

## Index

- 100-split task, 300–308
- A Linguistic Atlas of Early Middle English, 263  
 A Linguistic Atlas of Late Mediaeval English, 263  
 A Representative Corpus of Historical English Registers, 120–121, 129
- absolute vs. relative judgments, 99  
 abstract styles/registers, 433, 439, 451  
 abstractness vs. concreteness, 298  
 accents, 231–241, 421–429  
   nuclear accents, 254  
   phrase accents, 255  
   pitch accents, 253–255, 257
- acceptability, 99–117, *see also* grammaticality ratings, 22, 76, 78, 82–92, 114, 116  
 accessibility, 298, *see also* givenness  
   noun phrase ~ hierarchy, 321–323  
*accommodate*, ‘intransitive’ use, 142  
 accountability, principle of, 21, 382, 384–385  
 acoustic analysis, 232–236, 239–241, 244–248, 252, 255–259
- active vs. passive, 280, 419  
 AD *see* American Drama  
 adjectives, 402, 406, 409–410, 413, 417  
   attributive vs. predicative, 127, 130–134, 402, 406  
   modifying noun phrase heads, 202–204  
   negation, 130–133
- adverbs, 319, 328–329, *see also* intensifiers  
 African American English (AAE), 19  
 age, 1, 36, 45, 51, 69–70, 74–75, 79–80, 82, 91, 195, 337–338
- agent vs. patient, 213  
 aggregate data analysis, 433–451  
 agreement, 319, 452, 454  
*ain't*, 317, 319–320, 329, 331, 441, 452, 454  
 alliteration, 260–261, 267, 269–271, 274–275  
 American Drama, 122, 132, 144, 148  
 American English vs. British English, 137, 154–156  
   in the OED, 154–156  
   innovations, 154
- anacoluthic constructions, 192  
 anacrusis *see* intonation analysis, British School  
 Analysis of Variance (ANOVA), 111–112, 411
- analyticity, 314, 323–327, 433, 435–439,  
   *see also* Greenberg’s analyticity  
   -syntheticity metric
- angloversals, 328–329  
 animacy, 215–216, 219–225, 280–293, 297–298, 300–301, 305, 322, *see also* individuation
- anthropology, 19, 413–414, 424  
   anthropological linguistics, 7–8, 22  
 aphesis, 263  
 APiCS *see* Atlas of Pidgin and Creole Structures  
 apparent-time change, 74–75, 339  
 archaeology, 56, 414, 424  
 ARCHER *see* A Representative Corpus of Historical English Registers
- areoversals, 327–330  
 arithmetic mean, 83, 85, 87, 90, 371, 442  
 articulatory phonetic methods, 9  
   description, 229–236  
   measurements, 228–243
- association plot, 370  
*assuming* (conj.), 150  
 as-the-crow flies distance *see* distance  
 Atlas of Pidgin and Creole Structures, 315  
*attest* + direct vs. prepositional objects, 154  
 auditory phonetic methods, 9, 228–236  
 Australian English, 163, 232, 256, 297, 310  
 auxiliaries, 162–164, 326, 329, 455,  
   *see also* modals  
   *do/have*, 437, 439
- averaging of averages, 338, 342, 347–348, 357–358  
 a-verse, 269–270
- back-clipping, 264  
 Bangladeshi English, 174  
 Bayesian approach, 338, 342–343, 347–358  
*be like*, quotative, 382, 384–385  
 Bing Maps, Microsoft, 58  
 B-LOB Corpus, 183, 189, 191  
 BNC, BNCweb *see* British National Corpus  
 Boolean operators *see* corpus retrieval  
 bootstrap approach, 338, 348, 352, 357  
 boundary tones *see* intonation analysis, AM tradition
- box plots, 372–373  
 breaks, phrasal vs. metrical, 273, 276

