Developing LexO: a Collaborative Editor of Multilingual Lexica and Termino-Ontological Resources in the Humanities

Andrea Bellandi,
Emiliano Giovannetti, Silvia Piccini
Istituto di Linguistica Computazionale - CNR
{name.surname}@ilc.cnr.it

Anja Weingart
Georg-August-Universität Göttingen
aweinga@gwdg.de

Abstract

In this paper we present a first version of LexO, a collaborative editor of multilingual lexica and termino-ontological resources. It is based on the *lemon* model, and aims at supporting lexicographers and terminologists in their work. Although the development of LexO is still ongoing, the editor is already being used within two research projects in the field of Computational Linguistics applied to Humanities: DiTMAO and Totus Mundus. This allowed to test the functionalities of LexO and to prove its high degree of flexibility according to the different extensions of the *lemon* model needed to fulfill the needs of the involved scholars.

1 Introduction

This paper describes the ongoing development of LexO, a web collaborative editor of lexical and termino-ontological resources based on the *lemon* model[1]. As it will be described later, LexO provides some peculiar features (such as references to texts and extensibility) that make it particularly suited to be used in the Humanities.

Nowadays, well-founded lexico-semantic models designed during the last two decades enable to build lexical resources providing a rich description of word meaning with a view to retrieving and processing lexical data in texts. The main models are: WordNet (Fellbaum, 1998), Framenet (Fillmore et al., 2003), Pattern Dictionary (Hanks and Pustejovsky, 2005), SIMPLE (Lenci et al., 2000) and Brandeis Semantic Ontology (Pustejovsky, 2006). Strongly inspired by the lexical model SIMPLE, the metamodel Lexical Markup Framework (LMF) (Gil Francopoulo and Soria, 2006) was created to provide a common model to represent and encode lexical resources, and to ensure interoperability among them.

As far as the terminological perspective is concerned, the ISO standard TMF - Terminological Markup Framework - was created in 2003 (Romary, 2001). This abstract model for the representation of multilingual terminological data was introduced to cover two concurrent standards: MARTIF (Machine-readable terminology interchange format, also known as ISO (FDIS) 12200) and GENETER, which belong to SALT family of data models and formats. Over the last years, however, terminologists have started to adopt models developed within the field of lexicology, in order to describe the relationships between terms in a richer way. In fact, the traditional methodologies for describing terms, focused on the analysis of conceptual aspects (onomasiological perspective), have led terminologists to take into account only taxonomic and meronymic relationships. Differently, lexicographic models, based on a semasiological, word-oriented approach, take into account a richer set of relevant relationships. This is why, for example, (Dancette and L'Homme, 2004) propose to convert specialized dictionaries using a formal lexico-semantic framework called Explanatory and Combinatorial Lexicology (ECL), developed by (Mel'čuk et al., 1995) in the framework of the Meaning-Text approach.

[1]In this paper we assume that the reader is already familiar with *lemon*. For an exhaustive description of the model, the reader is reffered to: http://*lemon*-model.net/ (last access: 17/07/2017)

The editor we here present, called LexO, is being developed with the objective of supporting both lexicographers and terminologists in their work of building, respectively, lexica and termino-ontological resources. This is the primary reason we chose *lemon* as LexO's underlying lexical model: *lemon* is the most recent model proposed in the field of Computational Lexicography which displays some characteristics that were deemed suitable for both lexicographers and terminologists. Firstly, *lemon* is based on LMF, the ISO standard used for lexica supporting Natural Language Processing (NLP) tasks and Machine Readable Dictionaries (MRD) and which has already been used to model a number of other important lexical resources such as the Princeton Wordnet, Framenet and Verbnet. Secondly, *lemon* was proposed to provide a standard for representing multilingual lexical resources using Semantic web technologies such as RDF and OWL. Finally, in *lemon* the conceptual and linguistic dimensions are separated but interconnected. The link between lexical entries and ontological concepts is reified through the class Lexical sense.

In terminology distinguishing between lexical and conceptual dimension is proven to be fundamental, at least from a methodological point of view, especially when addressing very different languages. The theoretical necessity of distinguishing between these two levels has led to the development of new paradigms (Roche, Roche), and strategies (Reymonet et al., 2007). While, typically, a lexicon is the inventory of the words (or lexemes) of a certain language, a termino-ontological resource is composed of terms of a specific domain which are related to concepts structured in a formal ontology describing that domain.

The paper is structured as follows. In Section 2 an overview is given of existing tools designed to handle lexica and termino-ontological resources. Section 3 describes the key characteristics of LexO and its architecture. In Section 4 two projects in which the tool is being used are described. Finally, Section 5 draws some conclusion and outlines what we are currently working on to improve the editor.

2 Existing editors

Concerning lexicon and terminology editors, several tools have already been proposed.

