NAtural LOgic meets MAchine Learning (NALOMA) 2020
Workshop @ WeSSLLI 2020, July 12-17 2020

Authors: Mark Steedman

Title: Linguistic Semantics and Natural Language Inference

Abstract: One of the most significant obstacles to application of natural language processsing technology (NLP) is the lack of a semantics that supports common-sense inference of the kind that human readers do quite effortlessly---for instance, understanding that a question about ownership of one company by another is answered by text about the recent purchase of the former by the latter. Such a semantics must be compatible with logical operators of the kind that form the backbone of linguistic semantics, such as negation and (some) generalized quantifiers. It must be based on a directed graph of meaning-postulates expressing "entailments" between relations over entities such as buying and owning. Such graphs are too vast and ramifying to build by hand. Two methods for building them automatically have been proposed: "machine-reading", using parsers over huge amounts of text from many sources to identify relations grounded in typed named-entities ind infer meaning-postulates; representing relations as "embeddings" vectors in dimensionally-reduced representations of their collocations and using distributional inclusion to detect similar sorts of entailment. The paper discusses both methods, and reports some progress on building arge entailment graphs, using them to defining semantic primitives for a form-independent semantics, and using that semantics to build and interrogate a large Knowledge graph from text.