

Index

- academic vocabulary, 52, 130–134, 143
 accuracy (*see also* syntax, dimensions of), 3–4, 171
 analytic scales. *See* marking, types of
 annotation (*see also* natural language processing), 43
 lemmatisation, 118
 manual vs automated approaches to, 84, 144, 155–156, 157, 159
 methodological issues with, 108–113
 tagging and parsing, 84, 110, 118, 155–156, 157, 159
 theoretical importance of, 108–113
 appropriateness, 3, 20, 62, 65, 91, 113, 120, 126, 130, 134, 135, 148
 audience. *See* communicative context, dimensions of
 automated marking. *See* marking, types of
- British Academic Written English corpus (BAWE), 157, 172
 British National Corpus (BNC), 128, 179, 180
- Cambridge English Qualifications (CEQs), 197
 Cambridge Learner Corpus (CLC), 2, 32
 China's Standards of English Language Ability (CSE), 25–26
 chunks. *See* formulaic language, measures of
 cognition. *See* writing, proficiency in
 coherence. *See* cohesion
 cohesion
 coherence vs cohesion, 183–184, 192–193, 200
 definition of, 183
 L1 vs L2 development, summary of, 199–200
 major approaches to
 Coh-Metrix, 192–199
 Halliday & Hasan, 184–192
 Tool for the Automated Analysis of Cohesion (TAACO), 196, 199
 measures of
 combined measures, 190–192
 conjunction, 185–186, 189–190
 givenness, 195–196, 199
 latent semantic analysis, 195–196, 199
 lexical cohesion, 186–187, 190
 reference, 184, 188
 situation model, 192, 198–199
 substitution and ellipsis, 184–185, 188–189
 textbase, 192, 193–194, 197–198
 Coh-Metrix, 46, 54, 112, 184, 192–199
 collocations. *See* formulaic language, measures of
 Common European Framework of Reference (CEFR), 1–2, 25–26, 27, 164, 170, 204
 communicative competence (*see also* writing, proficiency in), 14, 17, 18, 21, 28, 30, 32, 36, 48, 62, 112
 communicative context
 dimensions of
 audience, 15, 21, 29, 41, 210, 213
 curricula/educational context, 3, 10, 22, 25–26, 29, 40, 41, 56, 57, 75, 134
 discipline, 12, 13, 41, 210
 genre, 11, 13, 20–21, 26, 29, 31, 34, 37, 39–42, 52, 53, 54, 74, 83, 94, 113, 130, 179, 192, 203, 210, 213
 marking. *See* marking
 prompt, 151, 168, 171
 register, 26, 79, 126, 130–134, 136, 138, 141, 142
 task, 15, 20, 32, 33, 38, 41, 55, 58, 65, 113, 171, 204
 topic, 20, 26, 29, 34, 37, 41, 55, 113, 126, 130, 134, 165, 179, 210, 213
 importance of, 14–15, 17, 20–21, 22, 25, 29, 33, 34, 37, 39, 40, 41, 52, 54, 55, 62, 209–211, 213
 competence. *See* proficiency
 complexity
 communicative, 3, 14, 38, 43, 50, 70
 lexical, 51, 124, 127, 128