Lexus[2] (Ringersma and Kemps-Snijders, 2007) is a collaborative Web-based lexicon tool developed at the Max Planck Institute for Psycholinguistics. It allows users to create lexica in LMF using the concept naming conventions of ISO data categories. It provides functionalities to include audio, video and still images to the lexicon. With Lexus, users can share lexica and define filters to visualize the entries. Lexus is freely available for use to registered users. Coldic (Núria Bel and Villegas, 2008) is a Web-based lexicographic platform. Similarly to Lexus, it manages LMF lexica. Coldic consists of a database, a graphical interface for the lexicographer and a web services interface. Among its features we cite the automatic generation of a graphical view of the lexical model that is used as a support in the query builder tool. Though released as open source, Coldic is no longer maintained. In addition, Coldic is a single-user tool, i.e. it cannot be used to create lexica in a collaborative way. On the contrary, Wordnet Editor (Szymanski, 2009) was conceived to be cooperative and graphical-oriented. The main goal of the project, carried out at the Gdansk University of Technology, was to create a system providing an easy-to-use interface for WordNet content navigation and editing in an interactive way. A demo version[3] should be available online, but at present the editing features are not accessible and the whole project seems discontinued. Another web editor is PoolParty (Schandl and Blumauer, 2010), a tool for the management of thesauri as Linked Data. PoolParty supports SKOS[4] and has an optional add-on for SKOS-XL. PoolParty allows users to model a vocabulary in RDFS or OWL, either locally or by importing it from external sources. *lemon* source is a Wiki-like site for manipulating and publishing *lemon* data aimed at the collaborative development of lexical resources. It makes it possible to upload a lexicon and share it with others. *lemon* source is an open source project, based on the *lemon* API, and it is freely available online for use. Regarding the *lemon* model, we also cite (Fiorelli et al., 2017), an editor

[2]http://tla.mpi.nl/tools/tla-tools/lexus (last access: 17/07/2017)

[3]http://wordventure.eti.pg.gda.pl/wne/wne.html (last access: 17/07/2017)

[4]https://www.w3.org/2004/02/skos/ (last access: 17/07/2017)

with custom forms to support in the construction of *lemon*. It is an extension of VocBench, a web-based collaborative thesaurus editing and workflow system, natively supporting Semantic Web standards such as RDF, OWL and SKOS(-XL).

Concerning terminologies, there are several commercial Computer-Assisted Translation softwares which integrate components dedicated to terminology management, such as, for example, Trados[5] and Multitrans[6]. It is worth mentioning also the LexGrid Editor (Johnson et al., 2005), a tool developed by the Division of Biomedical Informatics Research of the Mayo Clinic providing the capability to author, view, validate, maintain and extend terminologies defined on the basis of the LexGrid terminology model. An editor designed for constructing corpus-based lexica is CoBaLT (Kenter et al., 2012). This web-based tool has been used to compile a large lexicon of historical Slovene and it manages importing and exporting in TEI P5.

Existing tools allowing users to edit resources on both lexical and ontological levels are very few. The Neon Toolkit[7] has been exploited by LabelTranslator, a tool developed by (Espinoza et al., 2008) in the form of a plug-in to support the LIR (Linguistic Information Repository) model. The tool provides a set of linguistic elements for localizing ontological elements. TextViz (Reymonet et al., 2007) is an editor taking explicitly into consideration references to a textual corpus. It has been developed as another plug-in, this time for the Protégé-OWL framework. TextViz is a visual annotation environment for the construction of Ontological and Terminological Resources (OTR) in the OWL-DL model. TemaTres[8] is an open source web application for the management of controlled vocabularies. It adopts a series of Semantic Web technologies for the representation of controlled vocabularies, thesauri, taxonomies and formal representations of knowledge. Lastly, we cite Tedi[9] (ontoTerminology EDItor), a tool in development at the University Savoie Mont Blanc for the construction of so called "ontoterminologies", defined as terminologies whose conceptual system is a formal ontology.

3 Distinctive characteristics of LexO

As emerges from the previous overview, editors of lexical, terminological or termino-ontological resources are not so widespread and do not always display at the same time all the requirements scholars working in the humanities consider crucial. In many cases, scholars are forced to adopt ontology editors, such as Pinakes (Bozzi and Scotti, 2015) and *Protégé*, to formalize their lexical or terminological resources. As a result, LexO is conceived to have all characteristics we list below. These features were defined on the basis of the experience gained in the creation of lexica and terminological resources in the framework of several projects in the field of Digital Humanities, see (Piccini and Ruimy, 2015), and (Piccini et al., 2016). We do not claim that this list is exhaustive; more features can be added in the future, thanks to the flexible architecture of LexO.

- Ease of use: the editor is meant to be used mainly by humanists and, thus, hide all the technical complexities related to markup languages, language formalities and other technology issues. To make an example, the creation of a new (*lemon*) lexical entry requires a single press of a button: the system, "under the hood", creates a new instance of the LexicalEntry class of the specified lexicon, a new form, a new lexical sense, and all the necessary relationships holding among them.

