



Information Society
Technologies

METIS-II:

The German-to-English MT System (FP6-IST-003768)

Profile

METIS II, the continuation of the successful assessment project METIS I, is an IST Programme, with a 3-year duration (01/10/2004 – 30/09/2007).

The METIS II consortium comprises the following partners:

- Institute for Language & Speech Processing [ILSP] (co-ordinator)
- Katholieke Universiteit Leuven [KUL]
- Gesellschaft zur Förderung der Angewandten Informationsforschung [GFAI]
- Universitat Pompeu Fabra [UPF]

METIS-II translates as yet four languages:
Greek, Dutch, German & Spanish English.

The METIS Approach

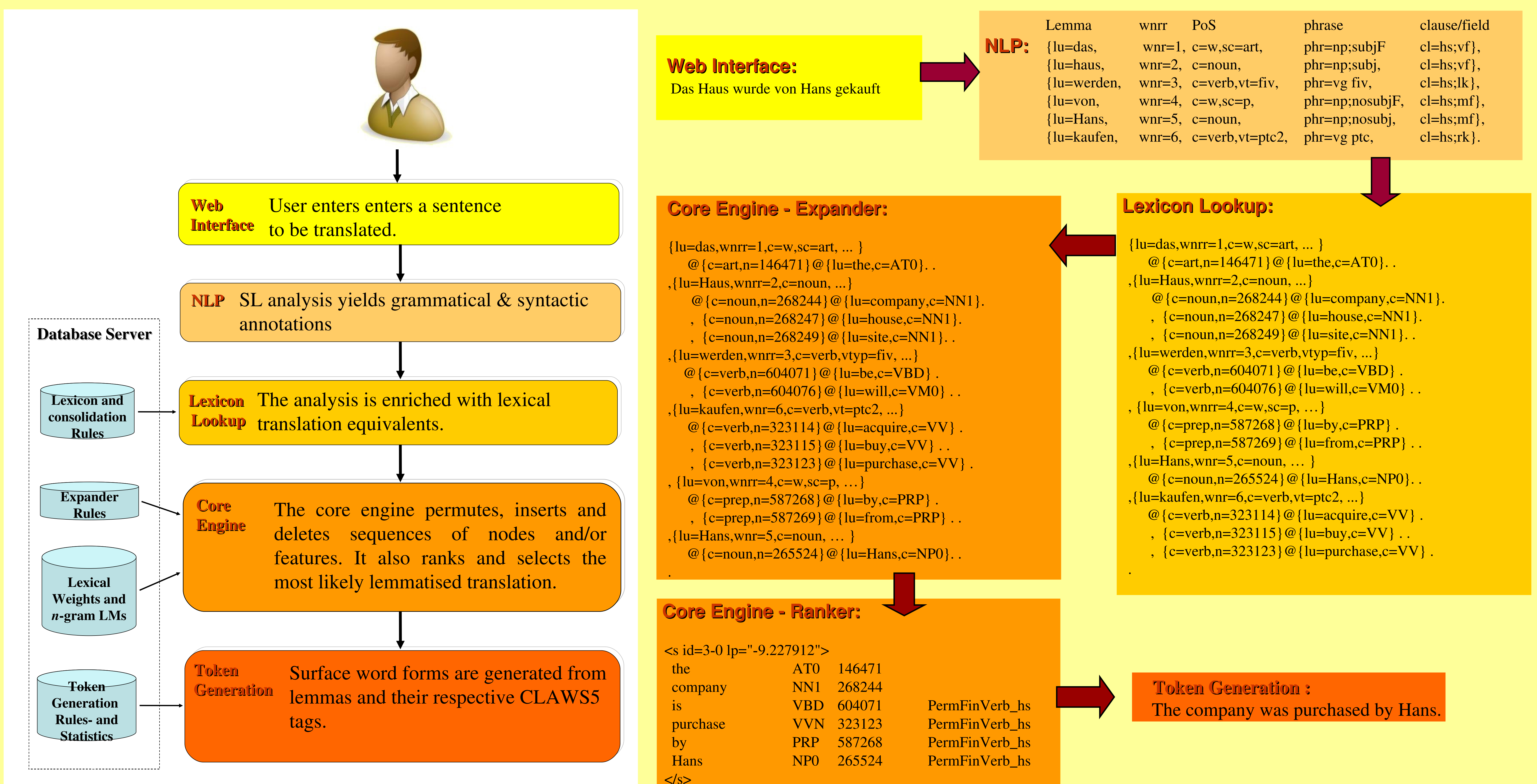
METIS II is a **hybrid** system, combining various approaches to machine translation (rule-based, statistical, pattern-matching techniques). It makes use of **readily available resources**, such as bilingual dictionaries or basic NLP tools, and it can be easily customised to handle different source (SL) and target language (TL) tags.

