Dissertation
My dissertation work was an interdisciplinary project in computer science and linguistics. The two major claims made in this work were that:

- There is a correlation between desirable formal properties of grammars and constraints on non-local dependencies in syntax. That is, for a certain family of formal systems, those that retain desirable formal properties also derive linguistic constraints on movement.

- The different syntactic constraints on various types of non-local dependencies can be reduced to independent details of the derivation of syntactic units of the size of a single clause.

These claims were made in the context of an investigation of reduced constructions in Romance (e.g., clitic climbing) and German (e.g., long distance scrambling) from the perspective of Tree Adjoining Grammar (TAG). TAG has been studied for many years, and is of interest for both linguistic reasons, for the way in which it enforces locality violations, and for computational reasons, for its nice mathematical and parsing properties. However, attempts to handle reduced constructions have led to revised systems that lost some of these attractive linguistic and computational properties.

In my work I showed that the problems caused by reduced constructions are the same as some previously unnoticed problems for TAG, leading to a solution that was able to retain the desirable properties. Within this solution I presented arguments for the two claims listed above.

Current Goals

- My current focus of research is information extraction, following from my work as part of the team that developed Penn’s first contribution to the Automatic Content Extraction (ACE) evaluation (http://www.nist.gov/speech/tests/ace/index.htm). My current research goal is to apply some of the conclusions from my earlier work to this area. In particular, I am testing the hypothesis that the notion of locality inherent in TAG will allow for an efficient approach to automatic extraction of predicate-argument structure.

  I am beginning this work with a more self-contained investigation - whether TAG can be used as a level more powerful and informative than simple sentence chunking, yet less computationally expensive than full parsing. This will test the claim that full parsing is required to recover the predicate-argument structure. I then plan on utilizing a TAG-based approach to some of the problems posed by work on the ACE project, concentrating on the task of recovering the information about relations between entities in the corpus. If this yields good results, I am very interested in trying to expand on this to build a more integrated system, rather than the pipelined statistical approach taken in the effort that I worked on. That is, the TAG elementary trees could be used as a type of parse forest and the different components of the system could all focus on those elementary trees, thereby overcoming the problems of a pipelined system. This long-term goal, however, is currently speculative and depends on the outcome of the current work.

- I am currently developing tools for labeling the parse trees in the Chinese Treebank with predicate argument tags. One reason for this work is to use the resulting labeled trees as a first step in the development of a proposition bank of “gold standard” predicate-argument structures for the sentences in the Chinese Treebank. Another reason is that as a result of this work we now have a corpus
of approximately 35,000 subcategorization frames for Chinese verbs, and my colleagues and I are interested in finding out how well such syntactic information can be used to group together the verbs into some sort of meaningful classes, along the lines of the Levin classes.

- Following my dissertation work, I am continuing the exploration of constrained mathematical formalisms for natural language analysis and the interaction between formal properties and linguistic constraints. A priority within this work is investigation of the other aspects of reduced constructions that I did not examine in my dissertation, such as German verb clusters. Particularly interesting is that other researchers have proposed analyses for these constructions using the mechanism of “remnant movement” that cannot be implemented within even a modified version of TAG, since it is a fundamentally different type of recursive operation, with quite different mathematical properties.