

This excerpt from

Foundations of Statistical Natural Language Processing.
Christopher D. Manning and Hinrich Schütze.
© 1999 The MIT Press.

is provided in screen-viewable form for personal use only by members of MIT CogNet.

Unauthorized use or dissemination of this information is expressly forbidden.

If you have any questions about this material, please contact cognetadmin@cognet.mit.edu.

Index

- χ^2 test, 169, 609
- σ -field, 40
- * as a marker of ungrammaticality, 9
- ? as a marker of marginal grammaticality, 9

- A* search, 442
- Abney (1991), 376, 414, 451
- Abney (1996a), 375, 376
- Abney (1996b), 34
- Abney (1997), 457
- absolute discounting, 216
- accuracy, 269, 577
 - in tagging, 371
- accusative, 85
- Ackley et al. (1985), 402
- active voice, 102
- ad-hoc retrieval problem, 530
- adding one, 202, 225
- additivity, 43
- adjective, 81, 87
- adjective phrases, 95
- adjunction, 107
- adjuncts, 103
- adnominal adjectives, 87
- adverb, 91
- adverbial nouns, 86
- agent, 101
- agglomerative clustering, 501
- agreement, 87

- Aho et al. (1986), 79
- algorithmic complexity, 506
- alignment, *see* text alignment and word alignment, 467, 470
- Allen (1995), 79, 402
- alphabet, 61
- Alshawi and Carter (1994), 261
- Alshawi et al. (1997), 492
- ambiguity, 110, 229, *see also* disambiguation and semantic similarity, 296
 - as a major difficulty for NLP, 17
 - of phrase attachment, 107
 - PP attachment, 278
 - widespread parsing ambiguity in language, 411
- analytic forms, 90
- anaphora resolution, 573
- anaphoric relations, 111
- anaphors, 85
- ancestor-free assumption (context-free grammars), 384
- Anderson (1983), 35
- Anderson (1990), 35
- annealing schedule, 479
- antonyms, 110
- Aone and McKee (1995), 313
- Appelt et al. (1993), 314
- Apresjan (1974), 258
- Apté et al. (1994), 579
- arc-emission HMM, 324

- Argamon et al. (1998), 608
- arguments, 103
- arrival vector, 477
- article, 87
- association, 185, *see also*
 - selectional association
 - in word alignment, 485
- attaching in left-corner parsing, 425
- attachment ambiguity, 107, 278, 312
 - indeterminacy, 286
 - information sources, 284
 - noun compounds, 286
- attributive adjectives, 87
- Atwell (1987), 344, 456
- AUTOCLASS, 514
- automata
 - and transformation-based learning, 367
- auxiliaries, 90
- average precision
 - interpolated, 536
 - uninterpolated, 535
- Baayen and Sproat (1996), 310
- back-off models, 219
- backpropagation, 603
- backward procedure, 328
- bag of words, 237
- bagging, 607
- bags, 495
- Bahl and Mercer (1976), 379
- Bahl et al. (1983), 198, 221, 338
- Baker (1975), 339, 379
- Baker (1979), 403
- balanced corpus, 19, 120
- Baldi and Brunak (1998), 340
- Barnbrook (1996), 146
- base form, 88
- baseline, 234
 - in tagging, *see* dumb tagger
- basic outcomes, 40
- Basili et al. (1996), 527
- Basili et al. (1997), 312
- Baum et al. (1970), 339
- Baum-Welch reestimation, *see*
 - forward-backward algorithm
- Bayes classifier, 236
- Bayes decision rule, 236
- Bayes error rate, 606
- Bayes optimal decision, 58
- Bayes' rule, *see also* Bayes' theorem, 236, 347, 473
- Bayes' theorem, 43, 44
- Bayesian decision theory, 57
- Bayesian model merging, 403
- Bayesian statistics, 54
- Bayesian updating, 54, 56
- bead, 468, *see also* word bead, 493
- beam search, 441
- Beeferman et al. (1997), 572
- bell curve, 52
- Bell et al. (1990), 222
- Benello et al. (1989), 370, 608
- Benson et al. (1993), 185
- Benson (1989), 184
- Berber Sardinha (1997), 572
- Berger et al. (1996), 591, 596, 597, 607
- Bernoulli trial, 45
- Berry and Young (1995), 565
- Berry et al. (1995), 565
- Berry (1992), 572
- best-first search, 441
- Bever (1970), 108
- bias, 34
- Biber et al. (1998), 146, 527
- Biber (1993), 146
- big oh notation, 506
- bigram, 31, 193
- bigram tagger, 346, 353
- bilingual corpus, 20
- bilingual dictionaries

- derivation from corpora, 484
- binary vectors, 298
- binomial distribution, 50, 274, 546
- bins, 192
- bitext, 466, 493
- bitext map, 477, 493
- Black et al. (1991), 456
- Black et al. (1993), 448, 449
- Black (1988), 244
- blending of parts of speech, 12
- Bod and Kaplan (1998), 457
- Bod et al. (1996), 457
- Bod (1995), 446, 448
- Bod (1996), 446
- Bod (1998), 446
- Boguraev and Briscoe (1989), 311, 314
- Boguraev and Pustejovsky (1995), 311, 312
- Boguraev (1993), 311
- Bonnema et al. (1997), 437
- Bonnema (1996), 437
- Bonzi and Liddy (1988), 571
- Bookstein and Swanson (1975), 548, 571
- Boolean queries, 530
- boosting, 607
- Booth and Thomson (1973), 403
- Booth (1969), 403
- Borthwick et al. (1998), 608
- boundary selector, 568
- Bourigault (1993), 312
- Box and Tiao (1973), 204
- bracketing, 98
- branching processes, 403
- Brants and Skut (1998), 376
- Brants (1998), 354
- Breiman et al. (1984), 241, 583
- Breiman (1994), 607
- Brent (1993), 271-274
- Brew (1995), 457
- Brill and Resnik (1994), 285, 312, 370, 457
- Brill et al. (1990), 371, 527
- Brill (1993a), 457
- Brill (1993b), 370
- Brill (1993c), 457
- Brill (1995a), 362, 365, 366, 370
- Brill (1995b), 365
- Briscoe and Carroll (1993), 428
- Britton (1992), 225
- Brown corpus, 19, 145
 - used for tagging, 378
- Brown et al. (1990), 485, 488, 489, 491
- Brown et al. (1991a), 239
- Brown et al. (1991b), 235, 239-241, 250, 253, 493
- Brown et al. (1991c), 470, 474, 478, 482, 493
- Brown et al. (1992a), 491
- Brown et al. (1992b), 80
- Brown et al. (1992c), 510-512, 528
- Brown et al. (1993), 486, 488-492
- Brown tag set, 139
- Bruce and Wiebe (1994), 261
- Bruce and Wiebe (1999), 591
- Brundage et al. (1992), 184
- Buckley et al. (1996), 566
- Buckshot, 518
- Buitelaar (1998), 258
- Burgess and Lund (1997), 303
- Burke et al. (1997), 377
- burstiness, 548
- c5 tag set, 140
- Canadian Hansards, 20
- canonical derivation, 422
- capacity, 69
- capitalization, in text processing, 123
- Caraballo and Charniak (1998), 409, 443