Most importantly, however, METIS II is **innovative** because it does not need bilingual corpora for the translation process, but **exclusively** relies on **monolingual TL corpora**.

METIS II handles sequences both at sentence and sub-sentential level, achieving thus to exploit the **recursive** property of natural language.

METIS II employs a series of **weights**, i.e. system parameters, in various phases of the translation process. Weights are associated with system resources and employed by the pattern-matching algorithm; they can be **automatically** adjusted to customise system performance.

METIS II Architecture



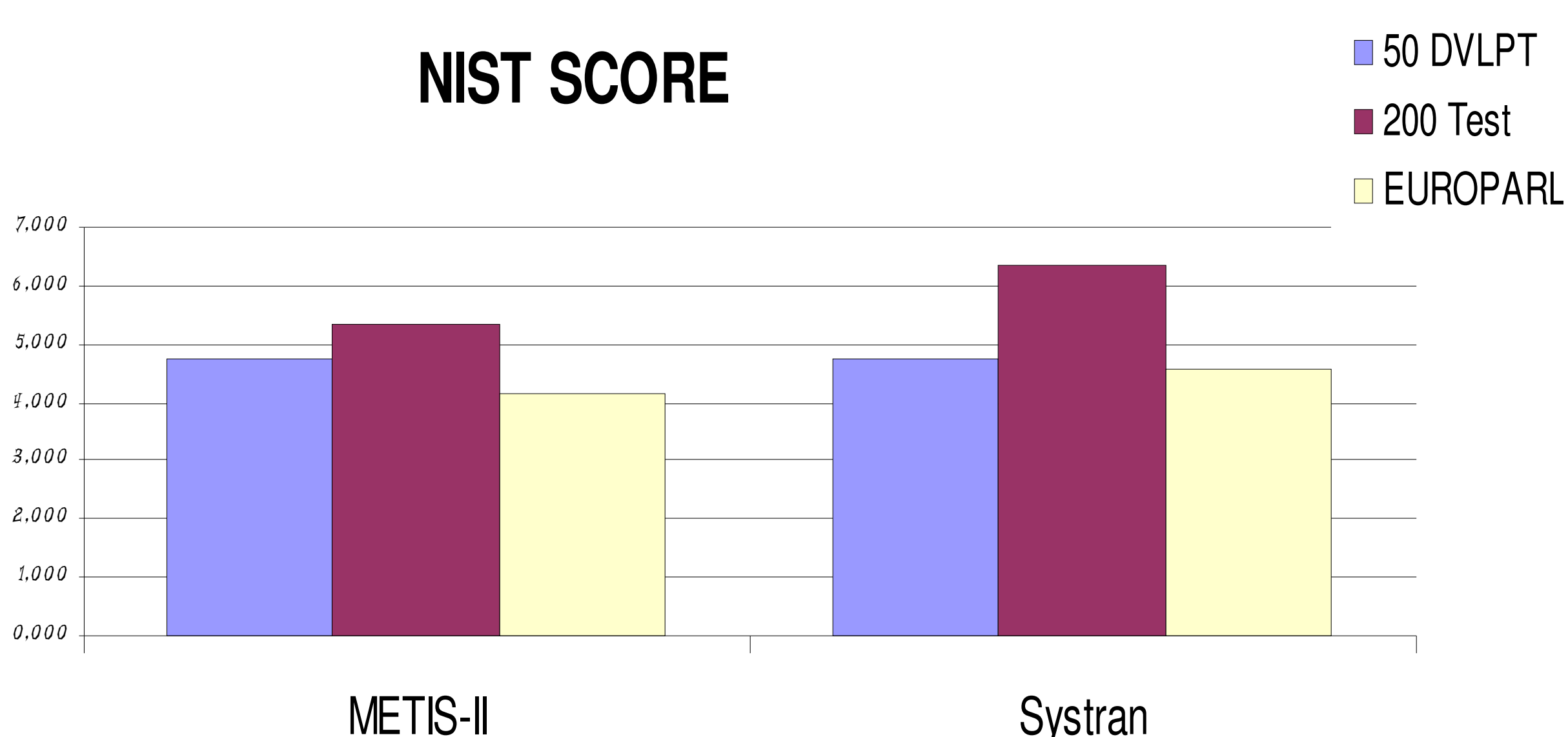
Evaluation Setup

Three experimental corpora were used for system evaluation:

