

Children acquire language spontaneously very rapidly, attaining an impressive mastery over an immensely complex means of conceptual thinking and communication within just a few years. The neural and cognitive mechanisms underlying this feat of learning have long been a mystery for linguists and natural language engineers. A few years ago, our lab has developed ADIOS – standing for Algorithm for Distillation Of Structure – in an attempt to elucidate the learning principles that may be involved in this sort of learning. Given a corpus of raw, unprocessed natural language sentences, ADIOS infers a probabilistic context free grammar that is able to generate novel grammatical sentences and identify unseen sentences from the same language. Its performance was assessed in a wide range of experiments, from learning corpora in English and Chinese, to bioinformatics applications.

In this lecture I will present some the main results of my own research on natural language grammar inference algorithms. After an overview of the drawbacks of ADIOS, some of which are shared by virtually all grammar inference algorithms that have been developed to date, I will present ConText, a new algorithm that sidesteps these drawbacks, and demonstrate how this allows it to attain much better performance. Being a much simpler algorithm, ConText makes possible an inquiry into the relation between grammar inference and a related task that is popular in NLP – that of unsupervised parsing, which aims to infer the syntactic parse trees of the sentences in the training corpus.

Another algorithmic development is COG, a scheme for Controlling Over-Generalizations, which, given a PCFG, prunes it to provide a better fit to the training corpus, leading to significantly better performance. As it turns out, the principle underlying COG is general and can be applied to other kinds of probabilistic models as well, by reducing probabilistic modeling to a set of discriminative classification problems. We demonstrate this by showing how COG can be applied to language modeling, yielding improved perplexity scores relative to an n-gram baseline.

The lecture will not presuppose any prior knowledge on natural language processing.