

Contents

<i>Illustrations</i>	ix
<i>Contributors</i>	xiii
<i>Acknowledgements</i>	xxvii

Section I 1

1	Historical perspective: what are corpora and how have they evolved?	3
	<i>Michael McCarthy and Anne O'Keeffe</i>	
2	Theoretical overview of the evolution of corpus linguistics	14
	<i>Elena Tognini Bonelli</i>	

Section II

Building and designing a corpus: what are the key considerations? 29

3	Building a corpus: what are the key considerations?	31
	<i>Randi Reppen</i>	
4	Building a spoken corpus: what are the basics?	38
	<i>Svenja Adolphs and Dawn Knight</i>	
5	Building a written corpus: what are the basics?	53
	<i>Mike Nelson</i>	
6	Building small specialised corpora	66
	<i>Almut Koester</i>	
7	Building a corpus to represent a variety of a language	80
	<i>Brian Clancy</i>	

8	Building a specialised audio-visual corpus <i>Paul Thompson</i>	93
Section III		
	Analysing a corpus: what are the basics?	105
9	What corpora are available? <i>David Y.W. Lee</i>	107
10	What are the basics of analysing a corpus? <i>Jane Evison</i>	122
11	What can corpus software do? <i>Mike Scott</i>	136
12	How can a corpus be used to explore patterns? <i>Susan Hunston</i>	152
13	What are concordances and how are they used? <i>Christopher Tribble</i>	167
14	What can corpus software reveal about language development? <i>Xiaofei Lu</i>	184
Section IV		
	Using a corpus for language research: what can a corpus tell us about language?	195
15	What can a corpus tell us about lexis? <i>Rosamund Moon</i>	197
16	What can a corpus tell us about multi-word units? <i>Chris Greaves and Martin Warren</i>	212
17	What can a corpus tell us about grammar? <i>Susan Conrad</i>	227
18	What can a corpus tell us about registers and genres? <i>Douglas Biber</i>	241
19	What can a corpus tell us about specialist genres? <i>Michael Handford</i>	255
20	What can a corpus tell us about discourse? <i>Scott Thornbury</i>	270
21	What can a corpus tell us about pragmatics? <i>Christoph Rühlmann</i>	288

22	What can a corpus tell us about creativity? <i>Thuc Anh Vo and Ronald Carter</i>	302
Section V		
	Using a corpus for language pedagogy and methodology	317
23	What can a corpus tell us about language teaching? <i>Winnie Cheng</i>	319
24	What features of spoken and written corpora can be exploited in creating language teaching materials and syllabuses? <i>Steve Walsh</i>	333
25	What is data-driven learning? <i>Angela Chambers</i>	345
26	How can data-driven learning be used in language teaching? <i>Gaëtanelle Gilquin and Sylviane Granger</i>	359
27	How can we prepare learners for using language corpora? <i>Passapong Sripicharn</i>	371
Section VI		
	Designing corpus-based materials for the language classroom	385
28	What can a corpus tell us about vocabulary teaching materials? <i>Martha Jones and Philip Durrant</i>	387
29	What a corpus tells us about grammar teaching materials <i>Rebecca Hughes</i>	401
30	Corpus-informed course book design <i>Jeanne McCarten</i>	413
31	Using corpora to write dictionaries <i>Elizabeth Walter</i>	428
32	Using corpora for writing instruction <i>Lynne Flowerdew</i>	444
33	What can corpora tell us about English for Academic Purposes? <i>Averil Coxhead</i>	458
34	How can teachers use a corpus for their own research? <i>Elaine Vaughan</i>	471

Section VII**Using corpora to study literature and translation** 485

- 35 What are parallel and comparable corpora and how can we use them? 487
Marie-Madeleine Kenning
- 36 Using corpora in translation 501
Natalie Kübler and Guy Aston
- 37 How can corpora be used to explore the language of poetry and drama? 516
Dan McIntyre and Brian Walker
- 38 How can corpora be used to explore literary speech representation? 531
Carolina P. Amador-Moreno

Section VIII**Applying corpus linguistics to other areas of research** 545

- 39 How to use corpus linguistics in sociolinguistics 547
Gisle Andersen
- 40 How to use corpus linguistics in the study of media discourse 563
Kieran O'Halloran
- 41 How to use corpus linguistics in forensic linguistics 578
Janet Cotterill
- 42 How to use corpus linguistics in the study of political discourse 591
Annelie Ädel
- 43 How to use corpus linguistics in the study of health communication 605
Sarah Atkins and Kevin Harvey
- 44 How can corpora be used in teacher education? 620
Fiona Farr
- 45 How can corpora be used in language testing? 633
Fiona Barker
- Index* 647

Index

- Aarts, J. 472
 Aarts, J. and Meijs, W. 5
 abstract word groups 388
 Academic Formulas List 460
 academic genres 259–62
 academic integrity 63–64
 academic journal articles 177, 394–95
 academic language: corpora 624; in use 458–60
 academic learning, corpus suitability for 391–92
 Academic Word List (AWL) 459; corpus-informed course book design 416; English for Academic Purposes (EAP), corpora and 459, 464–65; vocabulary teaching materials 391, 393
 academic writing 244
 access to corpora 107–9
 accuracy, measures of 190
 Ackerley, K. and Coccetta, F. 425, 466
 Adams, D. 136
 adaptation of electronic material 62
 Ädel, Annelie xiii, 11, 591–603
 Adolescent Health Email Corpus (AHEC) 605, 608, 609, 610, 611
 adolescent health language 607–8
 Adolphs, S. and Carter, R. 425
 Adolphs, S. and Knight, D. 6, 25, 38, 67, 93, 97, 99, 102, 258, 290, 299, 305, 326
 Adolphs, S. *et al.* 39, 262, 606, 607, 609
 Adolphs, Svenja xiii, 38–49, 73, 89, 98, 290, 296, 297, 409, 430, 466, 475, 533, 534, 548, 549, 565, 595, 627
 adverbial connectors 492
 African American Vernacular English: language development, corpus software in research 187; literary speech representation 534
 Aggleton, P. and Homans, H. 616
 Ahmad, K. 346
 Aijmer, K. 39, 244, 271, 410, 454, 481, 492, 493
 Aijmer, K. and Simon-Vandenberg, A.-M. 11
 Aktas, R. and Cortes, V. 460, 462
 Al-Qaeda 580
 Alderson, J.C. 637
 Alexopolou, T. 638, 639
Alice in Wonderland (Carroll, L.) 169–70, 171, 173
 alignment: of text pairs 490–91, 511
 Allan, R. 321, 337, 352, 425
 Allén, Sture 16
 Allen, W.S. 5
 alphabetical sorting 379
 Altenberg, B. 221, 245, 271, 492
 Amador-Moreno, Carolina P. xiii–xiv, 9, 309, 531–42
 Amador-Moreno, C.P. and O’Keeffe, A. 539–40
American Heritage Dictionary 428
 American National Corpus (ANC): corpora 110; corpus-informed course book design 414; research, teachers’ use of corpora in 473; variety of a language, corpus building to represent 87
 Ammerman, S. *et al.* 610
 analysis of: creativity 305–9; genre, methodological advantages in 258–59; ‘how X is talked about’ in political discourse 595–96; key words in poetry and drama 519–22; keywords in political discourse 597–98; legal casework 578; parallel and comparable corpora 491; spoken corpora 39, 48–49; vocabulary teaching materials 388, 391–92; *see also* authorship analysis; corpus analysis
 Andersen, G. and Stenström, A.B. 414
 Andersen, Gisle xiv, 11, 101, 229, 230, 232, 233, 295, 547–61

- Anderson, B. 403
 Anderson, J. *et al.* 39
 Anderson, W. and Beavan, D. 39
 annotation: audio-visual corpus building 98–99;
 written corpus building 63
 anonymity, issues of 43–44
 Anthony, L. 167, 502
 Anthony of Padua 3
 antonyms 207–8
Anvil 46, 98–99, 99–100
 application issues: in corpus linguistics (CL)
 6–12; writing instruction, use of corpora for
 450–52
 applied linguistics 401–2, 409
 appositive clause 158
 Aquinas, St. Thomas 4, 22
 Arabic English 488
 ARCHER Corpus 113
 archives 107
 Århus Corpus of Contract Law 56
 arithmetical power, qualitative change and 18
 Arndt, H. and Janney, R.W. 298
 Artificial Intelligence (AI) 141
 Ashby S. *et al.* 39
 Ashdown, Paddy 201
 assembly, audio-visual corpus building 99–101
 assessment, essential principles of 633
 Associated Press 594
 Association of Language Testers in Europe
 (ALTE) Members 633, 634, 635, 639
 Aston, G. and Burnard, L. 84, 177
 Aston, G. and Kübler, N. 367
 Aston, G. *et al.* 380
 Aston, Guy xiv, 8, 119, 319, 320, 321, 337, 360,
 364, 367, 380, 431, 444, 472, 477, 501–13,
 623
 Atkins, B.T.S. and Rundell, M. 442
 Atkins, S. and Harvey, K. 77
 Atkins, S. *et al.* 86
 Atkins, Sarah xiv–xv, 605–17
 audio-visual corpus building: annotation 98–99;
 Anvil 98–99, 99–100; assembly 99–101;
 Augmented Multi-party Interaction (AMI)
 Meeting Corpus 97, 99, 102; Bergen Corpus
 of London Teenage Language (COLT) 101;
 British Sign Language 94; Child Language
 Data Exchange System (CHILDES) 98;
 Computerised Language Analysis (CLAN) 98;
 Computers in the Human Interaction Loop
 (CHIL) 97, 102; COSMOROE framework
 99; data collection 96–97; description of
 audio-visual corpora 93–96; *EXMARaLDA*
 project 99, 102; further reading 103; future
 developments 101–3; German Sign Language
 94; GLOSSA Corpus 101; HTML codes 100;
 Human Speech Project (HSP) 102; interface
 99–101; *NITE Toolkit* 99, 102; Nottingham
 Multi-Modal Corpus 102; *Observer XT*
 99–100; *Praat* 98–99; purpose of audio-visual
 corpora 93–96; quality video recording 97;
 QuickTime 100, 101; Rhetorical Structure
 Theory (RST) 98–99; SACODEYL project
 94; SCOTS Corpus 100; sign languages 94;
 Singapore Corpus of Research in Education
 (SCoRE) 94–95, 101; *Soundscriber* 98; spoken
 language, sets of 93–94; transcription 98–99;
 video file formats 100; *Wordsmith Tools* (Scott,
 M.) 95; XML 96, 99, 102, 103
 Augmented Multi-party Interaction (AMI)
 Meeting Corpus: audio-visual corpus building
 97, 99, 102; spoken corpus building 39
 Auran, C. *et al.* 312
 Austin, J.L. 302
 Australian Corpus of English (ACE) 109
 authenticity 10, 42, 259, 320, 326, 340, 360–61,
 362, 489, 533, 635
 authorship analysis: potential for forensic corpus
 in 578–80; specialised individual corpora in
 583
 automatic rating software 640
 automatic speech recognition (ASR) 560
 availability of corpora 107–20
 average, concordances of 396–98
 Azar, B. 235
 Babel Chinese-English Parallel Corpus 120
 Bachman, L.F. 633, 635
 Baker, M. 24, 494, 495, 501
 Baker, P. 206, 208, 255, 275, 276, 283, 373,
 533, 565, 597, 599, 616
 Baker, P. *et al.* 372, 565, 624
 balance: assessment of 86–87; written corpus
 building 60
 Baldry, A. 466
 Baldry, A. and Thibault, P. 93, 299
 Baldwin, T. *et al.* 403, 404
 Bang, M. 163
 Bank of English: corpora 109–10; corpus
 linguistics (CL) 24; forensic linguistics 578,
 586; lexis 197–98, 199, 206; pattern
 exploration 153; research, teachers' use of
 corpora in 473, 476–77; specialist genres 256;
 written corpus building 55
 Barbieri, F. 238, 244, 373
 Barbieri, F. and Eckhart, S.E.B. 366, 404
 Bargiela-Chiappini, F. and Harris, S. 262
 Bargiela-Chiappini, F. and Nickerson, C. 262
 Barker, Fiona xv, 476, 633–43
 Barlow, M. 5, 167, 347, 350, 462, 467, 491,
 495, 502, 594
 Barni, M. and Gallina, F. 638
 Baroni, M. and Bernadini, S. 440
 Baroni, M. *et al.* 440, 508, 512
 Barry, P.J. *et al.* 140

- basic considerations: building a corpus 31;
corpus analysis 122–23; written corpus
building 53–54
- Bax, S. 623
- Bazerman, C. 257, 260
- BE06 Corpus 109
- Becker, Boris 429
- Becket, A. 3, 168, 169, 170
- Bedmar, D. 366
- Bednarek, M. 244, 246, 576
- Beeby, A. *et al.* 513
- behaviourism 14, 354
- Bender, T.K. 4, 169
- Bentley, Derek 581–82
- Berber Sardinha, T. 593
- Bergen Corpus of London Teenage Language
(COLT) 11; audio-visual corpus building 101;
corpora 111; corpus-informed course book
design 414; grammar 234; sociolinguistics 551,
555, 557
- Berkenkotter, C. and Huckin, T. 262
- Bernardi, S. 8
- Bernardini, S. 319, 326, 348, 349, 359, 360, 363,
365, 372, 377, 380, 462, 477
- Bernardini, S. and Castagnoli, S. 512
- Bernardini, S. and Zanettin, F. 496–97
- Bevitori, C. 594, 601
- BFSU Chinese-English Parallel Corpus 120
- Bhatia, V.K. 255, 258, 262, 267
- Bhatia, V.K. *et al.* 449–50
- Bianchi, F. and Pazzaglia, R. 450
- Biber, D. and Barbieri, F. 245
- Biber, D. and Burges, J. 247
- Biber, D. and Conrad, S. 148, 237, 241, 246,
247
- Biber, D. and Ebrary, I. 374
- Biber, D. and Finegan, E. 244, 247, 291
- Biber, D. and Jones, J.K. 247
- Biber, D. and Kurjian, J. 625
- Biber, D. and Reppen, R. 322, 323
- Biber, D., Conrad, S. and Cortes, V. 272
- Biber, D., Csomay, E. *et al.* 273, 276
- Biber, D. *et al.* 7, 22, 39, 54, 198, 260, 291, 387,
403, 430, 444, 536, 613, 623, 626, 637;
building for variety of a language 81, 87, 89;
corpus analysis 125, 126, 130; course book
design 413, 417, 419; creativity 303; discourse
271, 272, 279; English for Academic Purposes
(EAP) 459, 460, 461; grammar 228, 229, 230,
232, 233, 234, 235, 236, 237; language
corpora, preparation of learners for 374;
language teaching 323, 333; multi-word units
213, 214, 216, 217, 221; registers and genres
241, 243, 244, 245, 246, 247; teachers'
research 472, 473, 476, 477
- Biber, Douglas xv, 15, 21, 234, 235, 241–48,
323, 398, 430, 579; building for variety of a
language 82, 83, 85, 87, 89; concordances
176, 177; corpus analysis 123, 128, 130;
corpus building, key considerations 31, 32;
discourse 270, 272, 275; political discourse
595, 597; small specialist corpora building
69, 70; specialist genres 257, 260;
teachers' research 473, 474; written corpus
building 56, 58
- bilingual corpora 23–24, 360, 367, 496
- Bird, S. and Liberman, M. 47
- Birdwhistell, R.L. 298
- Birmingham Collection of English Texts 24
- Black English Sentence Scoring (BESS) 187
- Blair, Tony 311, 571, 573, 593, 596
- Blake, N. 532
- Blake, William 9, 516, 517, 521, 522
- Bloch, J. 446–47
- Blunkett, David 571
- Bolton, K. *et al.* 38
- Bondi, M. 272
- Boote, D.N. 622
- Bosch, A.V.D. and Daelemans, W. 312
- Bottrall, M. 516
- Boulton, A. 362, 363, 365, 366, 367, 452
- Bowker, L. and Barlow, M. 513
- Bowker, L. and Pearson, J. 324, 374, 375, 495,
498, 508, 513
- Bowra, C.M. 516, 521
- Bradley-Stevenson, C. and Mumford, J. 609
- brain scans 142
- Braun, S. 115, 320, 323, 324, 359, 360, 380
- Brazil, D. 298
- Breck, Eric 98
- Breen, M. 403
- Brentzen, K. and Vangsnes, Ø.A. 559
- Breyer, Y. 328, 361, 362, 471
- British Academic Spoken English (BASE):
corpora 114; English for Academic Purposes
(EAP), corpora and 459; language testing 637;
specialist genres 260; teacher education 624
- British Academic Written English (BAWE):
corpora 116, 117; English for Academic
Purposes (EAP), corpora and 459; language
testing 637; specialist genres 260; teacher
education 624; vocabulary teaching materials
391
- British Council 408
- British Medical Association (BMA) 609
- British National Bibliographic Cumulated Subject
Index (1960–1964)* 84
- British National Corpus (BNC) 637; Baby
Corpus 176, 177, 178; BNC-XML 502–3,
504, 506, 508; concordances 175, 176–77,
178–79; corpora 108, 109–10, 110–11, 112,
114, 118; corpus-informed course book design
414; dictionaries, use of corpora in writing
434; forensic linguistics 578, 583, 586;

- grammar 234, 238; healthcare communication 608–9, 613–14; language testing 637; lexis 198; media discourse 570, 572; parallel and comparable corpora 494; poetry and drama, language of 524; political discourse 591, 592, 597–98; research, teachers' use of corpora in 473, 476–77; small specialised corpora building 77; sociolinguistics 551, 552, 554–55, 555; specialist genres 256; spoken corpus building 38, 45, 48; teacher education 625; variety of a language, corpus building to represent 83, 84, 85, 86; written corpus building 55–56, 58, 59, 60
- British Received Pronunciation (RP) 111
- British Sign Language 94
- Bromiley, G.W. 3, 168
- Brook, G.L. 532
- Brooks, L. 641
- Brown, B. *et al.* 609
- Brown, G. and Yule, G. 271
- Brown, P. and Levinson, S. 266, 424
- Brown, R. 185
- Brown Corpus: corpora 109; corpus linguistics (CL) 15–16, 21, 22, 24; forensic linguistics 578; parallel and comparable corpora 488; registers, genres and 241; variety of a language, corpus building to represent 84, 85, 86; written corpus building 56
- Bruthiaux, P. 246
- building a corpus: basic considerations 31; data requirements 31–32; file formats 33–34; further reading 36; future considerations 36; headers 33; HTML codes 35; laughter, transcription of 35; mark-up requirements 35–36; parts-of-speech (POS) tags 35–36; recording files (spoken corpora) 34; text collection 32–35; text definition 32–33; text saving 33; transcription of files (spoken corpora) 34–35; *see also* spoken corpus building; written corpus building
- Bulgarian, Brown Corpus of 119
- Bunton, B. 247
- Burges, J. 233
- Burkette, A. 533, 534
- Burnard, L. 42, 45, 50, 51, 58, 60
- Burrows, J. 9
- Burstein, J. *et al.* 640
- Busa, Father Roberto 4, 15
- Bush, George W. 311, 594, 596
- Business English Corpus (BEC) 55–56, 58, 59, 62, 64
- Business English Lexis 173–75
- Business Letters Corpus (BLC) 114
- Business word list 459
- business writing 448–49
- Byrd, Pat 466
- C-ORAL-ROM project 25
- Caldas-Coulthard, C.R. 282
- Callen, M. 615–16
- Cambridge Advanced Learner's Dictionary* 438
- Cambridge and Nottingham Business English Corpus (CANBEC) 72, 74–75, 258, 266
- Cambridge and Nottingham Corpus of Discourse in English (CANCODE): corpora 111; corpus-informed course book design 414; English for Academic Purposes (EAP), corpora and 463; grammar teaching materials 463; multi-word units 217; registers, genres and 241; research, teachers' use of corpora in 477; sociolinguistics 559; specialist genres 258, 260, 264–66; spoken corpus building 38, 41; variety of a language, corpus building to represent 82, 84–85, 86, 87, 89
- Cambridge English for Speakers of Other Languages (ESOL)* 633, 634, 637
- Cambridge ESOL Certificate of Proficiency in English: corpus-informed course book design 415, 416; language testing 633, 634, 637, 639; research, teachers' use of corpora in 475–76
- Cambridge Grammar of English*: grammar 228, 234; registers, genres and 243
- Cambridge International Corpus (CIC): corpora 6; dictionaries, use of corpora in writing 437; English for Academic Purposes (EAP), corpora and 463; research, teachers' use of corpora in 473, 477; small specialised corpora building 66, 68; specialist genres 256; written corpus building 55
- Cambridge Learner Corpus (CLC): corpus-informed course book design 415, 416; English for Academic Purposes (EAP), corpora and 463; language testing 636, 639; small specialised corpora building 66, 68
- Cameron, D. 40, 41, 42, 44
- Cameron, D. and Kulick, D. 616
- Camlex New Words Service 437
- Campbell, S. 504
- Campoy-Cubillo, M.C. and Silvestre López, A.J. 327
- Canadian Hansard* 23
- Candlin, C. and Hyland, K. 259, 262
- Candlin, S. 606
- Canepari, L. 45
- Capel, A. and Sharp, W. 416
- Carleton, William 540
- Carletta, J. *et al.* 99
- Carroll, Lewis 168, 169
- Carter, C. 302, 303, 305, 308, 309, 310
- Carter, C. and McCarthy, M.J. 264
- Carter, R.A. and Adolphs, S. 36, 39, 298, 299, 425, 627
- Carter, R.A. and McCarthy, M.J. 7, 8, 11, 25, 32, 39, 66, 131, 228, 229, 234, 236, 238, 243,

