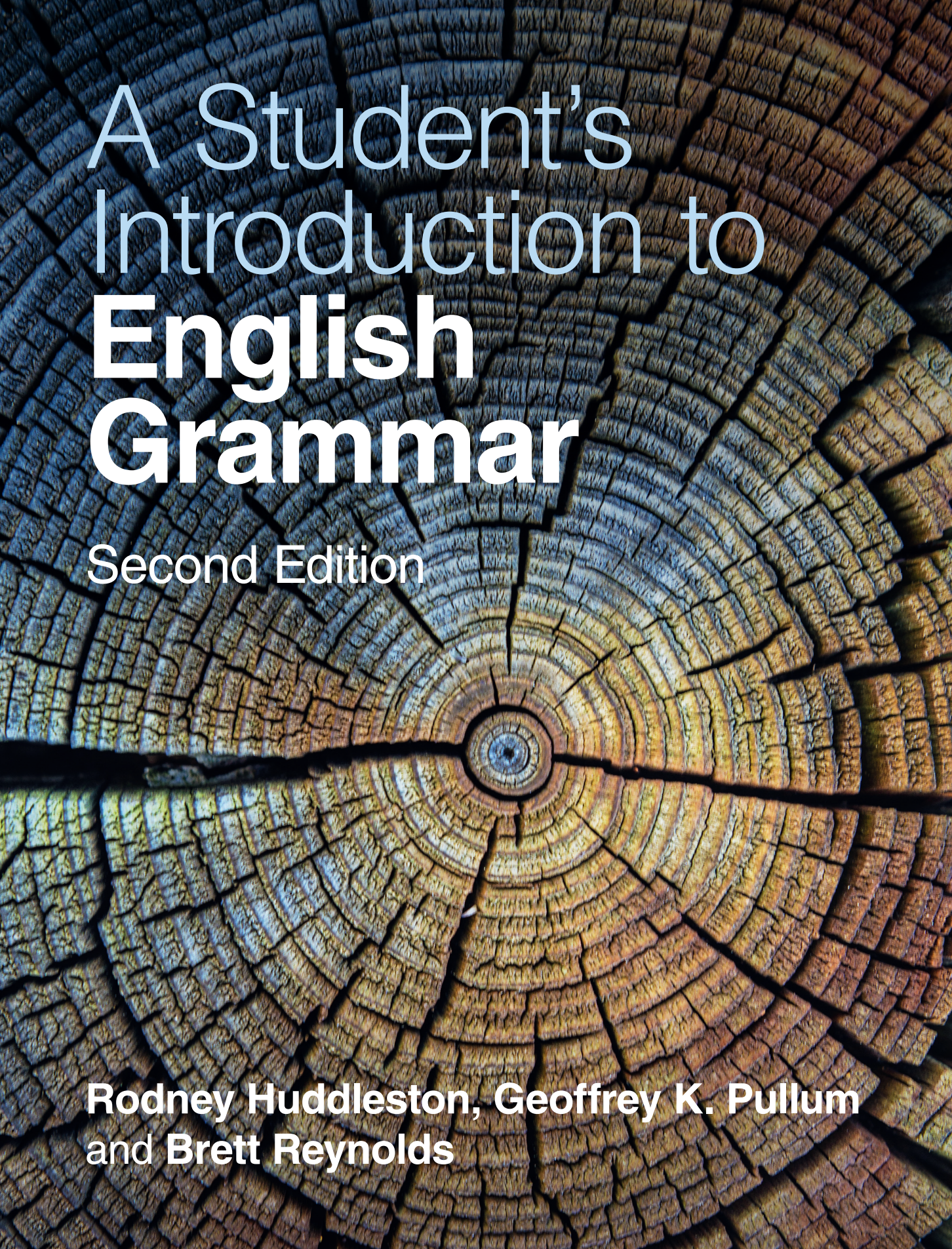

A Student's Introduction to
English Grammar
Second Edition
Rodney Huddleston, Geoffrey K. Pullum
and Brett Reynolds

A Student's Introduction to English Grammar

SECOND EDITION

A new edition of a successful undergraduate textbook on contemporary international Standard English grammar, based on Huddleston and Pullum's earlier award-winning work, *The Cambridge Grammar of the English Language* (2002). The analyses defended there are outlined here more briefly, in an engagingly accessible and informal style. Errors of the older tradition of English grammar are noted and corrected, and the excesses of prescriptive usage manuals are firmly rebutted in specially highlighted notes that explain what older authorities have called 'incorrect' and show why those authorities are mistaken. Intended for students in colleges or universities who have little or no background in grammar or linguistics, this teaching resource contains numerous exercises and online resources suitable for any course on the structure of English in either linguistics or English departments. A thoroughly modern undergraduate textbook, rewritten in an easy-to-read conversational style with a minimum of technical and theoretical terminology.

Rodney Huddleston was educated at Cambridge and Edinburgh, and taught English language at the University of Queensland for the majority of his career before beginning a decade of full-time work leading the team that produced *The Cambridge Grammar of the English Language* (*CGEL*) in 2002.

Geoffrey K. Pullum, a co-author of *CGEL*, was educated at York, Cambridge, and London, and has taught linguistics at University College London, the University of California, and the University of Edinburgh.

Brett Reynolds was educated at Temple University in Tokyo and is a professor at Humber College, Toronto, specializing in academic English and the teaching of English as a second language.

A Student's Introduction to English Grammar

SECOND EDITION

Rodney Huddleston
Geoffrey K. Pullum
Brett Reynolds

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/highereducation/isbn/9781316514641
DOI: 10.1017/9781009085748

First edition © Rodney Huddleston and Geoffrey K. Pullum 2005
Second edition © Rodney Huddleston, Geoffrey K. Pullum and Brett Reynolds 2022

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First edition 2005
19th printing 2020
Second edition 2022

Printed in the United Kingdom by TJ Books Limited, Padstow, Cornwall

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Huddleston, Rodney D., author. | Pullum, Geoffrey K. author. | Reynolds, Brett, 1969– author.
Title: A student's introduction to English grammar / Rodney Huddleston, Geoffrey K. Pullum, Brett Reynolds.
Description: Second edition. | [Cambridge, United Kingdom] ; [New York, N.Y.] : Cambridge University Press, 2021. | Includes index.
Identifiers: LCCN 2021025054 (print) | LCCN 2021025055 (ebook) | ISBN 9781316514641 (hardback) | ISBN 9781009088015 (paperback) | ISBN 9781009085748 (ebook)
Subjects: LCSH: English language–Grammar. | BISAC: LANGUAGE ARTS & DISCIPLINES / General
Classification: LCC PE1112 .H79 2021 (print) | LCC PE1112 (ebook) | DDC 428.2–dc22
LC record available at https://lccn.loc.gov/2021025054
LC ebook record available at https://lccn.loc.gov/2021025055

ISBN 978-1-316-51464-1 Hardback
ISBN 978-1-009-08801-5 Paperback

Additional resources for this publication at www.cambridge.org/SIEG2

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

Preface for the Student *page* xi
Preface for the Instructor xiii
List of Abbreviations xviii

1 Introduction 1
1.1 The English Language 1
1.2 Describing and Advising 7
1.3 The Structure of Sentences 11
1.4 Investigation and Disconfirmation 14
Exercises on Chapter 1 18

2 Overview of the Book 20
2.1 Word Forms and Lexemes 20
2.2 Phrases and Clauses 23
2.3 Verbs and Verb Phrases 25
2.4 Complements in the Clause 26
2.5 Nouns and Noun Phrases 27
2.6 Adjectives and Adverbs 29
2.7 Prepositions 31
2.8 Adjuncts 33
2.9 Negation 33
2.10 Clause Type 34
2.11 Subordinate Clauses 34
2.12 Relative Constructions 35
2.13 Comparative and Superlative Constructions 35
2.14 Non-Finite Clauses 35
2.15 Coordination 36
2.16 Information Packaging 37
Appendix: Notational Conventions 38
Exercises on Chapter 2 39

3 Verbs and Verb Phrases 42
3.1 Verb Inflection 42
3.2 Auxiliary Verbs 50

3.3 Perfective and Imperfective Interpretations 56
3.4 Primary Tense: The Present and Preterite 58
3.5 Secondary Tense: The Perfect 63
3.6 Progressive Aspect 67
3.7 Modality and the Modal System 69
Exercises on Chapter 3 80