- Collaborativeness: LexO, being a web application, makes collaborative editing possible. The collaborative construction process of lexical resources offers very promising research opportunities in the context of electronic lexicography. As a matter of fact, a team of users, each one with his/her own role (lexicographers, domain experts, scholars, etc.), can work on the same resource collaboratively. As a result, resources quickly increase in size and are constantly updated. In

[5] http://www.sdltrados.com/ (last access: 17/07/2017)
[6] https://www.multitranstms.com (last access: 17/07/2017)
[7] http://neon-toolkit.org/wiki/Main_Page (last access: 17/07/2017)
[8] http://www.vocabularyserver.com/ (last access: 17/07/2017)
[9] http://christophe-roche.fr/tedi (last access: 17/07/2017)

addition, the automatic consistency checking supported by OWL reasoners can play a crucial role when lexical resources are constructed collaboratively in order to avoid possible "conflicting" assertions.

- Sharing and linking: the editor adheres to international standards for representing lexica and ontologies in the Semantic Web (such as *lemon* and OWL), so that lexical resources can be shared easily or specific entities can be linked to existing datasets.

- Reference to texts: the linking of lexical entries to specific portions of texts (i.e. attestations) is a typical linguistic and philological requirement: lexicographers and terminologists may create their lexical (or terminological) resources from texts; although currently in progress, LexO intends to provide features to link each entity of the resource (being it a form, a term, a concept, etc.) to a text or to a very specific portion of a text, via canonical references mechanisms such as CTS (Tiepmar et al., 2014). Appropriate extensions of the *lemon* model are being developed to represent attestations.

- Extensibility: conceived to handle historical and ancient lexica and terminologies as well, the editor is flexible and extensible enough to formalize peculiar features of such linguistic resources. Among the first major extensions we are currently working on, we cite diachrony and attestation, the first to be implemented by starting from the already available *lemon*-DIA (Khan et al., 2014) and the second one from the work by (Bellandi et al., 2017). It is worth underlying that the process of extension in LexO is facilitated by the fact that also *lemon*, the lexical model of reference, is designed to be modular and to integrate new components easily.

These two latter features make LexO particularly suited to be applied in Humanities, although it may be used by lexicographers and terminologists in general.

With regard to the *lemon* lexical model, we adopted an in-memory persistency solution by exploiting the OWL-API 5.0, a Java API and reference implementation for creating, manipulating and serialising OWL ontologies.

Here we present a first version of LexO. Not all the characteristics listed at the beginning of this Section have already been developed. From the technical point of view, currently data consistency is implemented at user interface level and a reasoner has not yet been set up. In addition, it must be underlined that the in-memory persistence we adopted is not a scalable strategy in case the resource size increases considerably. However, we successfully tested this version of LexO within two research projects aiming at encoding multilingual lexica and termino-ontological resources. The usage of the tool is documented in the next Section, but here we provide an overview of the main interface. It is composed of 4 columns (see the center of Figure 1 and 2). The leftmost column allows scholars to browse lemmas, forms and senses, according to the *lemon* model. By clicking one of them, the system shows the lexical entry of reference in the second column alongside the lemma and its forms, and, in the third one, the relative lexical senses. A user can annotate linguistic and lexicographic properties concerning the lemmatization of terms, such as script types, transliterations and types of variants (see 4.1), and lexico-semantic relations between senses, such as synonymy and translation (see Section 4.1, and Section 4.2) or link a sense to the concept of an ontology of reference (see Section 4.2). The last column, which can be shown or hidden, is used to show the details of the lexical entry which is linked to another one by means of a specific relation.

4 Use Cases

In the next subsections, we show our tool in action within the framework of two projects: DiTMAO and Totus Mundus.

4.1 The DiTMAO Project

LexO is being developed in the context of the project "Dictionnaire de Termes Médico-botaniques de l'Ancien Occitan" (DiTMAO[10]), which aims at constructing an ontology-based information system for Old Occitan medico-botanical terminology. Old Occitan is the medieval stage of Occitan, the autochthonous Romance language spoken in Southern France, today regional minority language with several dialects. During the Middle Ages, the region and its language played a significant role in medical science due to the medical schools of Toulouse and Montpellier and the strong presence of Jewish physicians and scholars. For this reason, Old Occitan medico-botanical terminology is documented both in Latin, Hebrew and Arabic characters (ben Isaak et al., 2011).