- 244, 246, 272, 323, 403, 414, 415, 419, 471, 476, 623, 626
- Carter, R.A. *et al.* 404
- Carter, Ronald A. xv, 8, 44, 48, 132, 152, 154, 208, 213, 215, 216, 219, 228, 302–13
- case sensitivity tools 433
- case studies: corpus-based CDA 567–75; determination of corpus size 55–56; Irish family discourse 89–90; Limerick Corpus of Irish English (LCIE) 88–89; Published Materials Corpus (PMC) 60; turn-taking 132–33
- Castagnoli, S. *et al.* 511
- categorisation of corpora 107–9
- CEDEL2 (L1 English–L2 Spanish) 639
- Celce-Murcia, M. and Olshtain, E. 270, 274
- Center for Advanced Language Proficiency Education and Research (CALPER) 54, 188–89, 640
- Centre for English Corpus Linguistics (CECL) 636; corpus linguistics (CL) 25; language testing 636, 638; teacher education 625
- Certificate of Italian as a Foreign Language (CILS) 638–39
- Chafe, W. *et al.* 82
- Chambers, A. and Le Baron, F. 350
- Chambers, A. and O'Riordan, S. 624
- Chambers, A. and O'Sullivan, Á. 7, 320, 326, 340, 452
- Chambers, A. and Rostand, S. 353
- Chambers, A. and Sripicharn, P. 333
- Chambers, A. and Wynne, M. 356
- Chambers, Angela xv–xvi, 7, 9, 158, 319, 320, 333, 345–56, 360, 365, 453, 467, 626
- Chambers, J.K. 548
- Channell, J. 291
- Charles, M. 216, 244, 349, 363, 364, 374, 447, 449
- Charteris-Black, J. 565, 593, 599, 600
- Chaski, C. 587
- Chatwin, Bruce 506
- Chaucer, Geoffrey 532
- Chemnitz Corpus of Specialised and Popular Academic English (SPACE) 114
- Chen, Q. and Ge, G. 459
- Cheng, W. and Warren, M. 38, 324, 328, 413
- Cheng, W. *et al.* 8, 35, 73, 81, 218, 220, 320, 326, 348, 349
- Cheng, Winnie xvi, 9, 213, 221, 223, 319–29, 364
- Cheshire, J. and Fox, S. 559
- Cheshire, J. *et al.* 558, 559
- Chiaro, D. 310
- Child Language Data Exchange System (CHILDES): audio-visual corpus building 98; corpora 116; language development, corpus software in research 186, 187, 188
- Chilton, Paul 564
- Chinese Academic Written English (CAWE) 117
- choice of suitable corpus: research, teachers' use of corpora in 474; vocabulary teaching materials 387–88
- Chomsky, Noam 14, 303, 555
- Christie, F. and Martin, J. 262
- Christie, John 581
- CHRISTINE Corpus 112
- Chujo, K. *et al.* 451
- Chung-Ling, S. 511
- chunks of language 101, 137, 138, 213, 334, 336, 398, 419–20, 441, 586; lexical chunks 95, 363, 460, 622
- Church, K.W. and Hanks, P. 435
- Cignoni, L. and Coffey, S. 310
- City University Corpus of Academic Spoken English (CUCASE), Hong Kong 114
- Clancy, Brian xvi, 80–91, 595
- classrooms: and corpora, effective connection of 404; corpora inside and outside 476–78; language in, corpus of 623–24
- Clear, J. 23, 57, 473
- co-occurrences of language forms 154, 155, 212, 218, 273, 305, 307, 435, 752
- Coates, J. 281
- Cobb, Tom 337, 338, 348, 459, 464
- COBUILD project: corpora 5, 6, 109; corpus linguistics (CL) 8; forensic linguistics 583; grammar 229; lexis 197; spoken corpus building 44; written corpus building 56
- coding spoken data 47–48
- Coh-Metrix 642
- Cohen, B. 136
- Cohen Minnick, L. 533, 534
- cohesive devices 340–41
- Collins, H. 364
- Collins, P. 245
- collocation: creativity in 306–7; dictionaries, use of corpora in writing 435–36; and frames in media discourse 568–69; of keywords in media discourse 565–66; lexis 199–200; multi-word units 212; pattern exploration 162–63; queries in translation 504–5
- combinational functions 447
- common errors, identification of 439
- Common European Framework of Reference (CEFR): corpus-informed course book design 415; corpus linguistics (CL) 8; language testing 634, 635
- common sampling frame, importance of 488
- Communicative Language Ability 635
- COMPARA Corpus 120, 488, 511
- comparable corpora *see* parallel and comparable corpora
- Compleat Lexical Tutor 459, 464, 625

- complexity 186, 236, 237, 299, 366, 367, 434;
of legal discourse 449–50; measures of 190;
semantic complexity 558; simplicity-
complexity dimension 187; syntactic
complexity 190, 555; teacher-learner
dynamics 407
- computers: access to 473–74; Computer
Mediated Communication (CMC) 626;
computerised corpus analysis 387;
Computerised Language Analysis (CLAN) 98,
186, 187; computerising written texts 60–63;
Computers in the Human Interaction Loop
(CHIL) 97, 102; concordances generated by
169–75; corpus software and computer
learning 141; deficiencies of 139–42;
expectations of 141–42; functionality of
136–39; speech of 141
- ConcGram* 218, 223, 327
- concordances 167–82; *Alice in Wonderland*
(Carroll, L.) 169–70, 171, 173; of average,
sample task based on 396–98; British National
Corpus (BNC) 175, 176–77, 178–79; Business
English Lexis 173–75; collocation and
semantic significance 568–70; computer
generated 169–75; concordancing tools and
dictionaries, use of corpora in writing 431–32;
consolidation of data 180; corpus analysis 122;
corpus data, working with 175–82; corpus
software 147; definition 167; exploration of
lines 128–30; further reading 182; initiation
work on data 179; interpretation of data
179–80; Key Word in Context (KWIC) 167,
171–72; language corpora and preparation of
learners for using 377–78; lemmas 172–73;
limitations of 182; lines, reading and
interpretation of 380–81; output of, multiple
lines of 444; *Oxford English Dictionary* 170;
parts-of-speech (POS) tags 177; phrases
173–75; pre-computers 167–69; reading
concordances 178–82; recycling data 181;
repeating work on data 182; reporting on data
180–81; restricted searches 176–78; results of
work on data 181; sampling 176; simple
concordances 171–72; sorting 175–76;
Through the Looking Glass (Carroll, L.) 175–76;
use in vocabulary teaching 395–98; wild-card
multi-word searches 173–75; word-forms
172–73; *Wordsmith Tools* (Scott, M.) 171
- Concordances in the Classroom* (Tribble, C. and
Jones, G.) 463
- confidentiality 44, 63, 256
- Coniam, D. 622, 626
- Connor, U. 247
- Connor, U. and Mauranen, A. 246
- Connor, U. and Upton, T.A. 246, 247, 262
- Connor, U. *et al.* 68, 71
- Connor-Linton, J. 247
- connotation 205–6, 312, 353–54, 373, 506, 507,
593, 598, 622
- Conrad, Joseph 169, 274
- Conrad, S. and Biber, D. 154, 244
- Conrad, Susan xvi, 4, 9, 39, 227–39, 246,
247, 347, 356, 401, 404, 410, 437, 459–60,
463, 473
- consent 32, 42–43, 96, 102, 475
- consolidation of data 180, 323–24
- constraints: on corpora use in grammar teaching
406–8, 409–10; on corpus software 136–42;
on use of corpora in forensic linguistics 579
- contagion, metaphors for 614
- context: and corpus, links between 66–67, 74;
data-driven learning (DDL) 345–46; and
meaning 202–3; relevance in teacher
education 623, 624; understanding grammar
through 227–29; of use in corpus-informed
course book design 420
- contrastive corpora 21–22, 24
- contrastive linguistics: corpus use for 373;
examination of corpora 624–25; parallel and
comparable corpora 492–93
- conversation: Conversation Analysis (CA) 9;
conversational organisation 293–95;
management syllabus 419–20; natural-
sounding conversation, replication of 423–24;
registers, genres and 244; spontaneous
conversation 548–49, 550; unedited corpus
conversations, problems of 423
- conversion of texts: by keyboarding 62–63; by
optical scanning 62
- Cook, G. 8, 35, 44, 308, 309, 395
- Cook, V. 227, 415
- Corbett, G. and Rogers, M. 346
- Corpas Pastor, G. and Seghiri, M. 508, 513
- corpora: access to 107–9; American National
Corpus (ANC) 110; ARCHER Corpus 113;
archives 107; Australian Corpus of English
(ACE) 109; availability of 107–20; Babel
Chinese-English Parallel Corpus 120; Bank of
English 109–10; BE06 Corpus 109;
behaviourism 14; Bergen Corpus of London
Teenage English (COLT) 111; BFSU
Chinese-English Parallel Corpus 120; bilingual
corpora 23–24; British Academic Spoken
English (BASE) 114; British Academic
Written English (BAWE) 116, 117; British
National Corpus (BNC) 108, 109–10,
110–11, 112, 114, 118; British Received
Pronunciation (RP) 111; Brown Corpus 109;
building of, preparation of learners in 375;
Bulgarian, Brown Corpus of 119; Business
Letters Corpus (BLC) 114; Cambridge and
Nottingham Corpus of Discourse in English
(CANCODE) 111; Cambridge International
Corpus (CIC) 6; categorisation of 107–9;

Chemnitz Corpus of Specialised and Popular Academic English (SPACE) 114; Child Language Data Exchange System (CHILDES) 116; Chinese Academic Written English (CAWE) 117; CHRISTINE Corpus 112; City University Corpus of Academic Spoken English (CUCASE), Hong Kong 114; of classroom language 623–26; COBUILD project 5, 6, 109; COMPARA Corpus 120, 488, 511; comparable multilingual corpora 119–20; for comparison 21–22; contrastive corpora 21–22; contrastive corpus 24; corpus developments 6; Corpus of Contemporary American English (COCA) 110; Corpus of Early English Correspondence (CEEC) 113; Corpus of English Dialogues (CED) 113; Corpus of Historical American English (COHA) 113; Corpus of Late Eighteenth Century Prose 113; Corpus of late Modern English Prose 113; *CorpusSearch* 113; creativity and 303–5; Czech National Corpus 118; development of new types in political discourse 602; developmental language corpora 116; diachronic corpora 20, 22–23; and discourse in corpus analysis 130; distribution sites 107; Dutch Corpus Gesproken Nederlands 118; Early English Books On-line (EEBO) 113; Enabling Minority Language Engineering (EMILLE) Corpus 119; English as a Lingua Franca in Academic Settings (ELFA) 118; English for Academic Purposes (EAP) 114; English language corpora 109–16; English Language Interview Corpus as a Second-Language Application (ELISA) 115; English-Norwegian Parallel Corpus (ENPC) 120; English-Swedish Parallel Corpus (ESPC) 120; frameworks for grammar teaching 401–2; Freiburg-Brown Corpus of American English (FROWN) 16, 22, 109; Freiburg Corpus of English Dialects (FRED) 111; Freiburg-LOB Corpus of British English (FLOB) 109; further reading 120; ‘general language’ corpora (spoken, written and both) 109–10; geographical corpora 21; German COSMAS Corpus 118; German National Corpus 118; and grammar teaching, benefits and types of use 402–6, 409–10; Hellenic National Corpus 118; Helsinki Corpus of English 113; historical corpora 21, 113–14; historical origins of 3–4; HKIED English-Chinese Parallel Corpus 120; Hong Kong Corpus of Conversational English (HKCCE) 117; Hungarian National Corpus 118; IJS-ELAN Slovene-English Parallel Corpus 120; importance of role in writing instruction 445; Innsbruck Computer Archive of Machine-Readable English Texts

(ICAMET) 113; inside and outside the classroom 476–78; International Computer Archive of Modern and Medieval English (ICAME) 107–8; International Corpus of Crosslinguistic Interlanguage (ICCI) 117; International Corpus of English (ICE) 109, 112–13, 117, 118; International Corpus of Learner English (ICLE) 116–17; Intonational Variation In English (IVIE) 111; Ireland, New Corpus of (NCI) 119; issues in building and use of 472–74; Italian CORIS/CODIS Corpus 118; Japan, Balanced Corpus of Contemporary Written Japanese (BCCWJ) 119; Japanese EFL Learner Corpus (JEFL) 117; Jacek English-Czech Corpus 120; Key Word in Context (KWIC) 4; Kolhapur Corpus of Indian English 109; Korean National Corpus (Sejong Balanced Corpus) 118; KWICFinder 116; Lampeter Corpus 113; Lancaster Corpus of Academic Written English (LANCAWE) 117; Lancaster Corpus of Mandarin Chinese (LCMC) 119; Lancaster-Leeds Treebank 112; Lancaster-Los Angeles Spoken Chinese Corpus (LLSCC) 119; Lancaster Oslo-Bergen (LOB) corpus 109, 112, 114; Lancaster Parsed Corpus (LPC) 112; and language awareness in teacher education 626–27; and language learning 319–21; Learner Business Letters Corpus (LBLC) 117; Limerick-Belfast Corpus of Academic Spoken English (LIBEL CASE) 114; Limerick Corpus of Irish English (LCIE) 111; lingua franca corpora 118; Linguistic Data Consortium (LDC) 107–8; Literature Online (LION) 113; London-Lund Corpus of Spoken English (LLC) 111; Longman Corpus Network 112; Louvain Corpus of Native English Essays (LOCNESS) 116; Louvain International Database of Spoken English Interlanguage (LINDSEI) 116–17; LUCY Corpus 112, 116; Machine Readable Spoken English Corpus (MARSEC) 111; Michigan Corpus of Academic Spoken English (MICASE) 112, 114; Michigan Corpus of Upper-Level Student Papers (MICUSP) 116, 117; Modern Chinese Language Corpus (MCLC) 118, 119; modern corpora, creation of 4–5; monitor corpora 20, 22, 23; *Monoconc* (Barlow, M.) 5; monolingual non-English corpora 118–19; Monroe Corpus 112; Montclair Electronic Language Learners’ Database (MELD) 117; MULTEXT-East Corpus 120; multilingual corpora 19–20, 23–24, 118; multimedia corpora, multimodal texts 114–15; Mutual Information (MI) 6; National Sound Archive 111; native speaker intuition or? 428–29; Newcastle Electronic Corpus of Tyneside

- English (NECTE) 111; Newdigate Newsletters Corpus 113; non-English corpora 118–20; non-native speaker corpora 25; normativeness 24; Old Bailey Corpus 113; Open Language Archives Community (OLAC) 108; origins of 14–15; Oslo Corpus of Bosnian Texts 119; Oslo Multilingual Corpus (OMC) 120; Oxford Concordance Program (OUP, 1987) 5; *Oxford English Dictionary* quotation database 114; Oxford Text Archive (OTA) 107–8; parallel multilingual corpora 119–20; parsed corpora 112–13; Parsed Corpus of Early English Correspondence 113; Peking University corpora 119; Penn-Helsinki Parsed Corpus of Middle English (PPCME2) 113; Polish National Corpus 118; political discourse and 592–95; Polytechnic of Wales (POW) Corpus 116; Portuguese Corpus do Português 118; positivism 14; Preparatory Action for Linguistic Resources Organisation for Language Engineering (PAROLE) 118; Professional English Research Consortium (PERC) 114; and punning 312; quantitative and qualitative revolution in 17–18; Reading Academic Text (RAT) corpus 114; relevance in sociolinguistics 547–49; Russian Reference Corpus (BOKR) 118; Saarbrücken Corpus of Spoken English (SCoSE) 112; SACODEYL project 115; sample corpora 20; Santa Barbara Corpus 111–12, 115; SCOTS Corpus 111, 115, 119; Singapore Corpus of Research in Education (SCoRE) 94–95, 101; Slovak National Corpus 118; Spanish Corpus del Español 118; special corpora 22; specialised corpora 114; of specialised genres (CSGs) 255–58, 259, 264–65, 266–67; speech corpora 110–12; spoken corpora 25–26; Spoken English Corpus (SEC) 111; SUSANNE Corpus 112; Switchboard Corpus 111; Taiwanese Mandarin Balanced Corpus 119; technology, influence in development of 5, 15–17; text collections 107; text interpretation and evaluation 502–4; time dimension, corpora along 22–23; and tools for writing instruction 450–51; topic corpora 21; types of 107; UCLA Chinese Corpus (UCLACC) 119; Vienna-Oxford Corpus of International English (VOICE) 118; web as corpus 115–16; WebCorp 115; Wellington Corpus of Written New Zealand English 109; Welsh Cronfa Electroneg o Gymraeg 119; Wolverhampton Business English Corpus 114; *Wordsmith Tools* (Scott, M.) 5, 6; York-Helsinki Parsed Corpus of Old English Poetry 113; ZJU Corpus of Translational Chinese (ZCTC) 119; Zurich English Newspaper Corpus (ZEN) 113
- corpus analysis: basic issues 122–23; case study, turn-taking 132–33; concordance lines, exploration of 128–30; concordances 122; corpora and discourse 130; corpus software 123; corpus techniques, combination with other approaches 132–33; corpus techniques, exploitation of 130–32; data set size 122–23; discourse, exploration of 130–33; display of frequency data 123–25; exploitation of basic corpus techniques 130–32; exploitation of frequency data 125; frequency analysis 123; frequency lists 122; frequency lists, comparison of 125–26; further reading 133–34; key-word lists, exploration of 127–28; keyness 127; language corpora, tools of 377–80; lexicography, tools for 431–33; manipulation of corpus data 122; negative key words 128; normalisation 126; online concordancing 128–29; parallel and comparable corpora 491; positive key words 127–28; random sampling 123; searching concordance lines 129–30; of sexual health and adolescent language 610–16; size, considerations of 122–23; sorting concordance lines 129–30; sub-corpora 123; triangulation of findings 123; word frequency lists, exploration of 123–26
- corpus-based: approach in EGAP 445–46; approach in ESAP 448–49; case study – Blake's *Songs of Innocence and of Experience* 516–22; case study – blockbuster movie scripts 523–29; case study – stylistic analysis 536–41; Critical Discourse Analysis (CDA) 563–67; cultural keywords 565–66; discourse 271–75; materials for language teaching 333–34; pragmatics, restriction in study of 289–91; qualitative analysis 570–74; research on political discourse 598–601; sociolinguistics 558–60; studies in healthcare communication 606–7; study of specialist genres 255–56
- corpus choice: corpus-informed course book design 413–15; translation, use of corpora in 504–11
- corpus-comparative statistical keywords (CCSKs) 566, 570–71, 571–72, 575
- corpus compilation: existing corpora and 488–90; key issues in 489–90
- corpus data: preparation of 374–77; preparation of learners for use of 374–75; uses in corpus-informed course book design 422–24; working with concordances 175–82
- Corpus Encoding Standard (CES) 48
- corpus-informed course book design 413–26; Academic Wordlist 416; age of readers, considerations on 414; American National Corpus (ANC) 414; Bergen Corpus of London Teenage English (COLT) 414; British National Corpus (BNC) 414; Cambridge and

