

# **Extended Finite State Models of Language**

PROCEEDINGS OF THE ECAI'96 WORKSHOP

ANDRÁS KORNAI

*Editor*

EVA EJERHED  
FREDERICK JELINEK  
LAURI KARTTUNEN  
ANDRÁS KORNAI

*Program Committee*

## CONTENTS

<b>Eberhard Bertsch and Mark-Jan Nederhof:</b>	
An innovative finite-state concept for recognition and parsing of context-free languages	1
<b>Aravind Joshi:</b> Comments on Bertsch and Nederhof	10
<b>Jean-Pierre Chanod and Pasi Tapanainen:</b>	
A non-deterministic tokeniser for finite-state parsing	11
<b>Erzsébet Csuhaaj-Varjú:</b> COLONIES: a multi-agent approach to language generation	14
<b>Mark-Jan Nederhof:</b> Comments on Csuhaaj-Varjú	21
<b>Eva Ejerhed:</b> Finite state segmentation of discourse into clauses	24
<b>Gregory Grefenstette:</b> Light parsing as finite-state filtering	34
<b>Frederick Jelinek:</b> Language modeling for speech recognition	40
<b>Mehryar Mohri:</b> Comments on Jelinek	47
<b>Aravind Joshi:</b> A parser from antiquity:	
An early application of finite state transducers to natural language parsing	49
<b>Lauri Karttunen:</b> Comments on Joshi	51
<b>Lauri Karttunen:</b> Regular expressions for finite-state syntactic description	53
<b>András Kornai:</b> Vectorized finite state automata	54
<b>Emmanuel Roche:</b> Comments on Kornai	60
<b>Kimmo Koskenniemi:</b> Finite-state morphology and information retrieval	62
<b>Eva Ejerhed:</b> Comments on Koskenniemi	66
<b>Mehryar Mohri, Fernando Pereira and Michael Riley:</b>	
Weighted automata in text and speech processing	67
<b>András Kornai:</b> Comments on Mohri, Pereira and Riley	72
<b>Richard Oehrle:</b> Finite-state methods, binding, and anaphora	74
<b>Emmanuel Roche:</b> Finite-state transducers: parsing free and frozen sentences	82
<b>Richard Sproat:</b> Comments on Roche	89
<b>Catherine Rood:</b> Efficient finite-state approximation of context free grammars	91
<b>Anne Schiller:</b> Multilingual finite-state noun phrase extraction	98
<b>Wojciech Skut:</b> Finite automata for processing word order	103
<b>Richard Sproat:</b> Multilingual text analysis for text-to-speech synthesis	108
<b>Kimmo Koskenniemi:</b> Comments on Sproat	115
<b>B. Srinivas:</b> Explanation-based learning and finite state transducers:	
Applications to parsing lexicalized tree adjoining grammars	117
<b>Masakazu Tateno, Hiroshi Masuichi and Hiroshi Umemoto:</b>	
The Japanese lexical transducer based on stem-suffix style forms	123
<b>J.M. Vilar, E. Vidal and J.C. Amengual:</b>	
Learning extended finite state models for language translation	127
<b>Bruce Watson:</b> Implementing and using finite automata toolkits	132
<b>Erzsébet Csuhaaj-Varjú:</b> Comments on Watson	136