

Parsing incomplete sentences revisited ^{*}

M. Vilares¹, V.M. Darriba¹, and J. Vilares²

¹ Department of Computer Science, University of Vigo
Campus As Lagoas s/n, 32004 Ourense, Spain
{vilares,darriba}@uvigo.es

² Department of Computer Science, University of A Coruña
Campus de Elviña s/n, 15071 A Coruña, Spain
jvilares@mail2.udc.es

Abstract. We describe a context-free parsing algorithm to deal with incomplete sentences, including unknown parts of unknown length. It produces a finite shared-forest compiling all parses, often infinite in number, that could account for both the error and the missing parts. In contrast to previous works, we derive profit from a finer dynamic programming construction, leading to an improved computational behavior. We also introduce a deductive construction, which has the advantage of simplifying the task of description.

1 Introduction

An ongoing question in the design of dialogue systems is how to provide the maximal coverage and understanding of the language, finding the interpretations that have maximal thresholds, when the computational process must be prompted immediately at the onset of new input. This is largely due to the fact that the user often does not know the type of questions that the system answers. In this sense, it is often better to have a system that tries to guess a specific interpretation in case of ambiguity rather than ask the user for a clarification. As a consequence, analysis of the utterance should continuously anticipate the interaction with the user, based on the expectations of the system.

To comply with these requests, we need a parser which analyses the input simultaneously as it is entered, even when current data are only partially known. Two factors are at the origin of this behavior in natural language man-machine interfaces, whether text or speech-based. In the case of the former, the input language can only be approximately defined and individual inputs can vary widely from the norm [6] due to ungrammatical spontaneous phenomena. In the case of the latter [7], inputs can only often be considered as a distorted version of any of several possible patterns resulting from an erroneous recognition process.

In this context, our aim is computational. We restrict interaction types to only those necessary for immediate understanding using a predictive model

^{*} Research partially supported by the Spanish Government under projects TIC2000-0370-C02-01 and HP2002-0081, and the Autonomous Government of Galicia under projects PGIDT01PXI10506PN, PGIDIT02PXIB30501PR and PGIDIT02SIN01E.