Nottingham Corpus of Discourse in English (CANCODE) 414; Cambridge ESOL 415, 416; Cambridge Learner Corpus (CLC) 415, 416; *The Cobuild English Course* (Willis, D. and Willis, J.) 413, 416, 465; *Collins COBUILD English Language Dictionary* 413, 428, 465; Common European Framework of Reference (CEFR) 415; contexts of use 420; conversation management syllabus 419–20; corpus choice 413–15; corpus data, uses for 422–24; corpus information, book areas which can benefit from 415–22; corpus-informed course materials, expansion in use of 424–25; course books, characteristics of 413; data-driven learning (DDL) 422; discourse management 418–20; electronic delivery, print *versus* 425; English as a Lingua Franca (ELF) 415; English for Academic Purposes (EAP) 416; English Language Teaching (ELT) 414; frequency, uses of 416; further reading 425–26; future for 424–25; genre of corpora, choice of 414; grammar syllabus 417; information about language 422; information about language syllabus 415–16; language, information about 422; learner and non-native-speaker data 415; lexico-grammatical patterns 417; multi-modal corpora, development of 425; multi-word items, classification of 419; natural-sounding conversation, replication of 423–24; permissions, seeking 417–18; presentation, use of corpus texts for 423–24; presentation and practice, models for 421–22; print *versus* electronic delivery 425; punctuation, problem of 424; requests 417; size of corpus 415; unedited corpus conversations, problems of 423; variety of corpora, choice of 414; Vienna-Oxford Corpus of International English (VOICE) 415; vocabulary syllabus 416–15; World Englishes 415

Corpus Linguistics and Linguistic Theory 37

corpus linguistics (CL): applications of 6–12; arithmetical power, qualitative change and 18; Bank of English 24; Bergen Corpus of London Teenage English (COLT) 11; Birmingham Collection of English Texts 24; Brown Corpus 15–16, 21, 22, 24; C-ORAL-ROM project 25; Centre for English Corpus Linguistics (CECL) 25; COBUILD project 8; Common European Framework of Reference (CEFR) 8; Conversation Analysis (CA) 9; Corpus of Translated Finnish (CTF) 24; corpus typology 20–26; Critical Discourse Analysis (CDA) 9, 11; data and theory, converging paths of 18; data-driven learning (DDL) 7; data scarcity 16; Diachronic Corpus of Present-Day Spoken English (DCPSE) 23;

digitisation 15–16; Discourse Completion Tasks (DCTs) 10; English as a Lingua Franca 25; English–Norwegian Parallel Corpus 24; English Profile project 8; evolution of 14–26; field linguists 15; Helsinki corpus 22; Hong Kong Corpus of Spoken English (HKCSE) 8; *Index Thomisticus* (Busa) 15; International Corpus of English (ICE) 21, 23; internet language 26; Jiao Da English for Science and Technology (JDEST) 16; journalistic material, privileging of 16; Key Word in Context (KWIC) 7, 19; Lancaster Oslo-Bergen (LOB) corpus 22; language teacher education (LTE) 8; and language teaching 327–29; Leuven Drama Corpus and Frequency List 22; Limerick Corpus of Irish English (LCIE) 10; London-Lund Corpus 23; Michigan Corpus of Academic Spoken English (MICASE) 25; Oslo Multilingual Corpus 24; *parole*, linguistics of 14–15, 20; PAROLE Corpus 24; performance, focus on 15; pragmatic markers 11; Survey of English Usage (SEU) 15, 16, 21, 23; theoretical shift from text-linguistics to 18–20; theory and data, converging paths of 18; *Wmatrix* 9, 517, 523

Corpus of American and British Office Talk (ABOT): small specialised corpora building 68, 69, 70, 72, 75; specialist genres 258, 263

Corpus of Contemporary American English (COCA) 110

Corpus of Early English Correspondence (CEEC) 113

Corpus of English Dialogues (CED) 113

Corpus of European Parliament Proceedings (EUROPARL) 593–94

Corpus of Historical American English (COHA) 113

Corpus of Late Eighteenth Century Prose 113

Corpus of late Modern English Prose 113

Corpus of Meetings of English Language Teachers (C-MELT) 478–80

Corpus of Spoken Professional American-English (CSPA) 594

CORpus of tagged Political Speeches (CORPS) 593

Corpus of Translated Finnish (CTF) 24

Corpus Oral de Lenguaje Adolescente (COLA) 552

Corpus Presenter 534, 535

corpus query system (CQS) 431, 434, 435, 441
corpus size: case study – determination of corpus size 55–56; corpus analysis and size, considerations of 122–23; corpus-informed course book design 415; determination for small specialised corpora building 67–68; for lexicographers 429; variety of a language, corpus building to represent 82–83;

- vocabulary teaching materials 392; written corpus building 54–56
- corpus software: access to 473–74; Artificial Intelligence (AI) 141; brain scan predictions by computers 142; comparison of strings 139–40; computer deficiencies 139–42; computer functionality 136–39; computer learning 141; computer speech 141; concordances 147; constraints in general 136–42; corpus analysis 123; data re-formatting 142–47; deficiencies of computers 139–42; expectations of computers 141–42; feedback on operations 141; functionality of computers 136–39; further reading 150; graphics 141; human-centric computing 141–42; key word lists 149–50; organisation (and re-organisation) of data 142–47; pattern recognition 140–41; predictions by computers 142; sorting data 142–47; string comparison 139–40; word lists 147–48
- corpus techniques: for advanced-level students 451–52; combination with other approaches 132–33; development of language testing 636–38; exploitation of basic techniques 130–32; in exploration of political discourse 595–98; integration into translators' toolkits 512; popularity of 238
- corpus typology 20–26
- CorpusSearch* 113
- Cortes, V. 214, 216, 221, 245, 398
- COSMOROE framework 99
- Cotterill, J. and Grant, T. 580
- Cotterill, Janet xvi, 10, 39, 77, 578–88
- Coulthard, R.M. 10, 214, 271, 581–82, 585–86
- Coulthard, R.M. and Johnson, A. 582, 587
- Council of Europe 634
- Coupland, J. 265
- Couture, B. 257
- Coxhead, A. and Byrd, P. 391, 452, 459, 464
- Coxhead, A. and Hirsh, D. 459
- Coxhead, A. and Nation, I.S.P. 374
- Coxhead, Averil xvii, 77, 125, 387, 388, 389, 391, 393, 396, 416, 458–67, 477, 638
- Crabtree, A. *et al.* 41
- CRATER Project 488
- Crawford, P. *et al.* 606
- creativity 302–13; collocations, creativity in 306–7; concept in linguistics of 302–3; corpora and 303–5; corpora and punning 312; creative structures, corpora and 313; departure from patterns and 305–8; in discourse, corpora in analysis of 305–9; further reading 313; humour, creativity and corpora 310–11; idiomaticity, creativity in 307–8, 311–12; novel word formations 305–6; patterning and 308–9; punning 308, 312; semantic annotation and creative idiomaticity 311–12; spoken creativity 309–10; studies of, corpora and facilitation of 310–13; word formation, novelty in 305–6; written corpus building and creation of new corpus, rationale for 54; written creativity 309–10
- Cresti E. and Moneglia, M. 25
- Creswell, A. 452
- Criminal Cases Review Commission (CCRC) 579
- Critical Discourse Analysis (CDA) 9, 11, 563–67
- Crookes, G. 272
- cross-modal communication in pragmatics 298
- Crossley, S.A. *et al.* 641–42
- Crowdy, S. 83, 430
- Cruden, A. 3, 168
- Crystal, D. 80, 90, 472, 615
- Crystal, D. *et al.* 187
- Csomay, E. 246, 260
- Cullen, R. and Kuo, I.-C. 413, 425
- Culpeper, J. 534, 541, 542
- Culpeper, J. and Kytö, M. 245, 247
- cultural keywords in media discourse 565, 566–67
- cultural processes, relationship with language 564
- Curado Fuentes, A. 452
- Cutting, J. 87, 260
- Czech National Corpus 118
- Dahlmann I. and Adolphs, S. 36, 38, 46
- Danet, B. and Herring, S.C. 246
- Danielsson, P. and Ridings, D. 490–91
- data: amounts, preparation of learners in dealing with large quantities 375–76; collection for audio-visual corpus building 96–97; concordances and interpretation of 179–80; data-driven or data-informed grammar 401; dataset comparability in sociolinguistics 557–58; diachronic data 109, 113, 215, 306, 430, 534–35, 540; entry in written corpus building 62–63; interpretation, caution in 382; manipulation of corpus data 122; organisation (and re-organisation) of 142–47; private data 61; publicly available data 61; questioning the data, preparation of learners in 376–77; re-formatting by corpus software 142–47; recycling data 181; repeating work on 182; reporting on 180–81; representational quality of 608; requirements for building a corpus 31–32; requirements for study of discourse 282–84; results of work on 181; scarcity of 16; set size for corpus analysis 122–23; spoken data 430; storage and retrieval 63; synchronic data 430; and theory, converging paths of 18; written corpus building 61; written data 430; *see also* corpus data
- data-driven learning (DDL) 345–56; actual language use and 349–52; context of 345–46;

- corpus-informed course book design 422;
 corpus linguistics (CL) 7; effectiveness of, assessment of 365–66; English for Academic Purposes (EAP), corpora and 463; further reading 356; history of 346–49; language pedagogy, changing and 354–56; material for language teaching and 360–62; multi-word units 222; operationalisation of 362–65; second language acquisition theory and 352–54; writing instruction, use of corpora for 444, 445
- Daubert criteria in forensic linguistics 587–88
- Daubert v. Dow Merrill Pharmaceuticals, Inc.* (509 U.S. 579 1993) 587
- Davies, M. 373
- Dawson, J. 169
- de Beaugrande, D. 55
- de Beaugrande, R. 548, 596
- De Cock, S. 338
- De Cock, S. *et al.* 38
- de Haan, P. 245, 402
- de Klerk, V. 238
- Deane, P. and Gurevich, O. 640
- Debrock, M. *et al.* 353
- definitions: concordances 167; discourse 270–71; healthcare communication 605–6; language testing 633–36; multi-word units 213–15; parallel and comparable corpora 487–88; political discourse 591–92; pragmatics 288–89; sociolinguistics 547; specialist genres 257
- Deignan, A. 11
- DeKeyser, R. 185
- design: corpus building to represent variety of a language, design issues 82–86; language test design 634–35; parallel and comparable corpora design 489–90; small specialised corpora building 69–71; spoken corpus design 39–42; vocabulary teaching materials 394–98; *see also* corpus-informed course book design
- Deutschmann, M. 551
- Developmental Sentence Scoring (DSS) 185, 187
- diachronic corpora 20, 22–23, 87, 597, 599;
 Diachronic Corpus of Present-Day Spoken English (DCPSE) 23
- diachronic data 109, 113, 215, 306, 430, 534–35, 540
- dialectal forms, treatment of 553–54
- dialectal features 564, 581
- Diani, G. 290
- Dickens, Charles 144, 307, 487
- dictionaries, use of corpora in writing 428–42;
 British National Corpus (BNC) 434;
 Cambridge International Corpus tools 437;
 Camlex New Words Service 437; case sensitivity tools 433; collocation 435–36; common errors, identification of 439;
 concordancing tools 431–32; content of corpus for lexicographers 430; corpora, native speaker intuition or? 428–29; corpus analysis tools for lexicography 431–33; corpus make-up, limitation of 440; corpus query system (CQS) 431, 434, 435; corpus query systems 441; diachronic data 430; dictionary text, use of corpora in production of 433–39; errors, identification of common 439; expansion of context, tools for 433; filtering tools 433; frequencies 438–39; further reading 442; future developments 440–42; genre divisions for lexicographic corpora 430; grammar 437; instances, tools for exclusion of occurrences 433; language genres, differences between 437; learner corpora, use for lexicography 439–40; learner profiles 439–40; lemmatisation tools 432–33; lexicographer, role of 441–42; lexicographers' use of corpora 428–31; limitations 440–42; multi-word unit searches, tools for 433; multilingual aligned corpus 431; Mutual Information (MI) 435; native speaker intuition, corpora or? 428–29; native speaker use, comparisons with 440; neologisms and new senses, identification of 436–37; part-of-speech tagging tools 433; personal prejudice 429; random selection of lines, tools for 433; regional language variety 430; semantic division 433–35; sentences, examples of 438; size of corpus for lexicographers 429; speaker language 430; spoken corpora, limitation of 440–41; spoken data 430; synchronic data 430; syntax 437; written data 430
- digitisation 15–16; abundance of electronic texts 563; electronic delivery, print *versus* 425; *see also* computers; corpus software
- Ding, H. 247
- discipline-specificity 396–98
- discourse 270–84; corpus and 271–75; creativity in, corpora and analysis of 305–9; data requirements for study of 282–84; definition of 270–71; discourse analysis 9; Discourse Completion Tasks (DCTs) 10; discourse-level factors in grammatical patterns 232–33; exploration of, corpus analysis and 130–33; further reading 284; limitations in corpus use for study of 275–76; management for book design 418–20; markers in pragmatics 295–96; practice of corpus-based approach 276–82; teacher education, discourses of 627–29
- discovery learning (DL) 320, 326, 363, 372
- Divay, M. and Vitale, A.J. 312
- documentation: linguistic authority of 505–6, 508; translation, use of corpora in 501
- Dodd, A. 502
- Dörnyei, Z. 472

- Dos Santos, V.P.M.P. 247
 Douglas, F. 39, 54, 57
 Downing, A. 11
 Doyle, Roddy 536
 drafting for translation 502
 drama *see* poetry and drama, language of
 Dunning, T. 127, 149, 566
 Durán, P. *et al.* 187
 Duranti, A. and Goodwin, C. 547
 Durrant, Philip xvii, 9, 387–98, 638
 Dutch Corpus Gesproken Nederlands 118
 Dyvik, H. 493
- Eagleson, R. 583
 Early English Books On-line (EEBO) 113
 Ebeling, J. 492
 Eckert, P. and Rickford, J.R. 547
The Economist 497
 Edge, J. and Richards, K. 621
 Educational Testing Services (ETS) 637, 640
 Edwards, J. 44
 Eggins, S. and Slade, D. 264, 476
 Ekbal, A. and Bandyopadhyay, S. 61
 Ekman, P. and O'Sullivan, M. 299
ELAN 549–50
 Eliot, T.S. 4, 169
 Ellis, N. and Larsen-Freeman, D. 283
 Ellis, N.C. 354, 389
 Ellis, N.C. *et al.* 283, 391, 460
 Enabling Minority Language Engineering (EMILLE) Corpus 119
Encyclopaedia Britannica 408
 Engels, L.K. 22
 English as a Foreign Language (EFL) 620, 623, 624, 626; learner corpora 116–17
 English as a Lingua Franca (ELF): in Academic Settings (ELFA) 118; corpus-informed course book design 415; corpus linguistics (CL) 25; English for Academic Purposes (EAP), corpora and 459
 English as a Second Language (ESL) 235, 238, 283, 328, 423, 467, 478; learner corpora 116–17; pedagogy 462; textbooks 231, 323, 459, 460; writing samples 190
 English for Academic Purposes (EAP): Academic Formulas List 460; academic language in use, corpora and aspects of 458–60; Academic Word List (AWL) 459, 464–65; British Academic Spoken English (BASE) 459; British Academic Written English (BAWE) 459; Business word list 459; Cambridge and Nottingham Corpus of Discourse in English (CANCODE) 463; Cambridge International Corpus of written and spoken English 463; Cambridge Learner Corpus (CLC) 463; *Compleat Lexical Tutor* 459, 464; corpora and 114, 458–67; corpus-based studies 460; corpus-informed course book design 416; data-driven learning (DDL) 463; English as a Lingua Franca 459; further reading 466–67; future for corpora in EAP 466–67; Hong Kong University of Science and Technology corpus of essays 459; International Corpus of Learner English (ICLE) 459; learner independence, sustainability of 462; learner language, corpora and insights on 465–66; lexical bundles in corpus studies 461; materials of, use of corpora in 463–65; Michigan Corpus of Academic Spoken English (MICASE) 459, 460, 463; multi-media corpora for 466; multi-word units 221; parallel corpora for 466–67; pedagogy of, influence of corpora on 461–63; Science-specific EAP word list 459; small specialised corpora building 67, 68; TOEFL 2000 Spoken and Written Academic Language Corpus (T2K-SWAL) 459, 460; vocabulary teaching materials 391, 392–93, 394, 398; *Wordsmith Tools* (Scott, M.) 459
- English for Academic Purposes (EAP), corpora and: *Academic Vocabulary in Use* (McCarthy, M. and O'Dell, F.) 463–64; *Better: A Surgeon's Notes on Performance* (Gawande, A.) 458; *Focus on Vocabulary* (Schmitt, D. and Schmitt, N.) 464; *Longman Dictionary of Contemporary English (LDOCE)* 465; *Macmillan Dictionary for Advanced Learners* (Rundell, M.) 465
- English for General Academic Purposes (EGAP): corpus-based approach in 445–46; corpus-driven approach in EGAP 446–47; writing instruction, exploitation of corpora for 445–47
- English for Specific Academic Purposes (ESAP): corpus-based approach in 448–49; corpus-driven approach in 449–50; writing instruction, exploitation of corpora for 448–50
- English for Specific Purposes (ESP): multi-word units 221; small specialised corpora building 67, 68
- English language corpora 109–16
- English Language Interview Corpus as a Second-Language Application (ELISA) 115
- English Language Teaching (ELT): corpus-informed course book design 414; publications and grammar teaching materials 402, 403–4, 404–5, 406, 407, 408
- English-Norwegian Parallel Corpus (ENPC) 24, 120, 488, 493
- English Profile Programme 8, 638
- English-Swedish Parallel Corpus (ESPC) 120
- Environmental Impact Assessment, Corpus of (EIA) 68
- error correction: corpus use for 373; identification of common errors 439;

- investigation, materials for language teaching 341–43; materials for language teaching 340
- Ervin-Tripp, S.M. 527
- ethics: spoken corpus building 42–44; written corpus building 63
- euphemism in political discourse 598–99
- European Corpus Initiative 488
- European Union (EU) 23–24; expansion and resulting immigration 567, 568–69, 573–74
- evaluation in pragmatics 291–92
- Evans, Timothy 581
- evidence-based: design 608; language teaching 259; screening tools 616; sex education 616
- Evison, Jane xvii, 9, 48, 77, 122–34, 147, 162, 200, 260, 473, 502, 566, 595
- Examination for the Certificate of Competency in English (ECCE) 639
- Examination for the Certificate of Proficiency in English (ECPE) 641
- EXMARaLDA project 99, 102
- expansion: of context, tools for 433; of corpora for writing instruction 452–53
- Exploring Academic English* (Thurstun, J. and Candlin, C.N.) 445
- Fairclough, N. and Wodak, R. 564
- Fairclough, Norman 267, 270, 281, 564, 565, 569, 593, 598
- Farr, F. and O'Keeffe, A. 67, 88
- Farr, F. *et al.* 38, 88, 90, 355, 535, 626
- Farr, Fiona xvii, 8, 132, 356, 471, 478, 537, 620–29
- FBI, unabomber investigation 584–85
- feedback 88, 323, 325, 335, 361, 393, 452, 478; on corpus software operations 141; formative feedback 640; Initiation-Response-Feedback (IRF) 623–24; teaching practice (TP) 627–28
- Fellbaum, C. 312
- Ferguson, G. 241
- Fernando, C. 307, 309
- Ferraresi, A. 116
- field linguists 15
- file formats 33–34
- Fill, A. and Mühlhäusler, P. 283
- Fillmore, C.J. 335
- Film Studies 522
- filtering tools 433
- Financial Times* 497
- First Certificate in English (FCE) 641
- first language development: corpus deployment in revealing 186–88; theories of 184
- Firth, J.R. 203, 212, 215, 216
- Fitzgerald, J.R. 585
- Fitzmaurice, S.M. 238, 245, 246, 247
- fixed expressions 201, 202, 205, 221, 307, 308, 310, 334, 336
- Fletcher, J. *et al.* 558
- Fletcher, W.H. 213, 218, 279, 512
- Fligelstone, S. 319, 356
- Flowerdew, J. 255, 256, 259, 260, 262, 360, 374
- Flowerdew, J. and Dudley-Evans, T. 247
- Flowerdew, J. and Wan, A. 247
- Flowerdew, Lynne xviii, 9, 55, 66, 67, 68, 70, 74, 76, 77, 272, 274, 444–54, 459, 466, 608, 638
- fluency, measures of 190
- Forchini, P. 216
- forensic linguistics 10, 578–88; Al-Qaeda 580; authorship analysis, potential for forensic corpus in 578–80; authorship analysis, specialised individual corpora in 583; Bank of English 578, 586; Bentley, Coulthard's investigation of case of 581–82; British National Corpus (BNC) 578, 583, 586; Brown Corpus 578; Christie, Svartvik's investigation of Evans and 581; COBUILD project 583; constraints on use of corpora in 579; Criminal Cases Review Commission (CCRC) 579; Daubert criteria 587–88; FBI, unabomber investigation 584–85; forensic corpus, notion of 578–80; further reading 588; future challenges 587–88; genre, effect on text 580; idiolect 579; jurors, explanations to 587; legal casework, analysis of 578; limitations of corpus linguists 587–88; pre-existing linguistic corpora in 580–82; questioned texts 579–80; text types, range of 578–79; timespan, effect on text 580; unabomber 584–85; web as reference corpus 584–87
- form: corpus driven 321–22; and function, distinction between 557
- formulae, identification of 388–89
- formulaic language: registers, genres and 245; vocabulary materials and 389–91
- Fortanet, I. 244
- Foster, D. 584
- Foucou, P.-Y. and Kübler, N. 453
- Fowler, R. 531
- Fox, B.A. and Thompson, S.A. 243
- Fox, G. 581–82
- Fox, N. 605
- Francis, G. 213, 310
- Francis, G. and Kramer-Dahl, A. 606
- Francis, Nelson 15, 16
- Francis, W.N. 472
- Frankenberg-Garcia, A. 452, 474, 511
- Frankenberg-Garcia, A. and Santos, D. 508, 511
- Frazier, S. 231
- Freddi, M. 246
- Freeman, D. 622
- Freiburg-Brown Corpus of American English (FROWN) 16, 22, 109
- Freiburg Corpus of English Dialects (FRED) 111

