

# **Algebraic decoder specification: coupling formal-language theory and statistical machine translation**

Matthias Büchse  
matthias@buech.se

January 2015

**Dissertation zur Erlangung des akademischen Grades  
Doktor rerum naturalium (Dr. rer. nat.)**

vorgelegt an der Technischen Universität Dresden  
Fakultät Informatik

eingereicht von Dipl.-Inf. Matthias Büchse  
\* 1983-08-12 in Köthen (Anhalt)

eingereicht am 2014-08-05

verteidigt am 2014-12-18

begutachtet durch Prof. Dr.-Ing. habil. Heiko Vogler  
Technische Universität Dresden  
Prof. Dr. rer. nat. Alexander Koller  
Universität Potsdam

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Decoder specification . . . . .	3
1.2	Hierarchical phrases . . . . .	4
1.3	Explicit syntax . . . . .	9
1.4	The algebraic framework, preliminary version . . . . .	12
1.5	Main contributions . . . . .	17
1.6	Related work and bibliographic remarks . . . . .	21
<b>2</b>	<b>Preliminaries</b>	<b>25</b>
2.1	Mathematical foundations . . . . .	25
2.2	Trees . . . . .	29
2.3	Algebras and semirings . . . . .	32
2.4	Weighted tree automata . . . . .	36
<b>3</b>	<b>Input product and output product of a weighted synchronous context-free tree grammar and a weighted tree automaton</b>	<b>45</b>
3.1	Introduction . . . . .	45
3.2	Weighted synchronous context-free tree grammars . . . . .	47
3.3	Closure under input and output product . . . . .	52
3.4	An Earley-like algorithm for the input product . . . . .	61
3.5	Conclusion, discussion, and outlook . . . . .	81
<b>4</b>	<b>Generic binarization of weighted grammars</b>	<b>83</b>
4.1	Introduction . . . . .	83
4.2	Interpreted regular tree grammars . . . . .	85
4.3	Binarization mappings . . . . .	91
4.4	Constructing a binarization mapping . . . . .	96
4.5	Constructing a computable binarization mapping . . . . .	105
4.6	Application to established formalisms . . . . .	114
4.7	Conclusion, discussion, and outlook . . . . .	126

<b>5</b>	<b>Determinizing weighted tree automata using factorizations</b>	<b>131</b>
5.1	Introduction . . . . .	131
5.2	Preliminary notions and results . . . . .	134
5.3	Determinization of classical WTA . . . . .	143
5.4	Deciding the twins property . . . . .	151
5.5	The case of non-classical WTA . . . . .	161
5.6	Conclusion, discussion, and outlook . . . . .	164
<b>6</b>	<b>Conclusion</b>	<b>167</b>
6.1	The algebraic framework, full version . . . . .	167
6.2	Outlook . . . . .	173
	<b>Bibliography</b>	<b>183</b>
	<b>Index</b>	<b>203</b>