

ON H-LEXICALIZED RESTARTING LIST AUTOMATA

MARTIN PLÁTEK^(A) FRIEDRICH OTTO^(B) FRANTIŠEK MRÁZ^(A)

^(A) *Charles University, Faculty of Mathematics and Physics
Department of Computer Science, Malostranské nám. 25, 118 00 Praha 1, Czech Republic*

`martin.platek@mff.cuni.cz`
`frantisek.mraz@mff.cuni.cz`

^(B) *Fachbereich Elektrotechnik/Informatik, Universität Kassel
34109 Kassel, Germany*
`f.otto@uni-kassel.de`

ABSTRACT

Following some previous studies on restarting list automata (RLA), we concentrate on a generalized and refined model – the *h-lexicalized restarting list automaton* (hLxRLA), which is useful for expressing properties of lexicalized syntax in computational linguistics. We present several subclasses of hLxRLA and provide some variants and extensions of the Chomsky hierarchy – the *h-lexicalized variants of the Chomsky hierarchy*. We compare the input languages of RLA, which are the languages traditionally considered in automata theory, to the so-called *basic* and *h-proper languages* of hLxRLA, which are used to define *h-lexicalized syntactic analysis*. The h-lexicalized syntactic analysis allows us to stress several nice syntactic properties of h-lexicalized restarting automata (hRLWW). We present a transformation from monotone RLWW-automata that recognize the context-free languages (CFL) as their input languages to deterministic monotone hRLWW-automata that compute h-lexicalized syntactic analysis for the whole class CFL through their basic and h-proper languages. Through this transformation, we obtain several types of deterministic hRLWW-automata and hLxRLA-automata that cover h-lexicalized syntactic analyses of CFL and that satisfy the Complete Strong Correctness Preserving Property.

Keywords: restarting list automaton, h-lexicalized language, h-lexicalized syntactic analysis, complete strong correctness preserving property, extended Chomsky hierarchy

1. Introduction

Chomsky introduced his well-known hierarchy of grammars to formulate the phrase-structure (immediate constituents) syntax of natural languages [2]. On the other hand, the syntax of most European languages (including English) is very often considered as a lexicalized syntax. In other words, the syntactic categories of Chomsky are

Some of the results of this paper have been announced at AFL 2017 in Debrecen, September 2017, and at NCMA 2018 in Košice, August 2018. Extended abstracts appeared in the proceedings of these conferences [19, 24].