- Freiburg-LOB Corpus of British English (FLOB) 109
- French, A. *et al.* 46
- French Interlanguage Database (FRIDA) 625
- frequency: analysis of 123, 132, 133, 189, 228; approach to identification of vocabulary teaching materials 390; counts in grammar 228–29; display of frequency data 123–25; exploitation of frequency data 125; intuitions on 154–55; Leuven Drama Corpus and Frequency List 22; lists, preparation of learners for using language corpora 379; pattern exploration 164–65; uses of for corpus-informed course book design 416; word frequency in literary speech representation 537–38; word frequency lists, exploration of 122, 123–26; writing dictionaries, use of corpora in 438–39
- Fries, C.C. 4, 15, 243
- Friginal, E. 246
- Fuentes, A.C. 374
- function: computer functionality 136–39; corpus driven 321–22
- further reading: audio-visual corpus building 103; building a corpus 36; concordances 182; corpora 120; corpus analysis 133–34; corpus-informed course book design 425–26; corpus software 150; creativity 313; data-driven learning 356; dictionaries, use of corpora in writing 442; discourse 284; English for Academic Purposes (EAP), corpora and 466–67; evolution of corpus linguistics (CL) 26; forensic linguistics 588; grammar 239; healthcare communication 617; language corpora, preparation of learners for using 382; language development, corpus software in research 192; language teaching 329; language teaching and data-driven learning 368; language testing 643; lexis 210; literary speech representation 542; materials for language teaching 343; multi-word units 223–24; parallel and comparable corpora 498; pattern exploration 165–66; poetry and drama, language of 530; political discourse 603; pragmatics 299–300; registers, genres and 248; research, teachers' use of corpora in 481; small specialised corpora building 77–78; sociolinguistics 561; specialist genres 267; spoken corpus building 49; teacher education 629; translation, use of corpora in 513; variety of a language, corpus building to represent 91; vocabulary teaching materials 398; writing instruction, use of corpora for 454; written corpus building 64
- future developments: audio-visual corpus building 101–3; corpora in English for Academic Purposes (EAP) 466–67; corpus-informed course book design 424–25; dictionaries, use of corpora in writing 440–42; in grammar teaching based on corpus findings 408–9, 410; language development, corpus software in research 191–92; sociolinguistics 560–61
- future prospects: building a corpus 36; forensic linguistics 587–88; language testing 642–43; political discourse 601–2; pragmatics 298–99; translation, use of corpora in 512–13; writing instruction, use of corpora for 452–54
- future research, multi-word units 222–23
- Gabel, S. 361
- Gabrielatos, C. 359, 362, 363, 367, 459, 462
- Gabrielatos, C. *et al.* 555, 558
- Gains, J. 246
- Gallina, F. 638
- García-Bermejo Giner, M.F. and Montgomery, M. 532
- Gardner, D. 388
- Gardner, R. 271
- Garretson, G. and Ädel, A. 594, 601
- Gaskell, D. and Cobb, T. 348, 452
- Gastil, J. 598
- Gavioli, L. 363, 374, 376, 454, 477
- Gavioli, L. and Aston, G. 320, 321, 326, 416, 453
- Gawande, Atul 458, 467
- Gee, J.P. 256, 267
- Geisler, C. 245
- general corpora: general reference corpora 504–6; language corpora and 371–72; phenomena studied using 292–98; spoken, written and both 109–10
- general lexicon 197–98
- General Service List* 388
- A General Service List of English Words* (West, M.) 387, 388
- genres: academic genres 259–62; analysis of, methodological advantages in 258–59; context variables and 494–95; of corpora, choice of 414; corpus use for learning about 373–74; divisions for lexicographic corpora 430; effect on text 580; genre approach, rationale for 257–58; language genres, differences between 437; perspectives on registers and 241–42; of political discourse 592; registers and 241–48; studies of, corpus-based 247–48; for study, emergence of new 602; variation in, variety of language and 81; *see also* registers, genres and; specialist genres
- geographical corpora 21
- George III 311
- German COSMAS Corpus 118
- German National Corpus 118
- German Sign Language 94

- Ghadessy, M. 260, 272, 388
 Gibbons, J. 583
 Gill, S. 623
 Gilmore, A. 333, 403, 413, 423, 426
 Gilquin, G. *et al.* 453, 458, 465, 466, 467
 Gilquin, Gaëtanelle xviii, 8, 9, 25, 130, 320, 325, 333, 359–68, 401, 405, 422, 439, 466, 477, 625, 636
 Gledhill, C. 245, 398
 Glisan, E.W. and Dreschler, V. 403
 GLOSSA Corpus 101
 Goethals, M. *et al.* 22
 Goethals, P. 493, 495
 Goethe, Johann Wolfgang 22
 Goffman, Erving 479, 581
 Grabe, W. and Kaplan, R. 244
 Graber, D. 591
 Graddol, D. 45
 Graddol, D. *et al.* 45
 grammar 227–39; acceptability, new challenged in assessment of 237–39; Bergen Corpus of London Teenage English (COLT) 234; British National Corpus (BNC) 234, 238; *Cambridge Grammar of English* 228, 234; COBUILD project 229; contexts, understanding grammar through 227–29; *Corpus Linguistics and Linguistic Theory* 237; corpus techniques, popularity of 238; discourse-level factors in grammatical patterns 232–33; frequency counts 228–29; further reading 239; grammar-vocabulary associations 229–31; grammatical co-text 231–32; grammatical patterns 229–34; language choice, context and 227; lexico-grammar 229–31; *Longman Grammar of Spoken and Written English* 228, 233–34; methodological principles, corpus-based grammar analysis 228–29; Michigan Corpus of Academic Spoken English (MICASE) 234; multiple features and conditions, simultaneous investigation of 234–35; patterns, understanding grammar through 227–29; patterns of correspondence 228; situation, context of 233–34; of speech 236–37; of variety 235; vocabulary associations 229–31
 grammar syllabus, book design 417
 grammar teaching materials 401–10; applied linguistics surrounding grammar, corpus studies and 401–2, 409; benefits of using corpora 402–6, 409–10; British Council 403; Cambridge and Nottingham Corpus of Discourse in English (CANCODE) 463; classrooms and corpora, effective connection of 404; constraints on corpora use 406–8, 409–10; corpora and grammar teaching, benefits and types of use 402–6, 409–10; corpora frameworks 401–2; corpus findings, debate on grammar teaching based on 408–9, 410; data-driven or data-informed grammar 401; *English Grammar in Use* (Murphy, R.) 402; English Language Teaching (ELT) publications 402, 403–4, 404–5, 406, 407, 408; future developments in grammar teaching based on corpus findings 408–9, 410; grammar frameworks 401–2; grammaticality, question of 404; internet, interactions with 408; language description frameworks 401–2; limiting factors on corpora use 406–8, 409–10; Linguists' Search Engine 409; multi-modal corpora, growth of 409; natural language processing (NLP) 403; needs and expectations of students 407; networks of corpora 408–9; PENN-Treebank 404; performance *vs.* competence division 401; *Practical English Usage* (Swan, M.) 402; prescriptive *vs.* descriptive debate 401; Second Language Acquisition theory 404; Survey of English 403; teacher-learned dynamic 407; theoretical orientation to, issues of 402; Turbo Lingo 409; types of corpora use 402–6; unification grammar 403
 Granath, S. 471
 Granger, S. and Meunier, F. 451
 Granger, S. and Tribble, C. 361
 Granger, S. *et al.* 188, 189, 373, 381, 439, 466, 636
 Granger, Sylviane xviii, 8, 9, 25, 130, 213, 319, 320, 325, 333, 342, 359–68, 373, 401, 405, 422, 439, 453, 465, 466, 477, 492, 493, 625, 636
 Graphic Online Language Diagnostic (GOLD) 188–89
 graphics in corpus software 141
 Greaves, Chris xviii, 123, 125, 148, 199, 200, 212–24, 279, 327, 336, 388, 390, 419, 433, 502, 523
 Greenbaum, S. 84, 85
 Greenbaum, S. and Svartvik, J. 552
 Greene, Jennifer 460
 Grice, H.P. 289, 303
 Grover, J. 612
 Gu, Y.G. 326
 Guangzhou Petroleum English Corpus 56
 Gumpertz, J.J. 547
 Guy, G. and Vonwiller, J. 558
 Gwyn, R. 617
 Hadley, G. 364, 365, 366
 Hafner, C. and Candlin, C. 450
 Hall, S. 523
 Hallebeek, J. 431
 Halleck, G.B. and Connor, U.M. 247
 Halliday, M.A.K. 18, 44, 213, 241, 263
 Halliday, M.A.K. and Hasan, R. 271, 273, 275
 Halliwell, S. 621

- Handford, M., Atkins, S. and Harvey, K. 262
 Handford, Mike xviii–xix, 9, 70, 72, 74, 75,
 126, 255–67, 449
Hansard 24
Hansard Canada French/English 488
 Hansard Corpus 593
 Hargreaves, P. 637
 Harris, Marvin 4
 Harvey, K. *et al.* 609, 610
 Harvey, Kevin xix, 605–17
 Harwood, N. 392, 398
 Hasan, R. 279
 Hawkey, R. and Barker, F. 637, 639, 642
 Haywood, Sandra 464–65
 headers 33
 Health Protection Agency 609
 healthcare communication 605–17; Adolescent
 Health Email Corpus (AHEC) 605, 608, 609,
 610, 611; adolescent health language, building
 a corpus of 607–8; British National Corpus
 (BNC) 608–9, 613–14; communication in
 institutional healthcare settings 606; conflation
 of AIDS and HIV 611–15; contagion,
 metaphors for 614; corpus analysis of sexual
 health and adolescent language 610–16;
 corpus-based studies 606–7; data,
 representational quality of 608; definition of
 605–6; further reading 617; health language,
 corpus use in exploration of patterns 608–16;
 Health Protection Agency 609; HIV/AIDS
 610–11, 611–15, 615–16; HIV infection,
 perceived attributions of 615; key corpus-
 based studies 606–7; language, central
 importance in 605; lexical items for analysis,
 identification of 608–9; medicalisation of
 society 605–6; metaphorisation of AIDS and
 HIV 611–15; 'miasma' theory of disease 614–
 15; NHS Direct 607; practical relevance of
 corpus analysis 616; reasons for study of 605–
 6; sexual health and adolescent language 609–
 10; sexually transmitted infections (STIs) 609,
 610, 611; sexually transmitted infections
 (STIs), 'sex' and 610–11; social experience of
 health 605; Society for Adolescent Medicine
 609; Teenage Health Freak 607–8; themes for
 analysis, identification of 608–9
 Hedgcock, J.S. 622
 Hehir, G. 536
 Heift, T. 192
 Hellenic National Corpus 118
 Helman, C. 612, 614, 615, 616
 Helsinki Corpus of English 22, 113, 534
 Helt, M.E. 247
 Hendricks, H. 637
 Henry, A. and Roseberry, R.L. 247
 Heritage, J. 335
 Herring, S.C. 246
 Herring, S.C. and Paolillo, J.C. 246
 Hewings, M. 324, 453
 Hewings, M. and Hewings, A. 245, 448
 Hewitt, L.E. *et al.* 188
 Hickey, L. 501
 Hickey, R. 534, 540–41
 Hidalgo Tenorio, E. 565
 Higgins, J. 361
 high-rising terminal (HRT) 558
 Hines, T.C. *et al.* 4
 historical corpora 21, 113–14
 historical origins: corpora 3–4; data-driven
 learning (DIDL) 346–49
The Hitch-Hiker's Guide to the Galaxy (Adams, D.)
 136
 HIV/AIDS 610–11, 611–15, 615–16; HIV
 infection, perceived attributions of 615
 HKIEd English–Chinese Parallel Corpus 120
 HKUST Computer Science Corpus 56
 HM Revenue and Customs (HMRC) 503
 HMTL codes: audio-visual corpus building 100;
 building a corpus 35; written corpus building
 61
 Hobbs, P. 606
 Hoey, M. 213, 222, 256, 261, 264, 273, 276,
 284, 306, 389, 574, 586
 Hoey, M. *et al.* 530, 533, 586
 Hoffman, S. 557
 Hofland, Knut 16, 21, 101
 Holmes, J. 90, 208, 262, 281
 Holmes, R. 247
 Holt, H. 584
 Hong Kong Corpus of Conversational English
 (HKCCE) 117
 Hong Kong Corpus of Spoken English
 (HKCSE) 8; small specialised corpora building
 68, 71, 73; spoken corpus building 38
 Hong Kong Laws parallel text 488
 Hong Kong University of Science and
 Technology corpus of essays 459
 Hoon, H. and Hirvela, A. 395
 Hopkins, Gerard Manley 4
 Hopper, P. 283
 Hori, M. 307, 542
 Horner, D. and Strutt, P. 639–40, 641
 Horst, M. *et al.* 464
 Howard, Paul 537, 538
 Howard-Hill, T.H. 4
 Howarth, P. 55
 Howatt, A.P.R. and Widdowson, H.G. 387
 Hudson, R. 81
 Hughes, A. 633
 Hughes, Beverley 571
 Hughes, G. 641
 Hughes, R. and McCarthy, M. 236
 Hughes, Rebecca xix, 9, 401–10, 428, 437, 531
 Hugo, Cardinal of St. Caro 167

- Huizhong, Yang 16
 human-centric computing 141–42
 Human Speech Project (HSP) 102
 humour, creativity and corpora 310–11
 Hundt, M. and Mair, C. 245
 Hundt, M. *et al.* 246, 625
 Hungarian National Corpus 118
 Hunston, S. and Francis, G. 229, 622
 Hunston, S. and Thompson, G. 244
 Hunston, Susan xix, 8, 68, 71, 87, 122, 127, 129, 130, 152–66, 229, 238, 255, 256, 259, 372, 381, 449, 460, 466, 595, 622, 626; data-driven learning (DDL) 363, 365, 366, 367; lexis 198, 199, 200, 204, 205; multi-word units 213, 220, 221, 223
 Hunt, K.W. 186
 Hunter, J. and Coxhead, A. 461, 466
 Hutchins, J. 496
 Hyland, H. 243, 244
 Hyland, K. 214, 216, 221, 222, 247, 259, 260, 346, 395, 444–45, 459, 460
 Hyland, K. and Milton, J. 260
 Hyland, K. and Tse, P. 244, 391, 459
 Hymes, D. 547
 Hyon, S. 257
 hypothesis testing, corpus use for 372–73

 Iacobucci, C. 266
 ideological reproduction 569–70; interrelatedness of semantic patterns and 574–75
 ideology 206; power and 563–64; social ideology 302
 idiolects (and idiolectal features of language) 22, 80, 579, 581, 584, 585, 593, 596
 idioms: idiom principle 213; idiomaticity, creativity in 307–8, 311–12; lexis 201; semantic annotation and creative idiomaticity 311–12
 IJS-ELAN Slovene-English Parallel Corpus 120
 Ilie, C. 594, 596
 Ilse, W.-R. 365
 Independent Advisory Group on Sexual Health and HIV 616
Index Thomisticus (Busa) 15
 Indianapolis Business Learner Corpus (IBLC) 68, 71
Industrial Society and Its Future (Kaczynski, T.) 584
 inferencing in media discourse 569–70
 Initiation-Response-Feedback (IRF) 623–24
 Innsbruck Computer Archive of Machine-Readable English Texts (ICAMET) 113
 interactional sociolinguistics 547
 interface, audio-visual corpus building 99–101
 International Computer Archive of Modern and Medieval English (ICAME) 107–8
 International Corpus of Crosslinguistic Interlanguage (ICCI) 117
 International Corpus of English (ICE) 109, 112–13, 117, 118; corpus linguistics and 21, 23; literary speech representation 534; parallel and comparable corpora 489; sociolinguistics 552, 555, 558–59; spoken corpus building 38; variety of a language, corpus building to represent 82, 84, 85, 87
 International Corpus of Learner English (ICLE) 116–17; English for Academic Purposes (EAP), corpora and 459; language development, corpus software in research 188, 191; language testing 636; parallel and comparable corpora 489; small specialised corpora building 68; teacher education 624–25
 International English Language Testing System (IELTS) 637, 641
 International Phonetic Alphabet (IPA) 45
 internet: corpus construction from 512; interactions with 408; language 26
 INTERSECT Project 488
 Intonational Variation in English (IViE) 111; sociolinguistics 558
 IPSyn metric 185, 187, 188
 Ireland, New Corpus of (NCI) 119
 Irish drama 540–41
 Italian CORIS/CODIS Corpus 118
 IVACS Symposium, Nottingham 533
 Ives, S. 531, 532

 Jackson, S. 610
 Jagger, Mick 311
 Jakobson, R. 302, 303
 James, Henry 169
 James, Z. 18
 Japan: Balanced Corpus of Contemporary Written Japanese (BCCWJ) 119; Japanese EFL Learner Corpus (JEFL) 117
 Jauss, H.-R. 302
 Jeffords, S. 522
 Jenkins, J. 415
 Jiao Da English for Science and Technology (JDEST) 16
 Johansson, C. 245
 Johansson, J. and Oksefjell, S. 239
 Johansson, S. *et al.* 84
 Johansson, Stig 11, 14, 16, 21, 119, 476
 Johns, T. and King, P. 364
 Johns, T.F. 7, 222, 320, 326, 333, 337, 339, 346, 347, 348, 351, 359, 360, 362, 363, 365, 367, 376, 395, 444, 463, 477, 622
 Johnson, A. 214
 Johnson, D. and Murray, J. 613
 Johnson, Dr. Samuel 4, 15
 Johnstone, B. 532

