Versione PROVVISORIA del contributo presentato al Convegno Annuale

DISCLAIMER

Questa versione dell'abstract non è da considerarsi definitiva e viene pubblicata esclusivamente per facilitare la partecipazione del pubblico al convegno AIUCD 2021

Il Book of Abstract contenente le versioni definitive e dotato di ISBN sarà disponibile liberamente a partire dal 19 gennaio sul sito del convegno sotto licenza creative commons.

Fostering the collaborative creation of Linguistic Linked Open Data with LexO, an open source editor of multilingual lexicons and terminologies

Andrea Bellandi, Emiliano Giovannetti, Simone Marchi, Silvia Piccini, Flavia Sciolette
Istituto di Linguistica Computazionale "A. Zampolli", CNR, Italy - {name.surname}(«»)ilc.cnr.it

ABSTRACT
In the context of the Semantic Web and the Linguistic Linked Open Data paradigm, we propose LexO, a web editor of computational lexicons and terminologies based on the *OntoLex-Lemon* model.

KEYWORDS
LexO; Linguistic Linked Open Data; Semantic Web; Computational Lexicography; Terminology; CLARIN.

1. INTRODUCTION
The growing availability of digitized data has led to the increased diffusion of linguistic resources compliant to the Semantic Web (SW) [2]. In a framework oriented to new paradigms, like Linked Open Data (LOD), the research perspective has increasingly focussed on sharing standards and methods, connecting and reusing resources, encouraging collaborative approaches [3]. In this context, we conceived LexO [1], an open-source editor that, in line with the open science philosophy and the FAIR principles [5], allows lexicographers and terminologists to collaboratively create computational lexicons and terminologies according to the *OntoLex-Lemon* model[1] (*lemon*). This latter is, indeed, considered the *de facto* standard in the context of the SW and has been recently introduced in VocBench [4], born to manage OWL ontologies and SKOS(/XL) thesauri. However, unlike VocBench, LexO is tailored to exclusively manage *lemon* resources and is designed for users who are not familiar with the formalisms underlying the model, and in general the SW and LOD related technologies.

2. LEXO
LexO makes the creation of resources easy, by abstracting from the verbosity of the *lemon* model that is based on complex ontological design patterns[2]. In the LexO interface distinct views are devoted to the linguistic and the conceptual dimensions, thus allowing a twofold approach in the construction of resources. On the one hand, in accordance with the onomasiological approach traditionally adopted in terminology, it is possible to import a domain ontology and then to lexicalize the concepts; on the other hand, users can adopt the semasiological perspective typical of lexicography, focussing their work on the definition of lexical senses. Furthermore, LexO includes functions to improve collaborative work. A team of users can work simultaneously on the same resource: the system keeps it consistent by locking the entries in the works, and users can add notes about a specific entry compiled by other users. LexO provides three types of user-profiles: *viewer* (can only access the dictionary view); *editor* (can create/modify entries); *administrator* (can add/remove users with specific profiles, validate entries, import the ontology, access the statistics of the lexicon).
LexO GUI is depicted in figure 1. Through the select button at the centre of the interface, a lexicographer can switch to the "variation and translation" or to the "syntax and semantics" part, according to the *lemon* model. The former module 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, e.g. the morphological and orthographic variations of a word. On the other hand, the "syntax and semantics" module allows to describe the syntactic behaviour of a word and its government pattern. LexO allows to map the argument of a predicate defined in an ontology and the syntactic argument introduced in a given syntactic frame. LexO has been already applied and tested in several projects, among which we cite DiTMAO (Dictionary of Old Occitan medico-botanical terminology), FdS (Diachronic multilingual lexicon of Ferdinand De Saussure's terminology), Totus Mundus (Bilingual Chinese-Italian resource of Matteo Ricci's map), and the Project for the Translation of the Babylonian Talmud into Italian. Furthermore, LexO was recently presented in the context of the

[1] https://www.w3.org/2016/05/ontolex/

[2] http://ontologydesignpatterns.org/wiki/Main_Page

CLARIN-IT Tour[3]. The tool can be freely downloaded from GitHub[4] and installed following the instructions included therein.