The textual basis of DiTMAO lexicon, as described in (Corradini and Mensching, 2010) and (Bos, Corradini, and Mensching, Bos et al.), consists of medico-botanical texts in Latin and in Hebrew script. Among the sources in Hebrew script, the most prominent text type are so-called synonym lists. These lists can be described as ancient multilingual dictionaries, which contain a large amount of Old Occitan medical and botanical terms in Hebrew characters with equivalents or explanations in other languages (also spelled in Hebrew characters), mostly in (Judaeo-)Arabic, but also in Hebrew, Latin, and sometimes in Aramaic (Mensching, 2004), (Mensching, 2009) and (ben Isaak et al., 2011). A special difficulty of medieval texts in vernacular languages is that most terms are documented in a large number of variants (reflecting different spellings, dialects, or historical stages of the languages at issue). The particularities of the DiTMAO corpus (medieval, multilingual and multi-alphabetical) made the lemmatization a complex and intriguing issue (Corradini and Mensching, 2007), (Corradini and Mensching, 2010), and (Corradini, 2014). At date of submission, DiTMAO contains 1758 Old Occitan lemma forms and 1854 variants in Latin script, and 1378 variants in Hebrew script; 305 corresponding terms in Hebrew, 625 terms in Arabic, 77 terms in Latin, 29 terms in Aramaic and 21 mixed terms. Whenever, possible translations into modern French and English are provided.

The DiTMAO project aims at making this terminology accessible to several scientific communities, such as those of Romance and Semitic studies, as well as that of the history of medicine. In order to be useful for an interdisciplinary research community, the terminology should not only be accessible via the lemmata, but also via the meaning or conceptual side of the terms. In traditional Old Occitan lexicography, and in traditional lexicography in general, these two ways of accessing the terminology correspond to two main types of dictionaries: (i) alphabetically ordered dictionaries, such as (Stempel et al., 1997), and onomasiological dictionaries, such as (Baldinger et al., 2005). In onomasiological dictionaries, the terms are grouped according to their meaning and conceptual relations. The lemon model naturally combines these two types of dictionaries. The terminology can be classified according to formal, linguistic criteria and according to the semantics of the terms in an ontology. As the lemon model is designed for modern language lexica, several domain specific extensions had to be defined in order to be suitable for a historical dictionary. The extensions concern the linguistic and conceptual domain as well as the addition of an attestation domain (Weingart and Giovannetti, 2016).

In the following, an example is presented, showing how LexO satisfies the requirements and the workflow of historical (Old Occitan) lexicography, with focus on the lemmatization. The screenshot shows the lemma entry of mandragora, meaning "mandrake", in the red box and one (of many) variants in Hebrew script in the blue box. Due to space limitations, we will focus only on the formal properties of a lemma and its graphical, morpho-phonological or alphabetical variants. A lemma and a variant form have the following common properties, which are domain specific extensions (marked by*) or categories taken from the Lexinfo ontology[11], an extension of lemon that provides data categories for linguistic annotations. The common properties are: "Alphabet" [I*] with the option for Latin, Hebrew or Arabic, the "Transliteration" [II*], which is active in the variant box, showing the transliteration

[10]DiTMAO is a joint project of the PIs Gerrit Bos (Universität zu Köln), Emiliano Giovannetti (Istituto di Linguistica Computazionale "Antonio Zampolli" of the CNR), Maria Sofia Corradini (Università di Pisa) and Guido Mensching (Georg-August-Universität Göttingen). The project is funded by the Deutsche Forschungsgemeinschaft (DFG). Project web page: https://www.uni-goettingen.de/en/487498.html

[11]http://www.lexinfo.net/ontology/2.0/lexinfo.owl (last access: 17/07/2017)

MDRʾGWLH of the hebrew variant. Further, both can be annotated for “Part of speech”, “Number” and “Gender” [III]. The “Documented In” field [IV*] shows the corpus-internal attestation. The lemma form has, in addition, the property of “Other documentation” [V*] for a corpus-external attestation. This is particularly important for two reasons: first, there is additional evidence for the meaning of a term in corpus-external sources and secondly, many terms are only documented in Hebrew script and in this case a corpus-external lemma or a reconstructed form will be used. The type of lemma can be indicated at the “Info” [VI*] drop down menu. The variant types [VII*] are also a domain specific extension. The variant MDRʾGWLH is read as “*madragolha” and differs from the lemma with respect to grapho-phonetic properties, in addition to difference in alphabet. In the yellow box, the semantic relations (translations, corresponding terms), and the conceptual link to the ontology can be managed. Furthermore, for plant names the external sources often mention the binominal scientific name, here Mandragora officinarum L., which is conceived similar to a translation. The leftmost column shows the navigation, which eases the reviewing process by the listing, sorting and counting options.

