

## **Finding Syntactic Characteristics of Surinamese Dutch**

*Erik Tjong Kim Sang (Meertens Institute)*

### *Introduction*

Surinamese Dutch is a variant of Dutch spoken in Suriname, a former colony of The Netherlands in the north of South America. The lexical differences between Surinamese Dutch and standard Dutch have been studied, see for example Donselaar (2013), but we would like to know if there are striking syntactic differences between the two language variants. In this study we will use automatic methods to compare syntactic features of two texts, one written in Surinamese Dutch and one written in standard Dutch, and describe the differences.

### *Method*

We selected two novels from the Digital Library for Dutch Language (dbnl.nl): one written in Surinamese Dutch and one written in standard Dutch. We processed both texts with Alpino, the best available syntactic parser for Dutch (Van Noord, 2006) which identifies syntactic relations between pairs of words. We compare the texts by counting the different syntactic relations and comparing their frequencies in each text. For the comparison we use the t-test (Church et al., 1991) in combination with additive smoothing (add 0.5 smoothing). The t-test computes scores for pairs of related frequencies with the formula  $(f_1 - f_2) / \sqrt{(f_1 + f_2)}$  where  $f_1$  and  $f_2$  are the relative frequencies of a word in two texts. After sorting the resulting t-scores from high to low, the top of the resulting list gives an indication about what words were more frequent in the first text than could be expected based on the second reference text.

### *Results*

We tested the comparison method by comparing the frequencies of tokens (word plus punctuation) in the two texts. This worked well: eight of the top 30 words suggested by the t-test were words that were specific to Surinamese Dutch. Next, we applied the t-test to four different syntactic patterns, where the head and dependent word were represented as either part-of-speech tags or lemmas. We inspected all sentences related to the top twenty of each of these four syntactic patterns and found that five of the patterns were specific for Surinamese Dutch, among which the well-known free usage of "to go" as a modal verb expressing future tense.

### *Concluding remarks*

We used an automatic method for finding syntactic differences between Surinamese Dutch and standard Dutch which employs the t-test (Church et al., 1991). Although the method works reasonable for discovering lexical differences (25% real differences in the top thirty of the suggestions), finding syntactic differences proved to be harder. We inspected 80 syntactic relations suggested by the t-test and found five real differences between the two language variants (6%). From our work we draw the conclusion that the t-test is useful for finding lexical and syntactic differences between language variants and that the syntactic difference between Surinamese Dutch and standard Dutch is most likely smaller than the lexical difference between these two language variants.