

Supertagging: Using Complex Lexical Descriptions in Natural Language Processing

Edited by Srinivas Bangalore and Aravind K. Joshi

**A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England**

© 2010 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

For information about special quantity discounts, please email
special_sales@mitpress.mit.edu

This book was set in LaTeX by the authors.

Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Supertagging : using complex lexical descriptions in natural language processing
/ edited by Srinivas Bangalore and Aravind K. Joshi.

p. cm. — (A Bradford Book)

Includes bibliographical references and index.

ISBN 978-0-262-01387-1 (hardcover : alk. paper) 1. Natural language processing
(Computer science) 2. Computational linguistics—Statistical methods. 3.
Semantics. 4. Lexicology. I.

Bangalore, Srinivas, 1969– II. Joshi, Aravind K. (Aravind Krishna), 1929–

QA76.9.N38S86 2010

006.3'5—dc22

2009028194

10 9 8 7 6 5 4 3 2 1

Subject Index

- A* and *B*, 292
- A* over *B*, 292
- B* under *A*, 292
- $[\square^{\perp}E]$ rule, 294
- $[\diamond I]$ rule, 294
- 2-lexical context-free grammar (2-LCFG), 118
- 2-lexical tree-adjoining grammar (2-LTAG), 126

- A* search, 455
- aakaankshaa, 359
- adjectives, 393
- adjoining, 5, 408
- adjunct transition, 200
- adjunct site, 200
- adjunction, 5, 123
- AkKo, 464
- almost parse, 13
- ambiguous keyboard, 463
- anchor, 451
- approximate parsing, 337
- argument site, 199
- argument table, 42
- argument transition, 199
- assignment, 366
- auxiliary tree, 5, 123

- basic karaka frame, see under karaka frame, 357
- Baum-Welch algorithm, 452
- bipartite graph matching, 366
- Bootstrapping, 149

- Candito's system, 103
- CCG, 222

- CCGbank, 222
- chart parsing algorithm, 200
- Chomsky normal form, 115
- CLSG, 2
- Combinatory Categorical Grammar, 2, 196, 222
- combining labeled and unlabeled data, 151
- complex primitives, 1
- composite supertagger, 185
- Condition on Elementary Tree Minimality (CETM), 411
- constraint graph, 364
- Constraint-Based Lexicalist Theory, 16
- constraints, 358
- context-free grammar, 115
- context-free superset, 174
- context-free supertag, 181
- cost function, 366
- cotraining, 151

- Danish, 389
- Data Oriented Parsing, 15
- DATR system, 102
- decoration string, 175
- deduction rule, 116
- deep linguistic features, 434
- delexicalized
 - elementary tree, 126
 - nonterminal, 118
- demands, 359
- dependency coverage criterion, 467
- derivation
 - process, 413
 - trees, 45, 368
- derived tree, 123

- description selector, 89
- dialogue graph, 462
- discriminative learning, 162
- distributional similarity, 440
- domain of locality, 3
- Dutch examples, 292
- DVTAG, 410
- Earley
 - parser, 182
 - recognizer, 182
- Earley parser, 212
- Eiche, 318
 - HPSG treebank for German, 318
- elementary dependency tree, 198
- elementary tree, 123
- elementary trees, 8
- error detection in treebank annotation, 61
- factoring of recursion, 9
- finite transducer, 180
- foot node, 5, 123
- Forward-Backward algorithm, 452
- frame generator, 91
- free word order, 357
- Generative Dependency Grammar, 197
 - derivation, 199
- Good-Turing discounting, 453
- grammar
 - development, 75
 - extraction, 40
- greedy search, 455
- guided parsing, 182
- head, 125
- head automaton grammar, 130
- Head Grammars, 2
- head percolation table, 41
- head-corner parsing algorithm, 138
- Head-driven Phrase Structure Grammars (HPSG), 2
- Hidden Markov Model, 450
 - Supertagger, 211
- horizontal markovization, 197
- HPSG, 319
 - arg rules, 326
 - German Grammar, 320
 - head rules, 326
 - Head-driven Phrase Structure Grammar, 319
 - HPSG-DOP, 313
 - HPSG-Supertag, 325
 - human sentence processing, 407
 - HyTAG, 101
 - implementation of parsing algorithm, 138
 - increased domain of locality, 8
 - incrementality, 407
 - index, 115
 - initial tree, 8, 123
 - integer programming, 365
 - inverse adjoining, 416
 - from left, 416
 - from right, 416
 - inverse substitution, 414
 - Italian, 395
 - karaka, 363
 - chart, 357
 - frame, 357, 363
 - frame transformation, 357
 - level, 356
 - KoHDaS, 456
 - LDA, 440
 - lexical categories, 222
 - lexical element, 123
 - Lexical Functional Grammars, 2
 - lexicalization, 5
 - lexicalized
 - automaton, 198
 - context-free grammar, 117
 - grammars, 373
 - nonterminal, 120
 - tree-adjoining grammar (LTAG), 125
 - Lexicalized Tree-Adjoining Grammar, 8
 - LexOrg, 77
 - LexTract, 40
 - Lightweight Dependency Analyzer, 440, 456
 - Lightweight Dependency Analyzer (LDA), 193
 - Linear Indexed Grammars, 2
 - link grammar, 130
 - localizing arguments, 1
 - log-linear models, 224
 - long-distance dependencies, 9

