

The ENeL COST Action: European Network for eLexicography

Thierry Declerck¹, Carole Tiberius², Simon Krek³, Eveline Wandl-Vogt⁴

¹ DFKI GmbH, Saarbrücken, Germany
declerck@dfki.de

² INL, Leiden, Netherlands
Carole.Tiberius@inl.nl

³ JSI, Ljubljana, Slovenia
simon.krek@ijs.si

⁴ ACDH, Vienna, Austria
Eveline.Wandl-Vogt@oeaw.ac.at

Abstract. In this abstract, we document the interest and the actual state of work within the ENeL project consisting in mapping various European lexicographic resources into a Linked Open Data compliant format and how we aim at bridging eLexicography and Semantic Web via cooperative approaches between projects.

1 Introduction

The aim of the European Network of e-Lexicography (ENeL)¹ is to increase, coordinate and harmonize European research in the field of e-lexicography and to make authoritative dictionaries on the languages of Europe easily accessible. ENeL is therefore making lexical knowledge of small and large languages available in a European dictionary portal. This effort involves the exchange of resources, technologies and experience in e-lexicography and provides support for dictionaries which are not yet online. A focus point of ENeL consists in discussing and establishing standards for innovative e-dictionaries that fully exploit the possibilities of the digital medium. In doing so, ENeL explores new ways of representing the common heritage of European languages by developing shared editorial practices and by interconnecting already existing information.

For participants of the Working Group 3 “Innovative eDictionaries” of ENeL it rapidly seemed obvious that the expanding Linked Open Data (LOD) framework², and more specifically the emerging Linguistic Linked Open Data (LLOD)³, could offer a potential infrastructure for realizing some of its goals. At the core of the

¹ ENeL is supported by an ISCH COST Action, with number IS1305. See <http://www.elexicography.eu/>.

² See <http://linkeddata.org/> for more details.

³ See <http://linguistics.okfn.org/resources/llood/> for more details

LLOD we found the Ontolex model for the representation of linguistic knowledge in the LOD, and which we started to test for the purposes of ENeL.

2 The Ontolex Model

This model is being developed in the context of a W3C community group⁴. It describes a modular approach to lexicon specification, allowing thus the eLexicographer to depart from the “book” view that the headword is the (unique) entry point to information encoded in a dictionary. Senses, usages, concepts, etc. can be independently described, accessed and are all linked to what was considered the headword, and which now is encoded as a virtual entry in a RDF model.

With Ontolex, we can now advocate for the fact that all elements of a dictionary entry can be described independently of each other and connected by explicit relation markers. Now, the components of a dictionary entry can be distributed in a network and be linked together by RDF encoded relations/properties. An important aspect of this model is also the relation called “reference”. This represents a property that supports the linking of senses of lexicon entries to knowledge objects available in the LOD cloud. This reflects also our view that the meaning of a lexicon (or dictionary) entry is no longer necessarily encoded in the lexicon (or dictionary) but can be referred to in the Web of data. Practically, this means that a dictionary author does not need to describe all components or elements of an entry in details, but that she/he can also draw on existing elements (e.g. the etymology of a word), and can simply refer to it. We are convinced that these properties of the model can facilitate and support the cooperation between various scientific lexicographers, and that this can result in virtual and collaborative research environments in the lexicographical field.

3 Cooperation between ENeL and Semantic Web Projects

In the light of the development mentioned above, it was clear that cooperation between ENeL and Semantic Web projects and initiatives is necessary. Contacts and exchanges of data have already taken place between ENeL and the LIDER project⁵, and this has proven to be very fruitful, since LIDER is dealing mainly with the representation of language data at large in the LOD cloud. ENeL can then provide for the LOD compliant publishing of authoritative dictionaries, based on the formalization of their data provided by LIDER. Various aspects of this cooperative work will be presented at eLex 2015 and at the MLW4 Workshop at ESWC 2015.

Beyond, ENeL is seeking for cooperation on how to link the lexical and other information contained in the authoritative dictionaries to encyclopedic knowledge available in the LOD data sets. Since authoritative dictionaries in the past conveyed also a lot of world knowledge, it is important to make this information available in the digital world and to link it to related facts encoded elsewhere in the Web of data.

⁴ See <http://www.w3.org/community/ontolex/> for more details

⁵ See <http://lider-project.eu>