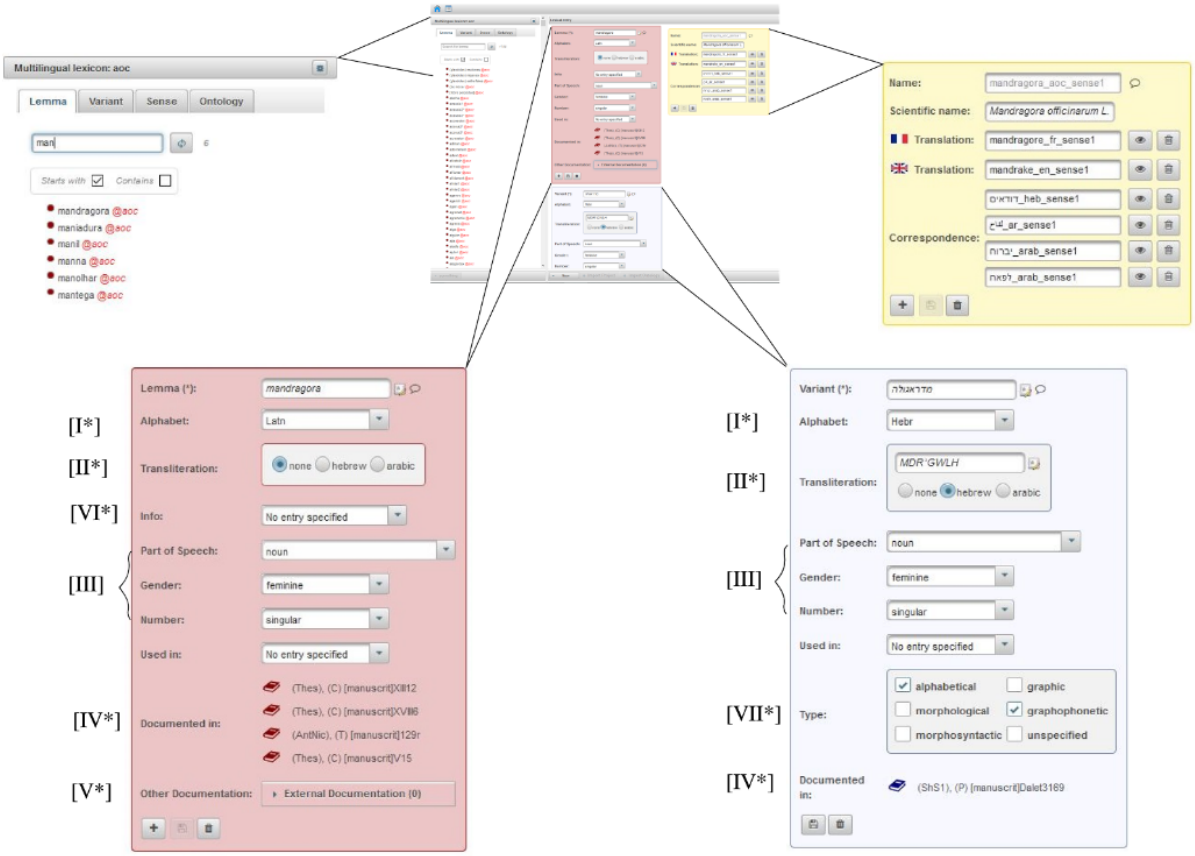

Figure 1: The components of LexO interface. Mandragora entry example.

4.2 The Totus Mundus Project

LexO has been adopted also in the framework of the Italian Project “Todo el mundo es nuestra casa. The World is Our Home. A virtual Journey Around the World Atlas by Matteo Ricci, SJ (1602)” (abbreviated in “Totus Mundus”), coordinated by Elisabetta Corsi, Chair Professor of Sinology at the University La Sapienza (Rome) and conducted in collaboration with the Historical Archives of the Pontifical Gregorian University (APUG) in Rome and the Institute of Informatics and Telematics (IIT) of the CNR in Pisa. The main objective of this project is to take users on a virtual journey through Matteo Ricci’s world map and through its translation into Italian made by the Jesuit sinologist Pasquale D’Elia in 1938 and preserved at APUG. D’Elia’s work is based on the third edition of the map created by Ricci in 1602 in Beijing in collaboration with the Chinese mathematician and astronomer Li Zizhao (1565-1630) and

titled *Kunyu Wanguo Quantu* ("A Map of the Myriad Countries of the World"). This third version, made to stand on six folding screens and to engulf its observer, is the earliest to survive and the first to have given the Chinese a larger cosmological and geographical vision of the earth.

As a matter of fact, the map includes images and annotations describing different regions of the world as well as explanations regarding conceptions of systems of the terrestrial and celestial world. In order to make it possible for scholars to access the Chinese and Italian texts on a semantic basis, a termino-ontological bilingual resource has been developed, where the conceptual and the linguistic layers are separated but intimately linked, in accordance with the paradigms and the methodologies developed over the last few years (see Section 1).

