

# Constraint-Based Parsing as an Efficient Solution: Results from the Parsing Evaluation Campaign EASy

Tristan Vanrullen, Philippe Blache, Jean-Marie Balfourier

Laboratoire Parole et Langage / CNRS UMR 6057  
23 Avenue Robert Schuman  
13100 Aix-en-Provence France  
{tristan.vanrullen, philippe.blache}@lpl.univ-aix.fr

## Abstract

This paper describes the unfolding of the EASy evaluation campaign for french parsers as well as the techniques employed for the participation of laboratory LPL to this campaign. Three symbolic parsers based on a same resource and a same formalism (Property Grammars) are described and evaluated. The first results of this evaluation are analyzed and lead to the conclusion that symbolic parsing in a constraint-based formalism is efficient and robust.

## 1. Introduction

Recent advances in parsing technologies make it possible to deal with different kinds of inputs: several shallow parsing systems are now available as well as deep parsers, providing a set of solutions according to the needs. Parsing unrestricted texts will make use of shallow parsers whereas deep techniques can be used when fine-grained parses are needed. However, these techniques do not make it possible to reuse the same technology and the same resources both for deep and shallow parsing, and offering the possibility to choose the parsing granularity (see workshop deep & shallow parsing). Such technology has been proposed in the framework of Property Grammars (see (Blache, 2001 and 2005)). PG is a fully constraint-based formalism in which linguistic information is entirely specified in terms of non hierarchical constraints. A grammar in this approach is a set of constraints and parsing an input consists in evaluating this set. The input is then described (we say characterized) by the set of satisfied and violated constraints after such an evaluation. This conception of parsing makes it possible to treat any kind of input, whatever its form. In terms of parsing, we know that symbolic techniques offer several well-known advantages in terms of grammar development as well as reusability of the components. We show in this paper that these techniques can also be robust and efficient. In a former study (see (Vanrullen and Blache, 2002)), due to the lack of French treebanks at this time, we have proposed an evaluation technique based on the comparison of the parsers outputs without needing a reference corpus. That kind of evaluation lead us to improve our parsers by detecting their differences and by widening our knowledge of their weaknesses in terms of robustness, precision and efficiency. The French evaluation campaign EASy (Evaluation des Analyseurs Syntaxiques, see, (Vilnat & al, 2003 and 2004)) gave us the possibility to evaluate our parsers among several participants and with a strict evaluation framework based on a guideline and a large reference corpus.

This paper gives an account of the evaluation within the EASy campaign of three Property Grammar parsers, reusing the same mechanisms and the same resources (lexicon and grammar).

After a brief introduction to the EASy evaluation frame-

work and a short presentation of Property Grammars, we will describe the three parsers based on Property Grammars build in *Laboratoire Parole et Langage* (hereafter LPL) to participate to the EASy campaign. Then, we will show our results and discuss them.

## 2. The evaluation framework

The EASy project (see <http://www.elda.org/easy>) aims at the evaluation of parsers for French. It proposes an evaluation methodology making it possible to compare syntactic analyzers and, as a side effect, produce a large validated linguistic resource by combining automatically the results of the campaign.

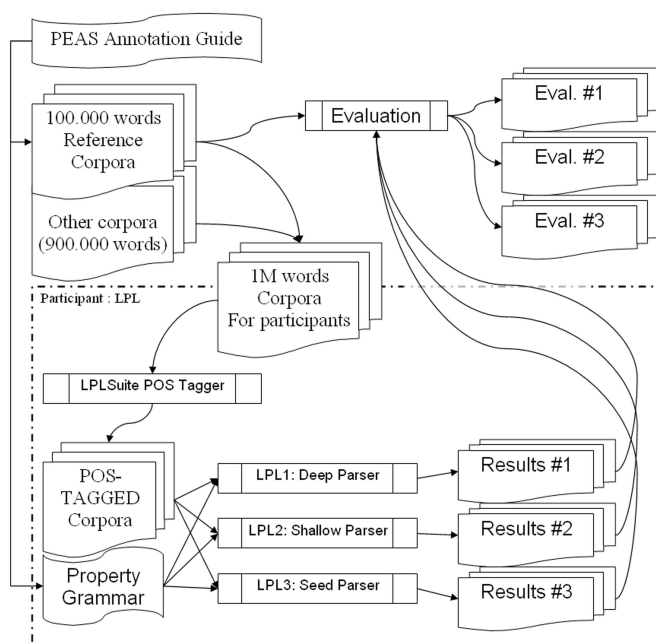


Figure 1: EASy campaign and three LPL parsers based on PGs

The corpus set (1 million words) comes from different sources newspapers, questions, websites, oral transcriptions, etc.), and contains different types: general corpora (21%), literary corpora: (23%), mail (15%), medical (6%),









