

LREC 2014 Workshop

CCURL 2014

**Collaboration and Computing for Under-Resourced Languages
in the Linked Open Data Era**

Reykjavík (Iceland), 26th May 2014

Poster Presentation

Title

Morphological Analysis for Less-Resourced Languages: Maximum Affix Overlap Applied to Zulu

Authors

Uwe Quasthoff, Sonja Bosch, Dirk Goldhahn

Abstract

The paper describes a collaboration approach in progress for morphological analysis of less-resourced languages. The approach is based on firstly, a language-independent machine learning algorithm, Maximum Affix Overlap, that generates candidates for morphological decompositions from an initial set of language-specific training data; and secondly, language-dependent post-processing using language specific patterns. In this paper, the Maximum Affix Overlap algorithm is applied to Zulu, a morphologically complex Bantu language. It can be assumed that the algorithm will work for other Bantu languages and possibly other language families as well. With limited training data and a ranking adapted to the language family, the effort for manual verification can be strongly reduced. The machine generated list is manually verified by humans via a web frontend.