- Cardie (1997), 145, 376
 Carletta (1996), 234
 Carrasco and Oncina (1994), 456
 Carroll and Charniak (1992), 456
 Carroll (1994), 442
 case, 84
 categorical, 7
 categorical perception, 10
 categorical view of language, 11
 categories, 575
 categorization, 575
 center of gravity, 499
 centroid, 499
 chain rule, 43
 chaining effect, 504
 Chang and Chen (1997), 494
 channel probability, 71
 Chanod and Tapanainen (1995), 373
 Charniak et al. (1993), 343, 345, 352, 355, 372, 379
 Charniak (1993), 401, 403
 Charniak (1996), 145, 435, 443, 444, 455
 Charniak (1997a), 377, 455, 457
 Charniak (1997b), 456
 Cheeseman et al. (1988), 514
 Chelba and Jelinek (1998), 409, 457
 Chen and Chang (1998), 261
 Chen and Goodman (1996), 207, 211, 218, 221–225
 Chen and Goodman (1998), 211, 222–225
 Chen (1993), 470, 480, 485
 Chen (1995), 403
 Chi and Geman (1998), 388
 Child Language Data Exchange System, 119
 CHILDES, 119
 Chitrao and Grishman (1990), 456
 Chomsky Normal Form, 389
 Chomsky (1957), 2, 16, 377, 421
 Chomsky (1965), 6, 34
 Chomsky (1980), 34
 Chomsky (1986), 5, 34
 Chomsky (1995), 421
 Choueka and Lusignan (1985), 259
 Choueka (1988), 183
 chunking, 407, 414
 Church and Gale (1991a), 202, 203, 212, 217, 224, 225
 Church and Gale (1991b), 171, 179, 485
 Church and Gale (1995), 549
 Church and Hanks (1989), 167, 168, 178, 187
 Church and Liberman (1991), 147
 Church and Mercer (1993), 34, 169, 312
 Church and Patil (1982), 18, 287
 Church et al. (1991), 168, 178, 187
 Church (1988), 353, 355, 375, 378, 379, 452, 456
 Church (1993), 470, 475, 476
 Church (1995), 571
 Clark and Clark (1979), 379
 class-based generalization, 295
 classes, 575
 classical pattern in HMM training, 360
 classification, 232, 498, 575
 classificatory features, 192
 clause, 93
 CLAWS1, 140
 Cleverdon and Mills (1963), 571
 CLIR, 571
 clitic, 85
 in text processing, 126
 closed word class, 82
 clustering, 232, 495–527
 agglomerative, 501
 as unsupervised learning, 498
 disjunctive, 499
 divisive, 501

- flat, 498
- hard, 499
- hierarchical, 498
- non-hierarchical, 498
- of terms in documents, 548
- soft, 499
- clusters, 495
- co-activation of senses, 258
- co-occurrence, 185, 554
- Coates-Stephens (1993), 188, 313
- cognates, 475
- cognition
 - as a probabilistic phenomenon, 15, 34
- cohesion scorer, 567
- collection frequency, 542
- Collins and Brooks (1995), 312
- Collins (1996), 416, 451–455
- Collins (1997), 416, 451, 453–455, 457
- collocation, 29, 94, 111, 151–189
- comparative, 87
- competence
 - linguistic, 6
- competence grammar, 8
- complementizers, 93
- complements, 103
- complete-link clustering, 503
- compositionality, 110, 151
- compounding, 83
- compression, 68
- computational lexicography, 187
- concordances, 31
- conditional entropy, 63
- conditional independence, 43
- conditional probability, 42
- confusion matrix, 374
- connectionism, *see* neural networks
- connectionist models, 603
- constituent, 93
- context-free, 101
- context-free assumption, 384
 - weakening of, 416
- context-free grammars, 97
- context-group discrimination, 253
- contextual interchangeability, 296
- Contextual Theory of Meaning, 152
- continuous sample space, 40
- conventionality, 9
- coordinating conjunction, 92
- coordination, 92
- Copestake and Briscoe (1995), 258
- Cormen et al. (1990), 79, 122, 327, 440, 471, 506, 527
- corpora, 6, 118
- corpus, 6
- corpus linguistics, 146
- correspondence, 470
- cosine, 299, 507, 540
 - and Euclidean distance, 301
- Cottrell (1989), 261
- count-counts, 213
- covariance matrix, 520
- Cover and Thomas (1991), 66, 72, 76, 77, 80, 182
- Cowart (1997), 34
- credit assignment, 488
- Croft and Harper (1979), 571
- cross entropy, 74, 510
- cross-language information retrieval, 571
- cross-validation, 210, 585
- crossing brackets, 434
- crossing dependencies, 468
- Crowley et al. (1995), 145
- Crystal (1987), 113
- cutoff, 535
- Cutting et al. (1991), 337
- Cutting et al. (1992), 508, 527
- Cutting et al. (1993), 527
- Daelemans and van den Bosch (1996), 608
- Daelemans et al. (1996), 370, 608

- Dagan and Itai (1994), 242, 247
 Dagan et al. (1991), 247
 Dagan et al. (1993), 485
 Dagan et al. (1994), 260
 Dagan et al. (1997a), 607
 Dagan et al. (1997b), 260, 304–306
 Damerau (1993), 175
 Darroch and Ratcliff (1972), 593, 607
 data representation model, 495, 576
 data-oriented parsing, 446
 database merging, 530
 de Saussure (1962), 113
 decision theory, *see* Bayesian decision theory
 decision trees, 578
 and transformation-based learning, 365
 declaratives, 96
 decoding, 71, 326, *see also* stack decoding algorithm, *see also* Viterbi algorithm, *see also* noisy channel model, 487
 in parsing, 440
 Deerwester et al. (1990), 564, 565, 572
 degree adverbs, 91
 DeGroot (1975), 79
 deleted estimation, 210
 deleted interpolation, 218
 Della Pietra et al. (1997), 607
 Demers (1977), 424, 459
 demonstratives, 87
 Dempster et al. (1977), 523, 524
 dendrogram, 495
 denominal verbs, 379
 dependence of events, 43
 dependency grammar, 428, 456
 vs. phrase structure grammar, 428
 dependency of arguments and adjuncts, 101
 dependency-based models for parsing, 451
 depth scorer, 567
 derivation, 83
 derivational probability, 421
 Dermatas and Kokkinakis (1995), 337, 371
 DeRose (1988), 378, 379
 Derouault and Merialdo (1986), 379
 determiner, 87
 development test set, 207
 dialog modeling, 573
 Dice coefficient, 299
 dictionary, 82
 dictionary-based disambiguation, 241
 Dietterich (1998), 209
 digram, 193
 dimensionality reduction, 556
 Dini et al. (1998), 261, 370
 direct object, 102
 disambiguation, 229
 based on a second-language corpus, 247
 dictionary-based, 241
 in parsing, 408
 selectional preferences, 290
 supervised, 235
 thesaurus-based, 244, 261
 thesaurus-based, adaptive, 245
 unsupervised, 252
 discounting, 199
 absolute, 216
 linear, 216
 discourse analysis, 111, 572
 discourse segmentation, 566
 discrete sample space, 40
 disjoint sets, 41
 disjunctive clustering, 499
 distortion, 489

