

Organisation

Organisers

John Carroll
Robert C. Moore
Stephan Oepen

Programme Committee

John Carroll, University of Sussex, UK
Gregor Erbach, Telecommunications Research Centre Vienna, Austria
Bernd Kiefer, DFKI Saarbruecken, Germany
Rob Malouf, Rijkuniversitet Groningen, The Netherlands
Robert C. Moore, Microsoft Research, USA
Gertjan van Noord, Rijkuniversitet Groningen, The Netherlands
Stephan Oepen, Saarland University, Germany
Gerald Penn, Bell Labs Research, USA
Hadar Shentov, Xerox Palo Alto Research Centre, USA
Kentaro Torisawa, Tokyo University, Japan

To obtain additional copies of these proceedings contact:

Dr. John Carroll
Cognitive and Computing Sciences
University of Sussex
Brighton BN1 9QH, UK
john.carroll@cogs.susx.ac.uk

Contents

Invited Talk

<i>Why not Cubic?</i> Ronald M. Kaplan	3
---	---

Papers

<i>Efficient Large-Scale Parsing — a Survey</i> John Carroll, Stephan Oepen	7
<i>Precompilation of HPSG in ALE into a CFG for Fast Parsing</i> John C. Brown, Suresh Manandhar	13
<i>Time as a Measure of Parsing Efficiency</i> Robert C. Moore	23
<i>Measuring Efficiency in High-accuracy, Broad-coverage Statistical Parsing</i> Brian Roark, Eugene Charniak	29
<i>Some Experiments on Indicators of Parsing Complexity for Lexicalized Grammars</i> Anoop Sarkar, Fei Xia, Aravind Joshi	37
<i>Large Scale Parsing of Czech</i> Pavel Smrž, Aleš Horák	43

Demos

<i>Cross-Platform, Cross-Grammar Comparison — Can it be Done?</i> Ulrich Callmeier, Stephan Oepen	53
<i>Tools for Large-Scale Parser Development</i> Hisami Suzuki, Jessie Pinkham	54

Invited Talk

Why not Cubic?

Ronald M. Kaplan
Xerox PARC
3333 Coyote Hill Road, Palo Alto
CA 94304, USA
kaplan@parc.xerox.com

It is well-established that the parsing problem for higher-level constraint-based formalisms such as LFG, PATR, and HPSG is in the NP-hard complexity class. Thus there are worst-case sentences and grammars for which there are no known polynomial algorithms. Unhappily, the most straightforward parsers for these formalisms tend to be exponential not only in the worst cases but also for the common cases of sentences and grammars that intuitively seem to be less complex. Research aimed at improving performance has typically accepted the inherently exponential nature of the problem and has then focused on implementation techniques that can lower the space/time computational resource curves but without actually changing their shape.

In this talk I will discuss an alternative strategy that we have been exploring in our work on LFG parsing. Instead of taking the exponential as a given for arbitrary grammars and asking how we can make it more palatable, we studied a restricted class of LFG grammars whose languages are obviously context-free. We then asked a different question: why doesn't a conventional LFG parser recognize these languages in cubic time, their theoretically obtainable bound? We developed a few key ideas that taken together provide for cubic performance for the special case of a completely context-free-equivalent LFG grammar, and provide nearly cubic performance for the less restricted set of LFG grammars that are linguistically relevant.