Research Statement

Tim Miller

October 22, 2009

Research Context

Current research in probabilistic models of natural language is heavily biased towards systems that do not take into account the fact that language is typically generated and recognized incrementally. Even systems designed to operate on spontaneous speech tend to take as input an entire utterance, already transcribed by a pre-process, and then perform downstream processing on the entire string at once. While these methods are effective on edited text and correctly transcribed fluent speech, a pipelined approach still separates uncertainty about different levels of the signal which might be useful when combined, and cannot operate in real time on streaming input.

My research, in contrast, focuses on developing *incremental processing* models for streaming input, such as speech or ambiguously punctuated text (e.g., unedited text in blogs or other digital media). This line of research offers contributions to both engineering and scientific research on models of human language. From an engineering perspective, these incremental models enable continuous interaction in human-computer interfaces, offer more realistic and useful representations of user input, and can potentially enable faster processing of language. From a scientific perspective, these models are attractive to psycholinguistics researchers, because their incremental nature means that they take advantage of the same information that would be available to a human listener, and can thus contribute to psycholinguistic models of language recognition.

Research Accomplishments

My thesis work concentrates on the modeling of speech repairs, a phenomenon in spontaneous speech in which speakers notice their own mistakes while speaking and then correct these mistakes with relatively little disruption. Speech repairs are an important area of study for natural language processing researchers, because they show where the human language generation process can go wrong, and how language generation and recognition systems can adapt to these errors. Human listeners often do not even consciously notice these corrections, yet they are unrecognizable or confounding to most current speech recognition systems. Other systems for detecting and correcting repair assume a pipelined approach, where a word string of the entire utterance is already available. This assumption is problematic, because it precludes real-time use, and because it is quite difficult to get the correct word string in the first place without a model of repair.

My work in this area takes an incremental approach, detecting and correcting repair online as input is received. This work is based on a syntactic model of repair which explicitly models incomplete syntactic categories, integrating the syntax of fluent and disfluent speech very naturally. This work began with the observation that a 'right corner transform,' a way of refactoring a standard context-free grammar, could potentially represent the partially finished phrases in a speech repair automatically. A grammar transformed in this way was used to train a standard statistical parser and resulted in improved performance over a baseline approach using a standard grammar in work presented at ACL '08 (Miller and Schuler, 2008b). This new representation was then mapped to a Hidden Markov Model (HMM), a time-series model which tracks the latent variable structure of a stream of input – in this case, the set of right-corner transformed grammar states for a sequence of observed words. This system was able to process input incrementally, and showed further performance improvements over a baseline, while improving its plausibility as a psycholinguistic model, in work presented at COLING '08 (Miller and Schuler, 2008a).

This work was extended in a paper presented at ACL '09 by modifying the original syntactic representation to better fit this new parsing model, by removing any *symbols* for disfluency from the grammar and instead making it an *operation* of the HMM parser (Miller, 2009a). Another extension models the short term memory recall of recent words by keeping a buffer of recent words as an extra hidden state. This buffer may then be used to easily detect repeats of short phrases (one or two words) which are a significant indicator in detecting speech repairs (Miller, 2009b). This result, presented at EMNLP '09, took information that had previously only been used in non-incremental approaches (Johnson and Charniak, 2004) and invented a new design that allowed it to function within an incremental system, leading to a large gain in accuracy detecting and correcting repair.

Finally, work in preparation now examines the tradeoff between parsing accuracy and "greediness" (technically beam search width) of the parsing algorithm. This work will show that even with a very greedy (and thus very fast) approach, the incremental HMM parser is more accurate than a baseline unoptimized parser. This speed advantage enables the incremental parser to expand its search space by including richer linguistic structure, such as variables for estimating a sequence of semantic roles along with the words at each time step.

Current and Future Directions

Current research and near-future research (3–5 years) is focused on the following three areas.

Integration of syntactic models and acoustic information

While my thesis has concentrated on deriving a plausible incremental *syntactic* model of speech repair, there is a great deal of information about repair in the *prosody*, or intonational contours of the speaker's voice in the audio signal. In addition, there is reason to believe that prosody contains more general information about the structure of the utterance that should be combined with lexical information. Leveraging this prosodic information in a syntactic processing model can increase the accuracy of the syntactic analysis, increase detection of repair, and moves one step closer to the goal of fully integrating syntax and speech. The first step in this approach is incrementally classifying prosodic cues using the standard ToBi annotation scheme, and conditioning the random variables in the model of disfluency developed in my thesis and in the parser on the state of the prosody.