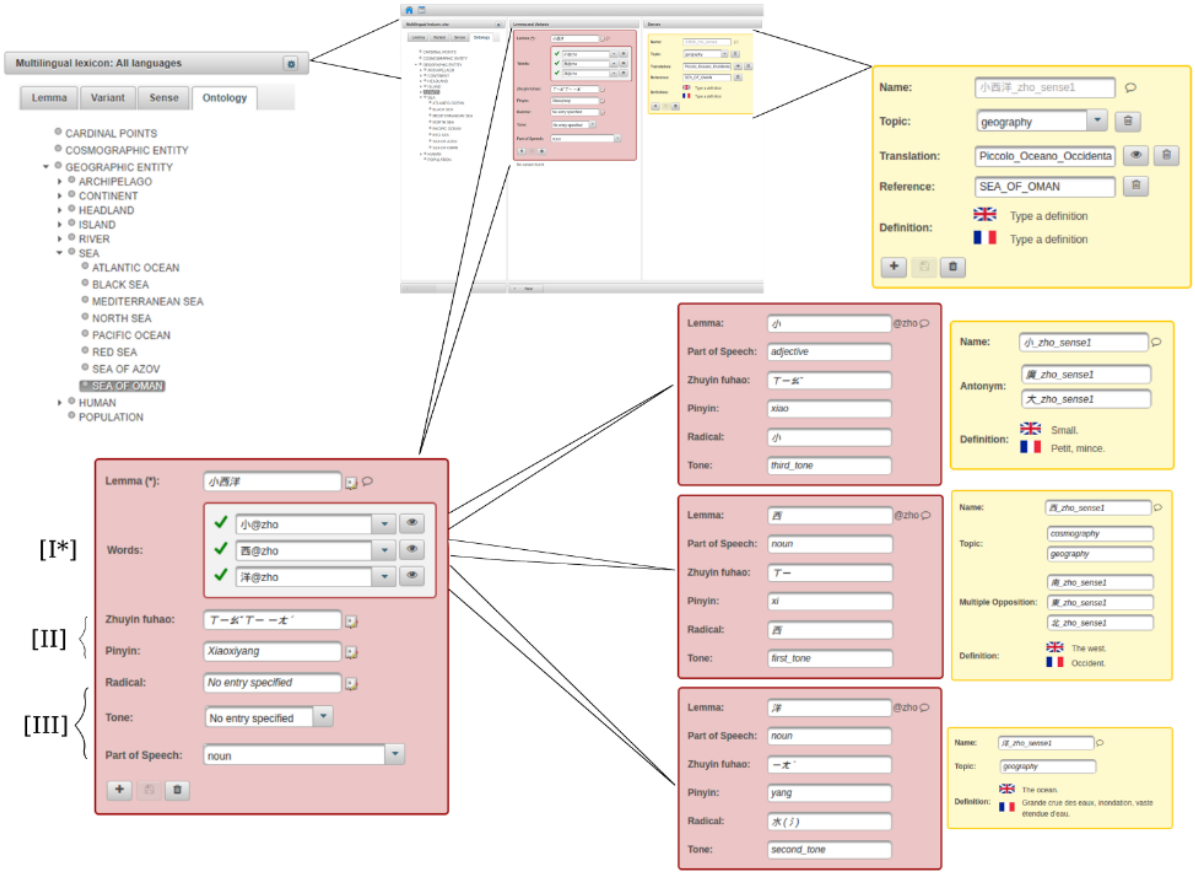

Figure 2: The components of LexO interface. The Little Western Ocean entry example.

The lexical component has been modeled in *lemon* and constructed with LexO, while the conceptual component has been structured into a formal ontology using *Protégé*. The taxonomy has been imported into LexO as well, in order to link each lexical sense to a concept of the ontology (see example in Figure 2). Compared to DitMAO lexicon, the size of Totus Mundus lexicon is smaller, as the project is still in its early stages. It currently contains 81 Chinese lexical entries (52 words and 29 multiwords) and 78 Italian lexical entries (61 words and 17 multiwords), which were extracted manually by the experts. Chinese terms are provided with French and English definitions, drawn respectively from the *Dictionnaire classique de la langue chinoise* by F. S. Couvreur S.J. and *the Chinese-English Dictionary* by the Australian Congregational Missionary R.H. Mathews.

Due to the flexible and modular architecture of the *lemon* model, classes and relationships have been easily customized in order to better meet the specific needs posed by the Chinese language. It must be emphasized that adapting *lemon* to seventeenth-century Chinese language constitutes a challenge and an interesting subject for reflection. Extensions were introduced, such as: i) the OWL class "ProsodicProperty", which subsumes the class "Tone", whose four different tones constitute the instances; ii) the Data

property “radical”, which refers to the graphical (and often semantic) component of Chinese characters, used to organise and list words in a Chinese dictionary; iii) two sub-properties of the Data Property “representation”, i.e. “pinyinTransliteration” and “zhuyinFuhaoTransliteration”. An example of Chinese lexical entry is illustrated in Figure 2, i.e. the multiword *Xiao xi yang* (litt. “The Little Western Ocean”). In the red box the three lexemes which compose the multiword are shown (I); *xiao* “little”, *xi* ”west”, *yang* “ocean”. By clicking on the eye near each lexical entry, on the rightmost column users can visualize the morphological properties as well as all the information concerning the lexical sense of the lexical entries the multiword is composed of. Specifically, in the red box Pinyin and Zhuyin Fuhao Transliterations are provided (II) and the morphological features are detailed such as Part of Speech, Tone, and Radical (III). The word sense is described in the yellow box: French and English definitions are given and lexical relations are also represented (for example synonymy, antonymy etc.) as well as the translation into Italian made by Pasquale D’Elia. The lexical sense is linked through the relation “reference” to an ontology concept. Geographic terminology has changed over time and ancient denominations result sometimes quite obscure especially for users who are not experts in this domain of knowledge. As a result, the ontology linking plays a crucial role, as it makes it possible to understand which geographic entity (sea, island, mountain, continent etc.) was designated by a certain term.

