

Interlinguas: Deep and Shallow

Aarne Ranta

University of Gothenburg
aarne@chalmers.se

Abstract

In 1629, Descartes proposed a “language of true philosophy” to serve as an interlingua for translating between languages. Over 300 years later, “semantic interlingua” appears on the top of the Vauquois triangle, as the deepest possible analysis guaranteeing the best possible translation. But the main stream of machine translation has considered the interlingua unrealistic and worked on lower levels of the Vauquois triangle, such as syntactic and lexical transfer.

However, the interlingua idea has advantages that do not depend on it being a deep semantic representation. An interlingua makes it possible to build highly multilingual systems without a quadratic blow-up of size. It also enables transfer of information, for instance, from high to low resourced languages; a related idea has recently been exploited in the Universal Dependencies project, which uses a shared set of labels and tags as a cross-lingual representation.

Grammatical Framework (GF) is a formalism that was designed for building interlingua-based multilingual grammars. Its original purpose was to enable special-purpose interlinguas precisely capturing the semantics of different domains, such as mathematics or touristic phrases. However, GF also enables interlinguas that are not so deep. They can be based on surface syntactic structures or just chunks of words. Recent developments of this idea have led to a translation system that currently works for all 182 pairs of 14 languages, ranging from English and German to Finnish and Chinese. This system has a stack of interlinguas, where a semantic layer produces high-quality translations whenever the input can be analysed by it, whereas the syntactic and chunk-based layers guarantee the robustness of the system. The interlingual grammar makes the system very compact in size, so that it can be run off-line on mobile devices.

Bio

Aarne Ranta is Professor of Computer Science at the University of Gothenburg and co-founder and CEO of the start-up company Digital Grammars AB. He defended his PhD in 1990 on the application of constructive type theory to natural language semantics, supervised by Per Martin-Löf. The theory developed in the thesis led to the idea of multilingual grammars, implemented as the system GF (Grammatical Framework) when Ranta worked at Xerox Research Centre Europe in 1997–1999. After Xerox, Ranta has led GF as an open-source project, which to date has had over 150 contributors working on over 30 languages. He has supervised ten PhDs and written three books, of which “Grammatical Framework: Programming with Multilingual Grammars” (CSLI 2011) has also appeared in Chinese. Ranta’s vision is to get linguistic knowledge formalized in a precise and efficient way and make it usable in practical applications.