- distribution, 50
- distribution-free, 49
- distributional noun clustering, 513
- ditto tags, 130
- divisive clustering, 501
- document frequency, 542
- document space, 296
- Document Type Definition, 138
- Dolan (1994), 261
- Dolin (1998), 572
- domination in a parse tree, 383
- Domingos and Pazzani (1997), 237, 607
- Doran et al. (1994), 376
- Dorr and Olsen (1997), 314
- dot-plot, 476
- Dras and Johnson (1996), 186
- DTD, 138
- Duda and Hart (1973), 232, 236, 523, 606, 608
- Dumais (1995), 564
- dumb tagger, 372
- Dunning (1993), 172
- Dunning (1994), 227
- Durbin et al. (1998), 340
- dynamic programming, 326

- E* measure, 269
- early maximum pattern in HMM training, 360
- Eeg-Olofsson (1985), 379
- Egan et al. (1989), 570
- Eisner (1996), 457
- Ellis (1969), 403
- Elman (1990), 608
- ELRA, 119
- Elworthy (1994), 359, 360, 372
- EM algorithm
 - for disambiguation, 253, 525
 - for Gaussian mixtures, 520
 - for HMMs, 333
 - for PCFGs, 398

- for text alignment, 474, 482, 488
- in parsing, 444
- emission probability, 321
- empirical expectation, 590
- empiricist, 5
- empty cept, 487
- empty nodes, 100
- English Constraint Grammar, 373
- entropy, 61
 - conditional, 63
 - joint, 63
 - of English, 76, 80
- entropy rate, 66
- envelope, 479
- epsilon transitions, 324, 338
- ergodic model, 76, 338
- error, 269
- estimation, 49
- estimators
 - combination of, 217
 - statistical, 196
- Estoup (1916), 24
- Euclidean distance, 301
- European Language Resources Association, 119
- evaluation, 267
 - in information retrieval, 534
 - of parsing, 431
- Evans and Zhai (1996), 189
- Evans et al. (1991), 188
- event, 40
- event space, 40
- exact match, 432, 530
- example-based, 493
- expectation, 46, 182
- expectation step, 522
- expected frequency estimates, 202
- expected likelihood estimation, 204
- experiment, 40
- exploratory data analysis, 497

- F* measure, 269, 536

- Fagan (1987), 377, 571
 Fagan (1989), 188
 fallout, 270
 false negatives, 268
 false positives, 268
 Fano (1961), 178, 182
 features in maximum entropy modeling, 591
 feed forward model, 339
 fertility, 489
 Fillmore and Atkins (1994), 258
 filtering, 530
 final test set, 207
 Finch and Chater (1994), 307
 Finch (1993), 527
 finite state transducer, 367
 finite-state automata, 314
 Firth, 34
 Firth (1957), 6, 151
 Fisher (1922), 225
 flat clustering, 498
 Flip-Flop algorithm, 239
 Fontenelle et al. (1994), 182, 187, 189
 Ford et al. (1982), 284
 forward procedure, 327
 forward-backward algorithm, 333, 524
 Foster (1991), 379
 four-gram, 193
 Frakes and Baeza-Yates (1992), 122, 571
 Francis and Kučera (1964), 145
 Francis and Kučera (1982), 119, 145, 146, 343
 Franz (1995), 374, 375
 Franz (1996), 285, 352, 608
 Franz (1997), 285, 312, 352, 608
 Frazier (1978), 387, 417
 free word order, 96
 Freedman et al. (1998), 79
 frequentist statistics, 54
 Friedl (1997), 121
 Fu (1974), 456
 full-form lexicon, 133
 function words, 20, 533
 functional categories, 82
 Fung and Church (1994), 477
 Fung and McKeown (1994), 470, 477
 Gale and Church (1990a), 205
 Gale and Church (1990b), 205
 Gale and Church (1991), 471
 Gale and Church (1993), 470-472, 474, 478, 481-484, 493
 Gale and Church (1994), 225
 Gale and Sampson (1995), 212, 213, 225
 Gale et al. (1992a), 233, 234, 257
 Gale et al. (1992b), 235, 236, 238, 253, 260
 Gale et al. (1992c), 238
 Gale et al. (1992d), 493
 Gale et al. (1992e), 233
 Gallager (1968), 182
 gaps, 567
 garden paths, 108
 Garside and Leech (1987), 456
 Garside et al. (1987), 145, 146, 378
 Garside (1995), 146
 Gaussian, 53
 multivariate, 520
 Gaussier (1998), 485
 Ge et al. (1998), 573
 generalization, 497
 class-based, 295
 semantic similarity, 295
 generalized iterative scaling, 591
 generation of phrases, 107
 generative linguistics, 6
 geometric mean
 for computing the probability of a derivation, 442

- gerund, 89
Ghahramani (1994), 523
Gibson and Pearlmutter (1994), 285
global optimum, 603
Gold (1967), 386
Goldszmidt and Sahami (1998), 301
Golub and van Loan (1989), 561
Good-Turing estimator, 212
Good (1953), 212
Good (1979), 225
Goodman (1996), 456
gradient descent, 576, 597
grammar induction, 386, 407
grammatical categories, 81
grammatical tagging, *see* tagging
grammatical zero, 221
grammaticality, 8, 15, 34
grammaticalization, 34
graphic word, 125
Greenbaum (1993), 144
Greene and Rubin (1971), 343, 373, 378
Grefenstette and Tapanainen (1994), 147
Grefenstette (1992a), 303
Grefenstette (1992b), 303
Grefenstette (1994), 376
Grefenstette (1996), 298
Grefenstette (1998), 571
Grenander (1967), 403
group-average clustering, 507
Guthrie et al. (1991), 261
Guthrie et al. (1996), 260
Günter et al. (1996), 29

Halliday (1966), 152
Halliday (1994), 34
Hansards, 20, 466
hapax legomena, 22, 199
haplology, 125
hard clustering, 499
Harman (1996), 570