Index

241

- linguistic, 4, 34, 36, 42, 43, 48, 50, 51, 60–62, 63–64, 65, 70, 81, 83, 91, 113–114, 116, 124, 127, 128, 171, 206
- phraseological, 171
- syntactic, 34, 36, 48, 51, 60–62, 63–64, 65, 70, 81, 83, 91, 113–114, 116, 206
- concurrent validity. *See* validity, types of
- conjunction. *See* cohesion, measures of
- construct validity. *See* validity, types of
- corpus linguistics (*see also* methodological issues)
 - and multidimensional analysis (MDA), 52–55
- core premises of, 28–32
- qualitative vs quantitative approaches, 33
- quantitative (QCL) approaches
 - definition of, 33
 - extralinguistic dimensions of, 38–42
 - limitations of, 34–38
 - linguistic dimensions of, 42–52
 - promise of, 33–34
- Corpus of Contemporary American English (COCA), 131, 135, 136, 155, 170, 180
- curricula/educational context. *See* communicative context, dimensions of
- demographic factors, types of
 - ethnicity, 38
 - gender, 38
 - L1, 29, 38, 145, 213
 - nationality, 38
 - socio-economic status (SES), 3
- discipline. *See* communicative context, dimensions of
- ellipsis. *See* cohesion, measures of
- fluency
 - communicative, 147
 - linguistic, 51, 60–62, 63–64, 65, 116, 147, 206
 - syntactic, 51, 60–62, 63–64, 65, 116, 206
- formulaic language
 - definition of, 147, 149
 - dimensions of
 - diversity, 150, 162, 163, 176
 - exclusivity, 153, 157
 - idiomaticity, 148, 155, 175, 176
 - sophistication, 171, 180
 - external vs internal measures of, 162–164, 168–171, 178–180
 - knowledge of, 149–152
 - L1 vs L2 development, summary of, 181–182
- measures of
 - bigrams, 168–170
 - chunks, 175–176
 - collocations, 147, 148, 153, 154–156, 157, 168, 169, 170, 171, 172, 174, 175, 176, 190, 195, 209
 - discourse bundles, 163
 - idioms, 147, 174, 175
 - lexical bundles, 150, 151, 153–154, 162–164, 165, 175, 181, 203, 206
 - n-grams, 153–154, 162–163, 164
 - p-frames, 154, 164
 - referential bundles, 163
 - syntactic bundles (NP-based, PP-based, VP-based), 167
 - trigrams, 168–169
- methodological issues, 177–181
- relevance of, 147–148
- statistical measures
 - collexme strength, 170
 - Delta-P, 90, 160, 161, 170
 - mutual information (MI), 157, 160, 169, 171, 173, 203
 - precision, 159
 - recall, 159
 - t-score, 160, 170
- GCSE, 143
- gender. *See* demographic factors, types of
- generic indices. *See* syntax, measures of
- genre. *See* communicative context, dimensions of
- Growth in Grammar corpus (GiG), 11–13, 32, 131, 148, 157
- holistic scales. *See* marking, types of
- human marking. *See* marking, types of
- IELTS, 9, 143, 148
- L1 influence. *See* demographic factors, types of
- L1 vs L2 development, 114–116, 144–146, 181–182, 199–200, 201–203
- latent semantic analysis. *See* cohesion, measures of
- LCA (Lexical Complexity Analyser), 128
- lexical bundles. *See* formulaic language, measures of
- lexical cohesion. *See* cohesion, measures of
- lexical density. *See* vocabulary, dimensions of
- lexical diversity. *See* vocabulary, dimensions of
- lexical richness. *See* vocabulary, dimensions of
- lexical sophistication. *See* vocabulary, dimensions of

242 Index

- marking
 descriptors, 25, 26, 56, 119
 frameworks, 25, 26, 27, 204
 rubrics, 2, 26, 39, 142, 143, 148, 164, 183, 204, 211, 213
 types of
 analytic scales, 26–27, 210–211
 automated marking, 31, 37–38, 57
 holistic scales, 26–27, 210–211
 human marking, 24, 26, 31, 39, 57, 119
 primary trait scales, 26–27
- methodological issues (*see also* corpus linguistics)
 causal vs correlational relationships, 19, 37–38, 66, 113
 construct specification, 47–52
 cross-sectional vs longitudinal approaches, 22, 23, 211
 dispersion, 44
 feature identification and operationalisation, 29–30, 32, 34–36, 42–52, 61–62, 63–64, 66, 69–70, 108–114, 115, 130, 144, 152–161, 177, 179, 205, 207, 210, 212
 L1 vs L2 proficiency, 42
 normalisation, 44, 111, 124, 125
 reliability, 4, 24, 29–30, 31–32, 38, 39, 43–45, 63, 66, 69, 109, 112, 115–116, 144, 156, 159, 178–179, 181
 representativity (*see also* communicative context, dimensions of), 4, 20, 33–42, 51, 57, 168, 179–180, 209–211
 sampling (*see also* communicative context, dimensions of), 4, 5, 29, 32, 33–42, 57, 179, 209–211
 validity. *See* validity
 variability in raters and rating systems, 24–27, 213
- multidimensional analysis (MDA), 52–55
- nationality. *See* demographic factors, types of natural language processing (NLP), 4, 5, 144, 155
 notable software, 128, 156, 157, 169, 179, 192–199
- n-grams. *See* formulaic language, measures of NLP. *See* natural language processing (NLP)
- parsing. *See* annotation
 p-frames. *See* formulaic language, measures of
 primary trait scales. *See* marking, types of
 proficiency. *See* writing, proficiency in
 prompt. *See* communicative context, dimensions of
 psycholinguistics. *See* writing, proficiency in
- reference. *See* cohesion, measures of
 register. *See* communicative context, dimensions of
 rubrics. *See* marking
- sampling. *See* methodological issues
 scales. *See* marking, types of
 situation model. *See* cohesion, measures of
 socio-economic status (SES). *See* demographic factors, types of
 Stanford CoreNLP, 156, 157, 169
 substitution. *See* cohesion, measures of
 syntax
 definition of, 56
 dimensions of
 accuracy, 60–62, 181
 appropriacy, 60–62
 complexity, 60–62, 63–64, 65, 70, 81, 83, 113–114, 116
 fluency, 60–62, 63–64, 116
 representational component vs usage component, 58–62
 interaction with vocabulary, 129, 131, 140
 L1 vs L2 development, summary of, 114–116
- measures of
 adjectives/adjective phrases, 55, 79–80, 129, 131
 adverbial clauses, 92–97, 107
 adverbials, placement of, 106–108
 adverbs/adverb phrases, 55, 75–76, 80, 129, 131
 canonical subject openings, 106
 clauses, 46, 64, 65, 66, 67, 68, 70–75, 88–101, 104–105
 clefts, 105
 comparative clauses, 80
 complement clauses, 97–100
 coordination, 101–103
 displacement structures, 105–106
 existential clauses, 105
 expletive structures, 105
 extraposition clauses, 105
 generic indices, 64–75
 Hunt's synopsis of clause to sentence length factors, 66–75
 information packaging, 103–108
 nouns/noun phrases, 76–78, 81–87, 129, 131
 passives, 104–105
 prepositions/preposition phrases, 78, 80–81
 relative clauses, 100–101
 sentence length, 67, 68, 70, 71, 72, 73