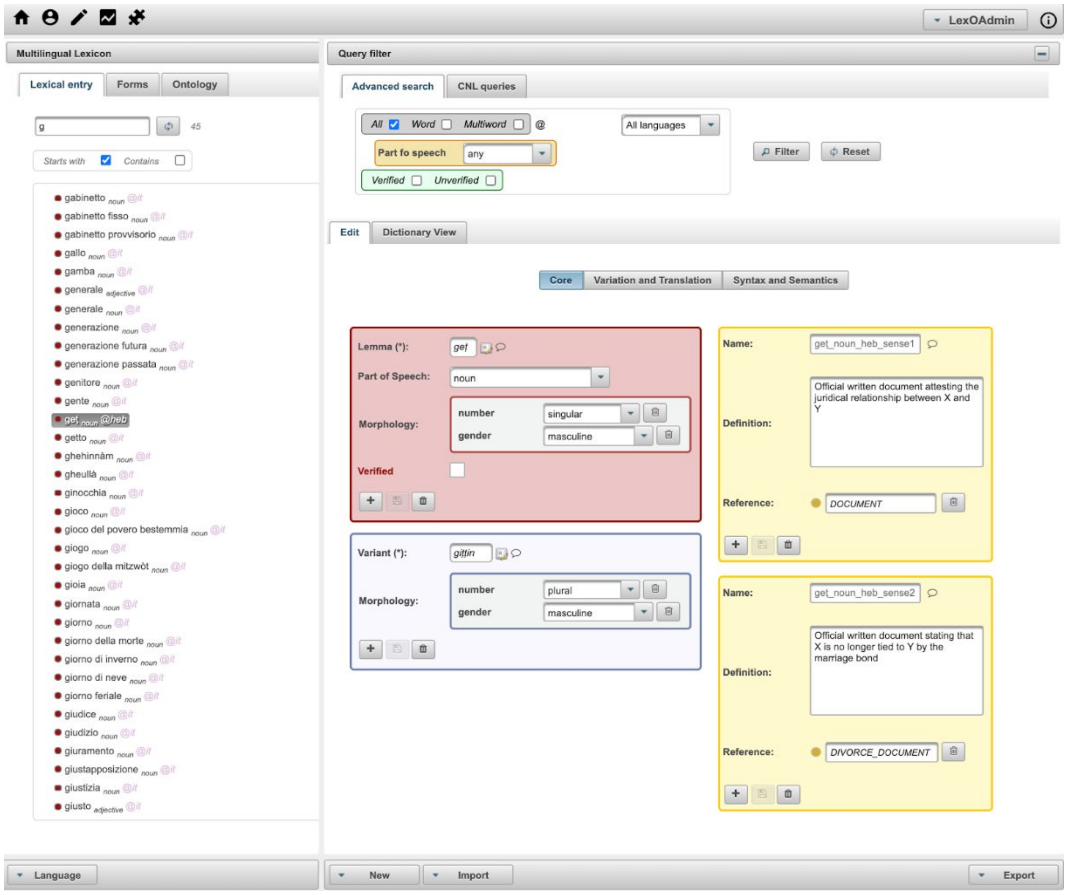

Figure 1. Lexical entries are listed on the left part of the GUI. Red, blue, and yellow boxes represent respectively the lemma, the forms, and the lexical senses related to the selected entry.

3. ACKNOWLEDGMENT

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).

BIBLIOGRAPHY

[1] Bellandi, Andrea, Emiliano Giovannetti, and Anja Weingart. «Multilingual and Multiword Phenomena in a lemon Old Occitan Medico-Botanical Lexicon». *Information*, 2018. **https://doi.org/10.3390/info9030052**

[2] Cimiano, Philipp, Christian Chiarcos, John P. McCrae, and Jorge Gracia. *Linguistic Linked Data*, Springer, 2020.

[3] McCrae, John, Guadalupe Aguado-de-Cea, Paul Buitelaar, Philipp Cimiano, Thierry Declerck, Asunción Gómez-Pérez, Jorge Gracia, *et al.* «Interchanging lexical resources on the Semantic Web». *Language Resources and Evaluation* 46, n. 4 (1 dicembre 2012): 701–19. **https://doi.org/10.1007/s10579-012-9182-3**.

[4] Stellato, Armando, Manuel Fiorelli, Andrea Turbati, Tiziano Lorenzetti, Willem Gemert, Denis Dechandon, Christine Laaboudi-Spoiden, *et al.* «VocBench 3: A collaborative Semantic Web editor for ontologies, thesauri and lexicons». *Semantic Web*, 11 n.5 (2020): 855-81. doi:10.3233/SW-200370.

[5] Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, *et al.* «The FAIR Guiding Principles for Scientific Data Management and Stewardship». *Scientific Data*, 2016. https://doi.org/10.1038/sdata.2016.18.

[3] https://www.clarin.eu/blog/clarin-it-presents-lexo-where-lexicography-meets-semantic-web

[4] https://github.com/andreabellandi/LexO-lite. The suffix "lite" refers to the current limitation to the management of large-sized resources. A full version of LexO overcoming this restriction is under development.