Large collections of translated texts – called *parallel corpora* – are often automatically aligned on word and sentence level to be used as training data for machine translation systems. We may also choose to syntactically analyze the sentences to produce syntax trees. If we do this on both sides and the nodes of the trees are also aligned, the end result is called a *parallel treebank*.

The best translation systems are statistically based, but in recent years there has been a shift to the incorporation of more linguistically motivated data, which includes the use of parallel treebanks. These are only useful on a very large scale because of the amount of information a system needs about how one language is to be translated into another in order to be effective. Because of this, we investigate techniques for the automatic and accurate alignment of these nodes. Another motive for our research is the fact that parallel treebanks are also useful for other techniques and that as a linguistic resource, remain scientifically interesting. This process is called *tree alignment*.

We find that a combination of statistical and rule-based techniques, using relatively small sets of training data and few features, is sufficient to produce very accurate alignments. Finally, we also find that when we apply alignments covering a relatively large set of nodes – even though some of them are wrong – on a syntax-based machine translation system, this leads to better translation results than applying alignments that are more accurate but fewer in number.