Statistical Processing of Linguistic Structure: Testing a Bayesian Model of Sentence Processing

Dan Jurafsky
University of Colorado, Boulder
(This talk describes joint work with Srini Narayanan.)

Language comprehension is a classic problem of reasoning under uncertainty. Language comes to us as a noisy, unsegmented, ambiguous mass of auditory waveforms or visual stimuli. Humans must somehow combine this input with other knowledge in order to come up with reasonable interpretations and actions. The natural model for handling this kind of decision-making under uncertainty is probability theory. Not only is probability theory the best normative model for making decisions given noisy, ambiguous input, but we also know that humans use probabilistic or statistical methods to resolve ambiguities in other cognitive arenas like vision, learning, and causal reasoning. But linguistic processing is known to involve rich complex structures at many levels. Can a coherent probabilistic interpretation be given for the problem of language interpretation at different linguistic levels? What kinds of conditional independence assumptions can we make in combining knowledge, and how can we represent these assumptions? And perhaps most crucially, how can sophisticated linguistic structural knowledge be combined with probabilistic augmentations?

In previous work [Jurafsky 1996, Narayanan and Jurafsky 1998, Narayanan and Jurafsky 2000], we have proposed that humans solve the problem of linguistic decision-making under uncertainty by acting as Bayesian, probabilistic reasoners. The advantage of a Bayesian approach to language processing is that it gives a model of what probability to assign to a particular belief or interpretation, and how these beliefs should be updated in the light of new evidence. We propose that in the course of processing a sentence, humans continuously compute the posterior probability of each partial interpretation, updating these probabilities as each piece of evidence arrives. Our probabilistic model is implemented via Bayesian networks, which afford this kind of on-line belief update, and also allow us to incorporate any kind of evidence; lexical, syntactic, semantic, pragmatic.

In this talk we describe an instantiation of the model based on three architectural principles and three kinds of probabilities. The principle of Parallelism states that multiple interpretations of a sentence are maintained simultaneously, that each interpretation is associated with a probability, and that this probability explains disambiguation preferences. The principle of Attention states that the comprehender places attentional focus on the most-probable interpretation, causing increased reading time when this interpretation drops in probability. The principle of Expectation states that the parser maintains probabilistic expectations about upcoming words and structure, and that unexpected words or structure also cause increased reading time. This instantiation of the model includes three kinds of probabilities: structural probabilities (probabilities of words or linguistic structure), valence probabilities (probabilities of structure given a lexical head), and N-gram or Markov probabilities (probabilities of words given neighboring words).

We then present three studies testing the model against behavioral results, and comparing against other probabilistic models of sentence processing (such as Crocker and Brants (2000), Hale (2001), and McRae, Spivey-Knowlton and Tanenhaus (1998).) In our first study we explore the ability of the model to predict human preferences in disambiguation. We show that the model is able to account for a wide variety of behavioral results on preference in lexical and syntactic ambiguities. In our second study we focus on the reading-time predictions of the model. We test the model on the reading time data from a behavioral study on the main clause/reduced relative ambiguity (McRae et al 1998). We show that the model is able to predict the effect of syntactic and semantic structure on reading time differences in ambiguous regions. In our third study we focus on the reading time results of Pickering, Traxler and Crocker (2000) on the direct object/sentential complement ambiguity. Pickering et al interpreted their results as an argument against frequency-based models. We show that, on the contrary, their results follows directly from the probabilistic expectations of our model.

Probabilistic models of syntax and semantics are only one factor in language comprehension. Obviously a complete model of sentence processing must also include other factors like short term memory limitations and locality (e.g. Gibson 1998/2000), prosody (e.g. Carlson, Clifton, & Frazier 2001; Watson and Gibson 2003), structural biases (e.g. Frazier and Clifton 1995, Townsend and Bever 2001), and the influence of many kinds of context (e.g. Altmann & Kamide 1999, Spivey et al 2003 inter alia). But we hope to have shown that many such factors could be embedded in a principled Bayesian cognitive architecture, and that it is possible to handle a wide variety of behavioral data by integrating rich models of linguistic structure with sophisticated probabilistic reasoning.