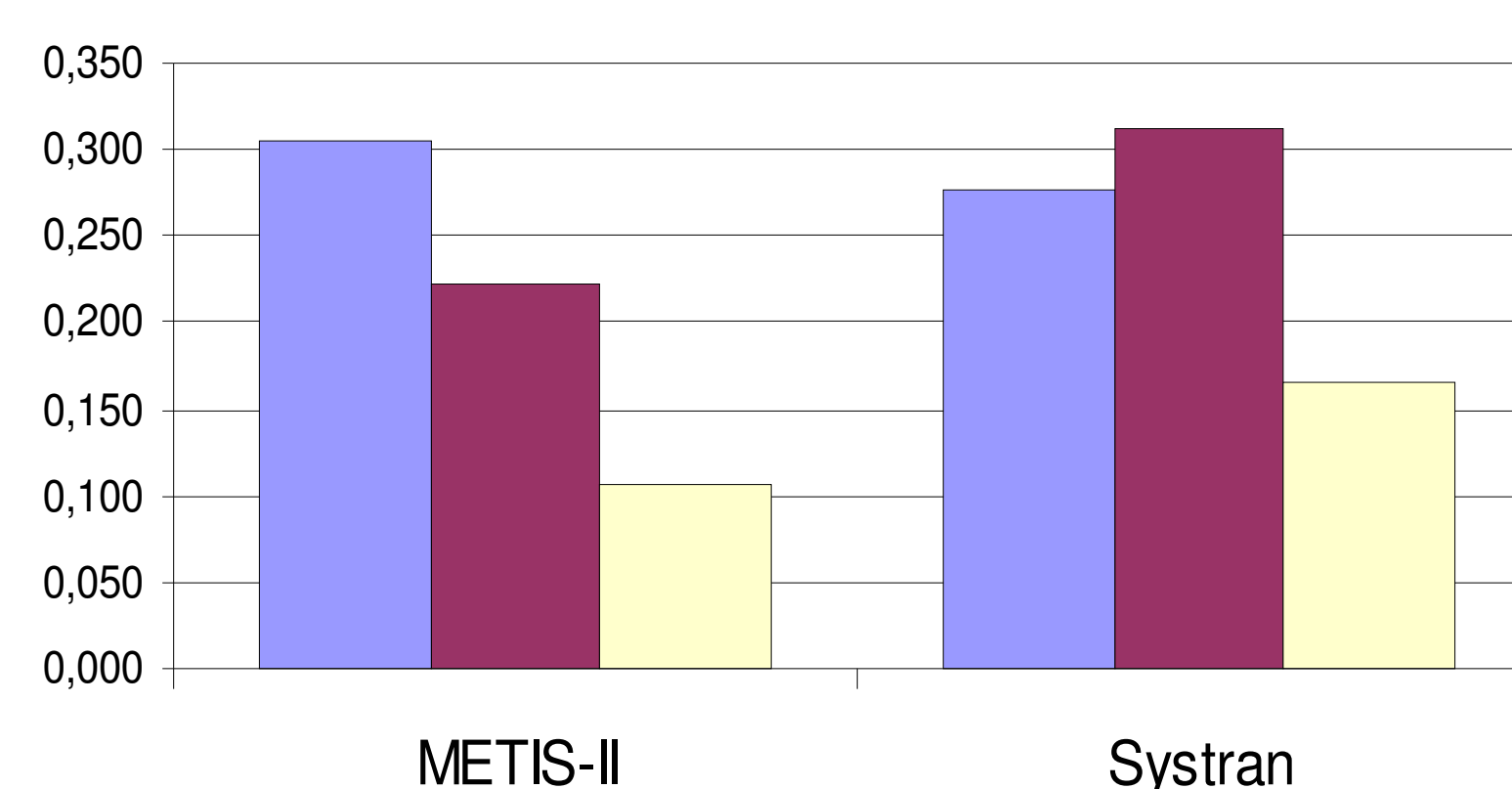
- Development set of 50 sentences (DVLPT)
- Test text of 200 sentences chosen to cover various translation divergencies
- Excerpt of 200 sentences from the EUROPARL corpus

The development set and the 200 test sentences were of relative complexity, containing one to two clauses each, covering various syntactic phenomena such as word-order variation, NP structure, negation, modification etc. The EUROPARL sentences were 32 words long at most. The graphs below show comparative results of BLEU and NIST scores for three corpora for METIS-II and Systran:

NIST SCORE



BLEU Score



Future Work

Future work involves further investigation of METIS II system architecture. More specifically, work towards the system optimization includes the following:

- Further system testing with a big number of test suites that will have more elaborate structures and deal with a wider range of phenomena
- Algorithm optimization in terms of accuracy
- Automatic fine tuning of weights
- Investigation of further feature functions
- Implementation of a post-editor module