

UROP Proposal — Supervised Tree-to-Tree Alignment for Machine Translation

Jason Katz-Brown, Edwin Chen

January 11, 2007

We will work with Professor Michael Collins on developing a supervised tree-to-tree alignment model in the context of his tree-to-tree statistical machine translation system.¹ A crucial step in this translation model is extracting alignments between the corresponding constituents given a parse tree in the source language and a parse tree in the target language. Collins et al.'s system currently uses a complex and occasionally error-prone set of heuristics to predict the alignments. Improving the tree-to-tree alignments produced in this step could improve translation results.

We propose to introduce a supervised tree-to-tree alignment step in the training of the machine translation system. This has the potential to improve the alignments beyond heuristic guesses and take advantage of gold-standard human annotations during the tree-to-tree alignment step. First we will build a graphical tool that shows a parse tree in the source language and a parse tree in the target language and allows human annotators to specify which constituents align to each other. Second we will gather human annotations from a sample of the training data using this tool. Third we will seek out patterns in the human annotations that provide insight into the alignment process between different languages and design a model that will facilitate stastically applying this knowledge to the tree-to-tree

¹Brooke Cowan, Ivona Kucerova, and Michael Collins. *A Discriminative Model for Tree-to-Tree Translation*. In proceedings of EMNLP 2006.

alignment step during training. Finally we will implement this alignment model and test it on different language pairs to see if improvements are made over the baseline heuristic system.

We will initially work with German-English data because parsers for both language are accurate and mature and Collins et al.'s German-English translation system offers a strong baseline for comparing tree-to-tree alignments and translation results. This work holds promise for a similar Chinese-English translation system using the same method. The proposed improvements to the tree-to-tree alignment step could be a significant boost to both translation systems.