

CONTENTS

CONTENTS	v
SERIES EDITOR'S INTRODUCTION	xi
FOREWORD	xiii
PREFACE	xv
<hr/>	
PART 1: Introducing the corpus	1
1. INTRODUCING ICE-GB	2
1.1 AIMS AND BACKGROUND	2
1.2 CORPUS DESIGN	4
1.3 EXTRA-CORPUS MATERIAL	8
1.4 COPYRIGHT	9
1.5 TRANSCRIPTION AND MARKUP	9
1.6 PART-OF-SPEECH TAGGING	13
1.7 SYNTACTIC PARSING	14
1.8 CROSS-SECTIONAL CHECKING	17
1.9 DIGITIZATION	17
1.10 EXAMINING ICE-GB TEXTS	18
2. THE ICE-GB GRAMMAR	22
2.1 INTRODUCTION	22
2.2 ICE WORD CLASSES	23
2.2.1 Adjective (ADJ)	24
2.2.2 Adverb (ADV)	25
2.2.3 Article (ART)	27
2.2.4 Auxiliary verb (AUX)	27
2.2.5 Cleft it (CLEFTIT)	29
2.2.6 Conjunction (CONJUNC)	29
2.2.7 Connective (CONNEC)	30
2.2.8 Existential there (EXTHERE)	30
2.2.9 Formulaic expression (FRM)	30
2.2.10 Genitive marker (GENM)	31
2.2.11 Interjection (INTERJEC)	31
2.2.12 Noun (N)	31
2.2.13 Nominal Adjective (NADJ)	33
2.2.14 Numeral (NUM)	33
2.2.15 Preposition (PREP)	34
2.2.16 Proform (PROFM)	35
2.2.17 Pronoun (PRON)	36
2.2.18 Particle (PRTCL)	38
2.2.19 Reaction signal (REACT)	38

2.2.20 Verb (V)	38
2.2.21 Miscellaneous tags	41
2.3 FUNCTIONS AND CATEGORIES	42
2.3.1 Adverbial (A) [Function]	42
2.3.2 Adjective Phrase (AJP) [Category]	43
2.3.3 Adjective Phrase Head (AJHD) [Function]	43
2.3.4 Adjective Phrase Postmodifier (AJPO) [Function]	43
2.3.5 Adjective Phrase Premodifier (AJPR) [Function]	43
2.3.6 Adverb Phrase Head (AVHD) [Function]	43
2.3.7 Adverb Phrase (AVP) [Category]	43
2.3.8 Adverb Phrase Postmodifier (AVPO) [Function]	44
2.3.9 Adverb Phrase Premodifier (AVPR) [Function]	44
2.3.10 Auxiliary Verb (AVB) [Function]	44
2.3.11 Central Determiner (DTCE) [Function]	44
2.3.12 Clause (CL) [Category]	45
2.3.13 Cleft Operator (CLOP) [Function]	45
2.3.14 Conjoin (CJ) [Function]	45
2.3.15 Coordinator (COOR) [Function]	45
2.3.16 Detached Function (DEFUNC) [Function]	45
2.3.17 Determiner (DT) [Function]	46
2.3.18 Determiner Phrase (DTP) [Category]	46
2.3.19 Determiner Postmodifier (DTPO) [Function]	46
2.3.20 Determiner Premodifier (DTPR) [Function]	47
2.3.21 Direct Object (OD) [Function]	47
2.3.22 Discourse Marker (DISMK) [Function]	47
2.3.23 Disparate (DISP) [Category]	47
2.3.24 Element (ELE) [Function]	47
2.3.25 Empty (EMPTY) [Category]	47
2.3.26 Existential Operator (EXOP) [Function]	48
2.3.27 Floating Noun Phrase Postmodifier (FNPPPO) [Function]	48
2.3.28 Focus (FOC) [Function]	48
2.3.29 Focus Complement (CF) [Function]	48
2.3.30 Genitive function (GENF) [Function]	48
2.3.31 Imperative Operator (IMPOP) [Function]	49
2.3.32 Indeterminate (INDET) [Function]	49
2.3.33 Indirect object (OI) [Function]	49
2.3.34 Interrogative Operator (INTOP) [Function]	49
2.3.35 Inverted Operator (INVOP) [Function]	49
2.3.36 Main Verb (MVB) [Function]	49
2.3.37 Nonclause (NONCL) [Category]	50
2.3.38 Notional Direct Object (NOOD) [Function]	50
2.3.39 Notional Subject (NOSU) [Function]	50
2.3.40 Noun Phrase (NP) [Category]	50
2.3.41 Noun Phrase Head (NPHD) [Function]	50
2.3.42 Noun Phrase Postmodifier (NPPPO) [Function]	51
2.3.43 Noun Phrase Premodifier (NPPR) [Function]	51
2.3.44 Object Complement (CO) [Function]	51
2.3.45 Operator (OP) [Function]	51
2.3.46 Parataxis (PARA) [Function]	51
2.3.47 Parsing Unit (PU) [Function]	52