4 Complements in Clauses 87
4.1 Introduction 87
4.2 The Subject 92
4.3 The Object 96
4.4 Predicative Complements 99
4.5 Overview of Complementation in VPs 104
Exercises on Chapter 4 106

5 Nouns and Determinatives 112
5.1 Introduction 112
5.2 Number and Countability 116
5.3 Determiners and Determinatives 123
5.4 Complements in NP Structure 127
5.5 Internal Modifiers in Nominals 129
5.6 External Modifiers 131
5.7 The Fused-Head Construction 133
5.8 Pronouns 137
5.9 Genitive Case 147
Exercises on Chapter 5 151

6 Adjectives and Adverbs 157
6.1 Adjectives 157
6.2 Adverbs 169
Exercises on Chapter 6 173

7 Prepositions and Particles 179
7.1 The Traditional Category of Prepositions 179
7.2 Extending the Preposition Category 181
7.3 Further Category Contrasts 186
7.4 Grammaticized Uses of Prepositions 189
7.5 Preposition Stranding 191
7.6 The Structure of PPs 194
7.7 PP Complements in Clause Structure 197
7.8 Prepositional Idioms and Fossilization 201
Exercises on Chapter 7 203

8 Adjuncts: Modifiers and Supplements 208
8.1 Introduction 208
8.2 Manner, Means, and Instrument 211
8.3 Act-Related Adjuncts 212
8.4 Space and Time 212
8.5 Degree 214
8.6 Purpose, Reason, and Result 214
8.7 Concessives 215
8.8 Conditionals 216
8.9 Four Other Clause-Modifying Adjuncts 217
8.10 Connective Adjuncts 220
8.11 Supplements 221
Exercises on Chapter 8 223

9 Negation 229
9.1 Negative and Positive Clauses 229
9.2 Subclausal Negation 230
9.3 Clausal Negation 232
9.4 Non-Affirmative Items 235
9.5 Scope of Negation 236
Exercises on Chapter 9 237

10 Clause Type 241
10.1 Speech Acts and Types of Clause 241
10.2 Interrogatives and Questions 244
10.3 Exclamatives 253
10.4 Imperatives and Directives 254
10.5 Performative Use of Speech Act Verbs 257
10.6 Minor Clause Types 258
Exercises on Chapter 10 258

11 Subordinate Clauses 262
11.1 Subordination 262
11.2 Clause Type in Content Clauses 263
11.3 Declarative Content Clauses 264
11.4 Interrogative Content Clauses 267
11.5 Exclamative Content Clauses 269
Exercises on Chapter 11 270

12 Relative Constructions 276
12.1 Relative Clauses as Modifiers in Nominals 276

12.2 Integrated versus Supplementary Relatives 282
12.3 Integrated and Supplementary Relative Words 287
12.4 Fused Relatives 289
12.5 A Relative Clause that Isn't a Modifier in a Nominal 292
Exercises on Chapter 12 292

13 Comparatives and Superlatives 297
13.1 Grade Inflection 297
13.2 *More* and *Most* 300
13.3 *Less* and *Least* 301
13.4 Comparison of Equality 302
13.5 Non-Scalar Comparison 303
13.6 Comparative Clauses 304
Exercises on Chapter 13 306

14 Non-Finite Clauses 311
14.1 Finite and Non-Finite Clauses 311
14.2 The Form and Meaning of Non-Finite Clauses 312
14.3 The Functions of Non-Finite Clauses 321
14.4 Transparent Verbs and Raised Subjects 324
14.5 Verbless Clauses 336
Exercises on Chapter 14 337

15 Coordinations 341
15.1 The Structure of Coordinations 341
15.2 Distinctive Syntactic Properties of Coordination 344
15.3 The Order of Coordinated Constituents 349
15.4 The Marking of Coordination 350
15.5 Layered Coordination 352
15.6 Main-Clause and Lower-Level Coordination 352
15.7 Joint versus Distributive Coordination 354
15.8 Non-Basic Coordination 354
Exercises on Chapter 15 357

16 Information Structure 360
16.1 Introduction 360
16.2 Passive Clauses 362
16.3 Extraposition 371
16.4 Existential Clauses 374
16.5 The *It*-Cleft Construction 377
16.6 Pseudo-Clefts 380

16.7 Dislocation: *He's clever, your dad* 381
16.8 Preposing and Postposing 383
16.9 Reduction 386
Exercises on Chapter 16 389

Index 395

Preface for the Student

In a sense, you already know English grammar. You must, in some unconscious way, if you're reading this. But being able to do something is different from understanding exactly what's being done. Knowing how to walk isn't the same as appreciating the anatomy of the human leg. The study of grammar involves developing an explicit account of how sentences are put together.

This involves investigation. The principles of English grammar are not set down in detail somewhere in some authoritative source, like etiquette rules or legal statutes. They have to be discovered through research, theory formation, and testing. Figuring out how to state the right set of principles and generalizations – even for English, which is almost certainly the most studied language in the world – is a deep and complex academic enterprise in which new results are still emerging.

That's not the impression you would get from the thousands of websites and popular books that claim to give advice on English grammar. They treat English grammar as known doctrine. And, to be frank about it, they repeat useless definitions formulated hundreds of years ago and propose baseless prohibitions and restrictions. Often they're inaccurate or even self-contradictory. This book breaks with a tradition going back centuries and presents a consistent analysis of English grammar that takes account of what's been discovered in modern linguistic research.

Studying English grammar is practically valuable, but it's also intellectually fascinating. It will give you a deeper insight into sentence structure, opening up new approaches to interpreting, appreciating, and using English effectively. That's not to say it will magically improve your writing or public speaking; but it will provide a solid basis for making progress toward that goal.

It will also provide you with some protection from the grammar bullies. Most English speakers have encountered nitpickers who seize upon sentences that they say are 'bad grammar'. Not just unintended slips of the tongue or typing errors; everybody makes those occasionally; they're making mystifying accusations about perfectly ordinary expressions, and calling them grammar errors. The beliefs held by these nitpickers are often grounded in myth rather than fact and refutable just by looking at examples of competent writing or speech. This book aims to help you resist bad advice from ill-informed error-spotters and usage snobs.

It means learning some new concepts and terms. If you're being advised to avoid passives, or get rid of adjectives, or shun split infinitives – familiar but misguided

warnings offered by thousands of writing tutors – you at least need to know what passives and adjectives and infinitives are, and what the facts are about how they are used. Then you can decide whether you want to follow the advice.

Along the way, this book will introduce you to some genuinely unexpected things about the English language. Too often grammar has been treated as a necessary set of dry instructions about fiddly details. That is nothing like what we present in this book. We've found it stimulating and intriguing to work on figuring out how sentence structure works and meaning is expressed. We hope it will be for you.

Preface for the Instructor

This book is designed to provide the basis for a one-semester course on the grammatical structure of the English language. It's aimed at departments of English as well as theoretical or applied linguistics. This new edition has been thoroughly revised throughout, with a view to increasing the accessibility of the book for students. We do draw extensively on discoveries made during the past century of linguistic research on English, but we don't presuppose prior acquaintance with linguistics, let alone a specific theoretical orientation or framework. We explain technical terms as they are introduced, and minimize arcane concepts and notations.

Despite thousands of changes, improvements, and updates, the structure of the book remains essentially the same as in the first edition, with one exception: we have added a new Chapter 8, surveying the types of adjuncts occurring in the structure of clauses. This has the happy consequence that the topics of the chapters now align exactly with the content of *The Cambridge Grammar of the English Language* (Rodney Huddleston and Geoffrey K. Pullum et al., Cambridge University Press, 2002; henceforth *CGEL*), the much larger reference grammar on which this book is based. So for anyone who wants greater detail about any topic in this book, student or instructor, the first course of action is simply to consult the chapter with the same number in *CGEL*.

