Grammar learning is a well-studied problem in NLP, but the task is particularly difficult for low-resource languages. In this talk, I will discuss our current work in learning combinatory categorial grammars from various forms of weak supervision. First, I will show how we can learn good sequence-based CCG supertaggers by encoding universal, inherent properties of the CCG formalism as priors over both the appearance of supertags and the transitions between supertags. These universal priors can, in turn, be combined with corpus-specific knowledge derived from available (partial) tag dictionaries and unannotated text to further improve tagging performance. Then, I will discuss our current efforts to extend these principles to tree grammars to learn CCG parsers. Finally, I will discuss how simple annotations—particularly those given in the Graph Fragment Language developed at CMU—may be used to help learn parsers under extremely tight annotation budgets.

This work is in collaboration with Jason Baldridge, Chris Dyer, and Noah Smith.

Dan is a Computer Science Ph.D. student at The University of Texas at Austin. His research focuses on Natural Language Processing and Machine Learning. He was a Best Talk Award nominee at NAACL this past year. He thinks slides should have less text and more animations.

http://www.cs.cmu.edu/~nlp-lunch/