Harnad (1987), 34
Harris (1951), 6
Harris (1963), 403
Harris (1988), 493
Harrison et al. (1991), 456
Harter (1975), 571
Haruno and Yamazaki (1996), 470, 483
Hatzivassiloglou and McKeown (1993), 527
Hawthorne (1994), 189
head, 95, 106, 430
Hearst and Plaunt (1993), 567, 568
Hearst and Schütze (1995), 313
Hearst (1991), 261
Hearst (1992), 313
Hearst (1994), 567
Hearst (1997), 567, 568, 570
held out data, 207
held out estimator, 206
Henderson and Lane (1998), 608
Hermjakob and Mooney (1997), 437, 457
Hertz et al. (1991), 608
Herwijnen (1994), 146
Hickey (1993), 146
Hidden Markov model, 317, 320
 vs. Markov model (terminology), 351
 predictive power compared to PCFGs, 387
 tagging, 356
hierarchical clustering, 498
hill climbing, 576, 597
Hindle and Rooth (1993), 279, 280, 283-287, 294
Hindle (1990), 178, 299, 527
Hindle (1994), 376
Hirst (1987), 261
historical linguistics, 113
history
 to predict next word, 192

- history-based grammars, 423, 448,
 see also SPATTER
- HMM, *see* Hidden Markov model
- Hodges et al. (1996), 182
- Holmes et al. (1989), 291
- holonym, 110
- homographs, in text processing,
 129
- homonyms, 110
- homophony, 110
- Honavar and Slutzki (1998), 456
- Hopcroft and Ullman (1979), 121,
 402, 405, 422
- Hopper and Traugott (1993), 34
- Hornby (1974), 272, 276, 310
- Horning (1969), 387, 403
- Huang and Fu (1971), 403
- Huddleston (1984), 379
- Hull and Grefenstette (1998), 188,
 571
- Hull and Oard (1997), 571
- Hull et al. (1996), 608
- Hull (1996), 132, 571
- Hull (1998), 485
- Hutchins (1970), 403
- hypernymy, 110
- hyponym, 110
- hyphenation, 126
- hyponymy, 110
- hypothesis testing, 162-176
 subcategorization, 273

- ICAME, 119
- Ide and Véronis (1998), 260, 261
- Ide and Véronis (1995), 146
- Ide and Walker (1992), 311, 312
- identification in the limit, 386
- idf, *see* inverse document
 frequency
- idiom, 111
- imperatives, 96
- implicit object alternation, 292

- improper distribution (probabilistic
 grammars), 388
- inconsistent distribution
 (probabilistic grammars), 388
- independence, 43
- independence assumptions, 192
 in machine translation, 490
- indicator random variable, 45
- indirect object, 102
- induction, 7
- inductive bias, 34
- infinitive, 89
- inflection, 83
- information extraction, 112, 131,
 376
- information gain, 583
- information need, 532
- information radius, 304
- information retrieval, 529
 as an application for tagging, 377
- information sources
 in tagging, 343
 used in disambiguation, 259
- initial maximum pattern in HMM
 training, 360
- initialization
 effect of in HMM training, 359
 of the Forward-Backward
 algorithm, 339
- inside algorithm, 392
- inside probabilities, 392
- Inside-Outside algorithm, 398, 401,
 525
- interactions, 352
- interlingua, 465
- International Computer Archive of
 Modern English, 119
- interpolation, 536
 in tagging, 353
- interrogative determiners, 88
- interrogative pronouns, 88
- interrogatives, 96

- intransitive, 103
- Inui et al. (1997), 428
- invariance
 - of place (grammars), 384
 - of time (Markov models), 345
- inverse document frequency, 543, 551
- inversion, 96
- inverted index, 532
- irregular
 - nouns, 84
 - verbs, 89
- Isabelle (1987), 463
- iterative, 498

- Jaccard coefficient, 299
- Jacquemin et al. (1997), 314, 377
- Jacquemin (1994), 314
- Jain and Dubes (1988), 501, 507, 527
- Jeffreys-Perks law, 204
- Jeffreys (1948), 225
- Jelinek and Lafferty (1991), 403
- Jelinek and Mercer (1985), 206, 210
- Jelinek et al. (1975), 339
- Jelinek et al. (1990), 403
- Jelinek et al. (1992a), 403
- Jelinek et al. (1994), 416, 450
- Jelinek (1969), 440, 441
- Jelinek (1976), 339
- Jelinek (1985), 357, 379
- Jelinek (1990), 225, 323
- Jelinek (1997), 113, 225, 339, 607
- Jensen et al. (1993), 314
- Johansson et al. (1978), 145
- Johnson (1932), 225
- Johnson (1998), 437, 448
- joint distributions, 48
- joint entropy, 63
- Joos (1936), 35
- Jorgensen (1990), 234, 257
- Joshi (1993), 266

- Justeson and Katz (1991), 313
- Justeson and Katz (1995a), 259
- Justeson and Katz (1995b), 153, 155, 314

- K mixture, 549
- k* nearest neighbors, 295, 604, 608
 - for disambiguation, 260
 - in tagging, 370
- K-means, 515, 525
- Kahneman et al. (1982), 257
- Kan et al. (1998), 572
- Kaplan and Bresnan (1982), 457
- Karlsson et al. (1995), 373
- Karov and Edelman (1998), 261
- Karttunen (1986), 266
- Katz (1987), 219, 223, 225
- Katz (1996), 551
- Kaufman and Rousseeuw (1990), 500, 527
- Kaufmann (1998), 572
- Kay and Röscheisen (1993), 146, 470, 478-480, 483
- Kehler (1997), 573, 608
- Kelly and Stone (1975), 261
- Kempe (1997), 369, 371
- Kennedy (1998), 146
- Kent (1930), 35
- Key Word In Context, 31, 35
- Kilgarriff and Rose (1998), 171
- Kilgarriff (1993), 258
- Kilgarriff (1997), 228, 258
- Kirkpatrick et al. (1983), 402
- KL divergence
 - as a measure of selectional preference strength, 289
 - as a measure of semantic similarity, 304
- Klavans and Kan (1998), 571
- Klavans and Tzoukermann (1995), 486
- Klein and Simmons (1963), 378

- Kneser and Ney (1995), 223
 Knight and Graehl (1997), 493
 Knight and Hatzivassiloglou (1995), 493
 Knight et al. (1995), 493
 Knight (1997), 464, 492
 Knill and Young (1997), 339
 KNN, 295, *see k* nearest neighbors
 knowledge sources, 232
 Kohonen (1997), 527
 Korfhage (1997), 571
 Krenn and Samuelsson (1997), 79
 Krovetz (1991), 188
 Kruskal (1964a), 527
 Kruskal (1964b), 527
 Kullback-Leibler divergence, 72
 in clustering, 513
 Kupiec et al. (1995), 155, 571
 Kupiec (1991), 403
 Kupiec (1992a), 403
 Kupiec (1992b), 354, 357, 373, 374, 381
 Kupiec (1993a), 485, 488
 Kupiec (1993b), 377
 Kučera and Francis (1967), 125, 145
 KWIC, *see* Key Word In Context, 35
 Kwok and Chan (1998), 566
- L_1 norm, 304
 L_p norm, 308
 Lafferty et al. (1992), 456
 Lakoff (1987), 18, 35, 258
 Lancaster-Oslo-Bergen corpus, 20, 130, 145
 Landauer and Dumais (1997), 572
 Langacker (1987), 35
 Langacker (1991), 35
 language
 as a probabilistic phenomenon, 15, 34
 language change, as a non-categorical linguistic phenomenon, 13
 language model, 71, 191, 509
 vs. parsing model, 415
 improvement through clustering, 509
 in machine translation, 486
 Laplace's law, 202
 Laplace (1814), 202
 Laplace (1995), 202
 Lari and Young (1990), 387, 402, 403
 Lari and Young (1991), 403
 Latent Semantic Indexing, 554, 564, *see also* Singular Value Decomposition, 572
 lattice, 327
 Lau et al. (1993), 225, 608
 Lau (1994), 592, 596, 607
 Lauer (1995a), 286, 428
 Lauer (1995b), 34, 428
 LDC, 119
 Leacock et al. (1998), 259
 leaf node, 583
 learning, 498
 learning curves, 586
 least squares, 557
 Leaving-One-Out, 211
 left corner parser, 424
 lemma, 132
 lemmatization, 132
 length of a vector, 300
 length-based methods
 for text alignment, 471
 Lesk (1969), 303
 Lesk (1986), 242
 levels of recall, 536
 Levin (1993), 272, 292
 Levine et al. (1992), 314
 Levinson et al. (1983), 337, 339
 Lewis and Jones (1996), 188, 571