The book is suitable for use with students of any English-speaking background, and the first edition has been used by many instructors at universities around the world. The authors, between them, have decades of experience teaching this material in Britain, Australia, Canada, and the USA. Although the two great dialect clusters of English – British and Australasian (BrE) and North American (AmE) – do differ syntactically in a few minor ways, the differences are almost always a matter of preferences rather than prohibitions, and they give rise to remarkably few difficulties for exposition. In a few clear cases we briefly discuss AmE/BrE divergence, but it's not a primary focus. The only thing that will look distinctively British to American readers is the spelling, where a dialect-defining binary choice had to be made. But seeing BrE spellings like *centre*, *colour*, and *signalling* (as opposed to AmE *center*, *color*, and *signaling*) shouldn't faze any literate American.

Chapter 1 is a brief introduction to some general issues relating to the linguistic study of English. Chapter 2 introduces some crucial concepts and provides a brief survey of the content of the entire book – a rapid summary of what will be

subsequently covered in more detail, the idea being to introduce much of the terminology and give a sense of where to look for particular phenomena. The detailed content of the book is in the fourteen subsequent chapters. Chapters 3, 5, 6, and 7 deal with the major lexical categories – verbs, nouns, determinatives, adjectives, adverbs, and prepositions – and the structure of the phrases that they head. (Chapters 3 and 5 are unavoidably lengthier than any others, because of the complexity of the verb system and the structure of NPs, and might need to be allotted double the class time spent on other chapters.) Chapter 4 describes the structure of simple, positive, active, declarative, non-coordinate clauses with no special stylistic reordering, which we call 'canonical clauses' – a term that we do not take to have theoretical significance, but use solely an expository convenience. (Those acquainted with the American linguistics of the 1950s may correctly see in it an echo of the notion of 'kernel sentence'.) The idea is to describe canonical clauses first, and then systematically tackle the ways in which other clauses diverge.

Chapters 9 to 16 present a step-by-step introduction to the ways in which clauses may diverge in structure from canonical ones: negation (Chapter 9), non-declarative clause types (Chapter 10), subordination (Chapter 11), relativization (Chapter 12), comparison (Chapter 13), non-finiteness (Chapter 14), coordination (Chapter 15), and the discourse-sensitive 'information-packaging' constructions that for many syntacticians are the most interesting part of English syntax (Chapter 16). We have relegated word-internal structure (inflectional morphology, the basics of lexical word formation, and the associated spelling rules) to an appendix published online.

This text is not advanced in the sense of needing prerequisite courses, but it is by no means a popularization. It is intended as the basis for a serious and detailed introduction to English grammar for undergraduates and masters-level students. Covering it all would be a substantial diet for a semester, because our coverage is unusually complete: rather than cherry-picking phenomena that highlight points of syntactic interest, we cover virtually the entire range of constructions found in sentences of contemporary English. Some instructors (especially those on ten-week rather than fifteen-week terms) have found it best to omit specific chapters from the course: it is possible to skip such topics as negation (Chapter 9), comparatives and superlatives (Chapter 13), or coordination (Chapter 15) without the gaps causing much trouble. And some adopters have told us they like to reorder things so that the interesting and important material of Chapter 16 on discourse-sensitive syntactic constructions is reached earlier.

The exposition is very deliberately informal in style. Too many scholars have felt obliged to write sentences like 'One should not imagine, however, that such locutions are unattested' when they could have said 'But you shouldn't think people never say that.' We've tried to lean toward the latter style. Syntax can be hard enough without dense expository prose making it harder. A modern textbook should employ what we call normal style – roughly the kind of conversational

language most instructors would use when explaining something in the classroom. We don't want to encourage one of the most damaging errors made by English teachers in the past: confusing formality with correctness, and consequently condemning features of ordinary conversational Standard English as 'bad grammar'.

We have also reduced the terminological burden a bit in this edition. Unfamiliar terms that did not do a lot of heavy lifting have been removed. For example, we have avoided the use of the term 'catenative' for the chained-together subjectless nonfinite clauses seen in *tends to try to seem to avoid being noticed*, though the coverage of important topics like raising in Chapter 14 is basically unchanged.

We have not attempted to cover topics like historical change, sociolinguistic variation, or the structure of non-standard dialects. Semantics is discussed only very informally, and we touch only lightly on anaphora, deixis, and ellipsis (*CGEL*'s Chapter 17); the interpretation of sentences with omitted or reduced parts connects to difficult questions in syntax, semantics, discourse, logic, and pragmatics that we feel are best covered in a second-level course. And we don't cover punctuation (*CGEL*'s Chapter 20), despite its critical importance for literacy, because its fixed and conventional nature – often stipulated in publishers' style sheets – makes it very different from the living and evolving phenomena of syntax that we are primarily concerned with.

A central aim of this text is to clarify the data that any theory of English syntax must be able to describe. In assessing the facts of usage we have regularly had recourse to standard data sources like the British National Corpus (BNC) and the enormously useful corpora made available by Mark Davies through **english-corpora.org**: the Corpus of Contemporary American English (COCA), News On the Web (NOW), the iWeb corpus, and others. But we don't try to illustrate with corpus examples throughout, or provide sources for the attested sentences we do use. The examples would nearly always have had to be adapted anyway, for brevity or to remove puzzling distractors. It could be a very useful exercise for students to test our descriptive generalizations by carrying out their own corpus investigations.

We have included more exercises than the first edition had. We would advise instructors to consider what they want to accomplish with exercises and make selective use of our suggestions. The exercises vary a great deal in depth and complexity. Some are quick and easy, but others are essay or research questions that may take the student many hours to complete, and students should be warned of that. Online multiple-choice exercises and other materials will be made available online using the Avallain e-learning platform. These will be accessible through the Cambridge University Press website at www.cambridge.org/SIEG2 and in some cases can be integrated directly into local learning management systems, permitting limited immediate feedback for students.

We provide a number of Usage Controversy Notes covering points where English speakers disagree with each other about what to call 'correct'. But we decided not to

include end-of-chapter bibliographic notes on sources or further reading in the book itself, given the open-ended nature of that enterprise; instead, we plan to provide such notes in an online document on the Cambridge University Press website at www.cambridge.org/SIEG2.[1]

Linguists will see that we reject some assumptions quite widely held in twentieth-century generative linguistics. The differences are sharp and explicit enough that they should provide grounds for discussion without causing confusion. For example, we do not believe subordinators ('complementizers') or coordinators ('conjunctions') are heads, and we treat *every day* as a noun phrase headed by *day* rather than a determinative phrase headed by *every*. For instructors in linguistics departments, our decisions on such points do not by any means preclude the possibility of their theoretical pros and cons being discussed as part of the course. For others, such stands on theoretically controversial points will matter little and may pass unnoticed. The important thing is that we are consistent: the assumptions we adopt are maintained throughout. That does not mean we are legislating a theoretical view: it is always possible to stop and ask whether certain facts about syntax are better explained under one theoretical conception rather than another.

One of the most important points we draw from our linguistics backgrounds, and try to impart to the reader, is that grammatical study is (or ought to be) a matter of discovery rather than legislation or pontification. Too often English grammar has been regarded as a collection of timeless edicts defining what is 'proper', or reduced to a short list of tips and hints for the insecure. This is unfortunate, because the grammars of English published over the last two or three centuries have created a descriptive tradition that we see as gravely flawed. In this book we reject much traditional dogma. Since the late eighteenth century, when Lindley Murray sold a million copies of his grammar textbook (largely plagiarized from Robert Lowth's three decades earlier), students have been told that nouns are words that name things; that pronouns are words that stand in place of nouns to avoid repeating them; that prepositions stand before nouns to relate them to other nouns; that the subject says what the sentence is about; that the object can be defined via the weird phrase 'receiver of the action'; that relative clauses are 'adjective clauses' because they 'describe nouns'; and so on. It is high time serious courses on the structure of English broke away from repeating such time-worn nonsense.

However, when we depart from the tradition, we do so in a strict and consistent way, following *CGEL*. Adopting even a minor change in our analyses can lead via a ripple of consequences to a contradiction. Working out the consequences of such potential revisions could lead to interesting ideas for final-year honours

[1] One other temptation we have resisted is adding subsidiary points in footnotes. Lovers of arcane footnotes, if there are any, will be saddened to discover that this is the only footnote in the book.

dissertations or master's theses that take the study of English syntax further. Nothing would please us more than seeing students working from this book improve on our description through their own research.

