

## ► **Tracing Semantic Change with Multilingual LLOD and Diachronic Word Embeddings**

**Florentina Armaselu,**

*University of Luxembourg, Luxembourg, florentina.armaselu@uni.lu*

**Elena-Simona Apostol,**

*Uppsala University, Sweden, elena-simona.apostol@it.uu.se*

**Christian Chiarcos,**

*University of Cologne / Goethe University Frankfurt, Germany, chiarcos@cs.uni-frankfurt.de*

**Anas Fahad Khan,**

*Istituto di Linguistica Computazionale “A. Zampolli”, Italy, fahad.khan@ilc.cnr.it*

**Chaya Liebeskind,**

*Jerusalem College of Technology, Israel, liebchaya@gmail.com*

**Barbara McGillivray,**

*King’s College London, United Kingdom, barbara.mcgillivray@kcl.ac.uk*

**Ciprian-Octavian Truică,**

*Uppsala University, Sweden, ciprian-octavian.truica@it.uu.se*

**Giedrė Valūnaitė-Oleškevičienė,**

*Mykolas Romeris University, Lithuania, gvalunaite@mruni.eu*

**Purpose:** The project will combine word embedding techniques and linguistic linked open data (LLOD) with theoretical aspects from lexical semantics, the history of concepts, and knowledge organization to trace the evolution of concepts in a collection of multilingual diachronic corpora of seven extinct and extant languages (Latin, Ancient Greek, Hebrew, French, Old Lithuanian, Romanian, German). The outcome will consist of a sample of diachronic ontologies to be published on the LLOD cloud. It will also comprise reflections on the potential interconnections across different languages that can be built through these knowledge structures.

**Design/methodology/approach:** The methodology will include the following steps: (1) Train diachronic static and contextual word embeddings (i.e., Word2Vec, FastText, ELMo, BERT) by time slice and sliding window on each dataset. Drawing on state-of-the-art methods on unsupervised lexical semantic change detection, we will model semantic change in the vector space by taking into account theoretical considerations such as semasiological and onomasiological mechanisms, and core – margin and intension – extension – label conceptualizations. (2) Create cross-lingual connections between concepts for overlapping and sequential intervals on the general timeline. (3) Devise transformation pipelines to move from the vector space to a representation of concept evolution through chains of label, core/intension, margin/extension at different time points using Semantic Web formalisms and vocabularies. For instance, we will use the OntoLex-Lemon model and its extension, the Frequency Attestations and Corpus

information (FrAC) module. We will employ the OWL-Time vocabulary to explicitly represent time in our ontologies, and possibly utilize the semantic data derived from our approach to enrich multilingual WordNets with diachronic information.

**Findings:** The project will contribute new forms of conceiving, detecting, and representing semantic change by combining theoretical, computational, corpus- and knowledge-based approaches for both high and low resource languages.

**Research limitations/implications:** The main limitations consist of potential data sparsity and the lack of existing LLOD mechanisms that fully support the representation of semantic change. The heterogeneous nature of the available datasets poses an additional challenge in processing the data through a common yet flexible methodology.

**Practical implications:** Given the complexity of the tasks, we will need to use multiple approaches and resources (corpus-, dictionary-, linked data-based), and to propose extensions to existing LLOD formalisms.

**Originality/Value:** The novelty of the proposal will consist in bridging research perspectives that have not yet been considered together (e.g., semasiological/onomasiological, core/margin, vector semantics, linked data) to trace concept evolution in a multilingual diachronic setting. It will also reside in broadening current LLOD models to accommodate spatio-temporal dimensions and the dynamics of change, and in reconciling the variety of the available datasets to build cross-lingual space-time connections.

**Keywords:** *diachronic word embedding, LLOD, semantic change, history of concepts, knowledge organization.*

**Research type:** Research paper.