Context

Use case initiated within the COST Action Nexus Linguarum, European network for Web centred linguistic data science (CA18209), 2019-2023. https://nexuslinguarum.eu/

Goals

Create a comparative methodological framework for tracing the "histories" or evolution of concepts in different languages and cultural fields (history, literature, philosophy, religion, etc.) and generate a sample of multilingual LLOD ontologies to represent semantic change and corresponding explanations, by using NLP and Semantic Web technologies.

Concepts

Domain: socio-cultural transformation.

Semantic fields: geo-political and cultural entities (Europe, West, East, etc.), education, sciences, technology and innovations, social and societal processes (migration, urbanisation, modernisation, globalisation), state and citizenship, beliefs, values and attitudes (e.g. religion, democracy, political participation), economy, health and well-being, everyday life, family and social relations, time and collective memory, work and leisure, customs and traditions, literature and philosophy.

Questions

Can the applied methodology inform us about the interrelation between linguistic, social and cultural innovation over time, and the socio-cultural roots of innovation?

What may be learned about the combination of human and machine agency in the process of construction and dissemination of knowledge through NLP and Semantic Web technologies, and of explaining the underlying mechanisms?

Workflow

Methodological starting points

Data preparation: conversions (XML to TXT), metadata extraction (language, date, genre), structuring by time slice (year, decade, century).

Theoretical modelling of semantic change: lexicostatistical (new meanings) vs. onomasiological (new lexical items) innovation mechanisms [1]; concept core and margin variability [2]; extension, and label that define the meaning of a concept, distance measures [3].

LLD modelling of semantic change: extensions of the OntoLex-Lemon model [4, 5] and the Lexical Markup Framework (LMF) to represent diachronic information [6].

Detection of lexical semantic change: word embeddings enriched with temporal-spatial information [7]; SemEval-2020 task 1 (unsupervised lexical semantic change detection) [8]; transformer-based semantic change detection [9]; https://github.com/towardsai/word2vec to Build a Simple Ontology Learning System.

Hypothesis testing

Can the applied methodology inform us about the interrelation between linguistic, social and cultural innovation over time, and the socio-cultural roots of innovation?

What may be learned about the combination of human and machine agency in the process of construction and dissemination of knowledge through NLP and Semantic Web technologies, and of explaining the underlying mechanisms?

References


Explaining AI (XAS): four principles of explainable AI systems: explanation (the system is "capable of providing an explanation"), meaningfulness (the recipient "understands the system’s explanations"), explanation accuracy (the explanation “may or may not accurately describe how the system came to its conclusion”), knowledge limits (the systems identify cases when they “were not designed or approved to operate”, or “their answers are not reliable”) [15].

Next steps

Hypothesis testing (can these various types of methods and tools be integrated into a coherent implementation?) and workflow of the implementation.