

# Automatic Detection of Grammar Errors in Primary School Children's Texts

## A Finite State Approach

### Abstract

This thesis concerns the analysis of grammar errors in Swedish texts written by primary school children and the development of a finite state system for finding such errors. Grammar errors are more frequent for this group of writers than for adults and the distribution of the error types is different in children's texts. In addition, other writing errors above word-level are discussed here, including punctuation and spelling errors resulting in existing words.

The method used in the implemented tool *FiniteCheck* involves subtraction of finite state automata that represent grammars with varying degrees of detail, creating a machine that classifies phrases in a text containing certain kinds of errors. The current version of the system handles errors concerning agreement in noun phrases, and verb selection of finite and non-finite forms. At the lexical level, we attach all lexical tags to words and do not use a tagger which could eliminate information in incorrect text that might be needed later to find the error. At higher levels, structural ambiguity is treated by parsing order, grammar extension and some other heuristics.

The simple finite state technique of subtraction has the advantage that the grammars one needs to write to find errors are always positive, describing the valid rules of Swedish rather than grammars describing the structure of errors. The rule sets remain quite small and practically no prediction of errors is necessary.

The linguistic performance of the system is promising and shows comparable results for the error types implemented to other Swedish grammar checking tools, when tested on a small adult text not previously analyzed by the system. The performance of the other Swedish tools was also tested on the children's data collected for this study, revealing quite low recall rates. This fact motivates the need for adaptation of grammar checking techniques to children, whose errors are different from those found in adult writers and pose more challenge to current grammar checkers, that are oriented towards texts written by adult writers.

The robustness and modularity of *FiniteCheck* makes it possible to perform both error detection and diagnostics. Moreover, the grammars can in principle be reused for other applications that do not necessarily have anything to do with error detection, such as extracting information in a given text or even parsing.

**KEY WORDS:** grammar errors, spelling errors, punctuation, children's writing, Swedish, language checking, light parsing, finite state technology