# Using regular tree grammars to optimise surface realisation

Abstract of the dissertation submitted to the Faculty of Computer Science, Free University of Bozen-Bolzano in fulfillment of the degree of

Masters of Science in Computer Science as part of the European Masters Program in Language and Communicaton Technologies

*Submitted by*:
Laura Haide Perez

*Supervisors*:
Dr. Rafaella Bernardi
Dr. Claire Gardent

## Abstract

Surface realisation is the task within Natural Language Generation responsible for mapping an *abstract linguistic structure* (e.g., a logical formula) into one or more sentence(s). Particularly, reversible surface realisers, i.e. surface realisers that use the same reversible grammar for both parsing and generation, generally take as input a flat logical formulae. This kind of realisers present a number of important advantages but at the same time require some optimisations to overcome the complexity introduced due to the generation from flat semantics. To remedy this shortcoming, various optimisations have been proposed. In particular, filtering techniques have been developed which, on some given criterion, filter out from the initial search space all candidate solutions which cannot possibly lead to valid ones.

This thesis work aims to optimise such a realiser, specifically, we aim to reduce the initial search space by reducing the effects of one of the sources of complexity namely, *lexical ambiguity*. The surface realiser we focus on, is based on a wide-coverage *Feature Based Lexicalized Tree Adjoining Grammar* extended with a compositional semantics. The realiser furthermore integrates a polarity filter. Although it was shown to drastically reduce the initial search space, polariy filtering still has some limitations. In this thesis, we investigate an alternative filtering technique based on a novel formal framework for modelling polarity counts, namely *Feature Based Regular Tree Grammar*. The underlying motivation for this work is that the derivation trees of a Tree Adjoining Grammar can be encoded as the language of a Regular Tree Grammar. Thus, we can model polarities (the syntactic resources and requirements associated with a given TAG tree) based on all the information provided by the elementary trees. Furthermore, detecting the initial candidate combinations of trees that can not possibly yield a full derivation tree boils down to parsing with an RTG.

In this thesis, we make the RTG filtering idea precise. Next, we provide it with an implementation by developing an RTG based module for generating a parse forest from the logical formulae input to the surface realiser. Third, we compare the impact of RTG-based filtering with polarity filtering on a number of carefully selected test cases and using different metrics.