Modularity in the METAL System

Jocelyn S. Liu

Linguistics Research Center
The University of Texas at Austin
Austin, Texas 78713-7247

METAL is a fully-automatic machine translation system which is under research and development at four sites in four countries. The METAL project, sponsored by Siemens AG of Munich, seeks to produce a commercially viable system capable of translating different language pairs. This paper presents a short discussion on the modularity of the METAL system, that makes the METAL system expandable to new source languages and target languages. The overall structure of the translation process is discussed.

Introduction

The METAL system has been under research and development at the Linguistics Research Center at the University of Texas since 1959. The project in its present configuration dates from 1978, when the USAF Rome Air Development Center provided underwriting for the development of an operational MT system. Until 1979 research was conducted on a theoretical basis. In 1979 when Siemens AG of Munich became a sponsor of the project, the commercial aspects were added and METAL became to be viewed as a possible commercial MT system for translation of technical texts.

MT projects are very often found to be tied to one specific language pair, and, in many cases, to unidirectional translation, with the result that the methodologies adopted in the research are aimed solely at translating from one specific language to another. As such, this type of MT system may be effective for translation of that specific language pair. However, expansion of such systems to include a new source or target language appears to be difficult, if not impossible.

In order to produce a multilingual MT system, research in the METAL project has been conducted in a highly modular way.

Interchangeable translation modules

The translation process in METAL consists of two modules - analysis of the source language and synthesis of the target language. The analysis module constructs sentence trees and handles anaphora resolution; the synthesis module transfers from the source language to the target language and generates the final output.

In the analysis module the source language is analyzed independent of the target language. Thus the analysis of a language can be used in translation from this language to other languages. For example, the analysis module for a German-to-English system can be used in systems that translate from German to, say, French.

The synthesis module generates text in the target language. The synthesis module for a language is designed to generate that language regardless of what the source language is. Hence, the synthesis module for a German-to-English system can be used in systems that translate into English.

The goal of having independent analysis and synthesis modules is for every analysis module to be compatible with every synthesis module and vice versa. For each language involved in the multilingual translation system, only one analysis module and one synthesis module are needed. The model is designed to have interchangeable modules, thus the analysis module from the German-to-
English system can be used with the synthesis module of the French-to-Dutch system to produce a German-to-Dutch system.

Having interchangeable modules does not imply that the analysis and synthesis of a language pair have to be done in exactly the same way as those of another pair. What it requires is that the output of the analysis modules and the input of the synthesis modules be compatible.

Lexicons

To maintain independent analysis and synthesis modules, the lexicon data base has to be constructed in a similar modular manner. The METAL system consists of two types of lexicons - monolingual lexicons and bilingual lexicons. Monolingual lexicons contain information about a language, coded as feature-value pairs which are needed for both analyzing the language as well as synthesizing the language. The monolingual lexicons in the METAL system are constructed to be reversible, i.e., the same monolingual lexicon of a language is used both in analysis (when translating from that language) and in synthesis (when translating to that language). Each monolingual lexicon is independent of the direction of the translation, though features used only in either analysis or synthesis, but not both, are also allowed. For each language in the system, only one monolingual lexicon is built. As new pairs of languages are coupled, the monolingual lexicons of both languages are updated.

The bilingual lexicons (also referred to as transfer lexicons) map the correspondence between two languages. Each bilingual lexical entry consists of two columns, the canonical form (i.e., the dictionary form) of the source language in one column and the corresponding canonical form of the target language in the other. For each direction of translation between two languages, one bilingual lexicon is constructed. Theoretically, the same bilingual lexicon can be used in both directions of translation in a given language pair. However, since the correspondence between two languages is not always one-to-one, a bi-directional bilingual lexicon will invariably be very complex and difficult to maintain. Since whether the bilingual lexicons are unidirectional or bi-directional does not affect the overall translation process in the METAL system, we chose to have separate bilingual lexicons for each direction of translation.

Conclusion

Based on the principle of modularity, research on METAL is conducted in Munich, Germany; Leuven, Belgium; Barcelona, Spain; and Austin, U.S.A.; on five languages, English, German, Spanish, French, and Dutch. The German-to-English system is the most mature part of the METAL system. The analysis module of the German-to-English system is currently being used in the development of a German-to-Spanish system at the METAL site in Barcelona. An English-to-German system is under development in Austin. The METAL site in Leuven, Belgium is developing the French-to-Dutch and Dutch-to-French systems. A German-to-French system will soon be underway.

Research at the four METAL sites is being conducted in an integrated way, acknowledging an ultimate goal that METAL will be capable of translation between any pair of these five languages. At present, METAL research is focused on these languages; however, the METAL system is not limited to them. The modularity in the translation process makes possible the inclusion of other languages to the system at a later stage.