2.3.48 Postdeterminer (DTPS) [Function]	52
2.3.49 Predeterminer (DTPE) [Function]	52
2.3.50 Predicate Element (PREDEL) [Category]	52
2.3.51 Predicate Group (PREDGP) [Function]	52
2.3.52 Prepositional (P) [Function]	52
2.3.53 Prepositional Complement (PC) [Function]	52
Prepositional Modifier (PMOD) [Function]	53
2.3.54 Prepositional Phrase (PP) [Category]	53
2.3.55 Provisional Direct Object (PROD) [Function]	53
2.3.56 Provisional Subject (PRSU) [Function]	53
2.3.57 Stranded Preposition (PS) [Function]	53
2.3.58 Subject (SU) [Function]	53
2.3.59 Subject Complement (CS) [Function]	54
2.3.60 Subordinator Phrase Head (SBHD) [Function]	54
2.3.61 Subordinator Phrase Modifier (SBMO) [Function]	54
2.3.62 Subordinator (SUB) [Function]	54
2.3.63 Subordinator Phrase (SUBP) [Category]	54
2.3.64 Tag Question (TAGQ) [Function]	55
2.3.65 Particle To (TO) [Function]	55
2.3.66 Transitive Complement (CT) [Function]	55
2.3.67 Verbal (VB) [Function]	55
2.3.68 Verb Phrase (VP) [Category]	55
2.4 FEATURE LABELS	55
2.5 SPECIAL TOPICS IN THE ICE-GB GRAMMAR	62
2.5.1 Inversion	62
2.5.2 Interrogative	63
2.5.3 Imperative	64
2.5.4 Coordination	65
2.5.5 Direct Speech	67

PART 2: Exploring the corpus

69

3. INTRODUCING THE ICE CORPUS UTILITY PROGRAM (ICECUP)	70
3.1 FIRST IMPRESSIONS	70
3.2 THE CORPUS MAP	71
3.3 BROWSING THE RESULTS OF QUERIES	72
3.4 VIEWING TREES IN THE CORPUS	74
3.5 VARIABLE QUERIES	74
3.6 'SINGLE GRAMMATICAL NODE' QUERIES	75
3.7 MARKUP QUERIES	76
3.8 RANDOM SAMPLING	77
3.9 TEXT FRAGMENT QUERIES	77
3.10 FUZZY TREE FRAGMENT SEARCHES	80
3.11 OPEN FILE	82
3.12 SAVE TO DISK	82
3.13 SEARCH OPTIONS	83

