

# **The Balancing Act**

**Combining Symbolic and Statistical  
Approaches to Language**

**Proceedings of the Workshop**

**1 July 1994  
New Mexico State University  
Las Cruces, New Mexico, USA**



# **The Balancing Act**

## **Combining Symbolic and Statistical Approaches to Language**

### **Proceedings of the Workshop**

**1 July 1994**

**New Mexico State University  
Las Cruces, New Mexico, USA**

**Published by the Association for Computational Linguistics**

© 1994, Association for Computational Linguistics

Order additional copies from:

ACL (Walker)  
C. N. 925  
Bernardsville, NJ 07924-0925, USA  
1-908-204-1337

## PREFACE

This volume contains the papers prepared for and presented at The Balancing Act Workshop: Combining Symbolic and Statistical Approaches to Language, held on 1 July 1994 at the New Mexico State University in Las Cruces, New Mexico, USA. The Workshop was sponsored by the Association for Computational Linguistics.

A renaissance of interest in corpus-based statistical methods has rekindled old controversies — rationalist vs. empiricist philosophies, theory-driven vs. data-driven methodologies, symbolic vs. statistical techniques. The aim of this workshop was to set aside a priori biases and explore the balancing act that must take place when symbolic and statistical approaches are brought together. We received an impressive number of high quality submissions, indicating the level of interest in these questions. Papers were accepted from authors having a wide range of perspectives in order to initiate discussion that includes philosophical, theoretical, and practical issues.

Research of this kind requires that the researcher make choices: What knowledge will be represented symbolically and how will it be obtained? What assumptions underlie the statistical model? What is the researcher gaining by combining approaches? Questions like these, and the metaphor of the balancing act, provide a unifying theme to these contributions from a wide spectrum of language researchers.

We wish to thank all those who submitted papers for consideration. We also acknowledge several anonymous reviewers who helped to make the program stronger. Finally, we particularly thank the authors who prepared their submissions under a very tight timetable.

Judith L. Klavans  
Columbia University  
Department of Computer Science  
500 W 120th Street  
New York, NY 10027, USA  
klavans@cs.columbia.edu  
Phone: (212) 939-7120  
Fax: (914) 478-1802

Philip Resnik  
Sun Microsystems Laboratories, Inc.  
Mailstop UCHL03-207  
Two Elizabeth Drive  
Chelmsford, MA 01824-4195 USA  
philip.resnik@east.sun.com  
Phone: (508) 442-0841  
Fax: (508) 250-5067



## Table of Contents

(papers listed alphabetically by first author)

INVITED TALK: Qualitative and Quantitative Designs for Speech Translation . . . . .	1
Hiyan Alshawi	
The Noisy Channel and the Braying Donkey . . . . .	11
Roberto Basili, Maria Teresa Paziienza, and Paola Velardi	
AMALGAM: Automatic Mapping Among Lexico-Grammatical Annotation Models . . . . .	21
Eric Atwell, John Hughes, and Clive Souter (Alternate Paper)	
Study and Implementation of Combined Techniques for Automatic Extraction of Terminology . . . . .	29
Béatrice Daille	
Parsing with Principles and Probabilities . . . . .	37
Andrew Fordham and Matthew Crocker	
Do we Need Linguistics When We Have Statistics? A Comparative Analysis of the Contributions of Linguistic Cues to a Statistical Word Grouping System . . . . .	43
Vasileios Hatzivassiloglou	
Complexity of Description of Primitives: Relevance to Local Statistical Computations . . . . .	53
Aravind K. Joshi and B. Srinivas	
The Automatic Construction of a Symbolic Parser Via Statistical Techniques . . . . .	60
Shyam Kapur and Robin Clark	
Integrating Symbolic and Statistical Approaches in Speech and Natural Language Applications . . . . .	69
Marie Meteer and Herbert Gish	
Combining Linguistic with Statistical Methods in Automatic Speech Understanding . . . . .	76
Patti Price	
Exploring the Statistical Derivation of Transformational Rule Sequences for Part-of-Speech Tagging . . . . .	86
Lance A. Ramshaw and Mitchell P. Marcus	
Bootstrapping Statistical Processing into a Rule-Based Natural Language Parser . . . . .	96
Stephen D. Richardson	
Recovering From Parser Failures: A Hybrid Statistical/Symbolic Approach . . . . .	104
Carolyn Penstein Rosé and Alex Waibel (Alternate Paper)	
Statistical versus Symbolic Parsing for Captioned-Information Retrieval . . . . .	112
Neil C. Rowe	
Learning a Radically Lexical Grammar . . . . .	122
Danny Solomon and Mary McGee Wood	

**Table of Contents**  
(papers listed alphabetically by title)

AMALGAM: Automatic Mapping Among Lexico-Grammatical Annotation Models .....	21
Eric Atwell, John Hughes, and Clive Souter (Alternate Paper)	
The Automatic Construction of a Symbolic Parser Via Statistical Techniques .....	60
Shyam Kapur and Robin Clark	
Bootstrapping Statistical Processing into a Rule-Based Natural Language Parser .....	94
Stephen D. Richardson	
Combining Linguistic with Statistical Methods in Automatic Speech Understanding .....	70
Patti Price	
Complexity of Description of Primitives: Relevance to Local Statistical Computations .....	53
Aravind K. Joshi and B. Srinivas	
Do we Need Linguistics When We Have Statistics? A Comparative Analysis of the Contributions of Linguistic Cues to a Statistical Word Grouping System .....	43
Vasileios Hatzivassiloglou	
Exploring the Statistical Derivation of Transformational Rule Sequences for Part-of-Speech Tagging .....	86
Lance A. Ramshaw and Mitchell P. Marcus	
Integrating Symbolic and Statistical Approaches in Speech and Natural Language Applications .....	69
Marie Meteer and Herbert Gish	
Learning a Radically Lexical Grammar .....	122
Danny Solomon and Mary McGee Wood	
The Noisy Channel and the Braying Donkey .....	21
Roberto Basili, Maria Teresa Paziienza, and Paola Velardi	
Parsing with Principles and Probabilities .....	37
Andrew Fordham and Matthew Crocker	
INVITED TALK: Qualitative and Quantitative Designs for Speech Translation .....	1
Hiyan Alshawi	
Recovering From Parser Failures: A Hybrid Statistical/Symbolic Approach .....	104
Carolyn Penstein Rosé and Alex Waibel (Alternate Paper)	
Statistical versus Symbolic Parsing for Captioned-Information Retrieval .....	112
Neil C. Rowe	
Study and Implementation of Combined Techniques for Automatic Extraction of Terminology .....	29
Béatrice Daille	