# Problems in Polish-to-German rule-based machine translation

Marcin Junczys-Dowmunt (Adam Mickiewicz University) and Filip Graliński (Poleng Sp. z o.o.)

The aim of this talk is to discuss some problems of Polish-to-German rule-based machine translation and to show their practical solutions. Here we present the research carried out in the process of developing a Polish-to-German module for the POLENG (commercial name: Translatica) MT system. POLENG is a rule-based system, and up to the present day modules for Polish-to-English, English-to-Polish, Polish-to-Russian and Russian-to-Polish translation have been created. The system is based on manually crafted grammar and transfer rules as well as on bilingual lexicons derived from human-readable dictionaries.

The translation process in POLENG can be divided into four stages: parsing the source text into a syntax tree, rewriting the tree into a target language tree, modifying the tree (e.g. for propagation of morphological features) and generating the target text from such a tree. As far as the first stage is concerned, we simply reuse the parser available from the Polish-to-English and Polish-to-Russian modules. The transfer rules (for the other stages) however need to be designed from scratch.

In this talk we present the POLENG system's formalism of transfer rules and discuss the development of transfer rules for the Polish-to-German translation process. As German is similar to English in some respects (e.g. definite and indefinite articles are used) and to Polish in some others (e.g. the inflectional system is much richer than that of English, German features a semi-free word order), we could adapt the basic design principles for the transfer rules partly from the Polish-to-English component and partly from the English-to-Polish component. There are, however, many aspects in which German is substantially different from Polish and English.

We will illustrate how a basic theory of topological fields – a widely accepted theory of word and phrase order of the German language – can be described with the presented formalism. Although the reordering of the words of a target language sentence is one of the final procedures during the translation and transfer process, its underlying principles have a strong impact on the way other phenomena are implemented throughout the entire process. Since, for instance, German finite verbs and their verbal complements are – contrary to Polish – located in different, often not adjacent fields, complex verb forms need to be broken up – a phenomenon that has so far not been met during the work on previous language pairs in the POLENG system. Examples how complex verbal phrases are treated by the syntactical transfer and how the topological sentence model allows to maintain thema-rhema relations of the original Polish sentence in the German translation will be given during the talk.

Another interesting problem we are going to address is the choice of the correct articles in the German target sentence, where morphological, syntactical, semantical, and pragmatical aspects are of importance. Whereas the pragmatical information of a sentence lies beyond the scope of our formalism, the three other aspects can be employed in a system of heuristics for article insertion implemented on all three levels of transfer.

**References**

Jeziorski, Jan. *Substantivische Nominalkomposita des Deutschen und ihre polnischen Entsprechungen.* Wydawnictwo Polskiej Akademii Nauk , Poland 1980.

Och, Franz J. *Maximum-Likelihood-Schätzung von Wortkategorien mit Verfahren der kombinatorischen Optimierung*. Studienarbeit, Universität Erlangen-Nürnberg, Germany 1995.

Och, Franz J. *Giza++: Training of statistical models*. Available at http://www.fjoch.com/GIZA++.html, 2000.

Och, Franz J., Ney Hermann. *A Systematic Comparison of Various Statistical Alignment Models*. Computational Linguistics, volume 29, number 1, pp. 19-51, March 2003.