- Lewis and Ringuette (1994), 607
- Lewis et al. (1996), 607
- Lewis (1992), 530
- lexeme, 83, 128, 132
- lexical, 265
- lexical acquisition, 265
- lexical categories, 82
- lexical chains, 572
- lexical coverage of words in a corpus, 310
- lexical entries, 266
- lexical methods of sentence alignment, 478
- lexical resources, 19
- lexical semantics, 110
- lexicalization (of grammars), 417,
see also non-lexicalized
- lexicalized dependency-based language models, 453
- lexicography
 - computational, 187
- lexicon, 82, 97, 265
- Li and Abe (1995), 312
- Li and Abe (1996), 313
- Li and Abe (1998), 527
- Li (1992), 28
- Lidstone's law, 204, 216
- Lidstone (1920), 225
- light verbs, 186
- Light (1996), 314
- likelihood function, 198
- likelihood ratio, 57, 172, 279
- limited horizon, 318
 - in tagging, 345
- linear discounting, 216
- linear interpolation
 - general, 220
 - of *n*-gram estimates, 322
 - simple, 218
- linear regression, 557, 596
- linear successive abstraction, 222
- linearly separable, 600
- linguistic
 - competence, 6
 - performance, 6
- Linguistic Data Consortium, 119
- link grammar, 456
- Littlestone (1995), 608
- Littman et al. (1998a), 571
- Littman et al. (1998b), 571
- LOB, 20
- LOB corpus, *see*
 - Lancaster-Oslo-Bergen corpus
- local coherence (in clustering), 503
- local extension, 368
- local maxima, 335, 401
- local optimum, 603
- local tree, 97
- loglinear models, 590
 - for disambiguation, 260
- long-distance dependencies, 100
- Losee (1998), 571
- Lovins stemmer, 534
- Lovins (1968), 534
- lower bound, 234
- lowercase, in text processing, 123
- LSI, *see* Latent Semantic Indexing
- Luhn (1960), 35
- Lyons (1968), 2, 145
- MacDonald et al. (1994), 109
- machine-readable dictionaries
 - in lexical acquisition, 314
- MacKay and Peto (1990), 222
- macro-averaging, 577
- Magerman and Marcus (1991), 387, 442, 456
- Magerman and Weir (1992), 387
- Magerman (1994), 450, 584, 607
- Magerman (1995), 416, 450, 454, 455
- main effects, 352
- Mandelbrot's formula, 25
- Mandelbrot (1954), 25, 35

- Mandelbrot (1983), 35
 Manhattan norm, 304
 Mani and MacMillan (1995), 188
 Manning and Carpenter (1997), 427
 Manning (1993), 275, 276
 Marchand (1969), 113
 Marcus et al. (1993), 146, 456
 Marcus et al. (1994), 456
 marginal distributions, 48
 marginal probability, 56
 Markov assumption, 77, 193, 318
 Markov chains, 77
 Markov model, 317, 318
 vs. Hidden Markov model
 (terminology), 351
 taggers, 345
 vs. PCFGs, 381
 Markov (1913), 317
 markup, 123, 136
 schemes for, 137
 Marr (1982), 462
 Marshall (1987), 378
 Martin et al. (1987), 18
 Martin (1991), 313
 Masand et al. (1992), 608
 matching coefficient, 299
 maximization step, 522
 maximum, *see* local maxima
 maximum entropy distribution, 590
 maximum entropy modeling, 607
 maximum entropy models
 in tagging, 370
 maximum likelihood estimate, 54,
 197
 maximum likelihood estimation
 sequences vs. tag by tag in
 tagging, 355
 Maxwell (1992), 179
 McClelland et al. (1986), 608
 McCullagh and Nelder (1989), 591
 McDonald (1995), 188
 McEnery and Wilson (1996), 146
 McGrath (1997), 146
 McMahan and Smith (1996), 371
 McQueen and Burnard (1994), 146
 McRoy (1992), 261
 mean, 46, 158
 medoids, 516
 Mel'čuk (1988), 266
 Melamed (1997a), 494
 Melamed (1997b), 485
 memoization, 326
 memory-based learning, *see k*
 nearest neighbors
 for disambiguation, 260
 Mercer (1993), 309
 Merialdo (1994), 356, 359
 meronymy, 110
 Miclet and de la Higuera (1996), 456
 micro-averaging, 577
 Miikkulainen (1993), 608
 Mikheev (1998), 136, 608
 Miller and Charles (1991), 257, 296
 Miller et al. (1996), 458
 Minimum Description Length, 403,
 514
 minimum spanning tree, 504, 527
 Minsky and Papert (1969), 603
 Minsky and Papert (1988), 608
 Mitchell (1980), 34
 Mitchell (1997), 80, 237, 523, 607
 Mitra et al. (1997), 188
 mixture models, 218
 modal auxiliaries, 90
 model, 55
 model class, 576
 model merging, 354
 modifier space, 298
 Moffat and Zobel (1998), 571
 monotonic, 502
 Monte Carlo simulation, 447
 Mood et al. (1974), 174, 571
 Mooney (1996), 209, 259
 Moore and McCabe (1989), 79, 187

- morphological processes, 82
- morphology
 - in lexical acquisition, 313
 - in machine translation, 491
 - in text processing, 131
- Morris and Hirst (1991), 572
- Mosteller and Wallace (1984), 549
- multi-layer perceptrons, *see* neural networks
- multinomial distribution, 51
- multiplication rule, 42
- multiword units, *see* collocations
- mutual information, 66
 - for disambiguation, 239