Finally, like all authors we have ended up with many debts, more than we can explain in detail. We particularly thank Jim Donaldson for many thoughtful comments; Aileen Bach, Michael DiPetta, and Peter Evans for comments, corrections, and proofreading; John Payne for advice and clarification on many syntactic points; John Joseph for discussion of the concept of Standard English; and Bryan Garner for useful discussion of controversial usage. We have naturally spent less time with our nearest and dearest than we could have done if we were not working on this book, and we thank them for their tolerance (they know who they are). And for obvious reasons, given that we completed this revision in the years 2020 and 2021, we thank everyone in the business of public health and vaccine production.

Abbreviations

CATEGORIES

AdjP	adjective phrase
AdvP	adverb phrase
$\text{Clause}_{\text{REL}}$	relative clause
Crd	coordinator
D	determinative
DP	determinative phrase
Intj	interjection
N	noun
N_{PRO}	pronoun
Nom	nominal
NP	noun phrase
P	preposition
PP	preposition phrase
Sbr	subordinator
V	verb
VP	verb phrase

FUNCTIONS

Comp	complement
Det	determiner
ExtMod	external modifier
Mod	modifier
O	object
Obj	object
O^{d}	direct object
O^{i}	indirect object
Pred	predicate
PredComp	predicative complement
S	subject
Subj	subject

OTHER

AmE	American English
BrE	British English
CGEL	*The Cambridge Grammar of the English Language*

do_{aux}	auxiliary verb ***do***
do_{lex}	lexical verb ***do***
if_{c}	conditional preposition *if*
if_{i}	interrogative subordinator *if*
sg	singular
too_{x}	*too* with the meaning "excessively" rather than "additionally"
we_{D}	the determinative *we* as in *We voters disagree.*
we_{P}	the pronoun *we* as in *We disagree.*
you_{D}	the determinative *you* as in *You politicians are all alike.*
you_{P}	the pronoun *you*
§	section

1 Introduction

1.1 The English Language

This book is about how sentences are constructed in English. Almost certainly, English is now the most important language in the world. This might not have been clear even just fifty years ago, but it cannot reasonably be denied today. English is used for government business in well over sixty countries and territories around the world. By international agreement it is the primary language for all air traffic control and maritime navigation. It's the uncontroversial choice of official language for almost all international academic conferences, and increasingly the main language of higher education globally. The European Union is increasingly using it for conducting business even though the United Kingdom is no longer a member.

The widespread use and enormous influence of English is not due to anyone having objectively judged it to be deserving. The unique status of the language is just a side-effect of various mutually reinforcing historical trends. From the eighteenth and nineteenth centuries onward, British exploration, missionary work, imperialism, and colonial policy began to spread English around the world. The whole North American continent ultimately adopted English as its dominant language (despite important early rivals like Spanish, French, and German). India's huge multilingual population recognized English as a subsidiary official language (alongside Hindi) for government business after independence. The twentieth century, especially after the outcomes of the first and second world wars, saw an extraordinary growth in the economic power of the USA and the British Commonwealth.

The growing importance of scientific research further strengthened the standing of English; it emerged as the language in which the majority of scientists publish their results. The economic importance of university education accentuated this effect: many universities in non-English-speaking countries allow or even require PhD dissertations to be submitted in English. And a hugely important additional factor was the global success of English-language broadcasting, publishing, songwriting, and above all film production: in Hollywood, the only place on earth where film-making budgets routinely run into the hundreds of millions of dollars, almost every film script is in English. It may not be a good thing for a single language to come so close to predominating all over the planet, but it seems to be happening.

Yet English may not have been the right language to pick as a global language. In fact a case could be made that it is quite ill-suited to its international role, for a number of linguistic reasons:

- The complex consonant clusters of spoken English are tongue-twisters for many foreign learners. The underlined portion of *Our strengths spring from our unity* can have either five, six, or seven consonantal sounds in a row, depending on how carefully the speaker struggles to pronounce it. (Many languages keep things nice and simple by avoiding adjacent consonant sounds in words.)
- Stress in words of more than one syllable is vital for intelligibility (and interacts in complex ways with rhythm and accent in sentences), but it is extraordinarily difficult for learners to get right: the syllable with heaviest stress is the first one in *cátapult*, the second in *carnívorous*, the third in *atavístic*, the fourth in *impressionístic*, the fifth in *amniocentésis*, and the sixth in *incomprehensibílity*. (In Hungarian things are easy: stress in a word is always on its first syllable.)
- English spelling is notoriously chaotic. Anyone who naively expected *bought*, *cough*, *dough*, *plough*, *rough*, *through*, and *borough* to have similar vowel sounds would be shocked to find that they all sound quite different and the spelling gives no clue. Some sounds in English have more than a dozen different spellings. (A language like Finnish has no such orthographic disorder; basic spelling can be learned in a few minutes.)
- Learners of English need to memorize the forms for about 200 irregular verbs: notice the strikingly different patterns for past tense formation in *bake* / *baked*; *break* / *broke*; *bring* / *brought*; *fall* / *fell*; *feel* / *felt*; *know* / *knew*; *have* / *had*; *make* / *made*; *sing* / *sang*; *take* / *took*; etc. (Swahili verbs, by contrast, have hardly any irregularity except in the verb meaning "be" in its present tense.)
- And the syntactic complexities of English – its intricate auxiliary verb system, varied patterns of verb/preposition pairings, profusion of distinct main and subordinate clause structures, and detailed principles of word order – will occupy the majority of this book.

Of course, other languages have their tricky bits too. English is not uniquely complicated. Nor is it any more logical, clear, regular, expressive, easily learnable, or otherwise perfect than any other language. The invented language Esperanto was specifically designed to be regular in structure and easy to learn, but despite being promoted by enthusiasts for more than a hundred years, no country has adopted it for official purposes, and the world has continued to gravitate toward English. Any other human language could in principle have served the same purposes, but none did. Today, given the extraordinary importance that has accidentally attached to English, with billions of people around the world using it every day, it is reasonable for any user to want not only a good grasp of its vocabulary but also a clear idea of its grammar. Providing it is the goal of this book.

1.1.1 Standard English

A language as widely distributed as English, socially and geographically, will exhibit a lot of social and regional variation. Linguists refer to varieties within a language as **dialects**. But the difference between being a dialect of some language and being a separate language is not clear-cut; to some extent it's tangled up with politics and identity. What's important is that linguists don't use the term 'dialect' just for rural, provincial, or uneducated varieties: they regard EVERYONE as speaking a dialect. The term 'English' actually covers a huge array of dialects, and this book is concerned with describing just one particularly tight-knit cluster of dialects usually referred to as **Standard English**.

There is no brief definition of Standard English; in a sense the whole of this book is a contribution to defining it. And the word 'standard' isn't supposed to imply that it has been approved by some committee, or that it's a model for everyone to aspire to or judge themselves by or fall short of. It's just a name for which we haven't got a good alternative. (People also talk about Standard Arabic or Standard Italian.)

One thing that could be said about the Standard English dialect cluster is that it has the property of being – in most contexts anyway – the variety of English least likely to draw attention to itself. It isn't from anywhere in particular. It's the variety of the language that most learners in English classes around the world aspire to learn, and the one in which English-language books, magazines, and newspapers are almost exclusively written. Broadly speaking, it is the kind of English least likely to be judged odd or inappropriate for any kind of public use, or to draw attention to the speaker's biography. In short, Standard English tends to convey less about a person speaking it than many other dialects do.

The word 'standard' does have the unfortunate property of suggesting some standard or level that has to be met, so that non-standard dialects are sub-standard. They are not, so resist that suggestion. Standard English owes its clout to historical accidents; things could have turned out differently. In saying that, we're not denying that it is used by the rich and powerful in the anglophone world: of course it is. But it's also used by millions of the poor and powerless. We're also not ignoring the existence of dialect prejudice: it is rampant, and millions of people suffer from being unfairly disrespected or ignored simply because of the dialect of English that they speak. We know that, and we deplore it. But the sociological facts of class bigotry and racial or national prejudice don't alter the facts of how sentences are structured in Standard English.

Misunderstandings about what 'Standard English' means will persist, we are sure of that. But keep in mind that on the occasions when we draw attention to features of non-standard dialects of English (as in 1.1.4 below), we are NEVER suggesting those dialects are **sub**-standard.

