Verification of Controlled Grammars

Gregor Thurmair
SIETEC Systemtechnik
Carl-Wery-Straße 22, D-81739 München

Abstract

The following paper describes the development of a tool which verifies controlled grammars. First, the development context and the application is described. Then, the definition of the task, and the architecture of the system are given: It consists of a syntactic analyser, a repository of ill-formed structures, a matcher, and an output formulator. Finally, system test results are presented, and some problems to be solved are explained; also, the application environment is sketched.

1 Context

The prototype described in the following paper was developed in the context of the "Translator's Workbench" Esprit Project, based on the METAL linguistic platform. This project aims at improving the documentation and translation process by offering a bundle of software tools, one of which is language checking. This covers spellers, grammar checkers, and controlled grammar verifiers.

The language checking TWB component is the result of trying to optimise the documentation process. Considerations how to improve the input of machine translation were compared with guidelines for technical authors, and large overlaps were detected. This resulted in a common effort to improve both the readability and the translatability of texts, by setting up styleguides for authors. As they define a sublanguage of their own, in that they restrict the grammar of a language they are called controlled languages.

The baseline of the project was the experience that in the case of grammar checking, it turned out that most existing grammar checkers are not reliable
and therefore are very restricted in their usability. This is due to the fact that most of them do not use a real grammar but are based on some more or less sophisticated pattern matching techniques. However, the fact that they sell shows that there is a need for those tools.

TWB again followed several approaches in grammar and style checking, described in /KUG92/. Among them is a small ATN based NP grammar for German in order to detect agreement errors, which turned out to be the most frequent in German texts (/HEL90/). A second approach has been followed for Spanish grammar checking: Here we used an existing grammar (the METAL analysis) and enriched it by a "peripheral" grammar on top of the core grammar which tries to identify the cases of ungrammaticality (agreement errors, wrong verb argument usage, etc.). During parsing, it can be detected if one of those special rules has fired, and if so, the appropriate diagnostic measure can be taken.

As a result, it turned out that grammar checking needs much more linguistic intelligence if it should be helpful and reliable. It needs fully developed lexicon and syntax components and some "heavy" machinery (in terms of computing power). The TWB tools are the better the more developed the underlying grammars are. However, this hampers their portability to other languages as it means considerable investment.

A last area of language checking was style checking, or better verification of controlled grammars. This is closer related to the documentation business as it tries to implement guidelines for good technical writing, conventions for style and layout, also implying language criteria.

2 Definition

The following paper does not intend to define what "good style" might be; there are many style checkers available (at least for English) which do not do more but formalise their author's intentions. Instead, the goal of the present paper is to present a software tool which allows for the verification of controlled grammars.

Controlled Grammars are subsets of a language, defined for certain purposes and for certain reasons: