*, pages 134–137, Leipzig.
- Bosque-Gil, J., Gracia, J., and Montiel-Ponsoda, E. (2017). Towards a module for lexicography in ontollex. In *Proceedings of the LDK 2017 Workshops: 1st Workshop on the OntoLex Model*, pages 74–84.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., and Suchomel, V. (2014). The sketch engine: Ten years on. *Lexicography*, 1(1).
- McCrae, J., Montiel-Ponsoda, E., and Cimiano, P. (2012a). Collaborative semantic editing of linked data lexica. In *In the Proceedings of the Language Resources and Evaluation*, pages 2619–2625, Istanbul.
- McCrae, J. P., Aguado de Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D., and Wunner, T., (2012b). *Interchanging Lexical Resources on the Semantic Web*, pages 701–709. Language Resources and Evaluation. Springer Netherlands.
- McCrae, J. P., Bosque-Gil, J., Gracia, J., Buitelaar, P., and Cimiano, P. (2017). The ontollex-lemon model: Development and applications. In *Proceedings of Electronic Lexicography in the 21st Century*, pages 19–21.
- McGuinness, D. L. and Harmelen, F. V. (2004). Owl web ontology language overview. W3C recommendation 10.
- Měchura, M. (2017). Introducing lexonomy: an open-source dictionary writing and publishing system. In *Proceedings of Electronic Lexicography in the 21st Century*, pages 19–21.
- Piccini, S., Bellandi, A., and Giovannetti, E. (2018). A semantic web approach to modelling and building a bilingual chinese-italian termino-ontological resource. In *XVIII EURALEX International Congress Lexicography in Global Contexts – Book of Abstracts*, pages 87–90, Ljubljana, Slovenia.
- Stellato, A., Turbati, A., Fiorelli, M., Lorenzetti, T., Costetchi, E., Laaboudi, C., Gemert, W. V., and Keizer, J. T. (2017). Vocbench 3: Pushing collaborative development of thesauri and ontologies further beyond. In *Proceedings of 17th European Networked Knowledge Organization Systems Workshop*, pages 39–52.