English dialects differ from each other most of all in **lexicon** (what words are in current use in typical speakers' vocabularies) and **phonology** (the system of pronunciation), and on these topics, this book says very little. Dialects differ much less in the structure of sentences, clauses, phrases, and words; and that will be our focus.

The grammatical variation within Standard English is trivial, but it tends to get a lot of attention from purists, trolls, snobs, critics, and bullies. They tend to obsess over a small set of well-known controversial points: how to use *hopefully*, where to use *whom*, choosing between *which* and *that*, saying *taller than me* or *taller than I*, and so on. These points of detail offer opportunities for alleging that someone has made a grammar error – and to imply on that basis that they're illiterate or stupid. Exaggerated emotions are provoked; angry letters to conservative newspapers are written. This sometimes makes it seem as if grammar is an area of major controversy. Yet it isn't. The largely unnoticed fact is that for the vast majority of questions about what's grammatically correct in Standard English, the answers are remarkably clear, as shown by huge masses of evidence. That widespread homogeneity is what makes a book like this possible.

1.1.2 Written and Spoken English

Human languages existed first in live-communication forms like speaking or hand-signing. Some languages, later in their history, came to be written down as well. The present writing system for English evolved slowly over nearly two millennia, and as we said in the previous section, it's illogical and irregular. But that's about the fit between spelling and sound. When we come to sentence structure, to a large extent the aspects of English we need to focus on tend to hold for both spoken and written English.

Sometimes we'll need to mention aspects of the language that are specific to speech (like intonation and stress) or limited to writing (like punctuation), but mostly we will be concentrating on grammatical facts that don't vary much between English as uttered orally and English as written or printed. The primacy of speech always tempts linguists to refer to 'the speaker' when talking about the person producing an utterance, but every time we say 'speakers' we intend to cover all users – writers as well as conversationalists.

1.1.3 British and American Subvarieties

Some minor points of grammatical difference can be found between two major subclusters within Standard English: the British, Australasian, and South African dialects which we will call **BrE**, and the Canadian and American dialects we will refer to as **AmE**.

One noteworthy example is that BrE users often say or write sentences like *I don't know if she's seen it yet, but she may have done*. That *done* on the end (we underline it

to draw attention to it) sounds quite odd to AmE speakers; they would say *I don't know if she's seen it yet, but she may have*.

Another such case is that singular nouns referring to groups of people with a common purpose, like companies or governments or teams, are treated as plural by most BrE speakers for purposes of verb agreement: in BrE a headline saying *England are collapsing* would probably be about a losing performance by the England cricket team, whereas *England is collapsing* would suggest an unprecedented geophysical catastrophe. AmE speakers tend to regard the first sentence as ungrammatical.

But most of the grammar differences between AmE and BrE are preferences rather than sharp distinctions. An AmE user will typically say *I did that already* where a BrE speaker would prefer *I've done that already*, though both versions would be understood by both communities of speakers. For the most part, it is hard to tell from grammatical features of a passage of written Standard English which side of the Atlantic it came from.

The most obvious sign of the difference is not in the syntax but in the well-known conventional spelling differences, which began to solidify when dictionaries were published by Samuel Johnson in England (1755) and Noah Webster in the USA (1828): a single occurrence of a spelling like *honor* tells you that you're reading AmE (because BrE would have *honour*). This book uses BrE spelling, but all literate users of English should be (and usually are) familiar with AmE spelling as well.

1.1.4 Other Dialects

There is dialect variation within Standard English: there are forms used by some speakers but not all. We indicate those, when it's relevant to the matter at hand, with a raised per cent sign (%). It'll mean that probably not everyone in a class of English speakers will agree on whether the sentence is grammatical.

One factor giving rise to dialect differences of this sort is that younger speakers don't always speak in exactly the same way as older speakers. In the BrE detective novels by P. D. James (born in 1920), characters say things like %*I hadn't a car back then* (where younger speakers would say *I hadn't got a car* or *I didn't have a car*), or %*If he had left I should have heard him* (the contemporary equivalent is *I would have heard him*), or %*Marcus was eleven years younger than I* (most contemporary speakers would say *younger than me*). The syntax of English changes only very slowly, but over a century the changes are perceptible. Baroness James's prose is not twenty-first century BrE. And AmE readers must occasionally find some of her expressions quite puzzling; even such simple phrases as *I was visiting my husband in hospital* are different (AmE speakers would always say *in the hospital*, without implying any prior mention of a specific hospital).

These minor dialect differences fall within the broad definition of Standard English, but in addition (as mentioned in 1.1.1) many non-standard dialects co-exist with Standard English, mostly in spoken form but also in dialogue passages written

by novelists. Standard English speakers in any moderately diverse community encounter these non-standard dialects every day in plays, films, songs, and conversations, so it is important not to be entirely ignorant of them. In the [b] cases of [1] the raised exclamation mark (ˡ) signals not that they are errors, but that they are correct form in several non-standard dialects; Standard English equivalents are given in [a].

[1]

		STANDARD ENGLISH DIALECT		NON-STANDARD ENGLISH DIALECTS
i	a.	*It doesn't matter what they did.*	b.	ˡ*It don't matter what they done.*
ii	a.	*I have never broken anything.*	b.	ˡ*I ain't never broke nothin'.*

- *Don't* as in [ib] is found in some non-standard dialects where Standard English would have *doesn't*, and *done* is found corresponding to Standard English *did*. The verb forms aren't accidental mistaken choices; it's just that not all dialects have the same verb morphology as Standard English.
- The word *ain't* in [iib] – an auxiliary verb form that famously signals a non-standard dialect – makes the clause negative, but the negation is also marked by *never*, and again by *nothin'*. People often call this 'double negation' (though if you count, you'll see that there are three marks of negation). What's actually going on is multiple marking of a single negation. This is normal in many languages, including Italian, Polish, and Russian. Standard English as in [iia] happens to lack this multiple-marking feature – though Standard English speakers know perfectly well how to interpret it when they hear it from a non-standard dialect speaker.

Features of the sort seen in the [b] examples of [1] would stand out as very surprising in a typical TV news broadcast or newspaper editorial – angry letters would flood in – because they are uncontroversially agreed to be non-standard. But people who naturally use the [b] sentences of [1] in conversation are not trying to utter the [a] versions. They said what they intended. And their speech is not deficient or illogical; they're using a dialect that has a different negation syntax from the Standard English one.

True errors in speech or writing are quite a different matter. When a newsreader misreads a word or a writer makes a typing error, it's a matter of intending one thing but accidentally producing another. That happens now and then to everybody, but it's sporadic and unpredictable – and even the person who made the error will usually agree, if it's pointed out, that it happened.

When our focus is entirely on Standard English, as it is throughout nearly all of the book, we'll use a prefixed asterisk to indicate sequences that are **ungrammatical**, in the sense of not conforming to the grammatical rules of the Standard English dialect as far as we can determine. So we might contrast *The dog ran away* (grammatical) with **Ran away the dog* (ungrammatical).

1.1.5 Formal and Informal Style

The distinction between standard and non-standard dialects of English is very different from the important distinction between different style levels within Standard English. What we mean by 'style' here is not literary style but just level of formality (sociolinguists often call it 'register'). There are many levels imaginable between (say) prepared text to be delivered as a speech on a very solemn occasion and casual email to a friend; we illustrate in [2]:

[2]

		LESS FORMAL (NORMAL)		MORE FORMAL
i	a.	*I'm the one she can rely on.*	b.	*I am the one on whom she may rely.*
ii	a.	*Who's it addressed to?*	b.	*To whom is it addressed?*

There's no call for the exclamation mark notation or the percent sign here. Standard English allows for plenty of variation in style depending on the context in which the language is being used. The [b] versions would generally be used only in very formal contexts indeed; in ordinary conversation they would seem pompous or just weird: uttering the [b] variants would be like putting on a bow tie and a top hat to take your dog for a walk. In most contexts, the [a] version would be overwhelmingly preferred. We will sometimes call the [a] versions **normal** rather than informal style. They're much more common; and though they're more relaxed, that doesn't mean they're inferior to the very formal [b] counterparts.

The distinction here isn't restricted to speech. Newspapers and magazines generally use a mixture of styles: a little less formal for some topics, a little more formal for others. We've chosen to use normal rather than formal style in this book, writing *something you'd be happy with* rather than *something with which one would be content.*

1.2 Describing and Advising

A book on English grammar can have either of two very different goals: the aim can be either to **describe**, which means trying to characterize the grammatical system, or to **advise**, which means trying to influence the way in which you use it.