- LTAG
 - derivation tree, 10
 - derived tree, 9
 - Lexicalized Tree-Adjoining Grammar, 8
 - Treebank grammar, 138
- LTIG, 317, 325
- mandatory karaka, 358
- match list reranking, 467
- matching, 367
- maximal matching, 367
- Maximum Entropy Supertagger, 211
- maximum entropy tagger, 224
- maximum likelihood estimation, 452
- meaning level, 356
- minimal construction, 9
- mixed associativity, 296
- mixed commutativity, 296
- modifier-modified trees, 368
- modifiers/adverbs, 382
- multitagging, 226
- n-best supertagging, 439
- negation, 384
- nonassociative Lambek calculus, 291
 - natural deduction rules, 293
- nonterminal, 116
- NP chunking, 159
- object shift, 384
- Paninian grammar, 357, 368
- parsed corpora, 418
- parsing efficiency, 148
- path backtracking, 454
- PCFG with latent annotations, 337
- PCFG-LA, 337
- Penn Treebank, 432
- Phonological Merge (PMerge), 378
- Phrasal Stress, 396
- polarity value, 457
- predicative auxiliary trees, 205
- Probabilistic Generative Dependency Grammars, 200
- probabilistic positional model, 202
- production, 115
- PropBank, 432
 - adjunct, 435
 - argument, 435
 - properties of German syntax, 318
 - prosodic hierarchy, 377
 - pseudo category, 452
- recurrent plausibility networks, 460
- Recursive Transition Networks, 197
- Redwoods Treebank, 318, 323
- regular superset, 174
- semantic
 - parsing, 441
 - role labeling, 437
 - roles, 439
- semantic computation, 11
- Semantic Merge, 376
- semantico-syntactic, 356
- sentence complexity, 143
- sentential arguments, 368
- set automaton, 179
 - mirror, 181
- shared parse forest, 200
- shift operation, 416
- significance vector, 459
- SLTG
 - stochastic lexicalized tree grammar, 313
- smoothing, 452
- Spanish, 393
- spell-out, 393
- split grammar, 130
- statement, 292
- statistical parsing, 149
- Strict Layer Hypothesis, 378
- strong connectivity hypothesis, 410
- strong lexicalization, 5
- Structural rules, 295
- SUBCAT list
 - scrambling of complements, 320
- subcategorization frame, 89
- substitution, 4, 123, 363, 414, 415
- suggestion list, 465
- supertag, 1, 435
- supertagger, 439
 - n-best, 455
- supertagging, 12, 439, 450
- supertags, 12
- surface level, 356
- surface syntactic features, 431
- Swedish, 384
- synchronous derivations, 376

syntactic lexical ambiguity, 141
Syntactic Lexicalization Hypothesis, 373
syntactico-semantic, 356

tabular parsing, 114
TAG, 368, 435
 extraction procedure, 435
 parser, 441
tagset table, 42
TAM, 357
 label, 357
task parameters, 460
terminal, 123
theta role, 363
transformation, 357
tree description, 78
 four classes, 78
tree frame, 435
tree generator, 83
Tree Insertion Grammars, 198
tree-adjoining grammar, 123, 435
 parsing, 138
treebank grammars, 35
trellis, 455
trigram supertagger, 13
type-lifting, 292

unary modality, 294
unknown events, 453
user-initiative dialogue system, 458

vertical markovization, 196
vibhakti for verbs, 356
vibhakti level, 356
Viterbi algorithm, 452

Waltz algorithm, 15
word groups, 459
yogyataa, 359