- British National Corpus, 148–151, 160, 162, 166, 182, 196, 199–210, 218, 221, 226, 296, 435, 437–439
- Brown Corpus, 155, 169, 182–194, 363
- but*, sentence-final, 82, 88, 90
- b-verse, 269–271
- by-agent* *see* passive
- cardinal vowels, 231
- CEEC *see* Corpus of Early English Correspondence
- cellular automata, 61
- Chadwyck-Healey collections *see* AD, EAF, EEPF, ECF, EPD, LION, NCF
- change *see* diachronic variation and
- chi-square test, 368–369
- Chomskyan linguistics, 2–3, 451
- CLAWS, 196, 199–200, *see also* corpora, tag-sets
- cleft constructions, 395–396, *see also* topicalization, *it*-cleft constructions, *wh*-cleft constructions
- clipping *see* back-clipping
- clique analysis, 43
- clitic groups, 265–266, 276
- cluster analysis *see* hierarchical agglomerative cluster analysis
- COBUILD *see* Collins Birmingham University International Database
- COCA *see* Corpus of Contemporary American English
- code-switching, 30
- cognates, 423
- cognitive linguistics, 7–8, 182, 362
- cognitive variables, 45, 47
- Collins Birmingham University International Database, 166
- community of practice (CofP), 44, 46–49, 51, 332
- complementation, 189–190, 210, 319, 452, 455 complement clauses, 406–407
- complexity principle, 139–141, 157
- compounds, 139, 265, 268–269, 276, 324, *see also* syntheticity stress, 265
- computer mapping *see* dialect maps
- concordancing software *see* corpus retrieval
- concrete styles/registers, 451
- concreteness vs. abstractness, 298
- conditional clauses, 415
- conditional inference trees, 399
- confidence intervals, 224, 343, 350–351
- conjunctions, 144, 150, 154, 192, 203, 436 coordinatng, 202–203 subordinating, 138, 144
- consonants, 228–229, 237–242, 423, 425, 428–429, *see also* /d/, /h/, /r/, /t/ articulatory features, 241–242 cluster simplification, 390, 392 gemination, 261 nasality, 423 shortening, 261 weakening, 261
- constituent order, 110, 213, 280, 319, 327, 455, *see also* SVC order, word order
- constraints *see* factors, variables constraint ranking, 388, 390–398
- contact linguistics, 313
- contact varieties, 69, 82, *see also* L1, L2
- content-focused styles/registers, 404, 414–417
- context, 181, 184, 187, 194, 215, 219–224, 382, 384–389, 393–400 phonological, 21, 392
- Continuous Lexical Decision (CLD) task, 306, 311
- contractions, 92, 149, 173, 266, 406, 410, 433 *isn't it*, 149
- co-occurrence patterns, 402–403, 407, 409, 419–420
- Corel Quattro, 58
- corpora, *see also* corpus retrieval annotation, 195–196 contemporary, 7 corpus linguistics, 9–10, 158, 177, 182, 193, 213–214, 224 customized grammatical tagger, 406–407 historical/diachronic, 7, 10, 119, 138, 276 meta-information, 161, 175–177, 195, 219, 226 parsing, 120–121, 134, 280 part-of-speech (POS) tagging, 120–121, 133–134, 174, 195–196, 199, 208–209, 211, 225, 406–407, 437 qualitative vs. quantitative analysis, 2, 4–5, 181, 212, 313 reference corpora, 119–121, 123, 126, 158–159, 163, 176–177, 183 representativeness, 177, 414, *see also* sampling size, 159, 166, 177 spoken corpora, 190–191 disfluencies, 197 sound-text alignment, 182 transcription of spoken data, 182 stucture, 177 tag-sets, 196, 199–200
- Corpus of Contemporary American English, 159
- Corpus of Early English Correspondence, 338–340, 343, 347, 360
- Corpus of Scottish Correspondence, 339
- corpus retrieval, 124–125, 134, 137–139, 174, 177, 195–197, 203–204, 206–208, 212 categorization and coding of hits, 212–226 Key Word In Context (KWIC), 122, 125 orthographic variants, 124 post-editing, 160, 165–167, 170–171, 175, 211 database approach, 218–219, 221, 225–226 in-corpus-tool approach, 218, 226 proximity searches, 124 search syntax, 124–125, 196, 198–199, 205, 211

- tag-based searches, 196, 198, 206, 208  
 wildcards, 124, 137
- correlations, 338, 362, 375–379, 438–439, 445  
 correlation coefficient, 370, 375, 377, 379,  
 444–445, 447
- co-text, 219
- counterbalancing in stimulus design, 108–111
- Cramer's *V* (correlation coefficient), 370–371
- creoles, 315, 318, 329–330, 398
- CSC *see* Corpus of Scottish Correspondence
- /d/*, variable, 389–393, 396, 398
- data sniffing vs. data testing, 159, 169
- data synthesis *see* aggregate data analysis
- dative alternation, 292, 295–311, 419
- DCPSE *see* Diachronic Corpus of Present-Day  
 Spoken English
- declarative clauses, 185
- deductive approaches, 4, 214
- definiteness, 216, 219, 221–223, 225, 284,  
 288–289, 298–299, 301, 305, 308
- deictic elements, 215
- Delaunay Triangulation, 65–66
- dendrogram, 448–449
- density estimation, 64
- dependent variables, 45, 379, 445
- derivation *see* syntheticity
- descriptive linguistics, 166
- desirability effects, 75, 81, 91
- determiners, 127, 133, 200, 203, 289, 437, 452  
 co-occurrence restriction, 289
- Diachronic Corpus of Present-Day Spoken  
 English, 183–194
- diachronic variation and change, 1–2, 9–12, 36,  
 51–52, 70, 83, 90, 156, 182, 212, 244,  
 332, 338, 362, 388, 398–400  
 apparent-time differences, 74–75, 339  
 change in genre-specific expectations, 193  
 in phonology, 9–10, 228–259  
 real-time differences, 74, 183–184, 188–192, 339  
 short-term drifts, 451  
*s*-shaped trajectory, 193  
 tempo of change, 191
- diacritics, 231, 240–241
- dialect atlases *see* dialect maps
- dialect geography, 53–54
- dialect maps, 9, 22, 53–68  
 tools, 56–67
- dialectology, 1, 9, 55, 313–318, 331–332
- dialectometry, 55, 57, 65–67, 433, 440, 442
- dialects, 53–56, 68, 76, 317, 321–322, 324, 328,  
 394, 397–398, 443, *see also* geographical  
 variation
- difficulty, have ~ (in) + gerund*, 150
- dimensions of variation, 402–420, 438–439  
 reduction technique, 435  
 scores, 403, 410–412, 415
- discourse, 181, 187–188, 191, 193–194, 382–383,  
 385–386, 397, 399–400, 404, 410, 414, 417  
 context *see* context  
 markers, 19, 73, 80, 92  
 organization, 319, 420, 455
- discriminant analysis, 64
- dispersion, measure of, 374
- distance, 57, 63, 66, 68, 421–431, 439–451  
 as-the-crow-flies, 442, 444–445, 447  
 Euclidean, 57, 439, 441–444, 446–448  
 geographic, 57, 439, 442–445  
 least-cost travel time, 442–445, 451  
 linguistic, 57, 68, 421, 435, 445  
 morphosyntactic, 442–445, 447  
 statistical, 439
- distractors in stimulus design, 105, 109
- do*, pro-verb, 185
- do* support, 85–87, 382
- do the V-ing*, 152
- don't* in the third person singular, 317, 441
- double object constructions *see* dative alternation
- doublet forms, 262, 272–274
- dual pathway program *see* NeighborNet
- dwarf*, plural form, 143
- EAF *see* Early American Fiction
- Early American Fiction, 122, 132, 144, 148
- Early English Prose Fiction, 122–124, 126–129,  
 133, 144, 148, 152
- ECF *see* Eighteenth-Century Fiction
- EPPF *see* Early English Prose Fiction
- effects, fixed vs. random *see* mixed effects models
- Eighteenth-Century Fiction, 122–123, 127,  
 129–133, 144, 148, 152, 153
- Electromagnetic Mid-sagittal Articulography  
 (EMA), 236–237
- electropalatograph, 241
- elicitation of data, 278–293
- elision, 267–268
- emic perspective, 37
- empirical cycle, 5, 70
- empiricism, 2–5
- English Prose Drama, 122–123, 127, 131–133,  
 144, 148
- entrenchment, 105, 117, 144, 147, 165, 273
- envelope game, 40–44, 50
- Environmental Systems Research Institute, 58, 61
- EPD *see* English Prose Drama
- ESRI *see* Environmental Systems Research Institute
- estimates and their uncertainties, 341–350
- eta-square parameter, 112
- ethical considerations, 7, 19–21, 34, 70–71
- ethnicity, 33, 69, 74, 80, 125, 338
- ethnography, 37–39, 44, 51, 70, 399
- ethnolect, 74
- etic perspective, 37
- Euclidean distance *see* distance

