**New ways to see the wood for the trees in Afrikaans syntax**

*Peter Dirix & Liesbeth Augustinus, KU Leuven*

Over the past few decades, corpus linguistics has become the basis of for both lexicographers compiling dictionaries and syntacticians compiling grammars. While a flat corpus without annotation is rather easy to collect and mostly sufficient for lexicographers, a syntactically annotated corpus or *treebank* is a must for syntax description. For smaller languages or non-standard variants, this type of resource is often lacking. Even Afrikaans with 7 million native speakers can be considered as a low-resource language in this respect.

We describe our efforts in creating resources (a manually verified morphosyntactic lexicon of 250,000 entries, a PoS-tagged large corpus of 50 million words, a small treebank of about 45,000 words) and tools (a simple search tool, the treebank-based GrETEL for Afrikaans search tool,[1] a tokenizer and a lemmatizer) for Afrikaans as well as the work we are starting on a gamification project in order to use crowdsourcing for syntactic annotation of dependency relations. We also show how we used these tools to carry out a quantitative corpus study for linguistic issues the literature does not really agree on, presenting a case study on the usage of substitute infinitives (*infinitivus pro participio*) in Afrikaans. We conclude with describing the issues we still have due to the quality and the size of the resources and propose solutions for future development which could also be applied to the corpus portal of the Virtual Institute for Afrikaans (VivA).[2]

[1] http://gretel.ccl.kuleuven.be/afribooms/
[2] https://www.viva-afrikaans.org/