# PaCo-MT: Parse and Corpus-based Machine Translation

## STEVIN: Spraak- en Taaltechnologische Essentiële Voorzieningen in het Nederlands STE-07007

http://www.ccl.kuleuven.be/Projects/PACO/

| List of partners |
|---|
| Centre for Computational Linguistics – KULeuven, Belgium |
| Alfa-Informatica – University of Groningen, Netherlands |
| OneLiner bvba - Belgium |

## Project duration: February 2008 — July 2011

## Summary

The PaCo-MT project is building a stochastic example-based transfer system translating from Dutch into English and French, and vice versa. It is a data-driven tree-to-tree based approach towards MT, transducing the input parse tree into a set of target language parse trees without node ordering. This Synchronous Tree Substitution Grammar (limited to regular subtrees) is induced from a subtree-aligned parallel treebank, using a discriminative model for tree alignment. Monolingual parses were created by pre-existing parsers, such as the Alpino parser for Dutch, the Stanford parser for English, and the Berkeley parser for French. A tree-based target language modeler using a probabilistic context-free grammar based on large monolingual treebanks decodes the output forest and determines node ordering.

By this approach we aim at combining the strengths of data-driven MT with the strengths of rule-based MT, avoiding the weaknesses of each of these approaches. Results show that although BLEU scores are not yet at par with Moses, long distance movements pose no problems for our approach, and we do not drop important words, yielding a more grammatical output than PBSMT systems.

Mikel L. Forcada, Heidi Depraetere, Vincent Vandeghinste (eds.)
*Proceedings of the 15th Conference of the European Association for Machine Translation*, p. 347
Leuven, Belgium, May 2011