

Computational grammars are computer programs which describe subsets of natural languages. They are used in applications such as machine translation and natural language generation, especially when the focus is on quality rather than coverage. Just like any other software, these grammars may contain bugs, and hence they need to be tested to ensure their quality. This thesis presents two contributions in testing computational grammars.

The first method ensures that individual grammar rules do not contradict each other. With just the grammar itself, and a wide-coverage word list for the language, we employ a set of logical constraints which reveal any internal inconsistencies in the grammar.

The second method ensures that the grammar produces correct language. We devise an interaction with grammar writers, who are experts in both linguistics and programming, and testers, who only need to be fluent in the language. Our solution is to generate a minimal and exhaustive sample of sentences, which the testers can read and point out any errors, or confirm that they are correct.

We present complete implementations of both methods, along with evaluation on languages from diverse language families.

*Inari Listenmaa* Formal Methods for Testing Grammars



UNIVERSITY OF GOTHENBURG

## Formal Methods for Testing Grammars

*Inari Listenmaa*

**Ph.D. thesis**

Department of Computer Science and Engineering  
University of Gothenburg, Sweden 2019



UNIVERSITY OF  
GOTHENBURG

IT Faculty

2019



ISBN 978-91-7833-322-6