Bayesian Tools for Natural Language Learning
Invited talk

Yee Whye Teh
Gatsby Computational Neuroscience Unit, UCL
`ywteh@gatsby.ucl.ac.uk`

In recent years Bayesian techniques have made good inroads in computational linguistics, due to their protection against overfitting and expressiveness of the Bayesian modeling language. However most Bayesian models proposed so far have used pretty simple prior distributions, chosen more for computational convenience than as reflections of real prior knowledge.

In this talk I will propose that prior distributions can be powerful ways to put computational linguistics knowledge into your models, and give two examples from my own work. Firstly, hierarchical priors can allow you to specify relationships among different components of your model so that the information learned in one component can be shared with the rest, improving the estimation of parameters for all. Secondly, newer distributions like Pitman-Yor processes have interesting power-law characteristics that if used as prior distributions can allow your linguistic models to express Zipf's Law and Heap's Law.

I will round up the talk with a discussion of the viability of the Bayesian approach, in a future where we have too much data, making the natural language learning problem more a computational rather than a statistical one.