

Generatory or: "How We Went beyond Sense Inventories and Learned to Gloss"

Roberto Navigli

Sapienza University of Rome

Sapienza NLP Group

Italy

`navigli@di.uniroma1.it`

Abstract

In this talk I present Generatory, an approach that goes beyond the mainstream assumption that word senses can be represented as discrete items of a predefined inventory, and put forward a unified model which produces contextualized definitions for arbitrary lexical items, from words to phrases and even sentences. Generatory employs a novel span-based encoding scheme to fine-tune an English pre-trained Encoder-Decoder system and generate new definitions. Our model outperforms previous approaches in the generative task of Definition Modeling in many settings, but it also matches or surpasses the state of the art in discriminative tasks such as Word Sense Disambiguation and Word-in-Context. I also show that Generatory benefits from training on definitions from multiple inventories, with strong gains across benchmarks, including a novel dataset of definitions for free adjective-noun phrases, and discuss interesting examples of generated definitions.

Joint work with Michele Bevilacqua and Marco Maru.