- Jones, M. and Haywood, S. 452, 461
 Jones, Martha xix–xx, 9, 387–98, 638
 Jonsson, C. 533
 JRC-Acquis Multilingual Parallel Corpus 488
 Jucker, A.H. 246
- Kacenska English-Czech Corpus 120
 Kachru, Y. 234
 Kaczynski, David 584
 Kaczynski, Theodore 584–85
 Kaeding, F.W. 387
 Kahlas-Tarkka, L. and Rissanen, M. 247
 Kaltenbock, G. and Mehlmauer-Larcher, B. 327
 Kanoksilapatham, B. 247
 Kaszubski, P. 451
 Keck, C.M. and Biber, D. 244
 Kendon, A. 40
 Kennedy, C. and Miceli, T. 320, 326, 355, 363, 365, 367, 380, 451, 453
 Kennedy, C. and Thorpe, D. 637
 Kennedy, G. 55, 58, 60, 122, 243, 320, 322, 324, 372, 547
 Kenning, M.-M. and Aston, G. 360, 373
 Kenning, Marie-Madeleine xx, 10, 21, 23, 24, 119, 431, 467, 487–98
 Kenny, D. 24, 497–98
 Kerry, John F. 594
 Kerswill, P., Cheshire, J. *et al.* 558
 Kerswill, P., Tøgersen, E. and Fox, S. 558
 Key Word in Context (KWIC) 95, 101, 116, 128, 147, 362, 378, 431, 434; concordances 167, 168, 171–72; corpora 4; corpus linguistics (CL) 7, 19; pattern exploration 156; sociolinguistics 555; vocabulary teaching materials 396
 Key Word in Context (KWIC) 396
 key-word lists: corpus software 149–50; exploration in corpus analysis 127–28; language corpora, preparation of learners for using 379–80; media discourse 565–67; media discourse dispersion plot 571–72; negative key words 128; poetry and drama, analysis of language of 519–22; political discourse analysis 597–98; positive key words 127–28; pronouns as key words in movies 524–25; relationship between cultural and corpus-comparative statistical keywords 566–67; vocabulary teaching materials 394–95
 keyboarding 62–63
 keyness: corpus analysis 127; healthcare communication, key corpus-based studies 606–7; keyword links in media discourse 572; poetry and drama, language of 517–19
 Kilgariff, A. 328, 451
 Kilgariff, A. and Greffentette, G. 115
 Kilgariff, A. *et al.* 54, 57, 62, 122, 328, 431
 King, P. and Woolls, D. 361
 Kinginger, C. 621, 622, 623
 Kipp, Michael 46, 98, 102
 Kirk, J.M. 532
 Kirkpatrick, A. 415
 Kjellmer, G. 389
 Knifton, C. 614
 Knight, D. and Adolphs, S. 36, 41, 425
 Knight, D. *et al.* 39, 40
 Knight, Dawn xx, 38–49, 409, 430, 466, 475, 549, 595, 627
 knowledge, construction of 622–23
 Knowles, G. 81
 Koestenbaum, W. 615
 Koester, Almut xx, 66–78, 96, 123, 258, 262, 263, 264, 430, 508, 595
 Kolhapur Corpus of Indian English 109, 488
 Koller, V. and Davidson, P. 565
 Konstantakis, N. 459
 Koosha, M. and Jafarpour, A.A. 367
 Korean National Corpus (Sejong Balanced Corpus) 118; variety of a language, corpus building to represent 87
 Kosem, I. 450–51
 Koteyko, N. 592, 599
 Koteyko, N. *et al.* 275
 Kozulin, A. 621
 Kramsch, C. 283
 Krashen, S. 333, 352
 Kretzschmar, W.A. *et al.* 234, 534
 Kretzschmar Jr., Willaim A. 534, 542
 Kreyer, R. 446
 Krishnamurthy, R. 16, 206
 Krishnamurthy, R. and Kosem, I. 391, 451, 459
 Kristoffersen, G. and Simonsen, H.G. 553
 Kryk-Kastovsky, B. 247
 Kübler, N. and Aston, G. 10, 23, 24, 119
 Kübler, Natalie xxi, 360, 373, 431, 501–13
 Kucera, Henry 15, 16
 Kuiper, K. 308, 389
 Kuiper, K. and Flindall, M. 274
 Kuo, C.-H. 244
 Kuo, C.-H. *et al.* 362
 Kwan, B.S.C. 247
 KWICFinder 116
 Kytö, M. 245
- Labour, New and Old in UK 598
 Labov, W. 81, 532, 547, 548
 Labov, W. and Waletzky, J. 279
 Lamb, S.M. 305, 390
 Lampeter Corpus 113; political discourse 593
 Lancashire, I. *et al.* 534
 Lancaster Corpus of Academic Written English (LANCAWE) 117
 Lancaster Corpus of Mandarin Chinese (LCMC) 119
 Lancaster-Leeds Treebank 112

- Lancaster-Los Angeles Spoken Chinese Corpus (LLSCC) 119
- Lancaster Oslo-Bergen (LOB) corpus 109, 112, 114; corpus linguistics (CL) 22; parallel and comparable corpora 488; political discourse 592; registers, genres and 241; sociolinguistics 548; variety of a language, corpus building to represent 84, 85, 86
- Lancaster Parsed Corpus (LPC) 112
- Langacker, G. 311
- Langford, B. 522
- language: awareness and corpora in teacher education 626–27; book design and information about 422; choice, context and grammar 227; description frameworks 401–2; genres of, differences between 437; healthcare communication, central importance in 605; Language Assessment, Remediation, and Screening Procedure (LARSP) 187; language chunks 101, 137, 138, 213, 334, 336, 398, 419–20, 441, 586; Language in the Workplace (LWP) 262; language technology, initiatives in 560; learning 7–9, 326–27, 495–96; pedagogy, data-driven learning and 354–56; teacher education context 620–21
- Language and Literature* 446, 534
- language corpora, preparation of learners for using 371–82; alphabetical sorting 379; concordance lines, reading and interpretation of 380–81; concordancing 377–78; contrastive studies, corpus use for 373; corpora building, preparation of learners in 375; corpus analysis tools 377–80; corpus data, preparation of 374–77; corpus data, preparation of learners for use of 374–75; corpus results, interpretation of 380–82; data amounts, preparation of learners in dealing with large quantities 375–76; data interpretation, caution in 382; discovery learning 372; error correction, corpus use for 373; frequency lists 379; further reading 382; general information, provision of 371–72; genre, corpus use for learning about 373–74; hypothesis testing, corpus use for 372–73; key word lists 379–80; learner as researcher 372; quantitative results, interpretation of 381–82; query types 378–79; questioning the data, preparation of learners in 376–77; reverse word sorting 379; sorting tools 379–80; specialised areas, corpus use and self-study in 374; student knowledge, finding out about 371–72; targets for self-study, corpus use for generation of 374; task objectives and corpora types, identification of 372–74; technical support 377; translation, corpus use for 373
- language development, corpus software in research 184–92; accuracy, measures of 190; adequacy of metrics 187–88; African American Vernacular English 187; Black English Sentence Scoring (BESS) 187; Center for Advanced Language Proficiency Education and Research (CALPER) 188–89; Child language Data Exchange System (CHILDES) 186, 187, 188; complexity, measures of 190; Computerised Language Analysis (CLAN) 186, 187; Developmental Sentence Scoring (DSS) 185, 187; first language development, corpus deployment in revealing 186–88; first language development, theories of 184; fluency, measures of 190; further reading 192; future developments 191–92; Graphic Online Language Diagnostic (GOLD) 188–89; International Corpus of Learner English (ICLE) 188, 191; IPSyn metric 185, 187, 188; Language Assessment, Remediation, and Screening Procedure (LARSP) 187; language development 184–85; mean length of utterance (MLU) 185, 188; measurement of language development 185–86; Michigan Corpus of Academic Spoken English (MICASE) 189–90; morphosyntactic structures 190; natural language processing (NLP) 191; number of different words (NDWs) 185, 188; order of learning 187; second language development, corpus deployment in revealing 188–91; second language development, literature on 186; second language development, theories of 184–85; sequence of learning 187; specific language impairment (SLI) 187; transcript attributes 190; validity of metrics 187–88; variability in language development 187; Written English Corpus of Chinese Learners (WECCL) 190; XML 189
- language teacher education (LTE) 8, 620–21; research, teachers' use of corpora in 471–81
- language teaching 7–9, 319–29; bridging corpus linguistics and language teaching 327–29; corpora and language learning 319–21, 478–80; corpus driven form and function 321–22; corpus evidence as teaching materials 322–26; corpus linguistics and language teaching, bridging between 327–29; form, corpus driven 321–22; function, corpus driven 321–22; further reading 329; language learning, tasks for 326–27; syllabuses for, exploitation of spoken and written corpora in 333–43; tasks for language learning 326–27; teaching materials, corpus evidence as 322–26
- language teaching and data-driven learning 359–68; data-driven learning material 360–62; effectiveness of DDL, assessment of 365–66; further reading 368; operationalisation of DDL 362–65; pedagogical functions of DDL

- 359–60; problems and limitations of IDL 366–68
- language testing 633–43; assessment, essential principles of 633; authenticity 635; automatic rating software 640; British Academic Spoken English (BASE) 637; British Academic Written English (BAWE) 637; British National Corpus (BNC) 637; Cambridge ESOL Certificate of Proficiency in English 633, 634, 637, 639; Cambridge Learner Corpus (CLC) 636, 639; CEDEL2 (L1 English–L2 Spanish) 639; Centre for English Corpus Linguistics (CEFL) 636, 638; Certificate of Italian as a Foreign Language (CILS) 638–39; Coh-Metrix 642; Common European Framework of Reference (CEFR) 634, 635; Communicative Language Ability 635; corpus use in, development of 636–38; definition of 633–36; English Profile Programme 638; Examination for the Certificate of Competency in English (ECCE) 639; Examination for the Certificate of Proficiency in English (ECPE) 641; First Certificate in English (FCE) 641; further reading 643; future prospects 642–43; inter-related aspects 635–36; International Corpus of Learner English (ICLE) 636; International English Language Testing System (IELTS) 637, 641; Language Testing and Assessment (LTA) 633, 636, 637, 638, 641, 642; learner corpus, use in informing 638–41; Lexicon of Spoken Italian by Foreigners (LIPS) 638; Longitudinal Database for Learner English (LONGDALE) 638, 639; Michigan Corpus of Academic Spoken English (MICASE) 637, 641; native speaker corpus, use in informing 641–42; Natural Language Processing (NLP) 640; psychometrics 634; purpose of, establishment of 633–34; Reference Level Descriptors (RLDs) 639; reliability 633; SAT Reasoning Test 642; task formats 634–35; test design 634–35; Test of English as a Foreign Language (TOEFL) 634, 637, 640; TOEFL 2000 Spoken and Written Academic Language Corpus (T2K-SWAL) 637; University of Cambridge Local Examinations Syndicate (UCLES EFL) 634, 636; University of Michigan English Language Institute (ELI-UM) 639, 641; validity 633; Varieties of English for Specific Purposes Database (VESPA) 638; WnCLE (L1 Spanish–L2 English) 639
- Lantolf, J.P. 185
- Lapadat, J.C. and Lindsay, A.C. 40, 44
- Larsen-Freeman, D. 227
- Larsen-Freeman, L. and Cameron, L. 283
- laughter, transcription of 35
- Lave, J. and Wenger, E. 479, 628
- Laver, J. 45
- Laviosa, S. 494, 497
- learner corpora: exploitation of 341–43; language testing, use in informing 638–41; Learner Business Letters Corpus (LBLC) 117; learner language, corpora and insights on 465–66; in lexicography 439–40; teacher education 623–26
- learners: independence of, sustainability of 462; non-native-speaker data, learner data and 415; profiles of 439–40; as researchers 372
- Lee, D. and Swales, J. 319, 320, 326, 447, 451, 454, 461, 462
- Lee, David Y.W. xxi, 6, 21, 25, 38, 107–20, 130, 238, 256, 257, 258, 259, 281, 308, 356, 391, 446, 473, 535, 625, 642
- Lee, J.J. 374
- Lee, L. 185
- Leech, G. 4, 17, 18, 55, 56, 86, 319, 345, 356, 444, 541, 547
- Leech, G. and Short, M.H. 531, 536
- Leech, G. and Smith, N. 238
- Leech, G. *et al.* 40, 43, 125, 198, 246, 424
- legal casework, analysis of 578
- legal writing 449–50; local grammar in 449
- Leitner, G. 407
- lemmas (and lemmatisation) 4, 116, 125, 158, 172–73, 197–98, 379, 388, 416, 585; healthcare communication 612–13; lemmatisation tools 432–33; parallel and comparable corpora 491, 497; political discourse 596, 601; sociolinguistics 552, 553, 557; translation, use of corpora in 504, 506
- Lenneberg, E.H. 187
- Lennon, P. 335
- Leung, C. *et al.* 624
- Leuven Drama Corpus and Frequency List 22
- Levinson, S.C. 288, 293, 296
- Levy, M. 362, 365
- Lewis, D.M. 11
- Lewis, M. 389
- lexical bundles: in EAP corpus studies 461; multi-word units 213, 214, 215, 217; registers, genres and 245
- lexicography: dictionaries, lexicographers' use of corpora in writing 428–31; lexical items for analysis, identification of 608–9; lexical priming, multi-word units 222; lexical repulsion, multi-word units 212; lexicogrammar 229–31; lexico-grammatical patterns 417; Lexicon of Spoken Italian by Foreigners (LIPS) 638; lexis and lexical sets 206; lexis and lexicon 197–99; role of lexicographer 441–42
- lexis 197–210; antonyms 207–8; Bank of English 197–98, 199, 206; British National Corpus (BNC) 198; COBUILD project 197;

- collocation 199–200; connotation 205;
context and meaning 202–3; fixed expressions
201; further reading 210; general lexicon
197–98; ideology 206; idioms 201; lexical sets
206; lexicon and lexis 197–99; meaning
202–6, 209–10; metaphor 204–5; multi-word
units 213, 221–22; opposites 207–8;
patterning 199–200; phraseology 208–9;
phraseology and phrases 199–201; polysemy
203–4; sets and synonyms 206–8; in spoken
language 208–10; synonyms 206–7; usage and
meaning 209–10; word formation 199
- LexisNexis Academic 594
- Li, Y. 219
- Limerick-Belfast Corpus of Academic Spoken
English (LIBEL CASE) 114
- Limerick Corpus of Irish English (LCIE) 10, 111;
case studies 88–89; literary speech
representation 535, 537, 538, 539; small
specialised corpora building 72–73; spoken
corpus building 38; teacher education 624,
626; variety of a language, corpus building to
represent 88–89, 89–90
- limitations: concordances 182; of corpus-
linguistic method in sociolinguistics 555–56;
of corpus linguists in forensic linguistics
587–88; in corpus use for study of discourse
275–76; dictionaries, use of corpora in writing
440–42; limiting factors on corpora use in
grammar teaching 406–8, 409–10; literary
speech representation, use of corpora in 541;
problems and limitations of data-driven
learning 366–68; sampling limitations, small
specialised corpora building 69
- Lindemann, S. and Mauranen, A. 244
- lingua franca corpora 118; English as a Lingua
Franca (ELF) 25, 238, 415, 459
- LinguaLinks 534
- LinguaLinks* 534
- linguistics: corpus studies with register as
predictor of linguistic variation 242–46;
features marking particular style of political
discourse, analysis of 597; Linguistic Atlas of
Middle and South Atlantic States (LAMSAS)
534; Linguistic Data Consortium (LDC)
107–8, 488; linguistic variability 69, 70;
literary data as source for linguistic evidence
532–33; and sociolinguistics, connection
between 548; systemic functional linguistics
(SFL) 257, 264–65; *see also* corpus linguistics;
forensic linguistics; sociolinguistics
- Linguists' Search Engine 409
- linking, signposting and 339
- listening: corpus-based materials in teaching
334–36; to data in sociolinguistics 556
- literary speech representation 531–42; African
American Vernacular English 534; Carleton's
18th-century stories 540; corpus-based stylistic
analysis, case study 536–41; corpus in
comparison of real and respresented speech
533–35; *Corpus Presenter* 534, 535;
*The Curious Incident of the Dog in the
Nightdress* (O'Carroll-Kelly, R.) 536–38, 541;
descriptive exposition 532; fictionalisation of
spoken language 531; further reading 542;
Helsinki Corpus 534; *Huckleberry Finn* (Twain,
M.) 534; International Corpus of English
(ICE) 534; Irish drama 540–41; IVACS
Symposium, Nottingham 533; Limerick
Corpus of Irish English (LCIE) 535, 537, 538,
539; limitations in use of corpora 541;
LinguaLinks 534; Linguistic Atlas of Middle
and South Atlantic States (LAMSAS) 534;
linguistic evidence, literary data as source for
532–33; manual annotation 536; non-standard
language, effectiveness of 532; Oxford Text
Archive 535; Project Gutenberg 535;
qualitative approaches 535–36; real and
fictional speech, similarities and differences
531–33; *The Reeve's Tale* (Chaucer, G.)
532; *Romeo and Juliet* (Shakespeare, W.) 534;
The Snapper (Doyle, R.) 536; speech
representation in literature 532–33; speech
representation in literature, corpus and
535–36; *TACT* 534; *Tom Sawyer* (Twain, M.)
534; *Uncle Tom's Cabin* (Beecher Stowe, H.)
533; word frequency 537–38; *WordSmith Tools*
(Scott M.) 534
- literary studies 9–10
- Literature Online (LION) 113
- London, J. 338
- London-Lund Corpus of Spoken English (LLC)
111; corpus linguistics (CL) 23; political
discourse 592; registers, genres and 241;
sociolinguistics 548, 551–52, 555; spoken
corpus building 38, 45; variety of a language,
corpus building to represent 85
- Long, S.H. et al. 186
- Longitudinal Database for Learner English
(LONGDALE) 638, 639
- Longman Corpus Network 112
- Longman Dictionary of Contemporary
English* 430
- Longman Exams Dictionary* (Mayor, M.) 465
- Longman Grammar of Spoken and Written English:*
grammar 228, 233–34; registers, genres and
243
- Longman Language Activator* 465
- Longman Spoken and Written English Corpus
(LSWE) 81, 87
- López-Rodríguez, C. and Tercedor-Sánchez, M.
495, 513
- Louvain Corpus of Native English Essays
(LOCNESS) 116

- Louvain International Database of Spoken English Interlanguage (LINDSEI) 116–17; teacher education 625
- Louw, B. 291, 292, 307, 353, 533
- Louw, W. 506
- Lu, Xiaofei xxi, 184–92, 283, 477
- LUCY Corpus 112, 116
- Lula da Silva, Luiz Ignácio 593
- Lupton, D. 614
- Ma, K.C. 55
- MA programmes 620–21
- McCafferty, K. 533, 540
- McCarten, Jeanne xxii, 9, 125, 325, 334, 336, 404, 413–26, 428, 459, 476
- McCarthy, Michael xxii, 3–12
- McCarthy, M.J. 8, 39, 44, 48, 228, 405, 424, 430, 471, 475, 477, 559–60, 620–21, 629, 640; building variety of a language 82, 84, 85, 86, 87; corpus analysis 125, 126, 132, 133, 201, 208; course book design 414, 416, 420; creativity 308, 309; data-driven learning (DDL) 345, 355; discourse 271, 272, 274; language teaching 325, 335; multi-word units 213, 214, 215, 216, 217, 219; registers and genres 241, 244, 246; specialist genres 262, 264, 265
- McCarthy, M.J. and Carter, R.A. 11, 123, 125, 236, 244, 324, 336, 415, 440, 602
- McCarthy, M.J. and Handford, M. 126, 246, 262, 425
- McCarthy, M.J. and McCarten, J. 423
- McCarthy, M.J. and O'Dell, F. 339, 391, 404, 463–64
- McCarthy, M.J. *et al.* 334, 335, 404, 419, 422, 474, 476
- McCarthy, M.J. McCarten, J. and Sandiford, H. 421, 422
- McEnery, A. 607
- McEnery, A. and Xiao, R. 541
- McEnery, A. *et al.* 243, 430, 459, 535
- McEnery, T. and Wilson, A. 345, 346, 356, 372, 391, 430, 548
- McEnery, T. *et al.* 5, 80, 81, 82, 85, 86, 87, 125, 127, 130, 132, 495, 497, 498
- Machine Readable Spoken English Corpus (MARSEC) 111
- machine translation: parallel and comparable corpora 496; translation, use of corpora in 503–4, 513
- McHoul, A. and Rapley, M. 606
- McIntyre, Dan xxii, 9, 309, 516–30
- McKenna, B. 246
- Macmillan English Dictionary* 436
- McNamara, T. 633, 634, 635
- MacWhinney, B. 186
- Mahlberg, M. 246, 283, 516, 535
- Maia, B. 489–90, 498, 507
- Maingay, P. 623
- Mair, Christian 22, 359
- Mäkinen, M. 247
- Malderez, A. and Bodóczy, C. 622
- Malmkjaer, K. 494, 511
- Manca, E. 23
- manual analysis 388, 601, 639
- manual annotation 536
- manual work in sociolinguistics, need for 557
- mapping: correspondences 493; grammatical features 233; semantic units 274; words in different languages 496
- Marco, M.J.L. 245
- mark-up requirements 35–36
- Marley, C. 244
- Marston, E. 581
- Martin, J.R. and Rose, D. 274
- Martin, M. 257
- Mason, L. 609
- materials for language teaching 333–43; chunks of language 101, 137, 138, 213, 334, 336, 398, 419–20, 441, 586; cohesive devices 340–41; corpus-based approaches, incorporation in a syllabus 333–34; error correction 340; errors, investigation of 341–43; further reading 343; language chunks 101, 137, 138, 213, 334, 336, 398, 419–20, 441, 586; learner corpora, exploitation of 341–43; linking, signposting and 339; listening, use of corpus-based materials to teach 334–36; oral fluency and spoken corpora 335–36; proof-reading 340; reading, use of corpus-based materials to teach 337–38; reporting verbs 339–40; signposting and linking 339; skills in reading, use of corpus-based materials to teach 337–38; skills in writing, use texts to develop 338–41; speaking, use of corpus-based materials to teach 334–36; spoken corpora, oral fluency and 335–36; spoken language features 334–35; verb reporting 339–40; writing, use of corpus-based materials to teach 337–38; writing, use texts to develop skills in 338–41
- Mattos, M. de A. 340
- Mauranen, A. 223, 235, 244, 366, 459, 494, 501
- Mauranen, A. and Bondi, M. 244
- Mautner, G. 565
- May, R. 494–95
- Maynard, C. and Leicher, S. 291
- Mayor, M. 465
- mean length of utterance (MLU) 185, 188
- meaning: in context 202–3; in context, multi-word units 212–13, 215–17; lexis and 202–6, 209–10; patterns and 212–13, 215–17; range of word meanings 394–95; usage and 209–10