4. BROWSING THE CORPUS	85
4.1 THE IDEA OF CORPUS EXPLORATION	85
4.2 NAVIGATING THE CORPUS MAP	87
4.3 BROWSING SINGLE TEXTS	90
4.4 THE TEXT BROWSER WINDOW	91
4.5 VIEWING WORD CLASS TAGS	99
4.6 CONCORDANCING A QUERY	100
4.7 DISPLAYING TREES IN THE TEXT	102
4.8 GRAMMATICAL CONCORDANCING IN ICECUP 3.1	104
4.9 DISPLAYING TREES IN A SEPARATE WINDOW	106
4.10 CONCORDANCING, MATCHING AND VIEWING TREES	112
4.11 LISTENING TO SPEAKERS IN THE CORPUS	114
4.12 SELECTING TEXT UNITS IN ICECUP 3.1	115
5. FUZZY TREE FRAGMENTS AND TEXT QUERIES	117
5.1 THE TEXT FRAGMENT QUERY WINDOW	117
5.2 SEARCHING FOR WORDS, TAGS AND TREE NODES	118
5.3 MISSING WORDS AND SPECIAL CHARACTERS	121
5.4 EXTENDING THE QUERY INTO THE TREE	123
5.5 INTRODUCING FUZZY TREE FRAGMENTS	125
5.6 AN OVERVIEW OF COMMANDS TO CONSTRUCT FTFs	128
5.7 CREATING A SIMPLE FTF	129
5.8 ADDING A FEATURE AND RELATING A WORD TO THE TREE	133
5.9 MOVING NODES AND BRANCHES	140
5.10 APPLYING A MULTIPLE SELECTION AND SETTING THE FOCUS OF AN FTF	143
5.11 TEXT-ORIENTED FTFs REVISITED	146
5.12 THE GEOMETRY OF FTFs	151
5.13 HOW FTFs MATCH AGAINST THE CORPUS	155
5.14 THE FTF CREATION WIZARD: A TOOL FOR MAKING FTFs FROM TREES	167
6. COMBINING QUERIES	177
6.1 A SIMPLE EXAMPLE	177
6.2 VIEWING THE QUERY EXPRESSION	178
6.3 MODIFYING THE LOGIC OF QUERY COMBINATIONS	180
6.4 USING DRAG AND DROP TO MANIPULATE QUERY EXPRESSIONS	185
6.5 REMOVING PARTS OF THE QUERY	189
6.6 LOGIC AND FUZZY TREE FRAGMENTS	189
6.7 EDITING QUERY ELEMENTS	191
6.8 MODIFYING THE FOCUS OF AN FTF DURING BROWSING	194
6.9 BACKGROUND FTF SEARCHES AND THE QUERY EDITOR	196
6.10 SIMPLIFYING THE QUERY	198
7. ADVANCED FACILITIES IN ICECUP 3.1	203
7.1 INTRODUCING ICECUP 3.1	203
7.2 THE LEXICON	206

7.3 THE GRAMMATICON	209
7.4 STATISTICAL TABLES	211
7.5 LEXICAL WILD CARDS	214
7.6 EXTENSIONS TO FUZZY TREE FRAGMENT NODES	217
7.6.1 Performing exact matching in FTFs	221
7.6.2 Specifying missing features and pseudo-features	221
7.6.3 Specifying sets of functions, categories and features	223
7.6.4 Specifying a logical formula	228

PART 3: Performing research with the corpus **232**

8. CASE STUDIES USING ICE-GB	233
8.1 CASE STUDY 1: <i>PRETTY</i> MUCH AN ADVERB	233
8.2 CASE STUDY 2: EXPLORING THE LEXEME <i>BOOK</i> WITH THE LEXICON	237
8.3 CASE STUDY 3: TRANSITIVITY AND CLAUSE TYPE	239
8.4 CASE STUDY 4: <i>WHAT SIZE FEET HAVE YOU GOT?</i> WH-DETERMINERS IN NOUN PHRASES	244
8.5 CASE STUDY 5: ACTIVE AND PASSIVE CLAUSES	249
8.6 CASE STUDY 6: THE POSITIONS OF <i>IF</i> -CLAUSES	252
9. PRINCIPLES OF EXPERIMENTAL DESIGN WITH A PARSED CORPUS	257
9.1 WHAT IS A SCIENTIFIC EXPERIMENT?	258
9.2 WHAT IS AN EXPERIMENTAL HYPOTHESIS?	259
9.3 THE BASIC APPROACH: CONSTRUCTING A CONTINGENCY TABLE	262
9.4 WHAT MIGHT SIGNIFICANT RESULTS MEAN?	266
9.5 HOW CAN WE MEASURE THE 'SIZE' OF A RESULT?	267
9.5.1 Relative size	267
9.5.2 Relative swing	268
9.5.3 Chi-square contribution	268
9.5.4 Cramer's phi	269
9.6 COMMON ISSUES IN EXPERIMENTAL DESIGN	269
9.6.1 Have we specified the null hypothesis incorrectly?	269
9.6.2 Are all the relevant values listed together?	270
9.6.3 Are we really dealing with the same linguistic choice?	272
9.6.4 Have we counted the same thing twice?	272
9.7 INVESTIGATING GRAMMATICAL INTERACTIONS	273
9.8 THREE STUDIES OF INTERACTION IN THE GRAMMAR	275
9.8.1 Two features within a single constituent	275
9.8.2 Two features in a structure	277
9.8.3 A feature and an optional constituent	280
9.8.4 Footnote: dealing with overlapping cases	281

