

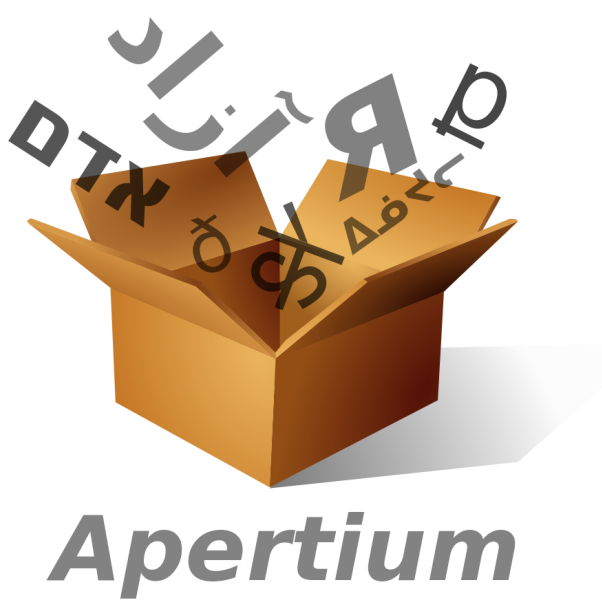
# Apertium: A free/open-source platform for machine translation and basic language technology

[www.apertium.org](http://www.apertium.org)

IRC: #apertium at [irc.freenode.net](http://irc.freenode.net)

Wiki: [wiki.apertium.org](http://wiki.apertium.org)

Contact: Mikel L. Forcada <[mlf@ua.es](mailto:mlf@ua.es)>, Francis M. Tyers <[ftyers@prompsit.com](mailto:ftyers@prompsit.com)>



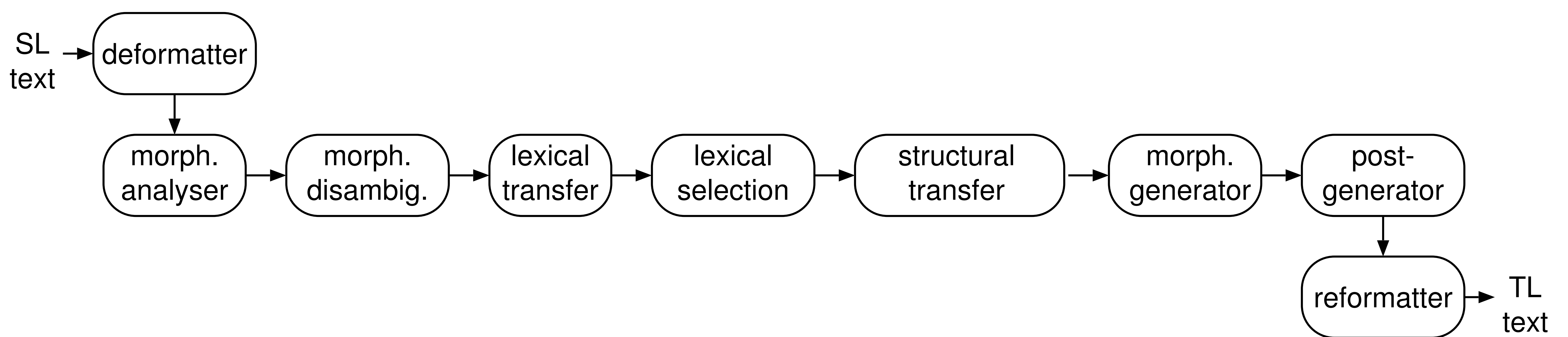
## Apertium components

Since 2005, Apertium provides:

1. A free/open-source, modular, shallow-transfer, language-independent machine translation **engine** with: text format management, finite-state lexical processing, statistical and constraint-based lexical disambiguation, and shallow structural transfer based on finite-state pattern matching
2. Free/open-source **linguistic data** in well-specified XML formats for a variety of languages and language pairs
3. Free/open-source tools: **compilers** to turn linguistic data into a fast and compact form used by the engine and software to learn disambiguation or translation rules from corpora.

Most of the engine developed inside Apertium but some external technologies used: Helsinki Finite-State toolkit (for some morphologically-rich languages), VISL CG-3 (constraint grammars for rule-based lexical disambiguation).

## Machine translation but not only!



- Apertium is a *rule-based machine translation system* but the pipeline contains many **monolingual** modules that can be used for other human-language technology tasks.
- Most modules are based on finite-state technology, with HMMs used for part-of-speech tagging and interpreted in the structural transfer.in the structural transfer.

## Licensing: free/open-source

Apertium language data and code are both licensed under the GNU General Public License:

- a free/open-source license allowing free distribution of unmodified and modified versions
- a *copylefted* license: it avoids private appropriation and encourages giving improvements back to the project (a *commons*) → **community**

## The Apertium community

- Very active group of hundreds of developers
- Wiki documentation ([wiki.apertium.org](http://wiki.apertium.org))
- Easy entry: Apertium linguistic modelling is simple, no need to program.
- IRC channel #apertium in [freenode.net](http://freenode.net)
- Mailing lists: [apertium-stuff@lists.sf.net](mailto:apertium-stuff@lists.sf.net) and other lists

## Research and business with Apertium

Apertium is already an active research and business platform:

- **Research:** 40+ publications, 2 PhD thesis, 4 master's theses
- **Business:** companies (Prompsit, Eleka, Imaxin Software, etc.) offering services to customers such as Autodesk, the Government of Catalonia, one of the main Basque banks, the daily newspaper *La Voz de Galicia*, etc.)

The free/open-source model creates a **community** which effectively connects **researchers**, **developers**, **vendors** and **users**.

## Ready-to-use Apertium products

- a stand-alone Java application for the desktop: `apertium-caffeine`
- an Android version for handhelds
- a stand-alone version (*Apertium Simpleton*) for Windows and MacOS.
- a plug-in for the OmegaT CAT platform `apertium-omegat`
- Available as an PPA repository for GNU/Linux users.



