

Automatic Determination of Setswana Noun Classes

Gabofetswe Malema, Boago Okgetheng, Moffat Motlhanka and Goaletsa Rammidi

Abstract

Setswana like other Bantu languages divides nouns into classes. The class determines among others the correct adjectival, possessive and relative concords. Therefore determination of noun class is useful in applications of grammar checking and parts of speech tagging. There are a few works that have been carried out with regards to automatic noun class determination of Setswana nouns. We have developed a rule based program that determines the class of a given Setswana noun. Although in most cases the prefix of a given noun determines its class, there are cases where this does not and there are also cases where 2 classes have the same prefix. In this study we have used Setswana morphological rules and exception files where morphology cannot resolve prefix conflicts. Setswana verb and noun morphological analyzers are used to determine if a given word is a verb or a noun. The analyzers are slightly modified to treat all words with an uppercase first letter as proper nouns. Results from experiments show that the tool gives a performance of 87%. Majority of errors are from conflicting noun class prefixes and failure of some rules. The tool is also lighter as it does not require a corpus to run compared to previous implementations.

Keywords: Setswana; Noun Class; Morphology