There must be a grammatical system of some kind: when speakers of the language compose sentences, they're doing something that non-speakers can't do. The aim of a description is to give an accurate account of the principles of sentence construction that guide them. Advising, on the other hand, is a matter of telling people how they ought to speak or write. That can be a respectable aim – helping them to improve their use of the language, on the assumption that their command of it might not be perfect and might need improvement.

There's nothing wrong with either goal, but this book definitely has description as its goal, not advice. We try to describe the principles defining Standard English, and mostly don't mean to imply anything about what sort of sentences you should use.

How exactly you want to speak or write will depend on many things we cannot know. You might be writing screenplay dialogue, or preparing a script for a solemn public declaration, or writing a friendly letter to someone you know well, or drafting a deliberately humorous editorial, or trying to craft something that will make the reader think it's from a foreigner. Sensible advice would have to be highly specific to a context that we couldn't possibly know about. Lots of how-to-write books seem to have already decided what you're going to be writing and what style you ought to adopt. We haven't made any such assumptions.

The two approaches – description and advice – wouldn't need to be in conflict if everyone agreed on the facts, and those facts were invariant across time and space. Descriptive grammar books would explain what the language is like, and advisory ones would tell you how to adhere to the description and avoid mistakes – and a mistake would be any failure to accord with the descriptive account. But the actual picture is complicated by the tendency of usage guides not to be based on accurate description, but instead to have a strongly reformist intent.

Some books on how to write, in short, don't seem to be interested in encouraging you to write the way expert current users do; instead, they urge you to avoid putting things in ways they disapprove of. They want you to write in ways that are now quite old-fashioned. They dismiss the practices of other writers as 'ignorant' or 'barbarous' and recommend usage that hardly any normal people follow.

There are exceptions: some books on usage are very good, basing their advice on carefully gathered evidence about what Standard English speakers actually say. But others are amazingly bad, peddling ridiculous myths about Standard English rather than useful advice, and often confusing informal style with grammatical error.

Although we hardly ever issue any advice about how you should speak or write, we do provide boxes headed '**Usage Controversy Note**'. These contain warnings or discussion about parts of the language where there is a dispute about what to treat as normal or correct. We warn you of a dispute and explain what usage manuals have said about it, and if they've gone wrong we point out where and how.

1.2.1 Technical Terms for Grammatical Description

Describing complex systems of any kind – car engines, legal statutes, musical scores, computer systems, human languages – calls for theoretical concepts and technical terms, like 'gasket', 'tort', 'crescendo', 'algorithm, or 'adverb'. It's unavoidable. We'll need to introduce and define many technical terms for grammatical concepts in this book.

What we mean by calling a word a technical term is simply that it's associated with a special meaning needed within a certain discipline. You can't guess the meaning of a technical term on the basis of how you've seen the word used before. For example, you may have already encountered the word 'imperative' in sentences

like *Rebuilding is our first imperative* (where it means "essential task") or *Diversification is imperative* (where it means "vitally necessary"); but seeing sentences like these gives you no clue as to the meaning of the word in the context of English grammar, where **imperative** identifies a certain type of clause which has a plain-form verb, often lacks a subject, and usually expresses a directive (see §10.4).

We'll generally give the explanation of a term just before we first use it, or sometimes immediately following that first use. There is no perfect order of introduction: the vocabulary of grammar can't all be explained at once, or in a perfectly logical order, because the meanings of grammatical terms are very tightly connected to each other: sometimes neither member of a pair of terms can be properly understood independently of the other one.

We'll use standard terms for three different areas within the linguistic study of form and meaning:

- **Morphology** deals with the internal form of words: *unopened* is a grammatically permitted combination of *un*, *open*, and *ed*: it complies with the morphological principles of English, whereas **openedun* does not.
- **Syntax** deals with the principles governing how words can be assembled into grammatically well-formed sentences: *I found an unopened bottle of wine* is grammatical but **I found a bottle unopened of wine* is not (*unopened* is in a position where the syntactic principles of English don't allow it).
- **Semantics** deals with the principles by which sentences are associated with their meanings. For example, it is a semantic fact that to say *I found an unopened bottle of wine* is to commit yourself to the claim that at a time in the past you came upon a wine bottle that had its original contents sealed inside.

We'll need technical terms in all three of these areas. And even fairly well-known terms (noun, verb, pronoun, subject, object, tense) will be defined here in ways you may not be familiar with, because the grammars of the past got a lot of things wrong. We won't assume any prior understanding of such terms, and we caution you in advance that we will depart from many mistaken definitions of the past. We'll devote just as much attention to familiar terms found in earlier grammars as we do to the occasional novel terms of our own.

Even a term as familiar as 'past tense' illustrates this point. Tense is a dimension on which verbs can differ in their form: *start* and *started* differ in tense, and *started* is the past tense form. For *say*, the past tense is *said*, and for *offend* the past tense is *offended*. The usual definition of 'past tense' found in grammar books and dictionaries says simply that past tense forms express or indicate a time in the past. But past tense verb forms DON'T always make reference to past time, and references to past time don't always use past tense forms. The following examples show this (the verbs we're interested in are underlined):

[3]

		STANDARD DEFINITION WORKS		STANDARD DEFINITION FAILS
i	a.	*The course started last week.*	b.	*I thought the course started next week.*
ii	a.	*If he said that, he was wrong.*	b.	*If he said that, she would leave him.*
iii	a.	*I offended my many fans.*	b.	*I deeply regret offending my many fans.*

The usual definition works for the [a] examples, but not for the [b] ones.

- In [i], the past tense *started* in the [a] case locates the start of the course in past time, but in [b] the same past tense form indicates an assumed starting time in the future. So not every past tense involves a past time reference.
- The pair in [ii] contrasts past time in [a] (which is conditional on whether he actually said something in the past) with future time in [b] (about what would follow from his saying it in a conjectured future utterance; §3.4.2 covers this use of the past tense with *if*). Again, there's a past tense, but no reference to past time.
- Finally, in [iii], the event of my offending my fans is located in past time in both cases, but whereas in [a] *offended* is a past tense form, in [b] *offending* is not. So not every past time reference involves a past tense.

The usual definition yields wrong answers for all the [b] examples: it would imply that *started* and *said* aren't past tense forms, but that *offending* is. Yet these aren't weird cases we concocted: they're perfectly ordinary. Sentences like them occur all the time. What's wrong is the traditional definition.

Why has the definition of 'past tense' as a form of the verb expressing past time been repeated in so many books, when it's clearly wrong? Perhaps because traditional grammars have tried to give a definition that would apply at least approximately in almost any language. However, if we need to reliably identify the English verbs that are in the past tense form, that won't do; the correlation between form and meaning isn't rigid or one-to-one. (This is all covered in greater detail in Chapter 3.)

What you can say is that reference to past time is the PRIMARY OR MOST CHARACTERISTIC USE of past tense forms in English. That might come closer to applying to other languages as well. But we can't determine whether some arbitrary word in English is a past tense verb form by simply asking whether it describes something happening in past time.

The definitions found in textbooks and dictionaries are often of very limited value in helping to identify items or kinds of expression within English. It's not that the ordinary meanings of words like 'past' are totally irrelevant: there are reasons why that term was adopted. But an account of verbs in English must specify the grammatical properties – the morphology and syntax, not just the semantics – that enable us to determine whether or not a specific expression should count as a past tense. Something similar holds for all the other grammatical terms we'll be using.

1.3 The Structure of Sentences

Non-linguists generally talk about language as if it's all about words. People say *Don't believe a word of it*, as if you could believe words; they say *Give me your word*, as if a single word could express a promise; they say *She has a way with words*, as if picking the right words was all it took to make a persuasive argument; they say *Thank you for your kind words* after being praised, as if it were the words that counted. But it's hardly ever mere words that people are talking about. It's **sentences**, not words, that can express beliefs, make promises, frame convincing arguments, or constitute praise. And it is a vital fact about sentences that they are not just collections of words, nor sequences of words. There is more to sentences than that. Sentences have **structure** that is governed by the principles of syntax.

