AltLang: an Automatic Converter between Varieties of English, Spanish, French and Portuguese

Gema RAMÍREZ-SÁNCHEZ

Prompsit Language Engineering, Av. de la Unversitat s/n, 03202, Elx, Alacant, Spain

gramirez@prompsit.com

Abstract. AltLang is a rule-based automatic converter for language varieties. It deals with differences in spelling, lexicon and local grammar along with numeric, style and punctuation conventions. It is available for varieties of English, Spanish, French and Portuguese. AltLang is based on the GNU GPL-based free/open-source technologies of the Apertium platform, and is offered as a service by Prompsit. It can be tested through an online application at www.altlang.net.

Description

AltLang is a rule-based automatic converter for language varieties. It aims at minimising the effort required to generate content in different varieties of the same language by automatically replacing controlled existing differences between them. AltLang can convert between American and British English, Canadian and European French, Latin American and European Spanish and Brazilian and European Portuguese with high accuracy. The technology behind AltLang comes from the free/open-source Apertium platform (Forcada et al. 2011). Data for dictionaries and rules have been semi-automatically adapted from two main sources: Apertium data and freely available or client-owned bilingual corpora. AltLang's knowledge grows through continuous contributions from clients and testers. AltLang gives priority to user-defined vocabulary and translation memories: vocabulary is input as a simple two-column spreadsheet and translation memories are supported through standard TMX files. AltLang reports basic statistics about the conversions (number of words and characters in source and target and number of substitution, shifts, insertions and deletions). It also shows the individual changes performed to a document with a convenient colouring. All these features allow the user to have full control over the behaviour of AltLang. It deals with a wide range of file formats including some interchange and localization ones (.xliff, .po). It can be used through an API, AltLang's web application and is available in other third-party tools such as the MateCat translation platform. Prompsit offers customisation services for AltLang to add new languages, new language varieties or domain-adaptation.

References

Forcada, M. L.; Ginestí-Rosell, M.; Nordfalk, J.; O'Regan, J.; Ortiz-Rojas, S.; Perez-Ortiz, J. A.; Sanchez-Martínez, F.; Ramírez.-Sánchez, G.; and Tyers, F. M. (2011). Apertium: a free/open-source platform for rule-based machine translation; Machine Translation, 25(2):127–144.