- media discourse 11–12, 563–76; British National Corpus (BNC) 570, 572; case study, corpus-based CDA 567–75; collocation and frames 568–69; collocation of keywords 565–66; concordancer collocation and semantic significance 568–70; corpus-based CDA 563–67; corpus-based cultural keywords 565–66; corpus-based qualitative analysis 570–74; corpus-comparative statistical keywords (CCSKs) 566, 570–71, 571–72, 575; critical discourse analysis (CDA) 563–64; cultural keywords 565, 566–67; cultural processes, relationship with language 564; electronic texts, abundance of 563; European Union (EU), expansion and resulting immigration 567, 568–69, 573–74; ideological reproduction and inferencing 569–70; ideological reproduction and interrelatedness of semantic patterns 574–75; inferencing 569–70; interrelatedness of semantic patterns 574–75; keyword dispersion plot 571–72; keyword links 572; keywords 565–67; orientation 564–65, 567; prosumer (producer-consumer) CDA 575–76; relationship between cultural and corpus-comparative statistical keywords 566–67; semantic patterns, ideological reproduction and interrelatedness of 574–75; semantic patterns, types via 'But' 572–74; semantic significance 568–70; social processes, relationship with language 564; *Sun* 567, 568, 569, 570–71, 571–72, 573, 574, 575; text structure, CCSKs and 570–71; *Wordsmith Tools* (Scott, M.) 566, 567, 569, 570, 571–72, 608, 628
- mediated acquisition in teacher education 623
- mediated political discourse 594–95
- Medical Anatomy Corpus (MAC) 62, 64
- medicalisation of society 605–6
- mega-corpora 66, 88, 255, 256–57
- Mendikoetxea, A. 639
- metadata 11, 42–44, 99, 189, 549, 551
- metaphor: lexis 204–5; metaphorisation of AIDS and HIV 611–15; political discourse 599–600
- methodology: corpus-based grammar analysis 228–29; genre analysis, advantages in 258–59; methods in sociolinguistics, use of corpora and 549–56; vocabulary teaching materials, ongoing developments 390–91
- metrics, adequacy of 187–88
- Meunier, F. 213, 410, 462
- Mey, J.L. 289
- Meyer, C. 61, 81, 82, 83, 84, 85, 87, 234
- Meyer, C. and Tao, H. 237
- Meyer, C.F. and Ebrary, I. 372
- 'miasma' theory of disease 614–15
- Michigan Corpus of Academic Spoken English (MICASE) 112, 114; corpus linguistics (CL) 25; English for Academic Purposes (EAP), corpora and 459, 460, 463; grammar 234; language development, corpus software in research 189–90; language testing 637, 641; registers, genres and 241, 244; research, teachers' use of corpora in 474; small specialised corpora building 68; specialist genres 259–60; spoken corpus building 38; teacher education 624; variety of a language, corpus building to represent 83, 84, 86
- Michigan Corpus of Spoken English (MICASE) 459, 641
- Michigan Corpus of Upper-Level Student Papers (MICUSP): corpora 116, 117; vocabulary teaching materials 391
- Milizia, D. 221
- Milizia, D. and Spinzi, C. 592, 596
- Milroy, L. 81
- Milton, J. 451, 453
- Mindt, D. 476
- Mindt, D. and Weber, C. 402
- Mishan, F. 326, 327, 423
- Mitchell, T.F. 102
- Mitchell, T.M. et al. 142
- modality: aspect and modality verb forms 272; deixis and 11; epistemic 493; multi-modal communication 298; multi-modal corpora, development of 425; multi-modal corpora, growth of 409; multi-modal corpora in pragmatics 298–99; multi-modality 5, 38–39; pragmatic functioning of 628; stance and 244–45
- Modern Chinese Language Corpus (MCLC) 118, 119
- modern corpora, creation of 4–5
- Mollin, S. 238
- monitor corpora 20, 22, 23, 436
- Monocanc* (Barlow, M.) 5, 122
- monolingual corpora 21, 108–9, 346, 352–53, 360, 373, 487–88; analysis of 491; compilation of 490; language learning and single corpora 495; learners' corpora building 375; non-English corpora 118–19; translation, use of corpora in 504–8
- Monroe Corpus 112
- Montclair Electronic Language Learners' Database (MELD) 117
- Montgomery, M. 532
- Monzó Nebot, E. 513
- Moon, R. and Greaves, C. 125
- Moon, Rosamund xxii–xxiii, 70, 197–210, 307, 308, 309, 310, 311, 441
- Moore, S. et al. 616
- Morell, T. 132
- Morley, J. 592, 596
- morphological information, tagging with 554
- morphosyntactic structures 190

- Morrison, A. et al. 41
 Morrow, P.R. 246
 Mostefa, D. et al. 97
 movie scripts, study of 522
 Mudraya, O. 448
 Mugford, G. 424
 Mukherjee, J. 361, 363, 365, 366, 367, 410, 452, 454, 474
 Mukherjee, J. and Rohrbach, J. 453
 MULTEXT-East Corpus 120
 multi-media corpora for EAP 466
 multi-world units 212–24; Cambridge and Nottingham Corpus of Discourse in English (CANCODE) 217; classification of 419; collocation 212; data-driven learning (DDL) 222; definition of 213–15; English for Academic Purposes (EAP) 221; English for Specific Purposes (ESP) 221; further reading 223–24; future research, implications and 222–23; idiom principle 213; lexical bundles 213, 214, 215, 217; lexical priming 222; lexical priming, notion of 222; lexical repulsion 212; lexis 213, 221–22; meaning in context 212–13, 215–17; n-grams 213–14, 215, 216, 217–19, 221–22; patterns and meaning 212–13, 215–17; phrase-frames 213, 218; phraseological variation 217–21, 221–22; searches of, tools for 433; studies of 221–22; study of 215–17
Multiconcord 491, 496
 Multicultural London English (MLE) 555, 558, 559
 multilingual corpora 19–20, 23–24, 118; multilingual aligned corpora 431; Multilingual Corpus Toolkit 523; North Germanic languages, multilingual corpus of 559; parallel multilingual corpora 119–20; translation, use of corpora in 509–11
 multimedia corpora, multimodal texts 114–15
 multiple translation corpora 511
 Munat, J. 306
 Munday, J. 497
 Murphy, B. and Cutting, J. 624
 Murphy, Raymond 402
 Murphy, A. 216
 Musolf, A. 600
 Mutual Information (MI): corpora 6; dictionaries, use of corpora in writing 435
 Nation, I.S.P. 387, 388, 395, 463, 464
 Nation, P. and Waring, R. 387, 398
 National Sound Archive 111
 native speaker corpora 641–42; intuition or corpora? 428–29
 Nattinger, J.R. and DeCarrico, J.S. 389, 419
 Natural Language Processing (NLP) 640; grammar teaching materials 403; language development, corpus software in research 191
 Neale, S. 522
 Nelson, G. 84, 234
 Nelson, Mike xxiii, 53–64, 122, 173, 187, 256, 258, 398, 430, 448–49, 459, 508, 535, 548, 595
 neologisms and new senses, identification of 436–37
 Nesi, H. 213, 216, 217, 459
 Nesi, H. and Basturkmen, H. 213, 216, 217, 245
 Nesi, H. et al. 391, 392
 Nesselhauf, N. 25, 359, 361, 363, 373, 453, 462, 465, 466, 467, 477
 Network of European Reference Corpora (NERC) 44
 networks of corpora 408–9
 Nevala, N. 247
 Nevalainen, T. and Kahlas-Tarkka, L. 245
 New Rhetoric School 257
New York Times 584
 Newcastle Electronic Corpus of Tyneside English (NECTE) 111
 Newdigate Newsletters Corpus 113
 Newmark, P. 501
 Newmeyer, F. 237
 news reports 594–95
 Newton, E.M. et al. 43
 n-grams 213–14, 215, 216, 217–19, 221–22
 NHS Direct 607
NITE Toolkit 99, 102
 Noguchi, J. 449
 non-English corpora 118–20
 non-institutional genres 264–67
 non-native speakers 55, 116–17, 190, 322, 346, 361, 416, 448, 624, 640; creativity 304, 308; data concerning 415; data-driven learning (DDL) 340, 342; language of 430; non-native speaker corpora 25, 415
 non-standard forms: effectiveness of language for literary speech representation 532; treatment in sociolinguistics of 553–54
 Nordic Dialect Corpus 554
 Norrick, N.R. 244
 North Germanic languages, multilingual corpus of 559
 Norwegian Spoken Corpus (NoTa) 552–53, 553
 Nostrand, H.L. 403
 Nottingham Multi-Modal Corpus (NMMC) 102; spoken corpus building 39, 45–46
 Nuttall, C. 337
 Nyikos, M. and Hashimoto, R. 621
 Oakes, M. 238
 Obama, Barack 152, 153, 602
Observer XT 99–100

- Ochs, E. 40, 44
 Ochs, E. and Taylor, C. 591
 O'Connell, D.C. and Kowal, S. 41, 44
 O'Connor Di Vito, N. 403
 O'Donnell, M.B. 99
 O'Donnell, M.B. *et al.* 221
 O'Halloran, Kieran A. xxiii, 9, 11, 206, 516, 530, 563–76, 592, 599, 600
 Ohmann, R. 303
 O'Keefe, A. and Adolphs, S. 294
 O'Keefe, A. and Amador-Moreno, C.P. 535
 O'Keefe, A. and Farr, F. 8, 452, 623, 624
 O'Keefe, A. *et al.* 8, 67, 69, 72, 74, 154, 377, 452, 471, 476, 481, 512, 578, 625; corpus analysis 125, 128, 130, 132; course book design 416, 419, 420, 426; creativity 308; discourse 271, 282; English for Academic Purposes (EAP) 459, 460; grammar 227, 228, 231; language teaching 324, 325, 326, 336; lexis 198, 199, 201; multi-word units 208, 212, 213, 215, 216, 217; specialist genres 258, 260, 264; vocabulary teaching materials 388–89, 392
 O'Keefe, Anne xxiii, 3–12, 9, 39, 67, 74, 88, 89, 123, 132, 246, 265, 291, 537, 565
 Old Bailey Corpus 113
 Olohan, M. 501
 Olohan, M. and Baker, M. 495, 497–98
 Olsson, J. 579
 online communication, language of 626–27
 online concordancing 128–29
 online interfaces 511
 Oostdijk, N. 58
 Open Language Archives Community (OLAC) 108
 opposites 207–8; oppositional stances, lexical differences between 599
 optical character recognition (OCR) software 54, 62
 optical scanning 62
 oral fluency and spoken corpora 335–36
 orientation in media discourse 564–65, 567
 origins of corpora 14–15
 Ortega, L. 184
 Orwell, George 598
 Osborne, J. 322, 325, 451
 Oslo Corpus of Bosnian Texts 119
 Oslo Multilingual Corpus (OMC) 24, 120
 O'Sullivan, Í 359
 O'Sullivan, Í and Chambers, A. 348, 365, 367, 373, 451, 452, 453
 Owen, C. 8, 238
 Oxford, R. 621, 623
 Oxford Advanced Learner's Dictionary (OALD) 434
 Oxford Concordance Program (OUP, 1987) 5
 Oxford English Dictionary 170, 565; quotation database 114
 Oxford Text Archive (OTA): corpora 107–8; literary speech representation 535
 Ozturk, I. 247
 Page, N. 531
 Paltridge, B. 392
 ParaConc 491
 parallel and comparable corpora 487–98; adverbial connectors 492; alignment of texts 490–91; analysis of 491; Arabic English parallel news 488; British National Corpus (BNC) 494; Brown Corpus 488; common sampling frame, importance of 488; comparable corpora, exploitation of 493; comparable multilingual corpora 119–20; as complementary resources 496–98; contrastive linguistics 492–93; corpus analysis 491; corpus compilation, existing corpora and 488–90; corpus compilation, key issues 489–90; corpus design 489–90; corpus preparation 490–91; CRATER Project 488; definition of 487–88; difference between 487; English for Academic Purposes (EAP) 466–67; English–Norwegian Parallel Corpus 488, 493; European Corpus Initiative 488; existing corpora, corpus compilation and 488–89; further reading 498; genres, context variables and 494–95; Hansard Canada French/English 488; Hong Kong Laws parallel text 488; International Corpus of English (ICE) 489; International Corpus of Learner English (ICLE) 489; INTERSECT Project 488; JRC-Acquis Multilingual Parallel Corpus 488; Kolhapur Corpus 488; Lancaster Oslo–Bergen corpus 488; language learning 495–96; Linguistic Data Consortium 488; machine translation 496; mapping of correspondences 493; *Multiconcord* 491, 496; *ParaConc* 491; political discourse 596–97; preparation of 490–91; text collection 490; textual systems 494; translation, use of corpora in 509–10, 510–11; translation comparatives 492–93; translation direction 488; translation studies 494–95; translation styles, investigation of 495; Translational English Corpus (TEC) 494; 'translationese' 496; Unesco Index Translationum 490; uses of 492–96; *Wordsmith Tools* (Scott, M.) 491, 497
 parliamentary debates 593–94
 parole, linguistics of 14–15, 20
 PAROLE Corpus 24
 Parrish, S.M. 4
 Parry, R. 606
 parsed corpora 112–13
 Parsed Corpus of Early English Correspondence 113
 Partington, A. 200, 207, 213, 245, 271, 282, 307, 309, 373, 381, 449, 594, 600, 602, 622

- Partington, A. and Morley, J. 245
- parts-of-speech (POS) tags: building a corpus 35–36; concordances 177; tagging tools 433; written corpus building 63
- Pastra, K. 99
- pattern exploration: appositive clause 158; Bank of English 153; co-occurrences of language forms 154, 155; collocation 162–63; creativity and patterning 308–9; frequency 164–65; frequency, intuitions on 154–55; further reading 165–66; grammar, understanding through 227–29; Key Word in Context (KWIC) 156; lexis and patterning 199–200; meaning and patterns 212–13, 215–17; pattern recognition software 140–41; patterns 152–54; patterns of correspondence 228; reading concordances 156–62; relative clause 158; repetition 153–54, 155; short cuts 162–64; similarities 155–56; spotting patterns 154–56; superlative adjectives, use of 153–54
- Paulussen, H. 374
- Pawley, A. and Syder, F.H. 389, 390
- Pearce, M. 208, 597
- Pearson, J. 507, 511
- pedagogy: data-driven learning, pedagogical functions of 359–60; of EAP, influence of corpora on 461–63; pedagogic corpora 623–26; teacher education, pedagogic applications 623–26
- Peking University corpora 119
- Pence, L.K. and Justice, L.M. 184, 188
- Penn-Helsinki Parsed Corpus of Middle English (PPCME2) 113
- PENN-Treebank 555; grammar teaching materials 404
- Peréz-Paredes, P. and Alcaraz-Calero, J.M. 115
- Perez-Parent, M. 96
- performance: *versus* competence division in grammar teaching 401; focus on 8, 15, 18; language knowledge and 413; learner errors and 190; machine translation 496, 513; writing performance 450, 452, 637
- Perl 534
- permissions: seeking in corpus-informed course book design 417–18; spoken corpus building 42–44
- personal prejudice 429
- Peterson, J.C. and Brereton, L.H. 598
- Philbin, P. 582
- Philip, G. 310, 513
- phonetic transcription 45
- phrase-frames 213, 218
- phraseology: lexis 208–9; phraseological variation 217–21, 221–22; and phrases in lexis 199–201; of research articles 398
- phrases in concordances 173–75
- Pilnick, A. 606
- poetry and drama, language of 516–30; British National Corpus (BNC) 524; corpus-based case study – Blake's *Songs of Innocence and of Experience* 516–22; corpus-based case study – blockbuster movie scripts 523–29; drama and corpora 516; Film Studies 522; further reading 530; gender imbalance in blockbuster movies 523–29; interpretive significance 518–19; key words, analysis of 519–22; keyness 517–19; movie scripts, study of 522; Multilingual Corpus Toolkit 523; poetry and corpora 516; pronouns as key words in movies 524–25; semantic domains 517–19, 526–27; *Songs of Innocence and of Experience* (Blake, W.) 516–22; speech patterns in blockbuster movies 523–29; *Wordsmith Tools* (Scott, M.) 523
- Polish National Corpus 118; variety of a language, corpus building to represent 87
- political discourse 12, 591–603; analysis of 'how X is talked about' 595–96; Associated Press 594; British National Corpus (BNC) 591, 592, 597–98; corpora and 592–95; corpora to compile, development of new types 602; corpus-based research on, examples of topics in 598–601; corpus comparisons 596–97; Corpus of European Parliament Proceedings (EUROPARL) 593–94; Corpus of Spoken Professional American-English (CSPA) 594; CORPUS of tagged Political Speeches (CORPS) 593; corpus techniques in exploration of 595–98; criteria, explicitness in 594–95; definition of 591–92; euphemism 598–99; further reading 603; future prospects 601–2; genres of 592; genres to study, emergence of new 602; Hansard Corpus 593; keyword analysis 597–98; Labour, New and Old in UK 598; Lampeter Corpus 593; Lancaster Oslo-Bergen (LOB) corpus 592; LexisNexis Academic 594; linguistic features marking particular style, analysis of 597; London-Lund Corpus 592; mediated political discourse 594–95; metaphor in 599–600; news reports 594–95; oppositional stances, lexical differences between 599; parliamentary debates 593–94; political speeches 593; press conferences 594; reported speech in 600–601; representation, favourable or unfavourable, in 598–99; topics to cover, emergence of new 602; White House briefings, metaphor in 600
- Politics and the English Language* (Orwell, G.) 598
- polysemy 203–4
- Polytechnic of Wales (POW) Corpus 116
- Pope, R. 302
- Portuguese Corpus do Português 118
- Portuguese-English Parallel Translation Corpus (COMPARA) 511
- positivism 14

- Post-Observation Trainer-Trainee Interactions (POTTI) Corpus: small specialised corpora building 67; teacher education 627–28; variety of a language, corpus building to represent 88–89
- Potter, Liz 438
- power, ideology and 563–64
- Praat* 98–99, 549
- pragmatics 10–11, 288–300; conversational organisation 293–95; corpus and restriction in study of 289–91; cross-modal communication 298; definition of 288–89; discourse markers 295–96; evaluation 291–92; further reading 299–300; future considerations 298–99; general corpora, phenomena studied using 292–98; multi-modal communication 298; multi-modal corpora 298–99; phenomena studied using general corpora 292–98; pragmatic markers in corpus linguistics 11; pragmatic translation 501; semantic prosody 291–92; sociolinguistics and 559; speech act expressions 296–98; turn-taking in conversation 293–95
- Pratt, M.L. 303
- Precht, K. 246
- Pretsky, M. 408, 625–26
- Preparatory Action for Linguistic Resources Organisation for Language Engineering (PAROLE) 118
- prescriptive *vs.* descriptive debate 401
- press conferences 594
- Prince, E.F. 243
- Prince, E.F. and Thompson, S.A. 243
- Prodromou, L. 8, 305, 415
- professional curiosity 471–72
- Professional English Research Consortium (PERC) 114
- professional genres, corpus and 262–64
- professional introspection and development, specialised corpora for 627–29
- Project Gutenberg 61, 107, 114, 170, 517, 535
- proof-reading 340
- prosumer (producer-consumer) CDA 575–76
- Proust, Marcel 22
- Psathas, G. and Anderson, T. 40
- psychology, writing on 450
- psychometrics 634
- Published Materials Corpus (PMC): case studies 60; written corpus building 60, 62, 64
- Pulcini, V. and Furiassi, C. 272
- punctuation, problem of 424
- punning 308, 312
- Purpura, J. 635
- Puurtinen, T. 495
- Python* 534
- Quaglio, P. 32, 246
- Quaglio, P. and Biber, D. 244, 246
- qualitative approaches 535–36
- quantitative and qualitative revolution 17–18
- quantitative results, interpretation of 381–82
- query types 378–79
- questioned texts 579–80
- QuickTime* 100, 101
- Quirk, R. *et al.* 403
- Quirk, Randolph 15, 80, 81, 402
- Ragan, P. 453
- Ramer, A.L.H. 187
- Rampton, B. *et al.* 624
- random sampling 123
- Raupach, M. 389
- Raymond, W.D. *et al.* 556
- Rayson, P. 517
- Rayson, P. *et al.* 9, 517
- reading, use of corpus-based materials in 337–38
- Reading Academic Text (RAT) Corpus 114; specialist genres 260
- reading concordances 178–82; pattern exploration 156–62
- Received Pronunciation (RP) 111
- recording: materials for 475; participation and 475–76; recording files (spoken corpora) 34; spoken corpus building 41–42
- recycling data 181
- Reference Level Descriptors (RLDs) 639
- regional language variety 430
- registers, genres and 241–48; academic writing 244; Brown Corpus 241; Cambridge and Nottingham Corpus of Discourse in English (CANCODE) 241; *Cambridge Grammar of English* 243; conversation 244; formulaic language 245; further reading 248; genre perspective 241–42; genre studies, corpus-based 247–48; Lancaster Oslo-Bergen (LOB) corpus 241; lexical bundles 245; linguistic variation, corpus studies with register as predictor 242–46; London-Lund Corpus 241; *Longman Grammar of Spoken and Written English* 243; Michigan Corpus of Academic Spoken English (MICASE) 241, 244; modality, stance and 244; multi-dimensional studies of register variation 246–47; register as predictor, corpus studies of linguistic variation with 242–46; register perspective 241–42; register variation, corpus studies when focus is describing registers 246; register variation, multi-dimensional studies of 246–47; register variation, variety of a language, corpus building to represent 81; synthetic negation 243; text category, centrality of 241
- relative clause 158
- reliability in language testing 633