PART 4: The future of the corpus **284**

10. FUTURE PROSPECTS	285
10.1 EXTENDING THE ANNOTATION IN THE CORPUS	285
10.2 EXTENDING THE EXPRESSIVITY OF FUZZY TREE FRAGMENTS	287
10.3 INCORPORATING EXPERIMENTS IN SOFTWARE	290

10.4 KNOWLEDGE DISCOVERY IN CORPORA	295
10.5 AIDING THE ANNOTATION OF CORPORA	297
10.6 TEACHING GRAMMAR WITH CORPORA	299
REFERENCES	301
APPENDIX 1. ICE TEXT CATEGORIES AND CODES	307
A1.1 SPOKEN CATEGORIES	307
A1.2 WRITTEN CATEGORIES	308
APPENDIX 2. SOURCES OF ICE-GB TEXTS	309
A2.1 S1A-001 TO S1A-090: DIRECT CONVERSATIONS	310
A2.2 S1A-091 TO S1A-100: TELEPHONE CALLS	312
A2.3 S1B-001 TO S1B-020: CLASSROOM LESSONS	312
A2.4 S1B-021 TO S1B-040: BROADCAST DISCUSSIONS	312
A2.5 S1B-041 TO S1B-050: BROADCAST INTERVIEWS	313
A2.6 S1B-051 TO S1B-060: PARLIAMENTARY DEBATES	313
A2.7 S1B-061 TO S1B-070: LEGAL CROSS-EXAMINATIONS	314
A2.8 S1B-071 TO S1B-080: BUSINESS TRANSACTIONS	314
A2.9 S2A-001 TO S2A-020: SPONTANEOUS COMMENTARIES	314
A2.10 S2A-021 TO S2A-050: UNSCRIPTED SPEECHES	315
A2.11 S2A-051 TO S2A-060: DEMONSTRATIONS	316
A2.12 S2A-061 TO S2A-070: LEGAL PRESENTATIONS	316
A2.13 S2B-001 TO S2B-020: NEWS BROADCASTS	317
A2.14 S2B-021 TO S2B-040: BROADCAST TALKS (SCRIPTED)	317
A2.15 S2B-041 TO S2B-050: NON-BROADCAST SPEECHES (SCRIPTED)	318
A2.16 W1A-001 TO W1A-010: UNTIMED STUDENT ESSAYS	319
A2.17 W1A-011 TO W1A-020: STUDENT EXAMINATION SCRIPTS	319
A2.18 W1B-001 TO W1B-015: SOCIAL LETTERS	319
A2.19 W1B-016 TO W1B-030: BUSINESS LETTERS	321
A2.20 W2A-001 TO W2A-040: ACADEMIC WRITING	324
A2.21 W2B-001 TO W2B-040: POPULAR WRITING	326
A2.22 W2C-001 TO W2C-020: NEWSPAPER REPORTS	327
A2.23 W2D-001 TO W2D-010: ADMINISTRATIVE/REGULATORY WRITING	329
A2.24 W2D-011 TO W2D-020: SKILLS AND HOBBIES	329
A2.25 W2E-001 TO W2E-010: PRESS EDITORIALS	330
A2.26 W2F-001 TO W2F-020: FICTION	330
APPENDIX 3. BIBLIOGRAPHICAL AND BIOGRAPHICAL VARIABLES	332
APPENDIX 4. STRUCTURAL MARKUP SYMBOLS	333
APPENDIX 5. A QUICK REFERENCE GUIDE TO THE ICE GRAMMAR	334
APPENDIX 6. SPECIAL CHARACTERS USED IN ICE-GB	337
INDEX	338