A semantic change time-lapse for Romance languages and English

Alina Maria Cristea,[1] Anca Dinu,[1,3] Liviu P. Dinu,[1,2]
Simona Georgescu,[1,3] Ana Sabina Uban,[1,2] Laurențiu Zoicaș[1,3]

[1] Human Languages Technologies Research Center, University of Bucharest
[2] Faculty of Mathematics and Computer Science, University of Bucharest
[3] Faculty of Foreign Languages and Literatures, University of Bucharest

{alina.cristea,ldinu,auban}@fmi.unibuc.ro,
{anca.dinu,simona.georgescu,laurentiu.zoicas}@lls.unibuc.ro

Abstract

In recent years, studies in computational linguistics have focused on the issue of lexical semantic change, tracking the shift in the meaning of words by looking at their usage across time in corpora dating from different time periods. Vector spaces and word embeddings have been used for tracking semantic shifts of words across different time periods. But all previous computational studies on lexical semantic change have, to our knowledge, only looked at the semantic change of the words within a single language; moreover, they are overwhelmingly focused on English. However, words do not evolve only in their own language in isolation, but are rather inherited and borrowed between and across languages.

We propose a multi-lingual and cross-lingual approach to automatically detecting and measuring semantic change, based on computational representations derived automatically from large corpora.

We introduce a novel way to measure semantic change synchronically across languages, by tracking the divergence of cognate words from their original etymon, for words in Latin and across languages, in the Romance language family [1]. We define a metric of cognate divergence and show how it can be used as a measure of language semantic divergence [2]. Additionally, we explore the statistical properties of the vectorial spaces where the word embeddings reside, in order to understand how semantic divergence affects word properties in the multi-lingual setting, and define statistical laws of cross-lingual semantic change [3].

Using the same technical framework, we develop an algorithm for detecting and correcting false friends as an application for natural language processing [4].

Secondly, we use diachronic data in order to track the evolution in word meaning over time for words in English, French, and Italian. We introduce the task of dating words in a language automatically [5], and propose a model to automatically conjecture the period when a word entered a language.

Keywords

Semantic change, cross-lingual, Romance languages.

Acknowledgements

This research is supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS/CCCDI UEFISCDI, project number 108, CoToHiLi, within PNCDI III.

References

[1] Cross-lingual Laws of Semantic Change, Ana-Sabina Uban, Alina Maria Ciobanu, Liviu P. Dinu, in Nina Tahmasebi, Lars Borin, Adam Jatowt, Yang Xu, Simon Hengchen (eds). Computational Approaches to Semantic Change. Berlin: Language Science Press., 2021.
[2] A Computational Approach to Measuring the Semantic Divergence of Cognates, Ana Sabina Uban, Liviu P. Dinu, Alina Ciobanu. In Springer LNCS, presented at the International Conference on Computational Linguistics and Intelligent Text Processing (CICLing), La Rochelle, France, 7-13 April 2019.
[3] Studying Laws of Semantic Divergence across Languages using Cognate Sets, Ana Sabina Uban, Alina Ciobanu, Liviu P. Dinu. In Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change 2019, Co-located with ACL 2019, Florence, Italy, 2 August 2019.
[4] Automatically Building a Multilingual Lexicon of False Friends with no Supervision, Ana Sabina Uban and Liviu P Dinu. In Proceedings of The 12th Language Resources and Evaluation Conference (LREC), p 3001–3007, 13-15 May 2020.
[5] Tracking Semantic Change in Cognate Sets for English and Romance Languages, Ana Sabina Uban, Alina Maria Cristea, Anca Dinu, Liviu P. Dinu, Simona Georgescu, Laurentiu Zoicas, In Proceedings of the 2nd International Workshop on Computational Approaches to Historical Language Change 2021, Co-located with ACL 2021, 5-6 August 2021.