- eWAVE, 317–320  
*excess, in ~ of + measure phrase*, 149  
 existential constructions, 329, 394, 396–398, 406, 436–438, 454  
 experimenter bias, 301  
 experiments, 2–3, 6–10, 99–105, 107–111, 301  
   artificiality of experimental data, 282  
   psycholinguistic standards, 100, 102, 109  
   stimulus design, 102, 105, *see also* stimuli  
 explicitness, 139–142, 149–152, 157
- f4 (software program), 21  
 factors, 1, 278–293, 382–420, *see also* cognitive variables, linguistic variables, social variables  
   factor analysis, 402–404, 407–409, 415  
   factor groups, 45–46, 396  
   factor interaction/correlation, 278–293, 382, 387  
   factor loadings, 408–410, 417  
   factor weight, 48, 50, 396, 398  
   operationalization, 283–284  
   relative strength, 395–396, 398, 445  
*fear, for ~ that*, 151  
 feet, metrical, 260–261, 267–269, 275  
 fiction databases *see* literature databases  
 fieldwork, 17–34, 37–39  
 fillers in stimulus design, 105, 109–110, 112–116  
 fixed-character displays, 58–61  
 F-LOB *see* Freiburg-LOB Corpus of British English  
 form/function asymmetry, 385–387  
 formality, 6, 72–74, 117, 149–152, 156, *see also* style  
 formants, 233–236, 247  
 formulaic language, 410  
 FoxBase database program, 58  
 FRED *see* Freiburg English Dialect Corpus  
 Freiburg English Dialect Corpus, 313, 440–442  
 Freiburg school of linguistics, 11  
 Freiburg-Brown corpus of American English, 155, 183–194  
 Freiburg-LOB Corpus of British English, 155, 183–194, 216–217, 219–223, 363  
 frequencies, 337–338, 341–359, 362–363, 367–371, 375–377, 387, 393, 410, 433, 435–445 *see under* quantitative analysis  
   expected/observed ~, 368–371  
   normalized ~, 146, 156, 377, 436–437, 441  
 friend-of-a-friend approach, 25, 70  
 friendship group, 39–52  
 fronting *see* topicalization  
 frontness of vowels, 423  
 Frown *see* Freiburg-Brown corpus of American English  
 functionalism, 313  
   functionalist explanations, 280  
   integrative ~, 313  
 future, different variants, 386
- gatekeepers, 38  
 gender *see* sex  
 gender diffusion, 321  
 generative linguistics, 3, *see also* theoretical linguistics  
 genitives, 278–293, 437, 439, 451, 453  
   's-genitive, 209, 278–280, 282–291, 437, 439  
   in Low German, 292  
   of-genitive, 278–279, 283–285, 287–291, 439  
 genre, 121, 123, 128–130, 133–134  
 geocoding, 56  
 Geographic Information Systems (GIS), 56–58, 61–68  
 geographical variation, 1, 9, 167, 191, 429, 431, *see also* dialects, varieties  
 geography *see* linguistic geographical, dialect geographical, word geographical  
 geolinguistics *see* dialectometry  
 gerunds, 140, 146, 150, 152, 165, 187, 339–340, 406  
   as prepositional complements, 339  
   modification of ~, 340  
   nominal vs. verbal uses, 340  
   objects of ~, 339  
 givenness, 215, 279, *see also* accessibility  
 Google Maps, 58–59, 67, 442  
 Google searches, 164, 171  
*got vs. gotten*, 153–154  
 grammatical markers, free vs. bound, 435–436, 439  
 grammaticality, 104–105, 299, *see also* acceptability  
   degrees of ~/ungrammaticality, 102, 104, 117  
   index, 324  
   judgments of ~, 295–296, 311  
 grammaticalization, 138, 330, 383, 399  
 graphic plotter grids, 59–60  
 graphical computer displays, 61–67  
 graphical exploration of data, 112  
 graphical user interface (GUI) *see* R (software)  
 GRASS GIS, 58  
 gravity, linguistic, 439, 442–445  
 Greenberg's analyticity-syntheticity metric, 323–327  
 groups, 396, 398, *see also* peer-groups, friendship groups
- /h/, variable, 127–129, 421, 427  
 harmonics, 233  
 HC *see* Helsinki Corpus of English Texts  
 headless verse lines, 268, 275  
 hedges, 406, 410  
 Helsinki Corpus, 120–121, 339  
 hiatus, 267  
 hierarchical agglomerative cluster analysis, 448–449  
 hierarchies, 27, 209, 214–216, 222, 314, 320–323, 328, 396, *see also* accessibility,

