Antonio Miranda García   amiranda@uma.es
Javier Calle Martín   jcalle@uma.es
David Moreno Olalla   dmoreno@uma.es
María Laura Esteban Segura   lauraesteban@uma.es
Nadia Obegi Gallardo   nobegi@uma.es
University of Málaga

Towards an automatic vowel length marker and tagger of Old English: a proposal

Several attempts have been made at the implementation of grammatical taggers in the last decade, both for ancient and modern languages. Even though each of them follows a different computational methodology, be it traditional or probabilistic, the most common performance criterion which all of them pursue is simply the *exact match*, that is to say, a one-to-one correspondence between the input and the output tag. The exact match criterion, however, is more than complicated when highly inflected languages like Old English are handled. The complex morphology of Anglo-Saxon, the number of phono-orthographic changes —such as vowel quantity, velarization, elision, assimilation, front mutation and the like—, along with the high level of inter and intradialectal variation, make it thorny to handle from a computational perspective. In the present paper, we intend to present a new tagger for Anglo-Saxon texts offered in two separate but at the same time inter-connected modules: a) a vowel length marker which allows to allocate the diacritic automatically; and b) the morphological tagger itself offering all the possible correct analyses of each word, irrespective of the context in which it is found. The overall accuracy of the tagger developed is quite respectable as its performance relies on two dictionaries, one for lemmas and the other for inflections. The former contains all the allomorphs of the lemma along with other plausible inter and intradialectal forms, thus solving the high level of variation which affects a language like Old English.