Sentences are the largest units that have syntactic structure. Paragraphs, chapters, and entire novels have structure too. But their structure is not of the sort that syntax defines. The structure of a novel is responsive to issues of plot and presentational strategy, not mere syntax. What to introduce first and what to leave till later in a story is entirely up to the author or storyteller; they're not under grammatical constraint. But English syntax dictates that you can't express *Pass me that hammer* with the word *me* shifted to the end: **Pass that hammer me* is just not grammatical (though *Pass that hammer to me* is).

Most of what we discuss in this book will in fact concern smaller units than sentences: for the most part we'll be examining the structure of **clauses**, which can be roughly characterized as the smallest units that can really say something – describe a situation (*Nobody cares*), express a claim (*Kim did it*), pose a question (*Are you crazy?*), or convey an instruction (*Go home*). Sentences generally consist either of a single clause or of a sequence of clauses linked by words like *and*. We'll also be paying close attention to the structure of **phrases**, which are sequences of crucially connected words in which one word is central (*the key to this door*).

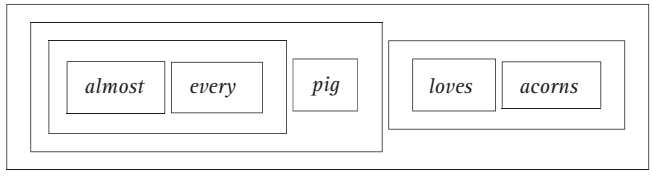
Three crucial assumptions about the structure of sentences, clauses, and phrases will help in understanding how we can describe them and how we can give diagrams of their structure. The first point is that sentences are composed of parts which we will call **constituents**, some of them being words and others being larger than words (a constituent can contain other constituents). The second is that the parts are classified into types, which we will call **categories**. And the third – very often confused with the second in traditional grammars – is that constituents of particular categories can have **functions** or grammatical roles within the larger constituents that contain them. All three of these crucial ideas need further discussion.

1.3.1 Constituents

Evidence interwoven throughout the rest of this book suggests that a sentence like *Almost every pig loves acorns* has two major syntactic parts: *almost every pig* is the first, and *loves acorns* is the second. *Almost every pig* also has two main parts: *almost every* is the first and *pig* is the second. So in addition to the five words in the sentence, which are its smallest units, it also has *almost every* and *almost every pig* and *loves acorns* among its **constituents**.

This fact – that there are parts containing parts, and so on – could be represented by putting the parts into boxes, and their subparts into boxes within boxes, like this:

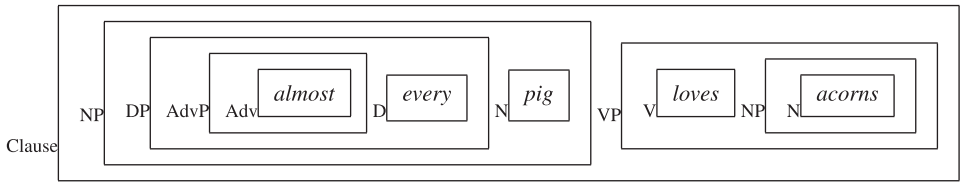
[4]



1.3.2 Categories

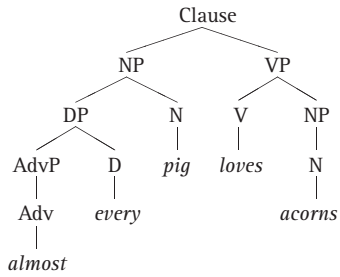
The second point is that these constituents are of different types: they are classified into **categories**. In the terms that we will outline in Chapter 2 and explain much more fully in the subsequent chapters, we'll say that *wall* and *city* are **nouns** (N), or more specifically, count nouns; *a* and *the* are **determinatives** (D), more specifically of the type known as **articles**; *a wall* and *the city* are **noun phrases** (NPs); *protected* is a **verb** (V), in a past tense form known as the **preterite**; *protected the city* is a **verb phrase** (VP); and the whole thing is a **clause**. We could try to represent some of these aspects of structure by giving the boxes labels, like this:

[5]



But this begins to look a bit cluttered and hard to read. It is clearer and cleaner to use what computer scientists call **trees**. These have become the standard way of representing sentence structures in modern works on syntax. We put the label of the whole thing at the top, draw downward-diverging lines to the labels of the next-largest boxes, and so on down to their contents. A tree for our example, corresponding exactly to [5], is shown in [6].

[6]



(We'll explain later why we show *almost* not just as an adverb – labelled Adv – but as an adverb phrase – AdvP – containing only one word, and why *acorns* is not just shown as a noun but as a one-word NP.)

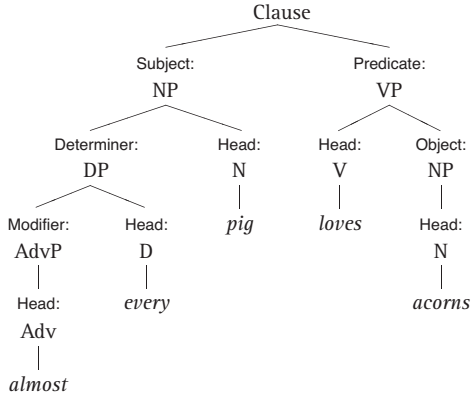
1.3.3 Functions

The third component of the description we will give for every sentence involves saying for each constituent not only what category it belongs to but also what function it has in the constituent that contains it. In our example, *almost every pig* is the **subject** of the clause; the noun *pig* is the **head** of that NP; the VP is the **predicate** of the clause; the verb *loves* is the **head** of the predicate (that is, of the VP *loves acorns*); the NP *acorns* is a **direct object** in that VP, required by the head verb *loves*; and so on. To include details of the grammatical functions, we will put the function name followed by a colon above each label, so the label of an NP functioning as the subject of a clause will look like this:

Subj:
NP

For our example, a suitable tree with function information included is shown in [7].

[7]



It should be obvious that the information summarized by the grammar is highly relevant to the way in which sentences convey meaning. While *pig* just refers to a type of animal, *every pig* makes reference to the entire species. *Almost* tampers with that, modifying the meaning of *every* so as to pick out nearly all pigs rather than all of them. *Loves acorns* names a property that some animals have and others don't. The sentence diagrammed in [7] attributes the *loves acorns* property to the members of a collection of animals that includes nearly all pigs, and the way the whole sentence is structured is crucial to that: it is THE WAY THE CONSTITUENTS ARE PUT TOGETHER, not just the words that are chosen, that allows this clause to have its specific meaning.

The task of a complete grammar is to make available, for each sentence of the language, an account of what the constituents are, and what category each constituent belongs to, and what the function is of each one in that sentence, right down to the individual words. Specifying the right categories for the words is the task of the dictionary for the language, so the design of a dictionary is intimately bound up with the task of the grammar. There is then more to be done to give a full description of the language: the phonological generalizations relevant to the pronunciation of each sentence, and all the semantic details defining the meanings of their parts, must also be specified. It is a huge project, even for very well-known languages.

1.4 Investigation and Disconfirmation

We've said that *Almost every pig loves acorns* has two major parts, an NP and a VP. It's an important point that we expect EVIDENCE to be relevant to such claims. We are not simply laying down (or repeating from earlier authorities) rules that you are supposed to follow. We are trying to identify the principles that correctly represent the way people use English – the way they construct sentences when they speak or write and the way they understand sentences in speech or writing and grasp their meaning.

When we say there is evidence that a sentence has two major constituents – an NP and a VP (a subject and a predicate) – we mean that other facts about the language tie in with this claim and support it as a general characterization. Suppose we consider a very simple clause like *This wall protected the city*. We can support the claim that it has two major constituents (the NP *this wall* and the VP *protected the city*) by an experiment where we add an adverb like *obviously* at various points in the sentence and see whether the result is grammatical. What we find is that the adverb is allowed only at certain points:

[8] i *Obviously this wall protected the city.*
ii **This obviously wall protected the city.*
iii *This wall obviously protected the city.*
iv **This wall protected obviously the city.*
v **This wall protected the obviously city.*
vi *This wall protected the city, obviously.*

And under the assumption we have made, we can account for this very neatly: adverbs of this sort (it is functioning as a modal adjunct: see §8.9) can't be positioned inside major constituents of the clause, but only before or between or after them. So *obviously* can fit before the first constituent as in [i]; between the two biggest constituents as in [iii]; or at the end as in [vi].