- animacy, definiteness, givenness,  
 individuation, referentiality  
 ~ in typology, 314, 321–322  
 ~ of prominence, 214
- historical linguistics, 1, 8, 10, 22, 119, 122, 134,  
 160, 260, 342, 425
- homophones, 421
- hope(s), in (the)* (Adj) ~, 146–147, 156
- hundredweight*, plural form, 143
- hypotheses, 394–396  
 alternative, 362, 375, 378  
 implicit, 105, 109  
 null, 362
- ICE *see* International Corpus of English
- ictus, 260, 268, 270–271
- identity markers, 82
- idiolects, 51, 74, 137, 140, 153–154, 157, 232
- implications, 4, 18–19, 90, 299, 328, 330, 400  
 biconditional, 328, 330–331  
 one-way, 328, 330–331
- in (the) event* governing NPs, gerunds and finite  
 clauses, 146
- independent variables, 45, 83, 279, 292, 372, 374,  
 376, 379, 445
- indexical fields, 47
- indexical layering, 50
- Indian English, 163, 173, 175, 239, 247
- individuation, 321–322, *see also* animacy
- inductive approaches, 2–4
- infinitives, 146, 188–191, 406  
 infinitival clauses, 152, 185  
 marked vs. bare/unmarked, 181, 184, 188–194
- informants, 19–20, 26–32, 39–44, 57, 59, 63–64,  
 70–92, 102–105, 108–110, 112–114,  
 341–358
- informational styles/registers, 411, 413–415,  
 438–439
- informed consent, 21, 30, 38, 52, 71
- ing* suffix, pronunciation, 382
- inner-circle varieties, 158, 162, 176
- intensifiers, 195, 197–198, 385–386
- interfaces, 58–59, 123–125, 182, 196, 365, 388  
 phonology-morphology, 260  
 syntax-intonation, 182
- inter-individual variation *see* idiolects
- International Corpus of English, 81, 106, 110,  
 115–117, 158–163, 175, 184, 186–187,  
 190–191, 196, 288, 290, 296, 326, 441
- International Phonetic Alphabet (IPA), 230–232,  
 234, 238–241
- Internet *see* web
- interrogative clauses, 105, 117  
 involving *the question + whether*, 151  
 non-canonical, 85  
*wh*-questions, 406, 410
- interval/ratio scales, 76–78, 83, 87–88, 90, 111
- interviews, 7–8, 10, 20, 22, 30–33, 69–92, 383  
 anonymity, 71  
 constraints, 39, 45, 51, 79  
 face-to-face, 71–72  
 metalinguistic, 6–7  
 modules, 31–32, 79  
 recording, 7, 17, 19–22, 24, 26–27, 29–30,  
 33–34  
 single vs. group, 30  
 sociolinguistic, 7, 22–33, 69, 71–72  
 structuredness, degree of, 31, 71–72
- intonation, 244–260, 276, 423  
 analysis, 252, 255  
 autosegmental-metrical (AM) tradition,  
 252, 255  
 British School, 252–255  
 nucleus, 251–255  
 phrase, 249, 251–254, 257
- introspective methods, 2–3, 22, 99–117
- inversions *see* topicalization
- involved styles/registers, 438
- isochrony of feet/syllables, 248
- isoglosses, 53, 67
- it*-cleft constructions, 210
- Janus-agents *see* passive
- jargon, 74
- join-count algorithm, 65
- judgment data *see* grammaticality
- juncture *see* breaks, phrasal vs. metrical
- Kendall's *tau*  $\tau$ , 377
- KWIC (Key Word in Context) *see* corpus retrieval
- L1 varieties, 318, 324–326, 445–450  
 acquisition, 362  
 high-contact L1, 318, 324  
 traditional/low contact L1, 318
- L2 varieties, 81–117  
 indigenized L2, 318, 445–446, 448, 450  
 interference from L1, 114
- Labov's Harlem study, 26
- LAE *see* Linguistic Atlas of England
- LAEME *see* A Linguistic Atlas of Early Middle  
 English
- LALME *see* A Linguistic Atlas of Late  
 Mediaeval English
- LAMSAS *see* Linguistic Atlas of the Middle  
 and South Atlantic States
- Lancaster-Oslo/Bergen Corpus, 149, 183–194, 363
- LANE *see* Linguistic Atlas of New England
- language contact, 7–8, 78, 80, 83, 85, 398–399,  
*see also* contact linguistics, L1, L2
- laryngograph, 241
- layer principle, 61–62
- layering, 383
- least-cost travel time *see* distance