- Renouf, A. 57, 212, 219, 306, 319, 346, 535
 Renouf, A. and Banerjee, J. 212
 Renouf, A. *et al.* 512, 625
 repetition in pattern exploration 153–54, 155
 reporting verbs 339–40
 Reppen, R. and Simpson, R. 39, 41, 127, 130
 Reppen, R. and Vásquez, C. 246, 478
 Reppen, R. *et al.* 561, 623
 Reppen, Randi xxiii, 31–36, 69, 82, 96, 98, 238, 247, 258, 290, 430, 459, 474–75, 508, 535, 541, 548, 549, 595
 representativeness: assessment of 86–87; of corpora in teachers' research 472–73; vocabulary teaching materials 390; written corpus building 60
 research, teachers' use of corpora in 471–81; American National Corpus (ANC) 473; audio 475; Bank of English 473, 476–77; British National Corpus (BNC) 473, 476–77; Cambridge and Nottingham Corpus of Discourse in English (CANCODE) 477; Cambridge ESOL 475–76; Cambridge International Corpus (CIC) 473, 477; choice of appropriate corpus 474; classroom, corpora inside and outside 476–78; computer access 473–74; corpora, issues in building and use of 472–74; corpora inside and outside the classroom 476–78; corpus building 474–76; Corpus of Meetings of English Language Teachers (C-MELT) 478–80; further reading 481; *An Introduction to Corpora in English Language Teaching* (McCarthy, M.J. *et al.*) 474; language teacher education (LTE) 478; language teaching and corpora 478–80; Michigan Corpus of Academic Spoken English (MICASE) 474; professional curiosity 471–72; recording, materials for 475; recording, participation and 475–76; representativeness of corpora 472–73; Self-Evaluation of Teacher Talk (SETT) 478; skills for use of corpora 473–74; software access 473–74; studies, proliferation of 477; teacher development tool, potential for corpus-based research 472; *TeleNex* (Hong Kong) 477; T2K-SWAL 477; transcription of spoken data 475–76; *Wordsmith Tools* (Scott, M.) 479
 Resnik, P. and Elkiss, A. 409
 Resnik, P. and Smith, N. 512
 reverse word sorting 379
 Rey, J.M. 246, 247
 Rézeau, J. 347
 rhetorical functions 447
 Rhetorical Structure Theory (RST) 98–99
 Rice, P. and Waugh, P. 302
 Richards, J.C. 388
 Rimé, B. and Schiaraturat, L. 299
 Rochester, E. 533, 534
 Rodriguez Inéz, P. 513
 Roe, P. 321, 346
 Rogers, M. *et al.* 513
 Roland, D. *et al.* 403
 Römer, U. 244, 324, 325, 328, 329, 451, 471, 476
 Rosenthal, D. and Moore, S. 610
 Roy, A.M. 288, 289
 Rua, L. 306
 Rühlemann, Christoph xxiv, 10, 48, 270, 288–300
 Rundell, M. 356, 436, 442, 465
 Rundell, M. and Granger, S. 439
 Rundell, M. and Stock, P. 620
 Russell, J. 184
 Russian Reference Corpus (BOKR) 118
 Saarbrücken Corpus of Spoken English (SCoSE) 112
 Sacks, H. *et al.* 9, 293
 SACODEYL project 115; audio-visual corpus building 94; teacher education 625, 627
 St. John, E. 321, 327
 St Caro, Cardinal Hugo of 3
 St Thomas Aquinas 4, 22
 Salager, F. 244
 Salager-Meyer, F. and Defives, G. 245
 sampling: concordances 176; definition of sampling frame 391–92; limitations for small specialised corpora building 69; random sampling 123; random selection of concordance lines, tools for 433; sample base definition 57; sample corpora 20; sample numbers, determination of 57–58; sample size 56–59
 Sampson, G. 237
 Samraj, B. 247
 Samuda, V. 392–93
 Santa Barbara Corpus: corpora 111–12, 115; variety of a language, corpus building to represent 82
 Santos, D. 493
 Santos, D. and Frankenberg-Garcia, A. 512
 Sarangi, S. 606
 Sarangi, S. and Roberts, C. 264
 SAT Reasoning Test 642
 Saussure, F. de 14, 18, 20
 Saville, N. 633, 637
 Sawyer, R.K. 303, 305
 scaffolded acquisition 623
 ScanDiaSyn project 559
 Schauer, G.A. and Adolphs, S. 10, 274
 Schelfhout, C. *et al.* 551
 Schiffrin, D. 270, 272, 282, 295, 420
 Schmidt, R.W. 333, 337, 353, 354, 622
 Schmitt, D. and Schmitt, N. 464
 Schmitt, N. 388, 389, 424
 Schmitt, N. and Zimmerman, C.B. 388

- Schneider, K.P. and Barron, A. 90, 561
 Scholfield, P. 47
 Schulze, M. 192
 Science-specific EAP word list 459
 Scollon, R. and Scollon, S.W. 257
 SCOTS Corpus 111, 115, 119; audio-visual
 corpus building 100; spoken corpus building
 39; written corpus building 57
 Scott, M. and Tribble, C. 95, 156, 273, 460
 Scott, Mike xxiv, 5, 6, 61, 77, 95, 122, 123, 127,
 130, 136–50, 158, 164, 167, 171, 213, 215,
 216, 260, 277, 279, 311, 347, 350, 361, 393,
 502, 534, 565, 566, 595, 597, 609, 628; *see*
 also *Wordsmith Tools*
 Scotton, C.M. and Bernsten, J. 403
 Seale, C. *et al.* 609
 searching concordance lines 129–30
 Searle, J.R. 302
 Second Language Acquisition theory: data-
 driven learning 352–54; grammar teaching
 materials 404
 second language development: corpus
 deployment in revealing 188–91; literature on
 186; theories of 184–85
 Seidlhofer, B. 8, 238, 337, 360, 415
 selection, process in spoken corpus building 40
 self-directed cognitive development 622
 Self-Evaluation of Teacher Talk (SETT):
 research, teachers' use of corpora in 478;
 teacher education 623–24
 self-study, targets for 374
 semantic annotation 311–12
 semantic division 433–35
 semantic domains 517–19, 526–27
 semantic patterns 572–74, 574–75
 semantic prosody 291–92
 semantic significance 568–70
 Semino, E. and Short, M. 246, 516, 536
 sentences, examples of 438
 sets and synonyms 206–8
 sexually transmitted infections (STIs) 609, 610,
 611; 'sex' and 610–11; sexual health and
 adolescent language 609–10
 SGML 47
 Shakespeare, William 3, 22
 Sharoff, S. 506, 513
 Shklovsky, V. 302
 Shorrocks, G. 532
 Short, M. *et al.* 532
 short cuts in pattern exploration 162–64
 Shortall, T. 642
 Sigley, R. 208
 sign languages 94
 signposting and linking 339
 Sikand, A. *et al.* 612
 similarities in pattern exploration 155–56
 Simon-Vandenberg, A.-M. 493
 simple concordances 171–72
 Simpson, M. 518
 Simpson, P. 531
 Simpson, R. and Mendis, D. 245
 Simpson, R. and Mendis, P. 460
 Simpson, R. *et al.* 38
 Simpson, R.C. 86, 214, 245, 260
 Simpson, R.C. *et al.* 459, 463, 641
 Simpson-Vlach, R.C. and Leicher, S. 241
 Sinclair, J. and Brazil, D. 335
 Sinclair, J. *et al.* 212, 213, 214, 219, 223
 Sinclair, J.M. and Coulthard, R.M. 335, 623
 Sinclair, J.M. and Kirby, D.M. 320
 Sinclair, John M. 8, 16, 17, 18, 20, 23, 24, 140,
 346, 507, 535, 608; concordances 167, 171,
 172, 173, 178, 179, 180; corpus analysis 123,
 125; creativity 304, 307; dictionary writing
 428–29, 432, 438, 442; English for Academic
 Purposes (EAP) 465, 467; forensic linguistics
 585, 586; grammar 229, 231; language
 teaching 319, 321, 322, 323, 324, 327;
 language variety, representation of 85, 86, 87;
 learners and language corpora 372, 380; lexis
 198, 199, 200, 204; multi-word units 213,
 214, 217, 218, 220, 221, 222, 223; pattern
 exploration 154, 156; pragmatics 291, 292;
 small specialised corpora building 66, 79;
 specialist genres 255, 260; spoken corpus
 building 39, 40, 42, 44, 45; teachers' research
 472, 473, 476, 481; vocabulary teaching
 materials 390; written corpus building 62, 63
 Sinclair, S. 245
 Singapore Corpus of Research in Education
 (SCoRE) 94–95, 101; audio-visual corpus
 building 94–95, 101
 size of corpus *see* corpus size
 Skandina, I. 496
 Skelton, J. 607
 Skelton, J. and Hobbs, F. 606, 607
 Skelton, J. *et al.* 607
 Sketch Engine (Kilgariff, A. *et al.*) 431, 435, 437
 skills: in reading, use of corpus-based materials to
 teach 337–38; for use of corpora 473–74; in
 writing, use texts to develop 338–41
 Slovak National Corpus 118
 small specialised corpora building: advantages of
 small corpora 73–77; British National Corpus
 (BNC) 77; Cambridge and Nottingham
 Business English Corpus (CANBEC) 72,
 74–75; Cambridge International Corpus
 (CIC) 66, 68; Cambridge Learner Corpus
 (CLC) 66, 68; compilation 71–73; context
 and corpus, links between 66–67, 74; Corpus
 of American and British Office Talk (ABOT)
 68, 69, 70, 72, 75; design of, important
 considerations 69–71; English for Academic
 Purposes (EAP) 67, 68; English for Specific

- Purposes (ESP) 67, 68; Environmental Impact Assessment, Corpus of (EIA) 68; further reading 77–78; Hong Kong Corpus of Spoken English (HKCSE) 68, 71, 73; Indianapolis Business Learner Corpus (IBLC) 68, 71; International Corpus of Learner English (ICLE) 68; Limerick Corpus of Irish English (LCIE) 72–73; linguistic variability 69, 70; local contexts of use 74–75; Michigan Corpus of Academic Spoken English (MICASE) 68; POTTI Corpus 67; sampling limitations 69; situational variability 69; size, determination of 67–68; specialisation 67–68; specialisation, degrees of 68; suitability of 66–67; transcription 71–73; transcription conventions 77; US–Flemish–Finnish writing project 71; Vienna–Oxford Corpus of International English (VOICE) 72–73; *Wordsmith Tools* (Scott, M.) 77
- Small Talk* (Coupland, J.) 265
- Smith, D. *et al.* 609
- Smith, N. *et al.* 6
- social experience of health 605
- social processes, relationship with language 564
- Society for Adolescent Medicine 609
- sociolinguistics 11–12, 547–61; automatic speech recognition (ASR) 560; Bergen Corpus of London Teenage Language (COLT) 551, 555, 557; British National Corpus (BNC) 551, 552, 554–55, 555; browsing and searching 557; Cambridge and Nottingham Corpus of Discourse in English (CANCODE) 559; corpora, relevance in 547–49; corpus-based sociolinguistics exemplified 558–60; corpus interfaces 553; corpus-linguistic method, problems and limitations 555–56; Corpus Oral de Lenguaje Adolescente (COLA) 552; dataset comparability 557–58; definition of 547; dialectical forms, treatment of 553–54; form and function, distinction between 557; further reading 561; high-rising terminal (HRT) 558; interactional sociolinguistics 547; International Corpus of English (ICE) 552, 555, 558–59; Intonational Variation in English (IViE) 558; Key Word in Context (KWIC) 555; Lancaster Oslo–Bergen (LOB) corpus 548; language technology, initiatives in 560; limitations of corpus-linguistic method 555–56; linguistics and, connection between 548; listening to data 556; London–Lund Corpus 548, 551–52, 555; manual work, requirement for 557; methods in, use of corpora and 549–56; morphological information, tagging with 554; Multicultural London English (MLE) 555, 558, 559; non-standard forms, treatment of 553–54; Nordic Dialect Corpus 554; North Germanic languages, multilingual corpus of 559; Norwegian Spoken Corpus (NoTa) 552, 553; Penn Treebank 555; possible future developments 560–61; pragmatics and 559; precision and recall 557; problems of corpus-linguistic method 555–56; researchers do's and don't's 556–58; ScanDiaSyn project 559; spoken corpora, use of 549–55; Spoken Dutch Corpus (CGN) 551; spontaneous conversation 548–49, 550; syntactic information, tagging with 554; technical opportunities inherent in corpora 560; transcriber judgements 556
- Solan, L. and Tiersma, P. 10, 587
- Somers, H. 513
- Someya, Y. 347, 367
- Songs of Innocence and of Experience* (Blake, W.) 516–22
- sorting: concordance lines 129–30; concordances 175–76; data, corpus software and 142–47; reverse word sorting 379; tools, language corpora and 379–80
- Soundscaper* 98
- Spanish Corpus del Español 118
- specialisation: areas of, corpus use and self-study in 374; degrees of 68; small specialised corpora building 67–68
- specialised corpora 22, 114; construction of 507–8; for professional introspection and development 627–29; specialist genres 256–57; translation, use of corpora in 507–8
- specialist genres 255–67; academic genres, corpus and 259–62; Bank of English 256; British Academic Spoken English (BASE) 260; British Academic Written English (BAWE) 260; British National Corpus (BNC) 256; Cambridge and Nottingham Corpus of Discourse in English (CANCODE) 258, 260, 264–66; Cambridge International Corpus (CIC) 256; CANBEC 258, 266; corpora of specialised genres (CSGs) 255–58, 259, 264–65, 266–67; corpus and study of genres 255–56; Corpus of American and British Office Talk (ABOT) 258, 263; definition of 257; evidence-based language teaching 259; further reading 267; genre analysis, methodological advantages in 258–59; genre approach, rationale for 257–58; Language in the Workplace (LWP) 262; mega-corpora 256–57; methodological advantages in genre analysis 258–59; Michigan Corpus of Academic Spoken English (MICASE) 259–60; New Rhetoric School 257; non-institutional genres, corpus and 264–67; professional genres, corpus and 262–64; Reading Academic Text (RAT) Corpus 260; Small Talk (Coupland, J.) 265; specialist corpora 256–57; systemic functional linguistics (SFL) 257, 264–65; transactional discourse 263, 266

- specific language impairment (SLI) 187
- speech: corpus-based materials in teaching 334–36; corpus in comparison of real and represented speech 533–35; grammar of 236–37; patterns in blockbuster movies 523–29; real and fictional speech, similarities and differences 531–33; reported speech in political discourse 600–601; representation in literature 532–33, 535–36; speaker language 430; speech act expressions 296–98; speech corpora 110–12
- Sperberg-McQueen, C.M. and Burnard, L. 45
- Spinzi, C. 221
- spoken corpora 25–26; limitation of 440–41; oral fluency and 335–36; use in sociolinguistics 549–55
- spoken corpus building: analysis of spoken corpora 39, 48–49; anonymity, issues of 43–44; Augmented Multi-Party Interaction Corpus (AMI) 39; British National Corpus (BNC) 38, 45, 48; Cambridge and Nottingham Corpus of Discourse in English (CANCODE) 38, 41; COBUILD project 44; coding spoken data 47–48; consent 43; corpus design 39–42; Corpus Encoding Standard (CES) 48; development of spoken corpora 38–39; ethics 42–44; further reading 49; Hong Kong Corpus of Spoken English (HKCSE) 38; International Corpus of English (ICE) 38; International Phonetic Alphabet (IPA) 45; layout of transcript 45–47; Limerick Corpus of Irish English (LCIE) 38; London-Lund Corpus 38, 45; metadata 42–44; Michigan Corpus of Academic Spoken English (MICASE) 38; multi-modality 38–39; Network of European Reference Corpora (NERC) 44; Nottingham Multi-Modal Corpus (NMMC) 39, 45–46; permission 42–44; phonetic transcription 45; recording 41–42; SCOTS Corpus 39; selection, process of 40; SGML 47; Text Encoding Initiative (TEI) 44–45; transcription 44–48; XCES 48; XML 47–48
- spoken creativity 309–10
- spoken data 430
- Spoken Dutch Corpus (CGN) 551
- Spoken English Corpus (SEC) 111
- spoken language: features for language teaching 334–35; lexis 208–10; sets in audio-visual corpus building 93–94
- Spolsky, B. 634
- spontaneous conversation 548–49, 550
- Sripicharn, Passapong xxiv, 7, 9, 130, 158, 172, 320, 355, 360, 365, 367, 371–82, 477
- Stanger, Kelly 530
- Starfield, S. 447, 462
- Stenström, A.B. 11, 271
- Stenström, A.B. *et al.* 234, 559
- Stevens, V. 348, 349, 361
- Stewart, D. 308, 504, 506
- Stewart, M. 610
- Stiles, W.B. 291
- Stockwell, P. and Wynne, M. 533
- Stockwell, Peter 533
- Stoller, F.L. *et al.* 393
- Stotesbury, H. 244
- Strassel, S. and Cole, A. 41
- strings, comparison of 139–40
- Strong, J. 168, 169
- Stuart, K. 39
- Stubbs, Michael 59, 198, 199, 205, 206, 207, 213, 255, 271, 273, 274, 276, 304, 307, 424, 533, 565–66, 569, 579, 602
- studies: corpora and facilitation of creativity 310–13; multi-word units 215–17, 221–22; proliferation in teaching research 477; *see also* research, teachers' use of corpora in
- sub-corpora 123
- sub-technical vocabulary 448
- Sullivan, J.P. 537
- Sum* 429
- Sun, Y.-C. 367
- superlative adjectives, use of 153–54
- Survey of English Usage (SEU): corpus linguistics (CL) 15, 16, 21, 23; grammar teaching materials 403; written corpus building 56
- SUSANNE Corpus 112
- Sutarsyah, C. *et al.* 459
- Suzuki, L. and Calzo, J. 616
- Svartvik, Jan 16, 45, 581
- Swales, J.M. 244, 255, 257, 260, 267, 272, 449, 463
- Swales, J.M. and Burke, A. 244
- Swales, J.M. and Lee, D. 256, 257, 258, 259, 260, 262
- Swales, J.M. and Malczewski, B. 131
- Swales, J.M. and Najjar, H. 272
- Swales, J.M. *et al.* 244
- Swan, M. and Walter, C. 424
- Swan, Michael 402
- Switchboard Corpus 111
- synchronic data 430
- Synge, John Millington 539–40
- synonyms 206–7
- syntactic information, tagging with 554
- syntax 437
- synthetic negation 243
- systemic functional linguistics (SFL) 257, 264–65
- Taavitsainen, I. 247
- Taavitsainen, I. and Pahta, P. 247
- Taavitsainen, I., Melchers, G. and Pahta, P. 532
- Tagliamonte, S. 561

