A Cascaded Syntactic Analyser for Basque

I. Aduriz *, M. J. Aranzabe, J. M. Arriola, A. Díaz de Ilarraza,
K. Gojenola, M. Oronoz, L. Uria

IXA Group (http://ixa.si.ehu.es)
Department of Computer Languages and Systems
University of the Basque Country
P.O. box 649, E-20080 Donostia
mailto:jiporanm@si.ehu.es

* Department of General Linguistics
University of Barcelona
Gran Via de las Corts Catalans, 585, 08007 Barcelona
itziar@fil.ub.es

Abstract. This article presents a robust syntactic analyser for Basque and the different modules it contains. Each module is structured in different analysis layers for which each layer takes the information provided by the previous layer as its input; thus creating a gradually deeper syntactic analysis in cascade. This analysis is carried out using the Constraint Grammar (CG) formalism. Moreover, the article describes the standardisation process of the parsing formats using XML.

1 Introduction

This article describes the steps we have followed for the construction of a robust cascaded syntactic analyser for Basque. Robust parsing is understood as "*the ability of a language analyser to provide useful analyses for real-world input texts. By useful analyses, we mean analyses that are (at least partially) correct and usable in some automatic task or application*" (Ait-Mokhtar *et al.*, 2002). The creation of the robust analyser is performed based on a shallow parser. In this approach, incomplete syntactic structures are produced and thus the process goes beyond shallow parsing to a deeper language analysis in an incremental fashion. This allows us to tackle unrestricted text parsing through descriptions that are organized in ordered modules, depending on the depth level of the analysis (see Fig. 1).

In agglutinative languages like Basque, it is difficult to separate morphology from syntax. That is why we consider morphosyntactic parsing for the first phase of the shallow syntactic analyser, which, in turn, will provide the basis for a deeper syntactic analysis.