- left-dislocations *see* topicalization  
 levels of measurement *see* scales  
 Levenshtein Distance (LD), 65–67, 431  
 lexical bundles, 410, 417  
 lexical effects in stimulus design, 107–108  
 lexical variation, 56, 79–90, 94, 402  
 lexicography, 158  
 lexicostatistics, 423  
 LibreOffice *see* OpenOffice  
*like* as discourse marker, 385  
 likelihood *see* probability of a variant  
 likelihood ratio test, 45  
 lingua franca, 19  
 Linguistic Atlas of England, 22  
 Linguistic Atlas of New England, 22  
 Linguistic Atlas of the Middle or South Atlantic States, 22, 53, 61–67  
 linguistic geography, 53  
 linguistic variables, 46–47, 69, 188  
 link maps, 442–443  
 LION *see* Literature Online  
 literate styles/registers, 410–414  
 literature databases, 119–134, 144, 148, 152, 156–157  
 Literature Online, 122, 133, 144, 148, 192  
 LLC *see* London-Lund Corpus  
 loanwords, 260–263, 268, 276  
 LOB *see* Lancaster-Oslo/Bergen Corpus  
 log odds of a variant, 298, 305, 307  
 log transformation, 441  
 logistic regression, 281, 292, 297–298, 303, 305, 382, 388, 394–396, 399–400  
  
 magnitude estimation, 99–117  
 manner of articulation, 238  
 MapInfo GIS software, 58, 64  
 maps *see* dialect maps, link maps, vector maps  
 mass data analysis *see* aggregate data analysis  
 matrix, 367–371  
   aggregate data, 41–42, 435–437, 442, 447  
   case-by-case data, 41  
   feature matrix, 435–436, 441, 446  
 median, 85–89, 342, 359, 371–374  
 mediopassives, 138  
 meta-linguistic statements, 163, 166, 169, 175, 260  
 metatheses, 423  
 meter *see* feet  
 microparametric variation, 314  
 Microsoft MapPoint, 58  
 minimal pairs, 269  
 mixed effects models, 111, 292, 296–297, 302, 305, 388, 399–400  
   fixed effects, 297, 305–306, 308, 310, 400  
   random effects, 297, 299, 302–303, 305, 400  
 mixed linear models *see* mixed effects models  
 mixed methods research, 4  
 modals, 162–163, 319, 406, 436, 439–440, 453  
   double modals, 19, 440, 453  
  
 monitoring *see* interview  
 monofactorial method, 379  
 morphosyntactic variation, 79–90, 94–97, 313, 318, 439–445  
 Multi-Dimensional (MD) analysis/scaling, 57, 67, 402–420, 447–450  
 multidimensional variation, 402–420, 433–434  
 multifactorial/multivariable methods  
   *see* multivariate methods  
 multivariate methods, 12, 49, 281, 292, 296, 314, 388  
*my/min(e)* and *h*-initial lexemes, 127–129  
  
 nasometer, 236–237  
 NCF *see* Nineteenth-Century Fiction  
 negation  
   and analyticity, 436  
   in a typological perspective, 316–320, 328–332  
   multiple *see* negative concord  
   of attributive adjectives, 127, 130–132  
   of sentential complements, 140  
 negative concord, 316–317, 319, 328–329, 452, 454  
 NeighborNet, 424–431  
 Neighbour Joining algorithm *see* NeighborNet  
 Neogrammarian approach, 53  
 neologisms, 160  
 network *see* social network analysis,  
   NeighborNet  
 networks *see* social networks  
 neurolinguistics, 7–8  
 neutralization, semantic, 386  
*never* as a negator in past tense contexts, 317, 319, 329, 452, 454  
 Newdigate Newsletters, 339  
 Nineteenth-Century Fiction, 122–123, 127, 129–133, 144, 147–148, 152  
 nominal scales, 77  
 nominalizations, 402, 406, 408, 410  
 normalization, 234, 251, 407  
 Northern Cities Shift, 235  
 noun phrases, 196–200, 205, 209–210, 216, 223, 267, 271, 309–311, 339  
   modification, 196, 200, 202–206, 209, 285, 289, 291  
 nouns, *see also* noun phrases  
   count vs. mass, 321–322, 328, 331  
  
 objectivity, 2–3, 37–38, 361  
 observer's paradox, 5, 18, 20, 26, 52  
 Occam's razor, 362  
 OED *see* Oxford English Dictionary  
 OpenOffice, 58, 364–365  
 operators *see* corpus retrieval  
 optimality theory, 399  
 optionality, 383  
 oral styles/registers, 407, 410, 412, 414, 436  
 order effects in stimulus design, 110