- n*-best list, 333, 409
- n*-gram models, 76, 192, 317, 381, 491
- Nagao (1984), 493
- Naive Bayes, 237, 596, 607
- Naive Bayes assumption, 237
- natural law of succession, 217
- nearest neighbor classification rule, 604
- Neff et al. (1993), 309
- negative binomial, 549
- negative evidence, 386
- neural networks, 603, 608
 - in tagging, 370
- Nevill-Manning et al. (1997), 188
- New York Times* corpus, 153
- Newmeyer (1988), 113
- Ney and Essen (1993), 216, 225
- Ney et al. (1994), 216
- Ney et al. (1997), 211, 225
- Ng and Lee (1996), 260, 608
- Ng and Zelle (1997), 457
- Nie et al. (1998), 571
- Nießen et al. (1998), 492
- noisy channel model, 68, 486
- nominative, 85

- non-categorical phenomena in language, 11
- non-hierarchical clustering, 498, 514
- non-lexicalized treebank grammars, 443
- non-local dependencies, 99
 - in machine translation, 491
- non-parametric, 49
- non-textual data
 - in lexical acquisition, 313
- nonterminal nodes, 97
- normal distribution, 52, *see also* Gaussian, 609
- normality assumption
 - for SVD, 565
- normalization
 - of vectors, 301
- normalization factor, 56
- normalized correlation coefficient, *see also* cosine, 300, 541
- normalizing constant, 43
- noun, 81
- noun compounds, 286
- noun phrase, 95
- null hypothesis, 162
- null transitions, 338
- Nunberg and Zaenen (1992), 258
- Nunberg (1990), 145

- $O(n)$, 506
- Oaksford and Chater (1998), 35
- Oard and DeClaris (1996), 565
- object, 101
- object case, 85
- objective criterion, 432
- observable, 520
- OCR, 123
- odds of relevance, 551
- one sense per collocation constraint, 250

- one sense per discourse constraint, 249
- open word class, 82
- optimally efficient, 442
- order, of a Markov model, 320
- orthonormal, 560
- Ostler and Atkins (1992), 258
- OTA, 119
- outside algorithm, 394
- outside probabilities, 394
- overfitting, 582
- Overlap coefficient, 299
- overtraining, 206
- Oxford Text Archive, 119

- Paik et al. (1995), 188, 313
- Palmer and Hearst (1994), 135
- Palmer and Hearst (1997), 135, 260, 608
- paradigm, 94
- paradigmatic relationship, 94
- parallel texts, 20, 466
- parameter tying, 338
- parameters, 50
 - of an n -gram model, 193
- parametric, 49
- parse, 107
- parse triangle, 393
- parser, 408
- PARSEVAL measures, 432, 456
- parsing, 107, *see also* PCFG, *see also*
 - semantic parsing
 - dependency-based models, 451
 - evaluation, 431
 - partial parsing, 375
 - probability of a tree vs. a derivation, 421
- parsing model, 415
- part of speech, 81
 - concept in tagging, 344
 - notion of, 144
- part-of-speech tag, part-of-speech tagging, *see* tag, tagging
- partial parsing, 375, *see also*
 - chunking
- participle, 88
- particle, 91, 374
- partition, 44
- passive voice, 102
- past perfect, 89
- patient, 101
- Paul (1990), 340
- PCFG, 382
 - as a language model, 387
 - estimated from a treebank, 443
 - predictive power compared to HMMs, 387
 - training, 398
- Pearlmutter and MacDonald (1992), 417
- Pedersen and Bruce (1997), 261
- Pedersen (1996), 175, 189
- Penn Treebank, 20, 412
- Penn Treebank tag set, 140
- perceptron convergence theorem, 600
- perceptrons, 597, 608
- Pereira and Schabes (1992), 444, 445, 449
- Pereira et al. (1993), 261, 293, 304, 513, 527
- performance
 - linguistic, 6
- periphrastic forms, 87, 90
- perplexity, 78, 510
- phone numbers, in text processing, 130
- phrasal verbs, 91, 186
 - in text processing, 130
- phrase structure grammar
 - vs. dependency grammar, 428
- phrases, 93, 485
 - in information retrieval, 532

- Pinker (1994), 113
- place invariance, 384
- plural, 82
- pointwise entropy, 73
- pointwise mutual information, 68, 178
- Poisson distribution, 545
- Pollard and Sag (1994), 457
- polyseme, 110
- polysemy
 - systematic, 258
- Pook and Catlett (1988), 244
- Porter stemmer, 534
- Porter (1980), 534
- position information, 532
- positive, 87
- possessive pronouns, 85
- posterior probability, 42, 236
- postings, 532
- poverty of the stimulus, 5
- power laws, 28, *see also* Zipf's law
- Poznański and Sanfilippo (1995), 313
- PP attachment ambiguity, *see* ambiguity, PP attachment
- pragmatics, 112
- precision, 268, 432
- precision-recall curves, 536
- predicative, 87
- prefixes, 83
- preposition, 91
- prepositional phrases, 95
- present perfect, 89
- present tense, 88
- Press et al. (1988), 218
- preterminal, 396
- priming, 416
 - and semantic similarity, 303
- principal component analysis, 527, 556
- prior belief, 54
- prior probability, 42, 236
- Probabilistic Context Free Grammar, 382
- probabilistic left-corner grammars, 424
- probabilistic models
 - and transformation-based learning, 366
- probabilistic regular grammars, 389
- Probabilistic Tree Substitution Grammar, 448
- probability distribution, 40
- probability function, 40
- probability mass function, 45
- probability mass of rules, 388
- probability ranking principle, 538
- probability space, 41
- probability theory, 40
- Procter (1978), 244, 261
- productivity
 - of language as motivation for lexical acquisition, 309
- progressive, 89
- projecting in left-corner parsing, 425
- Prokosch (1933), 35
- pronoun, 85
- proper names, 86, 124, 186
- proper nouns, *see* proper names
- pruning, 582
- pseudo-feedback, 566
- pseudowords, 233
- punctuation, 145
- Pustejovsky et al. (1993), 311
- Pustejovsky (1991), 258
- Qiu and Frei (1993), 303
- qualifier, 91
- quantifier, 88
- query expansion
 - using semantic similarity, 295
- question answering, 377

