Regional vs. Global Finite-State Error Repair*

M. Vilares[1], J. Otero[1], and J. Graña[2]

[1] Department of Computer Science, University of Vigo
Campus As Lagoas s/n, 32004 Ourense, Spain
{vilares,jop}@uvigo.es
[2] Department of Computer Science, University of A Coruña
Campus de Elviña s/n, 15071 A Coruña, Spain
grana@udc.es

Abstract. We focus on the domain of a regional least-cost strategy in order to illustrate the viability of non-global repair models over finite-state architectures. Our interest is justified by the difficulty, shared by all repair proposals, to determine how far to validate. A short validation may fail to gather sufficient information, and in a long one most of the effort can be wasted. The goal is to prove that our approach can provide, in practice, a performance and quality comparable to that attained by global criteria, with a significant saving in time and space. To the best of our knowledge, this is the first discussion of its kind.

1 Introduction

A classic problem in error repair is how far into the string to validate the process. Given that it is not possible to ensure that the correction and the programmer's intention are the same, the goal is to find the least-cost one. This can only be judged in the context of the entire input, and global methods [4, 5] are not necessarily the best option, due to their inefficiency, but are the most commonly used and for this reason considered to be the most appropriate. An alternative consists of examining the non-global context and attempting to validate repairs by tentatively recognizing ahead, following a successful approach on context-free grammars (CFGs) [7].

In this sense, although all proposals on error repair in the Chomsky's hierarchy are guided by some kind of linguistic data, whether grammar or automaton-based, each level strongly conditions the strategy to follow. So, requests on regular grammars (RGs) are different from those dealing with CFGs [8], where parses are not usually performed in depth, but breadth-wise; whilst the number of states in the associated push-down automaton is often small in practice. Our proposal takes this into account by limiting the search space associated to the repair. We explore the alternatives according to the topology of the corresponding finite automaton (FA). This allows us to restrict

* Research partially by the Spanish Government under projects TIN2004-07246-C03-01, TIN2004-07246-C03-02 and HP2002-0081, and the Autonomous Government of Galicia under projects PGIDIT03SIN30501PR and PGIDIT02SIN01E.