- orderly differentiation *see* structured heterogeneity
- ordinal scales, 76–77
- orthographic variants *see* corpus retrieval
- orthography vs. phonology, 260–261, 267–268, 272, 277
- overtones *see* harmonics
- Oxford English Dictionary, 22, 24, 136–157, 263–266  
 definitions, 137–138, 156  
 quotations, 136–157
- Pairwise Variability Index (PVI), 249
- Pakistani English, 158, 173–174
- palatograms, 242
- paraphonology, 268, 276
- parsing *see* corpora
- parsing theory of grammatical variation, 280
- participant observation, 19–21, 36–52, 70
- part-of-speech (POS) tagging *see* corpora
- parts-of-speech, 435–439, *see also* corpora, part-of-speech tagging
- passive, 162, 196, 212–226, 406–407, 410, 419  
 Janus-agents, 219  
 long passive (with *by* agent), 212–215, 217–220, 222–226  
 passival, 162  
 progressive passive, 159, 161–177, 182
- past participles, 127, 129–130, 133, 165, 197, 452, 455  
 modifying nouns, 407  
 variant inflections, 127, 129–130, 153–154
- past tense, 405–407  
 variant inflections, 127, 129–130  
 vs. present perfect, 363
- patient vs. agent, 213
- Pearson correlation, 444–445, 447
- Pearson residuals, 369
- peer-groups, 90
- Penn-Helsinki Parsed Corpus of Early Modern English, 127–128
- Penn-Helsinki Parsed Corpus of Middle English, 127
- philological approach, 181, 193
- phonemes, 228–229, 232, 241  
 inventory, 228  
 shifts, mergers, splits, 272
- phonetic variation, 228–259, *see also* consonants, vowels
- phonological variation, 10, 37, 228–259, 389–393, 402, *see also* consonants, vowels
- phonological weight, 284
- phonotactics, 228, 239, 393
- phylogenetic software *see* NeighborNet
- phyloversals, 327
- pidgins, 315, 318, 329–330, 398
- pitch, 233, 244, 246, 251–259  
 accents/movements/range, 246–259
- Pitney Bowes Business Insight, 58
- place of articulation, 238–239, 242, 423
- point-pattern analysis, 62–66
- polar questions, 316, *see also* interrogative clauses
- pooling of data, 338, 341–342, 347, 357
- population biology, 424
- possessive constructions, 282–293, 396
- possessive pronouns *see* my/min(e)
- possessive relations, 279–280, 282
- possessors *see* possessive constructions
- possessums *see* possessive constructions
- posterior interval *see* Bayesian approach
- power spectrum, 247
- precision, 165, 174, 186, 188, 197–199, 204, 206, 209–211
- predictors *see* factors, variables
- prehead *see* intonation analysis, British School
- prepositional phrases, 139, 402, 406, 410, 413  
 adjuncts, 106–109, 114–117  
 postmodifying noun phrase heads, 206–211  
 prepositional dative, 295, 297–299, 301–311
- prepositions, 56, 319, 326, 436–437,  
*see also* prepositional phrases
- pied piping vs. stranding, 105–108, 113–117
- prepositional verbs, 106–109, 114–116  
 variation in preposition usage, 150–152
- prescriptive linguistics, 103, 187
- present perfect, 162, 176, 182, 363, 452–453
- progressive passive, 159, 161–177, 182  
 vs. past tense, 363
- present tense, 75, 175, 182, 317, 319, 326, 329, 385, 388, 410, 413, 454  
 -s-inflection, 388
- Principle of End Weight *see* syntactic weight
- Principles and Parameters, 316, 451
- probabilistic linguistics, 362
- probability of a variant, 296–306
- procedural styles/registers, 404, 415–416
- production models, 11, 280
- professional stranger handlers, 25
- programming languages, 58, 61, 126, 365, 381  
 Java, 58–59, 110  
 Perl, 59, 174, 436, 441  
 Python, 58–59  
 Visual Basic, 58, 65
- progressive passive, 159, 161–177, 182
- Promax rotation, 409
- pronouns, 199, 208, 319, 321–322, *see also* you,  
 ye/you  
 demonstrative, 329  
 indefinite, 208, 395, 413  
 personal, 208, 406, 408, 410, 413, 415, 437  
 pronominality effects, 297–298, 301, 305  
 reflexive, 138, 142, 208, 437  
 relative, 386, 388–389, 394,  
*see also* relativization
- prosody, 260, 263, 267, 275, 277,  
*see also* suprasegmentals



- pseudo-cleft constructions *see* *wh*-cleft constructions  
 psycholinguistics, 8, 22, 295, 300, 311, 362  
 psychophysics, 99, 104  
 pulmonic consonants, 238
- quadrats, 62–63  
 quantization, 236  
 queries *see* corpus retrieval  
 questionnaires, 2, 6–10, 69–92, 110, 283, 314–317, 332  
 quotatives *see* *be* like
- /r/*, variable, 421, 428, *see also* rhoticity  
 R (software), 364–381  
 R commander *see* R (software)  
 $R^2$  values, 444–445  
 random forests, 399  
 randomization, 82, 110, 354  
 randomness, spatial, 57  
 raster images, 58  
 reaction time studies, 307  
 real-time change, 74, 339  
 recall, 165, 186, 189, 197–200, 209–211  
 recipient *see* semantic roles  
 reconstructed account of events, 415  
 reconstruction of stress, 260–277  
 recording data, 7, 17–34, *see also* interview, transcription of data  
   surreptitious recording, 7, 20–21, 27  
 referentiality, 279  
 reflexives, 452–453  
 regional variation *see* geographical variation  
 register, 6, 74, 83, 338, 383, 402–420, 433–435, 450–451, *see also* style  
   macro-register, 436  
 regression  
   linear, 376–377  
   logistic, 281, 292, 382, 388, 394–395, 399–400  
   probabilistic, 451  
   regression, 396  
   with breakpoints, 378  
 regularization, 129–130  
 relative vs. absolute judgments *see* absolute vs. relative judgments  
 relativization, 319, 323, 331  
   as postmodification of noun phrase heads, 196, 206  
   nominal relative clauses, 185  
   *than whom/which*, 147–148, 151  
   *that*-relative clauses, 106–107, 114–116, 417  
   *wh*-relative clauses, 106–107, 113–114, 117  
   zero relative clauses, 102, 322–323, 331, 389, 394–397, 441, 454  
 reliability, 2–3, 163, 171–173, 198, 231–232, 239, 247, 251, 255, 414  
   of coding, 213, 225–226  
 replicability, 160, 177, 361  
 representativeness, 70, 79, 177, *see also* corpora, sampling  
 rhoticity, 239, 421–430  
 rhymes, 271–272  
 rhythm, 235, 244, 248–251, 260, 267, 269, 275  
 Rhythm Ratio (RR), 250–251  
*risk* (n.), dependent clausal structures, 139–140  
 rotation techniques in factor analysis, 414, *see also* Promax rotation  
 rounding of vowels, 234–235, 423  
 Rule of Precedence, 270
- sampling  
   in corpus design, 414  
   in typology, 313–315, 324  
   judgement sampling, 70  
   random sampling, 70  
   sample size, 70, 337–338, 350–358  
   stratified random sampling, 70  
 SAS Business Analytics software, 58, 64  
*sat/stood, be ~*, 149–150  
 scales, 73, 76–78, 81, 103, 410  
   dichotomous scales, 77  
   polytomous scales, 77  
 scatter plot, 437, 445  
 schwa, */r/*-coloured, 421  
 scribal evidence *see* orthography vs. phonology  
 search engines *see* web crawlers  
 segments, phonetic/phonological, 10, 181, 260–261, 423  
   addition, 423  
   loss, 423  
 semantic roles  
   recipient, 295, 298–299, 301, 306–307, 309  
   theme, 295, 300–302, 305–306, 308–309  
 sentence completion task, 310  
 sex, 1, 33, 36, 45–46, 51, 69–70, 74, 80, 111, 195, 297, 310, 337  
 Sievers' Rule of Precedence, 270  
 similarity  
   morphosyntactic, 450–451  
   phonetic, 422  
 slang, 74  
 smoother, 376–377, 444  
 snowball technique, 70  
*so*, intensifier, 384–387  
 social desirability effect *see* desirability effects  
 social meaning of linguistic features, 37, 47–52  
 social network analysis (SNA), 36–37, 40–44, 50–52  
 social networks, 24–26, 33–34, 70, 74  
   aggregate networks, 41–44  
 social practice, 44, 51  
 social structure of a local community, 24–25, 36  
 social variables, 32, 45–47, 69, 74, 429, *see also* age, sex, socio-economic class



