Apertium: A Free/Open-Source Platform for Machine Translation and Basic Language Technology

Mikel L. FORCADA[1] and Francis M. TYERS[2]

[1] Departament de Llenguatges i Sistemes Informàtics, Universitat d'Alacant, Alacant
[2] HSL-fakultehta, UiT Norgga árktalas universitehta, Romsa

`mlf@dlsi.ua.es, francis.tyers@uit.no`

Abstract. Apertium is a free/open-source platform for rule-based machine translation which was started in 2005. The Apertium community is using it to build machine translation systems for a variety of language pairs, especially, but not only, for related-language pairs, where shallow transfer suffices to produce good quality translations. Apertium is also being used to develop monolingual language processors for these languages. We present the Apertium platform: the translation engine, the encoding of linguistic data, and the tools developed around the platform.

Description

Apertium (wiki.apertium.org, Forcada et al. 2011) is a free/open-source platform for machine translation (MT), providing a shallow-transfer rule-based MT engine, data (dictionaries, rules) for more than 40 language pairs, and a wide variety of software to manage, convert and exploit the data and run the engine in a variety of environments. Apertium is also the free/open-source community developing it since 2005.

Apertium is built as a pipeline in which the source text is gradually converted into target text. The pipeline contains, among others, filters to handle markup, finite-state morphological analysers and generators using both a native format and the Helsinki Finite-State Toolkit (HFST) one, part-of-speech taggers using Constraint Grammar, hidden Markov Models, and sliding-window classifiers, and shallow-transfer rules capable of identifying, transforming and reordering syntactic chunks. These components may also be used for other human-language technologies besides MT.

Apertium is also available in several end-user products: apps for Android smartphones and tablets, a plugin for the OmegaT computer-aided translation environment, and a number of stand-alone applications for desktops. It also powers the multilingual open-content management environment, Wikimedia Content Translation, used to translate Wikipedia articles.

Finally, Apertium is both a research and a business platform, enabling reproducibility and transferability of results and successful marketing of MT products and services.

References

Forcada, M. L., Ginestí-Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J. A., Sánchez-Martínez, F., Ramírez-Sánchez, G. and Tyers, F. M. (2011) Apertium: a free/open-source platform for rule-based machine translation. Machine Translation 24(1) 1–18