Abstract: The goal of this thesis is to implement and evaluate a software tool for automatic alignment of Czech and English tectogrammatical trees. The task is to find correspondent nodes between two trees that represent an English sentence and its Czech translation. Great amount of aligned trees acquired from parallel corpora can be used for training transfer models for machine translation systems. It is also useful for linguists in studying translation equivalents in two languages. In this thesis there is also described word alignment annotation process. The manual word alignment was necessary for evaluation of the aligner. The results of our experiments show that shifting the alignment task from the word layer to the tectogrammatical layer both (a) increases the interannotator agreement on the task and (b) allows to construct a feature-based algorithm which uses sentence structure and which outperforms the GIZA++ aligner in terms of f-measure on aligned tectogrammatical node pairs. This is probably caused by the fact that tectogrammatical representations of Czech and English sentences are much closer compared to the distance of their surface shapes.

Keywords: tectogrammatical trees, word alignment, machine translation