- socio-economic class, 1, 36, 195, 338  
 sociograms, 42  
 sociolects, 433  
 sociolinguistics, 1, 8, 10, 17, 21–22, 27, 33, 71, 362  
   first wave, 36, 45, 51–52  
   historical, 337  
   quantitative, 51, 337  
   second wave, 36  
   variationist, 21, 33, 53, 224, 331–332, 338–339, 382–400  
 sociology, 18  
 Sound Comparisons project, 423, 425, 430  
 SoundScriber, 21  
 Southern Plantation Overseers Corpus, 339  
 spatial statistics, 57, 62  
 specificational cleft constructions *see* wh-cleft constructions  
 spectral tilt, 235, 247  
 spectrograms, 233–234, 246, 255  
 speech acts, 251–252  
 speech analysis software, 233–234, 246–247, 257–258  
 spent N (*in/on*) + gerund, 150  
 Splitsree 4 package, 424  
 SPOC *see* Southern Plantation Overseers Corpus  
 spoken (vs. written) language, 80, 82, 84–90, 326–327, 332  
 spreadsheet software, 125, 362–363, 365  
 stance, 410–419  
 standard deviation, 103, 352–353, 357, 359, 410  
 standard language, 332  
 standard lexical sets *see* Wells's standard lexical sets  
 statistical methods, 12, 314, 362, *see also* multivariate methods, multifactorial methods  
 statistical modelling, 181, 387–388, 393, 398–399  
 statistical significance, 357, 379, 388, 395, 398  
 statistics *see* statistical methods, statistical significance, spatial statistics  
 stimuli, 99–105, 108–117, 288  
   experimental stimuli, 101–105  
   reference stimuli, 100–103  
 stress, 100–103, 229–244  
   auto-stressed suffixes, 269  
   clashes, 131–132  
   contrastive, 131, 252  
   fixed stress vs. free lexical stress, 244  
   in compounds/phrases, 265–266, 268–269, 276  
   primary vs. secondary vs. tertiary, 245–246, 262, 265, 276  
   root initial/left-edge (Germanic), 262–265, 272, 274  
   shifting in loan words, 261–265  
   weight-based (Latin), 265  
 stress-timing, 248  
 stricture, degrees of consonantal, 423  
 structured heterogeneity, 382–383, 398  
 style, 6, 72–74, 76, 78, 82, 84–85, 89, 91, 185, 191, 383, 399–400, *see also* register  
   elevated style, 149  
   formal style, 191  
   informal style, 149–150, 154, 185, 187, 191–192  
   literate style, 123, 133  
   oral style, 123, 133  
   style-switching, 30  
   stylistic layering, 191  
 subordination, 319, 406, *see also* relativization  
 superlatives, 151, 453  
 suprasegmentals, phonetic/phonological, 10, 228, 244–259  
   diachronic change, 260–277  
 survey techniques, 8, 62–65, 71, 314–319, 434, 445–447  
 SVC (Subject – Verb – Complement) order, 185  
 Swadesh basic vocabulary lists, 423  
 Switchboard Corpus, 296–302  
 syllables, 216, 229, 232, 244–257, 260–277  
   counting, 260  
   loss, 263–264  
   quantity, 247, 251, 253, 257, 276  
   structure, 393  
 syllable-timing, 248  
 synchronic variation, 9, 24, 188, 229, 273  
 syncope, 262, 268  
 synzesis, 268  
 synsemantic words *see* grammatical markers  
 syntactic weight, 11, 279–280, 284  
   Principle of End Weight, 216  
 syntheticity, 314, 323–328, 433, 438, *see also* Greenberg's analyticity-syntheticity metric  
 syntopic variation, 9  
 /t/, variable, 47–49, 239, 429  
 T2K-SWAL Corpus *see* TOEFL 2000 Spoken and Written Academic Language Corpus  
 tail *see* intonation analysis, British School  
 taxonomy, numerical *see* aggregate data analysis  
 technical vocabulary, 415  
 tense, mood and aspect (TMA), 19–20, 319  
 than, 147  
 that  
   complementizer, 141, 192, 386, 407, 419, 434  
   omission/deletion, 406–407, 410, 419, 434  
   pronoun, 386  
 that of NPI-ing, 153  
 theme *see* semantic roles  
 theoretical linguistics, 22, *see also* Principles and Parameters, Chomskyan linguistics, generative linguistics