As we can see in Figure 2, *Xiao xi yang* was the ancient denomination of the sea of Oman. The concept is formally described in an ontology which has been built in *Protégé*.

LexO offers also the opportunity to link each lexical sense to external resources such as dBpedia, Wikipedia etc., in accordance with the Semantic Web philosophy.

5 Conclusion

In this paper we presented a first version of LexO, a collaborative editor of multilingual lexica and termino-ontological resources, based on the lemon model. The editor has been created to support lexicographers and terminologists in their work. Despite the fact that the development of LexO is still ongoing, the editor is already being used within two research projects: DiTMAO and Totus Mundus. Adopting LexO in these projects has allowed us to prove its high flexibility, since extensions of the lemon model were introduced easily, to fulfill the needs of the involved scholars.

We are currently focusing our research in the inclusion of other characteristics, such as the diachronic and diatopic variation of both lexical and conceptual aspects as well as the reference to texts. Regarding the ontological level, we plan to enhance LexO with multiple ontology editing. As another major update we want to allow users to create their own extensions of the lemon schema directly inside LexO and to have the interface automatically adapting to the customized model, similarly to what has been done by the team working on VocBench. As soon as it will be stable and documented enough, we plan to release LexO for the community.

Acknowledgment

This research has been partially funded by the DFG in the context of the cooperation agreement between prof. Guido Mensching, director of the DiTMAO project at the Seminar für Romanische Philologie of the Georg-August-Universität Göttingen and the Istituto di Linguistica Computazionale “A. Zampolli” of the Italian National Research Council (August 29th, 2016).

References

Baldinger, K., N. Winkler, and T. Shabafrouz (2005). *Dictionnaire onomasiologique de l’ancien occitan: DAO*. Niemeyer.

Bellandi, A., F. Boschetti, F. Khan, A. M. Del Grosso, and M. Monachini (2017). Provando e riprovando

modelli di dizionario storico digitale: collegare voci, citazioni, interpretazioni. In *AIUCD 2017 Conference - Book of Abstracts*, pp. 119–125.

ben Isaak, S. T., G. Bos, and M. Hussein (2011). *Medical Synonym Lists from Medieval Provence: Shem Tov Ben Isaac of Tortosa: Sefer Ha-Shimmush. Book 29: Part1: Edition and Commentary of List 1 (Hebrew-Arabic-Romance/Latin)*, Volume 1. Brill.

Bos, G., M. Corradini, and G. Mensching. Le ditmao (dictionnaire des termes médico-botaniques de l'ancien occitan): caractères et organisation des données lexicales. *In Proceedings of the XIen Congrès de l'Asociacion Internacionala d'Estudis Occitans (AIEO)*.

Bozzi, A. and A. Scotti (2015). Pinakes e pinakes text : due strumenti per l'archiviazione, lo studio e l'interrogazione di documenti digitali di cultura. In *Reti in rete : per l'inventario e l'edizione dell'Archivio Vallisneri. S. Olschki, Firenze*, pp. 31–52.

Corradini, M. S. (2014). Lessico e tassonomia nell'organizzazione del dictionnaire de termes médico-botaniques de l'ancien occitan (ditmao). *In Revue de linguistique romane* (309-310), 87–132.

Corradini, M. S. and G. Mensching (2007). Les méthodologies et les outils pour la rédaction d'un lexique de la terminologie médico-botanique de l'occitan du moyen âge. *In Actes du XXVe Congrès International de Linguistique et de Philologie Romanes 6*, 87–96.

Corradini, M. S. and G. Mensching (2010). Nuovi aspetti relativi al dictionnaire de termes médico-botaniques de l'ancien occitan (ditmao): creazione di una base di dati integrata con organizzazione onomasiologica. In *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas*, Volume 8, pp. 113–124.

Dancette, J. and M.-C. L'Homme (2004). Building specialized dictionaries using lexical functions. *Linguistica Antverpiensia, New Series–Themes in Translation Studies* (3), 113–131.

Espinoza, M., A. Gómez-Pérez, and E. Mena (2008). *LabelTranslator - A Tool to Automatically Localize an Ontology*, pp. 792–796.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Language, Speech, and Communication. Cambridge, MA: MIT Press.

Fillmore, C. J., C. R. Johnson, and M. R. Petruck (2003). Background to FrameNet. *International Journal of Lexicography 16.3*, 235–250.

