Rule-based machine-translation between Finnish and German

Tommi Pirinen
Universität Hamburg

tommi.antero.pirinen@ uni-hamburg.de

With this poster I present a work-in-progress rule-based machine translation between German and Finnish based on the Apertium[1] machine translation system (Forcada et al. 2011). The system is composed of canonical NLP components for rule-based systems: morphological analysis, chunking, lexical selection, chunk re-ordering and structural translation and morphological generation. One of the points that I want to highlight in this poster is the workflow and the supporting infrastructure for it; unlike a typical coursework or research project, this machine translation has been modeled as a *lexicon and grammar engineering while language learning* type of project. On software engineering side I have developed tools to extend mono- and bilingual lexicons while learning the OOV words (notably, words that are OOV for RBMT are new words for the language learner) in texts, and I am in process to extend these tools for the grammar learning—strucutral transfer interaction. I have also modernised the build infrastructure from sf.net SVN to github with full support of continuous integration and automatic testing, providing an excellent platform for language learners to extend the lexicons and grammars without fear of breaking other existing systems that depend on these lexicons and grammars.

Apertium systems are modular pipelines combining basic NLP tools with machine translation specific modules. The three lexical modules that are the most important for the system development and language learning are the morpholgical analysers and the lexical translation, which correspond the vocabulary of the learner / MT system. The resources for Finnish and German morphological analysis were available at the start of the project, but to our knowledge, this is the first free and open source Finnish-German bilingual resource of its kind. For the other parts of the pipeline that are more specific to apertium, such as chunking (shallow syntax parsing), re-ordering and transfer rules.

References: • Forcada, M.L. et al. (2011): Apertium: a free/open-source platform for rule-based machine translation. *Machine translation* 25(2). 127–144.

---

[1] http://wiki.apertium.org/