- there*, dummy/existential *see* existential constructions
- thereby*, 149, 151–152
- th*-fronting, 45–50
- Thiessen neighbors, 65
- Thiessen polygon, 65
- TIME corpus, 159, 184, 375
- tiring effect, 82
- to do with*, 149
- to try and* (X) verb stem, 149
- ToBI (Tone and Break Indices) transcription, 252, 255–257
- TOEFL 2000 Spoken and Written Academic Language Corpus, 405, 410–411, 417
- tokens, 199, 386, 388–389, 394, 396–397, 400, 436–437
- tone units (TU), 252–253
- tones, 252–259, *see also* pitch movement
- topicalization, 105, 213, 215
- transcription of data, 19, 21, 31, 34, *see also* interview
- transparency index, 324
- travel time, least cost *see* distance
- tree diagrams *see* dendrograms
- t*-test, 374
- Tukey tests, 111
- tunes, 252, 256
- types, 273–274
- typology, 1, 7–8, 214, 225, 268, 313–332, 433  
 feature catalogues, 317  
 hierarchies, 214, 314  
 methods, 314–332  
 parameters, 314, 316–318
- typoversals, 327
- UCINET *see* social network analysis
- ultrasound imaging, 236
- uninformative prior *see* Bayesian approach
- universals, 132, 313–314, 317, 327–331, 393, 398, 433  
 implicational, 328, 330–331  
 unrestricted, 328–329
- usage ratings, 83, 85–90
- U*-test, 374–375, 380
- validity, 2–3, 103–104, 247–248, 251, 414
- Variability Index (VI), 251
- Variability-based Neighbor Clustering, 378
- Variable rule analysis (Varbrul), 45, 388–389
- variables, 1, 70, 74–75, 77, 83, 92, 296–297, 300–301, 304, 388–389, 393, 398,  
*see also* factors, cognitive variables,  
 linguistic variables, social variables  
 macro-variables, 338  
 micro-variables, 338
- variance, 408–409, 411, *see also* Analysis of Variance
- variation *see* diachronic, geographical, inter-individual, lexical, morphosyntactic, phonological, synchronic variation, idiolects
- varieties, 23, 305, 308–311, 393–397, 399, 434–435, 440, 445–450, *see also* dialects, African American, American, Australian, Bangladeshi, British, Indian, Pakistani English, L1, L2
- varioversals, 328–330, 450
- vector maps, 59
- vectors *see* R (software)
- verbs  
 causative, 415  
 degrees of transitivity, 206, 210  
 factual, 410  
 impersonal, 341  
 mental, 406–413  
 morphology, 319, 452, 454  
 of desire, 415  
 of effort, 407  
 of likelihood/probability, 410, 417  
 of occurrence, 417
- vernacular, 72, 75, 78–79, 81, 85, 382–383, 441
- vernacular principle, 20
- vernacular universals, 313, 327–330
- verse, 266–277  
 artificiality, 266  
 as source of evidence, 266
- view*, with *a ~ to/ toward(s) -ing*, 149, 152
- visualization, 57, 61, 379, 438
- vocal folds, 240–241
- voice onset time (VOT), 239–241
- voicing of segments, 239–242, 423
- vowels, 228–242  
 articulatory features, 228–242, 431  
 deletion, 247, 261, 263 *see also* aphesis, elision, syncope  
 diphthongization, 428  
 height, 423  
 length/lengthening, 261, 423–424, 428  
 mergers, 428  
 qualitative changes, 428  
 reduction, 247–248, 261
- WALS *see* World Atlas of Language Structures
- want*-constructions, 84–90
- way*, V *one's ~*, 151
- web, 22, 25, 110, 158–177, 187  
 source of meta-linguistic data, 175–177  
 web-as-corpus approach, 169–170, 296  
 web-based data, 158–177  
 web-for-corpus-building approach, 173–177
- web crawlers, 159–160, 170–172, 177,  
*see also* Google
- weight *see* syntactic weight, phonological, factor weight

- Wells' standard lexical sets, 229–230  
 West Fife High Pipe Band corpus, 39  
 WFHPB corpus *see* West Fife High Pipe Band corpus  
*wh*-cleft constructions, 183–193  
 Wilcoxon test *see* U-test  
 wildcards, 124, 137, *see also* corpus retrieval  
 word geography, 53–57  
 word order *see* constituent order, SVC order  
 World Atlas of Language Structures, 318  
 World Englishes, 318–332, 398, 425, 434, 445–452  
 world wide web *see* web  
*yall*, 329  
*ye/you* replacing *thou/thee*, 340  
*you guys*, 329  
*you* replacing *ye*, 339–340  
 zero variants, 384–389, 394–397  
 z-scores, 103