Fiorelli, M., T. Lorenzetti, M. T. Pazienza, and A. Stellato (2017). *Assessing VocBench Custom Forms in Supporting Editing of Lemon Datasets*, pp. 237–252.

Gil Francopoulo, Monte George, N. C. M. M. N. B. m. P. and C. Soria (2006). Lexical markup framework (lmf). In *Proceedings of LREC2006, Genoa, Italy*.

Hanks, P. and J. Pustejovsky (2005). A pattern dictionary for natural language processing. *Revue Française de linguistique appliquée 10*(2), 63–82.

Johnson, T. M., H. R. Solbrig, D. C. Armbrust, and C. G. Chute (2005). Lexgrid editor: Terminology authoring for the lexical grid. In *AMIA 2005, American Medical Informatics Association Annual Symposium, Washington, DC, USA, October 22-26, 2005*.

Kenter, T., T. Erjavec, M. v. Dulmin, and D. Fišer (2012). Lexicon construction and corpus annotation of historical language with the cobalt editor. In *Proceedings of the 6th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, LaTeCH '12, pp. 1–6.

Khan, F., F. Boschetti, and F. Frontini (2014). Using lemon to model lexical semantic shift in diachronic lexical resources. In *Proceedings of the Workshop on Linked Data in Linguistics 2014 (LDL-2014)*.

Lenci, A., N. Bel, F. Busa, N. Calzolari, E. Gola, M. Monachini, A. Ogonowski, I. Peters, W. Peters, N. Ruimy, M. Villegas, and A. Zampolli (2000). Simple: A general framework for the development of multilingual lexicons. *Journal of Lexicography 13(4)*, 249–263.

Mel'čuk, I., A. Clas, A. Polguère, et al. (1995). Introduction à la lexicologie explicative et combinatoire. Duculot, Paris/Louvain-la-Neuve.

Mensching, G. (2004). Per la terminologia medico-botanica occitana nei testi ebraici: le liste di sinonimi di shem tov ben isaac di tortosa. *Atti del Convegno internazionale Giornate di studio di lessicografia romanza: il linguaggio scientifico e tecnico (medico, botanico, farmaceutico e nautico) fra Medioevo e Rinascimento (Pisa, 7-8 novembre 2003), Pise, ETS*, 93–108.

Mensching, G. (2009). Listes de synonymes hébraïques-occitanes du domaine médico-botanique au moyen âge. *La voix occitane. Actes du VIIIe Congrès Internationale d'Études Occitanes 1*, 509–526.

Núria Bel, Sergio Espeja, M. M. and M. Villegas (2008, may). Coldic, a lexicographic platform for lmf compliant lexica. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Piccini, S., A. Bellandi, and G. Benotto (2016). Formalizing and querying a diachronic termino-ontological resource: the Clavius case study. In *Proceedings of the Workshop Digitization to Knowledge, July 11, 2016, Krakow, Poland*, Number 126, pp. 38–41.

Piccini, S. and N. Ruimy (2015). Plotiterm: Une ressource termino-ontologique du vocabulaire plotinien. In A. Bozzi (Ed.), *Digital Texts, Translations, Lexicons in the Web*. Firenze: L.S. Olschki.

Pustejovsky, J. (2006). Towards a generative lexical resource : The brandeis semantic ontology. *Proc. LREC 2006*, 385–388.

Reymonet, A., J. Thomas, and N. Aussenac-Gilles (2007). Modelling ontological and terminological resources in owl dl. In *Proceedings of ISWC*, Volume 7.

Ringersma, J. and M. Kemps-Snijders (2007). Creating multimedia dictionaries of endangered languages using LEXUS. In *Proceedings of Interspeech 2007. Baixas, France: ISCA-Int.Speech Communication Assoc.*, pp. 65–68.

Roche, C. Le terme et le concept : fondements d'une ontoterminologie. *TOTh 2007 : Terminologie et Ontologie : Théories et Applications, Jun 2007, Annecy, France*, 1–22.

Romary, L. (2001). An abstract model for the representation of multilingual terminological data: Tmf-terminological markup framework. In *TAMA 2001*.

Schandl, T. and A. Blumauer (2010). *PoolParty: SKOS Thesaurus Management Utilizing Linked Data*, pp. 421–425.

Stempel, W.-D., C. Kraus, R. Peter, and M. Tausend (1997). *Dictionnaire de l'occitan médiéval: DOM: Supplément*. Niemeyer.

Szymanski, J. (2009). Wordventure – developing wordnet in wikipedia-like style.

Tiepmar, J., C. Teichmann, G. Heyer, M. Berti, and G. Crane (2014). A new implementation for canonical text services. In *Proceedings of the 8th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*, pp. 1–8.

Weingart, A. and E. Giovannetti (2016). Extending the lemon model for a dictionary of old occitan medico-botanical terminology. In *International Semantic Web Conference*, pp. 408–421.