Hybrid Machine Translation

SMatxinT, a HMT Architecture Guided by Syntax

- Candidate chunk translations are calculated with an SMT system and used to enrich the RBMT tree-based source representation with more alternatives.
- The most probable combination among the available fragments is obtained with a monotone statistical decoding (m) following the order provided by the RBMT system.
- The evaluation on news (out-of-domain) shows a preference for SMatxinT. There is no necessity for re-ranking the RBMT choice (r).

Comparative Corpora from Wikipedia

- Define an in-domain vocabulary (MTFs in MFTs) used to evaluate the degree of similarity.
- Stopping criterion: levels with a minimum percentage of interlanguage links.
- Graph-based IR-based (ongoing work)

Wikipedia Enrichment

Prototypes for including relevant information into a Wikipedia article in language L₂ from the same article in L₁:
1. Paragraph alignment (similarity measures)
2. Relevant paragraph identification (from unaligned paragraphs)
3. Determination of the insertion position into the target
4. Preliminary translation of the relevant paragraphs by a translation engine