- Quinlan (1986), 583
Quinlan (1993), 583
Quinlan (1996), 607
Quirk et al. (1985), 113
- Rabiner and Juang (1993), 113, 337, 339, 478
Rabiner (1989), 339
Ramsey and Schafer (1997), 187
Ramshaw and Marcus (1994), 366
random variable, 45
rank, 23
rare events, 199
Rasmussen (1992), 527
rationalist, 4
Ratnaparkhi et al. (1994), 285
Ratnaparkhi (1996), 370, 452, 607
Ratnaparkhi (1997a), 457
Ratnaparkhi (1997b), 607
Ratnaparkhi (1998), 285, 608
Read and Cressie (1988), 175
reallocating, 514
recall, 268, 434
recipient, 102
recomputation of means, 515
recursivity, 98
redundancy, 68
reflexive pronouns, 85
regular expressions, 120
regular language, 121
relative clauses, 95
relative entropy, 72
relative frequency, 49, 175, 197
relevance feedback, 530
reliability, 192
renormalization, 213
representative sample, 119
residual inverse document frequency, 553
Resnik and Hearst (1993), 285, 286
Resnik and Yarowsky (1998), 258
Resnik (1992), 457
Resnik (1993), 288
Resnik (1996), 288-294
Reuters, 579
reversibility
 in tagging, 355
rewrite rules, 96
Reynar and Ratnaparkhi (1997), 136, 607
R IDF, 553
Riley (1989), 134, 135
Riloff and Shepherd (1997), 313
Ristad and Thomas (1997), 354, 377
Ristad (1995), 217, 225
Ristad (1996), 596
Roark and Charniak (1998), 313
Robertson and Sparck Jones (1976), 571
ROC curve, 270
Rocchio (1971), 607
Roche and Schabes (1995), 367, 368
Roche and Schabes (1997), 314
Roget (1946), 244
root form, 83
Rosenblatt (1962), 603
Rosenfeld and Huang (1992), 220
Rosenfeld (1994), 225, 608
Rosenfeld (1996), 225, 608
Rosenkrantz and Lewis (1970), 424
Ross and Tukey (1975), 155
Roth (1998), 606
routing, 530
rules, 3
Rumelhart and McClelland (1986), 608
Rumelhart and Zipser (1985), 603
Rumelhart et al. (1986), 608
Russell and Norvig (1995), 442
- Sakakibara et al. (1994), 404
Salton and Allen (1993), 568
Salton and Buckley (1991), 572

- Salton and McGill (1983), 571
Salton and Thorpe (1962), 378
Salton et al. (1983), 308
Salton et al. (1994), 571
Salton (1971a), 303
Salton (1971b), 571
Salton (1989), 132
sample deviation, 159
sample space, 40
Sampson (1989), 310
Sampson (1995), 145
Sampson (1997), 34
Samuel et al. (1998), 573
Samuelsson and Voutilainen (1997),
372, 373
Samuelsson (1993), 352
Samuelsson (1996), 222
Sanderson and van Rijsbergen
(1998), 257
Sankoff (1971), 403
Santorini (1990), 146
Sapir (1921), 3, 150
Sato (1992), 493
Saund (1994), 499
scaling coefficients, 337
Schabes et al. (1988), 266
Schabes et al. (1993), 445
Schabes (1992), 457
Schapire et al. (1998), 607
Schmid (1994), 370, 607
Schütze and Pedersen (1995), 255,
259
Schütze and Pedersen (1997), 298,
572
Schütze and Singer (1994), 354
Schütze et al. (1995), 607, 608
Schütze (1992a), 233
Schütze (1992b), 303
Schütze (1993), 608
Schütze (1995), 307, 371, 527
Schütze (1996), 34
Schütze (1997), 35, 258
Schütze (1998), 253, 256, 261, 263
scope, 111
selectional association, 289
selectional preference strength, 289
selectional preferences, 105, 288,
312
selectional restrictions, 18, 105,
288
self-information, 61
semantic domain, 296
semantic hierarchies, automatic
acquisition, 313
semantic parsing, 457
semantic roles, 101
semantic similarity, 295, *see also*
topical similarity
semantic transfer approach, 465
sense tagging, 253
senses, 110, 229
Senseval, 259
sentence boundary identification,
260
in text processing, 134
sentences
distribution of lengths in a
corpus, 136
SGML, 137
shallow parsing, *see* partial parsing
Shannon game, 191
Shannon (1948), 60
Shannon (1951), 191
Shemtov (1993), 472, 493
Sheridan and Smeaton (1992), 307,
571
Sheridan et al. (1997), 571
shifting in left-corner parsing, 425
Shimohata et al. (1997), 189
Siegel and Castellan (1988), 79, 234
significance level, 163
Silverstein and Pedersen (1997),
527
Sima'an et al. (1994), 446

- Sima'an (1996), 447
 Simard and Plamondon (1996), 482
 Simard et al. (1992), 475
 Sinclair (1995), 187
 Singhal et al. (1996), 571
 single-link clustering, 503
 singular, 82
 singular value decomposition, 558,
 572
 normality assumption, 565
 sink state, 390
 Sipser (1996), 121
 Siskind (1996), 313
 skewed distribution of senses, 257
 Skut and Brants (1998), 376, 608
 Smadja and McKeown (1990), 162
 Smadja et al. (1996), 188
 Smadja (1993), 162, 177, 188
 Smeaton (1992), 377, 571
 Smith and Cleary (1997), 457
 smoothing, 199
 in tagging, 354
 Snedecor and Cochran (1989), 172,
 187, 209
 sociolinguistics, 113
 soft clustering, 499
 Sparck Jones and Willett (1998),
 571
 Sparck Jones (1972), 571
 sparse data problem, *see also* rare
 events
 in machine translation, 491
 SPATTER, 450
 specifier, 106
 speech corpora, 131
 speech recognition, 457
 splitting criterion, 582
 Sproat et al. (1996), 314
 Sproat (1992), 146
 St. Laurent (1998), 146
 stack decoding algorithm, 440
 standard deviation, 47
 Standard Generalized Markup
 Language, 137
 standard normal distribution, 53
 Stanfill and Waltz (1986), 608
 start symbol, 96
 state-emission HMM, 324
 stationary, 76, 318
 statistical estimators, 196
 statistical inference, 191
 Statistical Natural Language
 Processing, xxxi
 Steier and Belew (1993), 188
 stemming, 132, 194, 534
 stochastic process, 45
 Stolcke and Omohundro (1993),
 340
 Stolcke and Omohundro (1994a),
 354
 Stolcke and Omohundro (1994b),
 403
 Stolcke et al. (1998), 573
 Stolcke (1995), 403
 Stolz et al. (1965), 378
 stop list, 533
 stopping criterion, 582
 Strang (1988), 79
 strongly equivalent, 389
 structuralists, American, 6
 Strzalkowski (1995), 188, 377, 571
 Stubbs (1996), 34, 146, 152, 187
 Student's *t* test, *see t* test
 subcategorization, 104
 subcategorization frame, 105, 271,
 313
 subcategorize for, 271
 subject, 101
 subject case, 85
 subject-verb agreement, 99
 subordinate clauses, 104
 subordinating conjunction, 93
 substitution test, 81
 subtopic boundaries

- in text segmentation, 567
- suffixes, 83
- superlative, 87
- supervised disambiguation, 235
- supervised learning, 232
- Suppes et al. (1996), 313
- Suppes (1970), 403
- Suppes (1984), 35
- surprise, 73
- Susanne corpus, 20
- SVD, *see* singular value decomposition
- synecdoche, 230
- synonyms, 110
- synset, 20
- syntactic ambiguity, 107
- syntactic categories, 81
- syntactic transfer approach, 464
- syntagma, 94
- syntagmatic information for tagging, 343
- syntagmatic relationship, 94
- syntax, 93
- synthetic forms, 89
- systematic polysemy, 258

