

Roberto Aranovich  
University of Pittsburgh

## **HANDLING OF MORPHOSYNTACTIC DIVERGENCES IN THE MAPUDUNGUN/SPANISH AVENUE TRANSFER GRAMMAR AND LEXICON**

### **Abstract**

AVENUE (Language Technologies Institute, Carnegie Mellon University) is a research project dedicated to the development of Machine Translation (MT) systems for languages lacking Language Technologies resources, such as lexical databases in electronic format, morphological analyzers, or previous developments in Machine Translation. The AVENUE Project has developed Language Technologies for Mapudungun and Quechua.

The aim of this paper is to discuss the handling of morphosyntactic divergences in the implementation of the Mapudungun/Spanish Transfer Grammar developed as a component of the AVENUE Mapudungun/Spanish Machine Translation System.

The system is composed of a series of programs and databases. The input to the system is a Mapudungun sentence, phrase or word, which is processed in different stages until transformed into a Spanish output.

Machine Translation (MT) consists of the use of computers to automate some or all the process of translating from one language to another (Jurafsky & Martin, 2000). The AVENUE MT System makes use of a transfer-based strategy (Peterson, 2002), which consists of the application of three steps: analysis, transfer, and generation. Analysis is the parsing of the source language into a syntactic representation, transfer is the mapping of the source language parse into a corresponding syntactic structure in the target language, and generation is the production of a target language output. These processes are performed according to the Transfer Rules contained in the Transfer Grammar. The function of the transfer rules is, first, to build a parse of a Mapudungun expression. Next, the rules transfer the relevant lexical and grammatical features to the target language in order to build equivalent Spanish expressions, copying, modifying, or rearranging grammatical values according to the requirements of the Spanish grammar and lexicon.

A translation divergence occurs when an adequate translation should take into account differences in the syntactic structure and the expression of the semantic components by the syntactic constituents (Trujillo, 1999). The types of divergences that have been most widely discussed in the Machine Translation literature are lexical, syntactic, and morphological divergences (Trujillo, 1999; Dorr, 1993; Arnold et al., 1994).

Although Mapudungun and Spanish also present lexical, syntactic, and morphological divergences, it is our understanding that the most important divergence between them is morphosyntactic in nature. By morphosyntactic divergence we mean a systematic divergence in the degree of synthesis (Comrie, 1989) of different grammatical structures, due to the properties of each language in terms of morphological typology.