Domain-Specific Multilingual Translation for Producers of Information

Aarne RANTA, Krasimir ANGELOV, Markus FORSBERG,
Thomas HALLGREN

Digital Grammars AB, Framnäsgatan 23, 41264, Gothenburg, Sweden

`{aarne,krasimir,markus,thomas}@digitalgrammars.com`

Abstract. Producers of multilingual information (webpages, manuals, etc) often need not solve the general problem of "translating anything", but can stay content with a more limited system, which however must be accurate. One way to do this is to define a *semantic interlingua*, which exactly fits the content to be translated. It has turned out that *algebraic datatypes*, and the way they are used in compilers to encode *abstract syntax*, are useful for this purpose. The grammar formalism Grammatical Framework (GF) was originally designed to support this kind of systems. While much of the current research in GF is focused on scaling it up to general-purpose translation, the creation of domain-specific systems has become a routine task, which is commercially exploited by the start-up company Digital Grammars AB. In the conference, we plan to show a demo of the tools for building such systems, as well as some actual systems built for customers.

Description

If we divide translation into the phases of *analysis* (of the source) and *generation* (of the target), we can identify some of the hardest problems as having to do with the analysis phase: in particular, *ambiguity* and *unexpected input*. The most radical solution to these problems is to *eliminate* them whenever possible. This is what *natural language generation* (NLG) is about: the source of documents in all target languages is *structured data*, such as databases or logical formulas. However, in real applications, such data is not always available in a clean form, but must be extracted from unstructured sources, typically from text. *Translation*, in this perspective, is NLG fed by parsing a source language. In GF, generation and parsing are defined simultaneously, since GF grammars are reversible.[1]

The mission of Digital Grammars AB[2] is to build high-quality documentation systems (either pure NLG or with translation) tailored for customers' needs. The technique enables both automatic and interactive translation, and is available for over 30 languages. Depending on the domain and on the work invested in building the system, we can reach publication quality (which is needed in real-time documentation) or come close to it (in which case we provide tools that support interaction; the system indicates confidence levels and displays translation alternatives with their semantic analyses). The product can be delivered in the form of batch translation jobs, as a web service, and as mobile (speech-enabled) applications.

---

[1] GF: http://www.grammaticalframework.org/

[2] Digital Grammars AB: http://www.digitalgrammars.com/