- t* test, 163, 187, 609
 - for system comparison, 209
- tableau, 439
- Tabor (1994), 34
- tag sets, 139, 146
 - design of, 144
 - effect on accuracy in tagging, 372
- tagging, 231, 341
 - accuracy, 371
 - applications, 374
 - baseline, *see* dumb tagger
 - early work, 377
 - Hidden Markov model taggers, 356
 - interpolation of models, 353
 - maximum likelihood: sequences
 - vs. tag by tag, 355
 - overview of tag sets, 139
 - reversibility of Markov chains, 355
 - sense tagging, 253
 - smoothing of estimates, 354
 - transformation-based learning, 361
 - trigram tagger, 346, 353
 - unknown words, 351
 - variable memory models, 353
- TAGGIT, 378
- tags, 82
 - blending of parts of speech, 12
- Tague-Sutcliffe (1992), 571
- Talmy (1985), 465
- Tanenhaus and Trueswell (1995), 109, 417
- Tanimoto coefficient, 299
- target feature, 192
- technical term, 152
- term, 152, 540
- term clustering, 548
- term distribution models, 544, 571
- term frequency, 542
- term weighting, 540, 541
- terminal nodes, 97
- terminological
 - expression, 186
 - phrase, 152
- terminology databases
 - derivation from corpora, 484
- terminology extraction, 152
- Tesnière (1959), 456
- test data, 206
- test set, *see also* test data, 577
 - development, 207
 - final, 207
- text alignment, 466
 - using a small bilingual dictionary, 477

- text categorization, 530, 575
- text corpora, 118
- text segmentation, 566, 572
- TextTiling, 567
- tf.idf, 543
- thesaurus-based disambiguation,
 - see disambiguation
- tied arcs, 338
- tied states, 338
- time invariance, 318
 - in tagging, 345
- token sequences, 567
- tokenization, 124
- tokens, 22, 124
- Tomita (1991), 428
- top-down clustering, 512
- top-down parsing, 424
- topic, 296
- topic-independent distinctions, 247
- topical similarity, 298
- Towell and Voorhees (1998), 259
- training, 333
- training data, 206, 333, 345
 - sensitivity to in machine translation, 490
 - sensitivity to in tagging, 372
- training procedure, 576
- training set, *see also* training data, 575
- transformation-based learning, 361
 - and automata, 367
 - and decision trees, 365
 - and probabilistic models, 366
 - for disambiguation, 261
 - in parsing, 457
- transitive, 103
- translation probabilities, 488
- translation probability, 487
- transliteration, 493
- transpose, 297
- Trask (1993), 113, 265
- tree
 - for representing phrase structure, 97
 - tree accuracy, 432
 - tree probability, 421
 - Tree-Adjoining Grammars, 457
 - treebank, 412
 - to estimate a PCFG, 443
 - trellis, 327
 - trial, 40
 - triangle inequality, 72
 - trigram, 193
 - trigram tagger, 353
 - true negatives, 268
 - true positives, 268
 - Twenty Questions, 62
 - two-Poisson Model, 548
 - Type I errors, 268
 - Type II errors, 268
 - types, 22
 - ungrammatical, 109
 - uniform distribution, 41
 - uniform-cost search, 440
 - unknown words in tagging, 351
 - unobservable, 520
 - unspecified object alternation, *see* implicit object alternation
 - unsupervised learning, 232
 - upper bound, 233
 - uppercase, in text processing, 123
 - use theory of meaning, 17
 - validation, 207, *see also* cross-validation, 584
 - validation data, 207
 - validation set, 584
 - van Halteren et al. (1998), 608
 - van Riemsdijk and Williams (1986), 10
 - van Rijsbergen (1979), 269, 298, 503, 527, 538, 570, 571
 - variable memory
 - in tagging, 353

- variance, 47, 158, 539
 - of system performance, 208
- vector length, 300
- vector space, 300
 - binary vectors, 298
 - for representing documents, 296
 - for representing head nouns, 297
 - for representing words, 296
 - vector space model, 539
- Velardi and Pazienza (1989), 313
- verb, 81
- verb group, 90
- verb particle constructions, 186
- verb phrase, 95
- Viegas et al. (1996), 314
- Visible Markov Models, 317
- Viterbi algorithm, 332
 - for PCFGs, 396
 - in parsing, 439
 - in tagging, 349
- Viterbi translation, 492
- Viterbi (1967), 339
- Vogel et al. (1996), 485
- Voutilainen (1995), 314, 373

- Waibel and Lee (1990), 113
- Walker and Amsler (1986), 311
- Walker and Moore (1997), 572
- Walker et al. (1998), 573
- Walker (1987), 244
- Wang and Waibel (1997), 492
- Wang and Waibel (1998), 492
- Waterman (1995), 527
- weakly equivalent, 389
- Weaver (1955), 462
- website, xxxvii
- Webster and Marcus (1989), 313
- Weinberg and Goldberg (1990), 187
- Weischedel et al. (1993), 352, 353
- wh-extraction, 100
- whitespace, 125
 - not indicating a word break, 129

- Wiener et al. (1995), 608
- Wilks and Stevenson (1998), 261
- Willett and Winterman (1986), 299
- Willett (1988), 527
- Witten and Bell (1991), 222
- Witten-Bell smoothing, 222
- Wittgenstein (1968), 2, 17
- Wong and Yao (1992), 572
- Wood (1993), 266
- Woolf (1973), 229
- word
 - graphic word, 125
 - in text processing, 124
- word alignment, 484
- word bead, 481
- word counts, 20
- word for word translation, 463
- word lattice, 409
- word order, 93
- word segmentation, 129
- word sense
 - discussion of concept, 256
- word space, 296
- word tokens, 21
- word types, 21
- WordNet, 20
- Wu and Wong (1998), 492
- Wu (1994), 470, 474
- Wu (1995), 485
- Wu (1996), 492

- X' theory, 106
- XML, 138

- Yamamoto and Church (1998), 572
- Yang (1994), 608
- Yang (1995), 608
- Yang (1998), 607
- Yarowsky (1992), 242, 245, 247, 261
- Yarowsky (1994), 251, 607
- Yarowsky (1995), 249, 250, 262

Youmans (1991), 310, 568
Younger (1967), 405

z score, 189
Zavrel and Daelemans (1997), 260,
447, 608
Zavrel et al. (1997), 285, 608
Zernik (1991a), 312
Zernik (1991b), 261
Zipf's law, 23, 199, 544
for intervals, 28
for senses, 27
Zipf (1929), 35
Zipf (1935), 35
Zipf (1949), 23, 35

This excerpt from

Foundations of Statistical Natural Language Processing.
Christopher D. Manning and Hinrich Schütze.
© 1999 The MIT Press.

is provided in screen-viewable form for personal use only by members of MIT CogNet.

Unauthorized use or dissemination of this information is expressly forbidden.

If you have any questions about this material, please contact cognetadmin@cognet.mit.edu.