NAACL-ANLP 2000 Workshop

Syntactic and Semantic Complexity in Natural Language Processing Systems

Editors/Organizing Committee
Amit Bagga, James Pustejovsky, Wlodek Zadrozny

April 30, 2000
Seattle, Washington

© 2000, Association for Computational Linguistics

Order additional copies from:

Association for Computational Linguistics
75 Paterson Street, Suite 9
New Brunswick, NJ 08901 USA
+1-732-342-9100 phone
+1-732-342-9339 fax
acl@aclweb.org

PREFACE

The last decade has seen an explosion in the work done in the development of robust natural language processing systems. A common methodology used in building these systems has been to analyze a sample of the data available (either manually, or automatically for training statistical systems), build statistical/heuristical schemas based upon the analysis, and test the system on a blind sample of the data. Due to this commonly used paradigm, an important area of research that has not been given the attention it deserves is the estimation of syntactic and semantic complexity faced by these systems in the tasks they perform.

The Workshop on Syntactic and Semantic Complexity in Natural Language Processing Systems, held on April 30th, 2000 at the Language Technology Joint Conference on Applied Natural Language Processing and the North American Chapter of the Association of Computational Linguistics (ANLP-NAACL2000) was organized around the goals of discussing, promoting, and presenting new research results regarding the question of complexity as it pertains to the syntax and semantics of natural language. In particular, the goal of the workshop was to focus on:

- estimation of the syntactic and semantic complexity of specific NLP tasks
- semantic complexity and world knowledge
- role of syntactic and semantic complexity in system design and testing
- syntactic and semantic complexity and its role in the evaluation of NLP systems
- use of syntactic and semantic complexity as a performance predictor
- relationship between syntactic and semantic complexity

We would like to thank all authors who showed their interest by submitting papers to the workshop. We would also like to thank the members of the program committee: Branimir Boguraev (IBM Research), J-P Chanod (Xerox, Grenoble), Shalom Lappin (Kings College, London), Aravind Joshi (University of Pennsylvania), Larry Moss (Indiana), Rohit Parikh (CUNY), and Adam Pease (Teknowledge).

Amit Bagga
James Pustejovsky
Wlodek Zadrozny

TABLE OF CONTENTS

Sentences vs. Phrases: Syntactic Complexity in Multimedia Information Retrieval 1
Sharon Flank

Using Long Runs as Predictors of Semantic Coherence in a Partial Document Retrieval System 6
Hyopil Shin and Jerrold F. Stach

Reducing Lexical Semantic Complexity with Systematic Polysemous Classes and Underspecification 14
Paul Buitelaar

Automatic Extraction of Systematic Polysemy Using Tree-cut 20
Noriko Tomuro

Dependency of context-based Word Sense Disambiguation from representation and domain complexity 28
Paola Velardi and Alessandro Cucchiarelli

Analyzing the Reading Comprehension Task 35
Amit Bagga

A Measure of Semantic Complexity for Natural Language Systems 42
Shannon Pollard and Alan W. Biermann

Example-based Complexity--Syntax and Semantics as the Production of Ad-hoc Arrangements of Examples 47
Robert John Freeman

Partially Saturated Referents as a Source of Complexity in Semantic Interpretation 51
David D. McDonald

Similarities and Differences among Semantic Behaviors of Japanese Adnominal Constituents 59
Kyoko Kanzaki, Qing Ma, and Hitoshi Isahara