Effective information extraction and question answering require modeling of the deep syntactic and semantic structures of natural language. At the same time, acquiring training data specifying these full structures is prohibitively expensive. Our goal is therefore to learn models of language that can induce latent structures (e.g., parse trees) from raw observations (e.g., sentences) in an unsupervised way. First, I will present spectral methods for learning latent parse tree models. In contrast to existing algorithms for learning latent-variable models such as EM, our method has global convergence guarantees. Second, I will present a semantic model that maps natural language questions to answers via a latent database query, and show results on large-scale question.

Bio: Dr. Percy Liang is an Assistant Professor of Computer Science at Stanford University (B.S. from MIT, 2004; Ph.D. from UC Berkeley, 2011). His research focuses on methods for learning richly-structured statistical models from limited supervision, most recently in the context of semantic parsing in natural language processing. He won a best student paper at the International Conference on Machine Learning in 2008, received the NSF, GAANN, and NDSEG fellowships, and is also a 2010 Siebel Scholar.