

# Using a Parallel Treebank for Detecting Source Language Influence on Target Language in Student Translations

---

Kerstin Lindmark<sup>1</sup>

## Abstract

The present pilot study deals with the detection of doubtful translations using parallel treebanks. The main purpose of my larger project is to examine patterns of source-language (L2) influence on target language (L1) involving deviations from target-language norms, particularly in translations made by neophyte translators. To find such patterns, a bilingual, multi-parallel corpus of translations made by translation students is used (each source sentence is aligned with multiple versions of the target sentence).

However, for the corpus to be really useful in this context, it would need to be aligned on a more fine-grained level than the sentence level. Since translations are seldom made word-for-word, but rather chunk-for-chunk, that more fine-grained alignment level would not necessarily have to be that of the word; preferably, perhaps, a level reflecting units which are translated in one go.

A treebank is described as a syntactically annotated corpus. A parallel treebank consists of two (or more) monolingual treebanks linked together on word or node level, i.e. an intermediate level that could be useful in this context.

In this poster, I will investigate the use and usefulness of treebanks for solving problems in the construction of the multi-parallel corpus.

---

<sup>1</sup> *e-mail*: kerstin.lindmark@ling.su.se