Index

243

- subordination, 46, 66, 67, 70–75, 80, 91–101, 105
- T-units, 44, 46, 48, 51, 63, 64, 65, 66, 67, 70–75
- verbs/verb phrases, 78–79, 88–91, 129, 131
- methodological issues, 57, 61–62, 108–114
- TAACO (Tool for the Automated Analysis of Cohesion), 196, 199
- TAALES (Tool for the Automated Analysis of Lexical Sophistication), 179
- tagging. *See* annotation
- task. *See* communicative context, dimensions of
- textbase. *See* cohesion, measures of
- TOEFL, 9, 54, 148, 168, 197, 198
- topic. *See* communicative context, dimensions of
- T-units. *See* syntax, measures of
- usage-based models of language acquisition, 147, 150–151, 179–180, 182
- validity
 - relevance of, 23, 31–32, 37–40, 55, 65, 149, 209
 - types of
 - concurrent validity, 32, 45, 47, 51, 111, 114, 209
 - construct validity, 31, 32, 33, 45–52, 65, 123
- vocabulary
 - dimensions of
 - lexical density, 51, 123, 137, 140–141, 145
 - lexical diversity, 19, 39, 50, 89, 123, 124–126, 129, 133, 145, 194, 203, 207, 212
 - lexical richness, 123, 140
 - lexical sophistication, 123, 126–140, 145, 207, 212
 - diversity measures
 - corrected type-token ratio (CTTR), 125, 126, 141, 163
 - Guiraud's index/Root TTR, 125, 128, 141
 - measure of textual lexical diversity (MTLD), 125, 126, 141
 - type-token ratio (TTR), 141
 - voc-D, 125–126, 141
 - interaction with syntax, 129, 131, 140
 - knowledge of
 - breadth vs depth, 119–120
 - lexical organisation, 123
 - productive vs receptive, 120–123
 - L1 vs L2 development, summary of, 144–146
 - methodological issues, 141–144
 - sophistication measures
 - word frequency, 126–140
 - word length, 134–137
 - word meaning, 138–140
- writing
 - general definition of, 14
 - process vs product, 5, 14–15, 18–19, 27, 35, 214
 - proficiency in
 - cognitive dimensions of, 19, 28, 36, 48, 82, 113, 123, 151
 - core analytical issues, 19–21
 - core model of, 18
 - definition and nature of, 15–21
 - development as both over time and in quality, 21
 - development in quality, 23–27
 - development over time, 22–23
 - psycholinguistic aspects of, 36, 122, 138, 147, 151, 152, 153, 155, 214
 - relationship to communicative competence, 14, 17, 18, 21, 28, 30, 32, 36, 48, 62, 112
 - status of linguistic features within, 27–32