Unsupervised learning of syntactic structure

The 'right-corner transform' mentioned above results in a set of non-traditional constituent categories that is richer than the default set of categories. This richer structure requires more training data to learn reliably, and labeled training data is expensive to obtain. As a result, unsupervised approaches to learning the grammar are desirable, since they can be trained on any large unlabeled dataset. Recent work in unsupervised grammar induction (Klein and Manning, 2004) has shown that unsupervised approaches can be very accurate. Recent work in non-parametric Bayesian methods (Teh et al., 2006; Finkel et al., 2007) has described methods for unsupervised learning that do not require specifying the number of categories in advance. Current work in progress takes advantage of these advances to attempt to learn a simpler version of the HMM parser mentioned above.

Generalized model of incremental semantics

Recent work in computational lexical semantics (Mitchell and Lapata, 2009) attempts to represent word meanings in vectors of probabilities representing co-occurrence statistics. Current work by colleagues at the University of Minnesota (Stephen Wu, William Schuler) attempts to integrate these models of lexical semantics with syntax in compositional models that operate incrementally, building on previous work with referential semantic models (Miller et al., 2007). Future planned work will take advantage of these semantic models and improve them by applying my expertise about the semantics of disfluency to produce more realistic models of the semantics of spontaneous speech. For example, the disfluent utterance *John went ho-... uh, he went to the store and then he went home* contains an error (*John went ho-*) which the speaker then corrected. The first thing to note about this sentence is that simply removing errors from a disfluent utterance is problematic, because they may contain essential semantic information – in this case, the fact that the pronoun *he* refers to John. But further, this example combines with the syntactic model of repair in my thesis to suggest a combined model for repair that is similar in structure to a disjunctive coordination (as in the word *or*), with a semantics that is similar to logical-OR in that the first clause (the error) may or may not be true if the second clause (the correction) is true.

Finally, all of these aspects of computational research are relevant to psycholinguistic research as well. While the general approaches outlined above are all done with an eye towards maintaining psycholinguistic plausibility, there is a thread of recent research which is empirically evaluating computational models in terms of psycholinguistic measurements (Roark et al., 2009; Boston et al., 2008; Hale, 2001). I have had personal communications with some of these researchers along these lines and intend to pursue evaluations of the models I develop to quantify their fidelity to psycholinguistic constraints.

References

Marisa F. Boston, John Hale, Shravan Vasishth, Umesh Patil, and Reinhold Liegl. 2008. Parsing costs as predictors of reading difficulty: An evaluation using the potsdam sentence corpus. *Journal of Eye Movement Research*, 2:1–12.

Jenny R. Finkel, Trond Grenager, and Christopher D. Manning. 2007. The infinite tree. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 272–279, Prague, Czech Republic, June.

John Hale. 2001. A probabilistic earley parser as a psycholinguistic model. In *In Proceedings of NAACL*, pages 159–166.

Mark Johnson and Eugene Charniak. 2004. A tag-based noisy channel model of speech repairs. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL '04)*, pages 33–39, Barcelona, Spain.

Dan Klein and Christopher D. Manning. 2004. Corpus-based induction of syntactic structure: models of dependency and constituency. In *ACL '04: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 478, Morristown, NJ, USA. Association for Computational Linguistics.

Tim Miller and William Schuler. 2008a. A syntactic time-series model for parsing fluent and disfluent speech. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING'08)*.

Tim Miller and William Schuler. 2008b. A unified syntactic model for parsing fluent and disfluent speech. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL '08), short papers*, pages 105–108.

Tim Miller, William Schuler, and Andy Exley. 2007. Elements of a spoken language programming interface for robots. In *Proceedings of HRI 2007*, Washington, D.C.

Tim Miller. 2009a. Improved syntactic models for parsing speech with repairs. In *Proceedings of the North American Association for Computational Linguistics*, Boulder, CO.

Tim Miller. 2009b. Word buffering models for improved speech repair parsing. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, Singapore.

Jeff Mitchell and Mirella Lapata. 2009. Language models based on semantic composition. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 430–439, Singapore, August. Association for Computational Linguistics.

Brian Roark, Asaf Bachrach, Carlos Cardenas, and Christophe Pallier. 2009. Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 324–333, Singapore, August. Association for Computational Linguistics.

Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. 2006. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.