- Taiwanese Mandarin Balanced Corpus 119
 Tamasi, S. 533, 534
 Tankó, G. 340, 341
 Tannen, D. 152, 281, 308, 309
 Tao, H. 132, 133, 244, 272, 294, 295, 323, 335, 424
 Tao, H. and McCarthy, M.J. 132, 244
 Tao, T. and Zhai, C.X. 496
 Tapscott, D. and Williams, A. 576
 Tasker, Y. 522
 tasks: concordances of average, sample tasks based on 396–98; Discourse Completion Tasks (DCTs) 10; language learning, tasks for 326–27; language testing task formats 634–35; task objectives and corpora types, identification of 372–74
 Taylor, L. and Barker, F. 476, 636
 Taylor, L. *et al.* 402, 403
 teacher education 620–29; academic language corpora 624; British Academic Spoken English (BASE) 624; British Academic Written English (BAWE) 624; British National Corpus (BNC) 625; Centre for English Corpus Linguistics (CECL) 625; classroom language, corpus of 623–24; Compleat Lexical Tutor 625; Computer Mediated Communication (CMC) 626; construction of knowledge 622–23; context, relevance of 623, 624; contrastive linguistic examination of corpora 624–25; corpora and language awareness 626–27; corpora of classroom language 623–26; current models and place of corpus linguistics 621–23; discourses of 627–29; educational change, implications for 622; English as a Foreign Language (EFL) 620, 623, 624, 626; French Interlanguage Database (FRIDA) 625; further reading 629; Initiation-Response-Feedback (IRF) 623–24; International Corpus of Learner English (ICLE) 624–25; knowledge, construction of 622–23; language awareness and corpora 626–27; language teacher education context 620–21; learner corpora 623–26; Limerick Corpus of Irish English (LCIE) 624, 626; Louvain International Database of Spoken English Interlanguage (LINDSEI) 625; MA programmes 620–21; mediated acquisition 623; Michigan Corpus of Academic Spoken English (MICASE) 624; online communication, language of 626–27; pedagogic applications 623–26; pedagogic corpora 623–26; Post-Observation Trainer-Trainee Interactions (POTTI) 627–28; professional introspection and development, specialised corpora for 627–29; SACODEYL project 625, 627; scaffolded acquisition 623; self-directed cognitive development 622; Self-Evaluating Teacher Talk (SETT) 623–24; specialised corpora for professional introspection and development 627–29; teaching practice 627–28; terminological debates 622; Virtual Learning Environments (VLE) 626; WebCorp 625
 technology: influence in development of corpora 5, 15–17; language corpora, technical support for 377; strategic engagement with 447; technical opportunities inherent in corpora 560
 Teenage Health Freak 607–8
 TeleNex (Hong Kong) 477
 Tennent, P. *et al.* 41
 terminological debates 622
 Terrence Higgins Trust 612, 614
 test design in language testing 634–35
 Test of English as a Foreign Language (TOEFL) 634, 637, 640
 Tetreault, J. and Chodorow, M. 640
 Teubert, W. 156, 213, 274, 493, 592, 597–98
 Teubert, W. and Cermakova, A. 372
 Teubert, W. and Krishnamurthy, R. 467
 texts: abundance of electronic texts 563; alignment of text pairs 490–91, 511; Birmingham Collection of English Texts 24; category of, centrality of 241; collections of 32–35, 107, 490; computerising written texts 60–63; conversion by keyboarding 62–63; conversion by optical scanning 62; definition of 32–33; dictionary text, use of corpora in production of 433–39; diversity of 83–85; gathering written texts 60–63; genre, effect on text 580; grammatical co-text 231–32; Hong Kong Laws parallel text 488; Innsbruck Computer Archive of Machine-Readable English Texts (ICAMET) 113; interpretation and evaluation of 502–4; length and number of 85–86; multimedia corpora, multimodal texts 114–15; organising written texts 60–63; Oslo Corpus of Bosnian Texts 119; Oxford Text Archive (OTA) 107–8; presentation, use of corpus texts for 423–24; questioned texts 579–80; Reading Academic Text (RAT) corpus 114; saving of 33; skills in writing, use of texts to develop 338–41; structure of, CCSKs and 570–71; Text Encoding Initiative (TEI) 44–45; text structure, CCSKs and 570–71; textual systems 494; theoretical shift from text-linguistics to corpus linguistics 18–20; timespan, effect on text 580; types of, range of 578–79; writing skills, texts to develop 338–41
 Thomas, J. and Wilson, A. 606–7
 Thomas, S. and Hawes, T.P. 244
 Thompson, G. 348
 Thompson, G. and Hunston, S. 291

- Thompson, P. and Sealey, A. 246
 Thompson, P. and Tribble, C. 179, 395, 446, 447, 453, 460
 Thompson, Paul xxiv, 34, 39, 40, 41, 43, 44, 46, 93–103, 326, 459, 467, 595
 Thompson, S.A. and Mulac, A. 243
 Thornbury, Scott xxiv–xxv, 9, 130, 270–84
 Thorne, S.L. 185
Through the Looking Glass (Carroll, L.) 175–76
 Thurston, J. and Candlin, C.N. 366, 395, 445, 446, 447, 463
 Tiedemann, J. 511
 time: dimension of, corpora along 22–23; timespan, effect on text 580
The Times 23
 Timmis, I. 415
The Tinker's Wedding (Synge, J.M.) 539–40
 T2K-SWAL 477
 Todd, W. 373, 376
 TOEFL 2000 Spoken and Written Academic Language Corpus (T2K-SWAL): English for Academic Purposes (EAP), corpora and 459, 460; language testing 637
 Tognini Bonelli, E. and Camiciotti, G. 246
 Tognini Bonelli, Elena xxv, 4, 14–26, 69, 130, 213, 218, 221, 223, 258, 262, 263, 319, 413, 415, 472, 493, 547
 Tomblin, S.D. 585
 Toolan, M. 531
 topic corpora 21
 Tottie, G. 133, 243
 Towell, R. *et al.* 389
 transactional discourse 263, 266
Transcriber 549
 transcription: audio-visual corpus building 98–99; conventions for small specialised corpora building 77; of files (spoken corpora) 34–35; layout of transcript for spoken corpus building 45–47; small specialised corpora building 71–73; spoken corpus building 44–48; of spoken data 475–76; transcriber judgements 556; transcript attributes 190
 translation, use of corpora in 501–13; alignment of text pairs 511; *AntConc* 502, 509; British National Corpus (BNC-XML) 502–3, 504, 506, 508; collocation queries 504–5; comparable corpora 509–10; construction of specialised corpora 507–8; corpora, text interpretation and evaluation 502–4; corpus choice 504–11; corpus tools, integration into translators' toolkits 512; documentation 501; documents, linguistic authority of 505–6, 508; drafting 502; further reading 513; future prospects 512–13; general reference corpora 504–6; HM Revenue and Customs (HMRC) 503; internet, corpus construction from 512; language corpora, preparation of learners for using 373; machine translation 503–4, 513; monolingual corpora 504–8; multilingual corpora 509–11; multiple translation corpora 511; online interfaces 511; *OPUS* 511; *ParaConc* 491, 497, 502; parallel corpora 510–11; Portuguese-English Parallel Translation Corpus (COMPARA) 511; pragmatic translation 501; specialised corpora 507–8; specialised translations 503; TEI Consortium 508; translating, processes and purposes 501–2; translation comparatives 492–93; translation direction 488; translation hypotheses 505–6; 'translationese' 511; translator training, use of corpora in 512; *Utz* (Chatwin, B.) 506; *WebBootCaT* (Baroni, M. *et al.*) 440; *Wordsmith Tools* (Scott, M.) 502, 511; *XAIRA* 502, 504
 translation studies 9–10; parallel and comparable corpora 494–95
 translation styles 495
 Translational English Corpus (TEC) 494
 'translationese': parallel and comparable corpora 496; translation, use of corpora in 511
 Trappes-Lomax, H. 623
 triangulation of findings 123
 Tribble, C. and Jones, G. 326, 346, 366, 372, 380, 395, 444, 463
 Tribble, C. and Scott, M. 378
 Tribble, Christopher xxv, 3, 7, 8, 55, 67, 123, 129, 130, 158, 167–82, 200, 213, 215, 216, 221, 260, 261, 262, 263, 264, 360, 362, 378, 395, 401, 431, 447, 448, 461, 473, 477, 502, 566, 569, 595
 'trickle down' from research to teaching 444
 Trudgill, W. 81
 Tse, P. 222
 Tsui, A.B.M. 462, 471, 477
 Tucker, P. 244
 Turbo Lingo 409
 turn-taking: case studies 132–33; in conversation 293–95
 Turnbull, J. and Burston, J. 326
 Twain, Mark 534
 types: of corpora 107; of corpora use 402–6
 UCLA Chinese Corpus (UCLACC) 119
 unabomber 584–85
 UNESCO 611–12, 613; Unesco Index Translationum 490
 unification grammar 403
 University of Cambridge Local Examinations Syndicate (UCLES EFL) 634, 636
 University of Michigan English Language Institute (ELI-UM) 639, 641
 Upton, T.A. 247
 Upton, T.A. and Connor, U. 68, 71, 247

- Ure, J. 424
- US-Flemish-Finnish writing project 71
- validity: in language testing 633; of metrics 187–88
- van Dijk, Teun 564
- van Lier, L. 283
- van Rij-Heyligers, J. 453
- VanderVeen, A. *et al.* 642
- Vangsnes, Ø.A. 554
- Vannestål, M. and Lindquist, H. 461, 462
- VanPatten, B. 185
- Varantola, K. 337, 501–2
- variational pragmatics 89, 90
- Varieties of English for Specific Purposes Database (VESPA) 638
- variety of a language, corpus building to represent: American National Corpus (ANC) 87; balance, assessment of 86–87; British National Bibliographic Cumulated Subject Index (1960–64) 84; British National Corpus (BNC) 83, 84, 85, 86; Brown Corpus 84, 85, 86; Cambridge and Nottingham Corpus of Discourse in English (CANCODE) 82, 84–85, 86, 87, 89; case study, Irish family discourse 89–90; case study, Limerick Corpus of Irish English (LCIE) 88–89; corpus size 82–83; design issues 82–86; further reading 91; genre variation 81; International Corpus of English (ICE) 82, 84, 85, 87; Korean National Corpus (Sejong Balanced Corpus) 87; Lancaster Oslo-Bergen (LOB) corpus 84, 85, 86; Limerick Corpus of Irish English (LCIE) 88–89, 89–90; London-Lund Corpus 85; Longman Spoken and Written English Corpus (LSWE) 81, 87; Michigan Corpus of Academic Spoken English (MICASE) 83, 84, 86; Polish National Corpus 87; POTTI Corpus 88–89; register variation 81; representativeness, assessment of 86–87; Santa Barbara Corpus 82; texts, consideration of diversity to include 83–85; texts, length and number 85–86; variational pragmatics 90; variety of a language 80–82; Willings Press Guide (1961) 84
- variety of corpora, choice of 414
- Vásquez, C. and Reppen, R. 478, 628
- Vaughan, Elaine xxv, 9, 32, 471–81, 622, 628
- Ventola, E. 89
- verb reporting 339–40
- Viberg, Å 492
- video file formats 100
- Vienna-Oxford Corpus of International English (VOICE) 118; corpus-informed course book design 415; small specialised corpora building 72–73
- Vilha, M. 244
- Vine, B. 262
- Vintar, Š and Hansen-Schirra, S. 495
- Virtual Learning Environments (VLE) 626
- Vo, Thuc Anh xxv, 48, 237, 302–13
- vocabulary: associations in grammar 229–31; concordances use in vocabulary teaching 395–98; importance of 387–89; input of teaching materials 392–93; materials and formulaic language 389–91; sub-technical vocabulary 448; syllabus for corpus-informed course book design 416–15
- vocabulary teaching materials 387–98; abstract word groups 388; academic journal articles 177, 394–95; academic learning, corpus suitability for 391–92; Academic Word List (AWL) 391, 393; average, sample task based on concordances of 396–98; British Academic Written English (BAWE) 391; choice of suitable corpus 387–88; computerised corpus analysis 387; concordances, use of 395–98; concordances of average, sample task based on 396–98; corpus size 392; corpus suitability for academic learning 391–92; design of teaching materials 394–98; discipline-specificity 396–98; English for Academic Purposes (EAP) 391, 392–93, 394, 398; formulaic language, vocabulary materials and 389–91; frequency approach to identification of 390; further reading 398; *General Service List* 388; identification of formulae 388–89; Key Word in Context (KWIC) 396; keyword lists 394–95; manual analysis 388; methodological developments, ongoing nature of 390–91; Michigan Corpus of Upper-Level Student Papers (MICUSP) 391; needs analysis 391–92; phraseology of research articles 398; range of word meanings 394–95; representativeness 390; sampling frame, definition of 391–92; teaching materials, design of 394–98; teaching materials, vocabulary input of 392–93; vocabulary, importance of 387–89; vocabulary input of teaching materials 392–93; vocabulary materials and formulaic language 389–91; word definition 388; word-list compilation 388; *Wordsmith Tools* (Scott, M.) 393, 394
- Vongpumivitcha, V. *et al.* 374
- Vygotsky, L.S. 621, 623
- Walker, Brian xxv, 9, 309, 516–30
- Wall Street Journal* 574
- Wallace, M. 621, 622
- Walpole, J.R. 531
- Walsh, S. and O'Keeffe, A. 132, 624
- Walsh, Steve xxvi, 9, 333–43, 478, 481, 623–24
- Walter, E. and Moon, R. 7
- Walter, Elizabeth xxvi, 7, 66, 83, 123, 125, 158, 428–42, 465, 638

- Ward, J. 397
- Warren, Martin xxvi, 71, 72, 123, 125, 148, 200, 212–24, 279, 325, 327, 336, 388, 390, 419, 433, 502, 523
- Warwick, I. *et al.* 612, 613, 615
- Washington Post* 584
- Waters, A. 472
- Watney, S. 612, 613, 615
- Watson-Todd, R. 452
- Web Concordancer* 512
- WebCorp* 115, 512; in teacher education 625
- Weber, J.-J. 449–50
- websites: evolution of corpus linguistics (CL) 26;
TalkBank 115; web as reference corpus 115–16, 584–87; *WebKWic* 116
- Weir, C.J. 633, 634
- Weiss, M. 614
- Wellington Corpus of Written New Zealand English 109
- Welsh Cronfa Electroneg o Gymraeg 119
- Wen, Q. *et al.* 190
- Wenger, E. 257, 479
- Wennerstrom, A. 299
- West, M. 387, 388
- Westin, I. and Geisler, C. 592, 597
- Whistle, J. 360, 363, 365, 367
- White, L. 185
- White House briefings, metaphor in 600
- Wible, D. *et al.* 451
- Wichmann, A. 290, 558–59
- Widdowson, H.G. 8, 255, 259, 270, 276, 320, 321, 349, 359, 360, 395, 415, 564, 621, 622
- Wilcox, C. 308
- Wilkinson, M. 508
- Wilks, Y. 213, 218
- Williams, I. 502, 511
- Williams, J. 185
- Williams, Raymond 206, 565, 567, 597
- Willings Press Guide* (1961) 84
- Willis, D. 360, 366, 416
- Willis, D. and Willis, J. 326, 413, 416
- Willis, J. 337, 348, 349, 380
- Wilson, A. and Thomas, J. 312
- Wilson, A. *et al.* 239
- Wilson, J. 591, 596, 598
- Winalign* 511
- Wines, J. and Taggart, C. 404
- Wmatrix* 9, 517, 523
- WnCLE (L1 Spanish-L2 English) 639
- Wodak, R. *et al.* 564
- Wodak, Ruth 564
- Wolfe-Quintero, K. 186, 189, 190
- Wolfram, W. and Schilling-Estes, N. 81
- Wolverhampton Business English Corpus 114
- Wong, D. and Peters, P. 294
- Wood, D. 389
- Woollett, A. *et al.* 611
- Wools, D. 588
- Word Sketch* (Kilgariff, A. *et al.*) 435–36
- words: definition of 388; formation of 199; frequency lists, exploration of 123–26; frequency of use in literary speech 537–38; novel word formations 305–6; novelty in formation of 305–6; number of different words (NDWs) 185, 188; range of word meanings 394–95; usage patterns of 444–45; wild-card multi-word searches 173–75; word-forms 172–73; word-list compilation 147–48, 388
- WordSmith Tools* (Scott, M.): audio-visual corpus building 95; concordances 171; corpora 5, 6; English for Academic Purposes (EAP) 459; media discourse 566, 567, 569, 570, 571–72, 608, 628; parallel and comparable corpora 491, 497; poetry and drama, language of 523; research, teachers' use of corpora in 479; small specialised corpora building 77; translation, use of corpora in 502, 511; vocabulary teaching materials 393, 394
- WordSmith Tools* (Scott M.): literary speech representation 534
- World Englishes 415
- Wray, A. 389–90, 398, 586
- Wray, A. *et al.* 43, 45
- writing: corpus-based materials in teaching 337–38; performance in, evaluation of corpora and 452; texts to develop skills in 338–41; written creativity 309–10
- writing instruction, use of corpora for 444–54; accommodationist perspective 447; application issues 450–52; business writing 448–49; combinational functions 447; concordance output, multiple lines of 444; corpora, importance of role 445; corpora and tools 450–51; corpus-based approach in EGAP 445–46; corpus-based approach in ESAP 448–49; corpus-driven approach in EGAP 446–47; corpus-driven approach in ESAP 449–50; corpus research to teaching, 'trickle-down' effect 444; corpus tools for advanced-level students 451–52; data-driven learning (DDL) 444, 445; EGAP writing instruction, exploitation of corpora for 445–47; ESAP writing instruction, exploitation of corpora for 448–50; expansion of corpora, other varieties 452–53; extension of corpora, out of classroom 453–54; further reading 454; future pathways, expansion and extension 452–54; legal writing 449–50; local grammar in legal writing 449; psychology, writing on 450; rhetorical functions 447; strategy training for learners and teachers 451–52; sub-technical vocabulary 448; technology, strategic engagement with 447; tools, corpora and

- 450–51; 'trickle down' from research to teaching 444; words, usage patterns of 444–45; writing performance, evaluation of corpora in 452
- written corpus building: academic integrity, balance with practical reality 63–64; adaptation of material already in electronic form 62; annotation 63; Århus Corpus of Contract Law 56; balance 60; Bank of English 55; basic considerations 53–54; British National Corpus (BNC) 55–56, 58, 59, 60; Brown Corpus 56; Business English Corpus (BEC) 55–56, 58, 59, 62, 64; Cambridge International Corpus (CIC) 55; case study, determination of size 55–56; case study, Published Materials Corpus (PMC) 60; COBUILD project 56; computerising written texts 60–63; confidentiality 63; conversion by keyboarding 62–63; conversion by optical scanning 62; corpus exploitation 55; creation of new corpus, rationale for 54; data 61; data entry 62–63; data storage and retrieval 63; ethics 63; exemplar testing 55; further reading 64; gathering written texts 60–63; Guangzhou Petroleum English Corpus 56; HKUST Computer Science Corpus 56; HTML codes 61; keyboarding 62–63; make-up and size of samples 58–59; Medical Anatomy Corpus (MAC) 62, 64; optical character recognition (OCR) software 54, 62; optical scanning 62; organising written texts 60–63; parts-of-speech (POS) tags 63; planning 54–56; private data 61; publicly available data 61; Published Materials Corpus (PMC) 60, 62, 64; representativeness 60; sample base definition 57; sample numbers, determination of 57–58; sample size 56–59; SCOTS Corpus 57; size, considerations of 54–56; Survey of English Usage (SEU) 56
- Written English Corpus of Chinese Learners (WECCL) 190
- Wynne, Martin 9, 39, 237, 533, 535
- XCES 48
- Xiao, R. and McEnery, T. 353
- Xiao, R. *et al.* 624
- XML: audio-visual corpus building 96, 99, 102, 103; language development, corpus software in research 189; spoken corpus building 47–48
- Yates, S.J. 272
- Yoon, H. 452, 461, 462
- Yoon, H. and Hirvela, A. 321, 348, 452, 461
- York-Helsinki Parsed Corpus of Old English Poetry 113
- Yule, G. 288
- Zanettin, F. 327, 495
- Zanettin, F. *et al.* 495, 512, 513
- ZJU Corpus of Translational Chinese (ZCTC) 119
- Zurich English Newspaper Corpus (ZEN) 113
- Zwier, L.J. 373