

Towards Zulu corpus clean-up, lexicon development and corpus annotation by means of computational morphological analysis.

[Download Here](#)



Institutional Repository

[UnisaIR Home](#) [College of Human Sciences](#) [School of Arts](#) [Department of African Languages](#) [F](#)

Towards Zulu corpus clean-up, lexicon development and corpus annotation by means of computational morphological analysis

Bosch, Sonja E.; Pretorius, Laurette

URI: <http://hdl.handle.net/10500/5539>

Date: 2011

Type: Article

Abstract:

This article reports on a practical, semi-automated procedure towards creating a clean, morphologically annotated Zulu corpus of tractable size that could eventually serve both as a gold standard for Zulu computational morphology and as basis for further linguistic annotation. A corpus development architecture is proposed which includes the corpus in various stages of development, a pre-processing module, the Zulu morphological analyser and its guesser variant, the machine-readable lexicon that serves as comprehensive lexical database for Zulu, and a human elicitation function for ensuring the integrity of the lexical data. This approach is novel in the sense that an existing rule-based, finite state Zulu computational morphological analyser is used as a core technology in this procedure to facilitate the complex, agglutinative nature of Zulu morphology. The corpus, at present consisting of the Zulu version of the South African Constitution, will undergo morphological analysis and tagging as a first level of annotation.

Citation: Bosch, SE; Pretorius, L. 2011. Towards Zulu corpus clean-up, lexicon development and corpus annotation by means of computational morphological analysis. *South African Journal of African Languages*, 31, no. 1 2011. pp.138-158

[Show full item record](#)

Files in this item



Name: bosch_sajal_v31_n ...
Size: 1.345Mb
Format: PDF

[View/Open](#)

Copyright Statement

Items in UNISA Institutional Repository are protected by copyright, with all rights reserved, unless otherwise indicated. Items may only be viewed and downloaded for private research and study purposes. Please acknowledge publications according to acceptable standards and norms.

This item appears in the following Collection(s)

- [Open Research Collection](#) [3606]
- [Research Outputs \(African Languages\)](#) [64]

[Contact Us](#) | [Send Feedback](#) Towards Zulu corpus clean-up, lexicon development and corpus annotation by means of computational morphological analysis, database in a timely manner takes the free gap function.