

Regional vs. global robust spelling correction

M. Vilares¹, J. Otero¹, and V.M. Darriba¹

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

Abstract. We explore the practical viability of a regional architecture to deal with robust spelling correction, a process including both unknown sequences recognition and spelling correction. Our goal is to reconcile these techniques from both the topological and the operational point of view. In contrast to the global strategy of most spelling correction algorithms, and local ones associated with the completion of unknown sequences, our proposal seems to provide an unified framework allowing us to maintain the advantages in each case, and avoid the drawbacks.

1 Introduction

In describing human performance in spelling correction, as compared to machine performance, we should try to take into account both the computational efficiency and the quality achieved in order to equal, or even do better than humans. This translates into a trade-off between the study of the often complex linguistic phenomena involved and the efficiency of the operational mechanisms available for implementation. In order to attain this goal, simple proposals can be sufficient to overcome most limits to providing an efficient strategy, even in the case of interactive applications. In fact, most approaches are oriented to improving first-guess accuracy [1] and/or to considering filter-based solutions to speed up the process [8]. So, system developers expect to reduce the time needed to ensure an adequate coverage of the problem, before taking into account more sophisticated linguistic information.

The state of the art techniques mainly focus on approximate string matching proposals, often firstly oriented to searching [2], although they can be easily adapted to robust spelling correction tasks [4]. Essentially, these algorithms apply a metric [5] to measure the minimum number of unit operations necessary to convert one string into another, sometimes embedding this task in the recognizer [10] in order to improve the computational efficiency. In this context, we identify a set of objective parameters in order to evaluate different approaches and algorithms in dealing with robust spelling correction.

* This research has been partially supported by the Spanish Government under projects TIN2004-07246-C03-01, and the Autonomous Government of Galicia under projects PGIDIT05PXIC30501PN, PGIDIT05SIN059E, PGIDIT05SIN044E, PGIDIT04SIN065E and PGIDIT03SIN30501PR.