Training Plan

Overview

The research plan that I have proposed requires me to have a good understanding of Semantic parsing in computational land and of human language learning in linguistics land. I intend to acquire these by taking a combination of relevant courses from the Computer Science and the Linguistics department. Along with these, I will require a deep understanding of the Abstract Meaning Representation paradigm some of which I hope to acquire through my summer internship this year (2015) at ISI USC working with the very creators of AMR. My graduate research and training will be conducted under the joint advising of my two advisors - Dr. Hal Daume III (from Computer Science) and Dr. Philip Resnik (from Linguistics department).

Coursework

In the first two years of my PhD program I completed eight out of the 10 course requirements of my home department (Computer Science). Out of the eight courses, five of them were cross-listed in the Linguistics department. Computational Linguistics I & II, seminar on Semantics in Computational Linguistics and Multi-lingual Natural Language Processing are some of these courses that will help lay the foundation of my training. In future, along with completing the two course requirements for my home department, I intend to take a few additional courses from the Linguistics department (like Syntax and Semantics) that will help me understand better some of the linguistic intuitions that humans use during parsing.

Science Policy Experience

Having been born and brought up in India, I have been exposed to a lot of different Indian languages. However I have seen very little computational work being done in any of them. Most of the Indian languages easily fall under the category low resource languages since digital data for them is almost non-existent. For my policy internship I would be very much interested in looking into what are the challenges that one would face if one were to start doing computational work on some of these low resource Indian languages. Some examples could be - digitization of data available in hard copy texts, putting together some of the existing research in these languages (if any), etc.