

PaCo-MT

Parse & Corpus-based Machine Translation

(STEVIN project 2008-2011)

- Centre for Computational Linguistics - KULeuven
- Alfa-Informatica - Rijksuniversiteit Groningen
- OneLiner BVBA

Purpose

- Building a machine translation system for language pairs
 - DUTCH - ENGLISH (+ ENGLISH - DUTCH)
 - DUTCH - FRENCH (+ FRENCH - DUTCH)
- Making use of
 - full parses
 - large corpora and treebanks
 - translation memories and parallel corpora

Workflow

- Source language analysis: Full Parse
(Alpino / PET+ERG / Alpage)
- From Source to Target:
 - dictionaries / translation memories / subsententially aligned parallel treebanks
 - transfer rules automatically derived from aligned parallel data
- Target language generation: From dependency tree to surface form

Translation system

- Post-editing interface for translators
- Edited translations are fed back into the data resulting in
 - adapted weights
 - updated translation memory
- Separate translation memories for different users depending on user profile