Since we're assuming that we're investigating a real system, one that actually exists and is in use by everyone who knows English, there is always the possibility that WE COULD BE WRONG ABOUT IT. We could have missed some evidence showing that some principle is not the one English users follow, or that there is a better way to formulate the principle we're talking about.

This point about evidence and investigation is worth illustrating with a small case study. Let's look at a fairly minor point of detail, touched on only very briefly later in the book (§5.8.2): how English speakers use pronouns to refer to ships. In this sentence from the *Encyclopedia Britannica* article about RMS Titanic we underline the pronoun that interests us:

[9] *On April 10, 1912, the Titanic set sail on <u>its</u> maiden voyage.*

The sentence uses *its maiden voyage* to mean "the Titanic's maiden voyage". So we have evidence that when using a pronoun rather than a full name to refer to a single ship, people use the appropriate form of the pronoun ***it*** (*it*, *its*, or *itself*). And when referring to more than one, writers employ ***they***; we know that because the same article happens to mention two other earlier contemporaries of the Titanic, the Lusitania and the Mauretania, and says:

[10] *The two passenger liners were garnering much attention for <u>their</u> expected speed.*

The fact that you can't use ***they*** for a single ship is very clear from the fact that sentences like [11] never seem to occur:

[11] **On April 10, 1912, the Titanic set sail on <u>their</u> maiden voyage.*

Any English speaker you ask will tell you that [11] sounds completely wrong. It's also clear that ***it*** is never used for more than one ship; [12] also sounds like a mistake to speakers of English:

[12] **The two passenger liners were garnering much attention for <u>its</u> expected speed.*

So our initial generalization can be put like this:

[13] ***It*** is the pronoun used for referring to a single ship, and ***they*** is the one used for referring to more than one ship.

But now some further evidence comes in. This sentence appears in the Wikipedia entry about the ship (as retrieved in early 2021):

[14] *RMS Titanic was the largest ship afloat at the time she entered service.*

Here the pronoun *she* has been used to refer to the ship, so the generalization in [13] is incomplete. What do we do now? Should we try to decide which of the two is 'correct'? Does either [9] or [14] have to be condemned as wrong? That would not be our assumption. Both articles seem to be written in perfectly respectable Standard English, and they agree with vast quantities of English prose found elsewhere. The rational assumption is that there are two different options within Standard English, and we need to revise our generalization.

But what's the new version? Do you get a free choice of pronouns between ***he***, ***she***, and ***it***? Could you just as well use ***he***? Well, it doesn't look like it. Nobody writes anything like [15]:

[15] **RMS Titanic was the largest ship afloat at the time he entered service.*

We mark this as ungrammatical on the assumption that the underlined *he* is supposed to refer to the underlined ship name. (Naturally it would be fine if *he* referred to the entry into service of some previously mentioned sailor, and not to the ship.) The right generalization seems to be [16]:

[16] For a single ship, you have a free choice between ***it*** (*it* or *its* or *itself* as appropriate) and ***she*** (or *her*, *hers*, or *herself*). For more than one ship, you have to use ***they*** (or *them* or *their* or *theirs*).

Does that settle it? You might assume so. You could read and write English for years without running into evidence that it was inaccurate. But remember, all such generalizations are just hypotheses. Any one of them could prove to be mistaken. We're not just trying to lay down simple rules to follow – that would be too easy. Our aim is to work out what the ACTUAL rules are – the tacitly followed principles that competent speakers of English really are using, whether they could say so or not.

In this case there is evidence that the tentative generalization in [16] is inadequate. The evidence comes from sentences with *itself* or *herself*. We can use these in a sentence about a ship righting itself (ceasing to list, or lean to one side). We predict that these examples should be correct:

[17] i *The ship was listing for a while, but eventually it righted itself.*
ii *The ship was listing for a while, but eventually she righted herself.*

So far so good: both of those seem fine. But we've picked the two examples that will occur to most people precisely because they sound natural. Our task, which is much less intuitive, is to test the hypothesis by seeing whether we have accounted for all the possibilities, whether they sound natural or not. In this case, we have two other cases to consider:

[18] i **The ship was listing for a while, but eventually it righted herself.*
ii **The ship was listing for a while, but eventually she righted itself.*

These definitely don't seem right: nobody mixes ***it*** and ***she*** like this within a clause. And we haven't ruled these out. So do we perhaps have to say that there are *it*-style users and *she*-style users, but any given piece of writing has to stick with one or the other? No, it turns out that isn't quite right either. Later on in the Wikipedia article we find this:

[19] *Although Titanic had advanced safety features, such as watertight compartments and remotely activated watertight doors, it only carried enough lifeboats for 1,178 people.*

This shows us that the Wikipedia article sometimes uses *she* and sometimes uses *it*. Yet **It righted herself* and **She righted itself* still seem wrong and are never encountered. How can we describe what's going on?

In [20] we give a third generalization that comes a lot closer to being fully correct and complete:

[20] For a single ship, you have a free choice between ***it*** and ***she***, but within any single sentence you have to stick with a consistent choice. For more than one ship, there is no choice and you have to use ***they***.

That can stand for now as our best hypothesis about the principle English users follow. But the point we are making is not that we have now arrived at the truth, or that you have to obey [20]. What we are illustrating here is that in order to do grammar you have to INVESTIGATE. Judgements about the rules of English grammar have to be based on evidence, and any given judgement could be mistaken.

At every point in this book we are offering our best attempt at a clear, simple, and accurate statement of the rules English speakers actually follow. But we could always be wrong. You could spot a piece of evidence that shows something we said is false. We would expect that to happen sometimes. And if your evidence is genuinely relevant – if it's not simply based on a stray typographical error or an irrelevant similarity that relates to a different type of sentence, but really represents what we find when Standard English users speak and write – then maybe a better description than ours could be constructed.

English grammar as we conceive of it is not a body of doctrine handed down from some infallible authority. The enterprise of figuring out what the grammar of English should say is a matter of live empirical investigation. Some areas are drastically under-explored and little understood. One example is the topic of the parenthetical interpolations and interruptions that we call **supplements** (see §8.11); another is coordination (see Chapter 15). You can participate in the investigation. We hope we have mostly done our work well, but we aren't necessarily always right. Keep a critical eye on our generalizations. Some of them can probably be revised and improved.

Exercises on Chapter 1

Note: When you complete an exercise, it's very important to check your answers. Those with errors should be marked and reattempted at a later time. Mastering the exercises will typically lead to better learning than simply re-reading the chapter. Further multiple-choice exercises are accessible through the Cambridge University Press website at www.cambridge.org/SIEG2.

We also include a number of supplementary research and challenge exercises, which will extend and deepen your learning.

1. Consider the words *the*, *dog*, *ran*, and *away*. Which are the three grammatical orders of those words? Discuss any possible grounds for doubt or disagreement that you see.

2. Consider what features of the following sentences mark them as belonging to formal style in Standard English. Rewrite them in a more informal style, keeping the meaning as close as possible to the original. Add notes on what you changed and why.

- **i** *To whom am I speaking?*
- **ii** *It would be a pity if he were to give up now.*
- **iii** *We hid the documents, lest they be confiscated.*
- **iv** *That which but twenty years ago was a mystery now seems routine.*
- **v** *One should always try to do one's best.*

3. For each of the following statements, say whether it is a **morphological**, **syntactic**, or **semantic** fact about English (see §1.2.1).

- **i** A completed grammatical sentence of Standard English that begins '*I believe that we …* ' must continue in a way that includes at least one verb.
- **ii** A witness who used to be a smoker would be wrong to answer *No* to the question *Did you ever smoke?*
- **iii** *Of* can occur as the last word in a Standard English sentence.
- **iv** The string of words **He it saw* can be made grammatical by positioning the word *it* after the word *saw*.

v The verb *enable* is formed from *able* by adding the *en·* prefix.
vi We find *fell* rather than **falled*, because *fall* doesn't take the *·ed* suffix.
vii When someone says *I was going to walk but I decided not to*, the sense is exactly the same as if they had said *I was going to walk but I decided not to walk*.
viii Wherever we can truthfully say *Montmorency bought a red Porsche*, we can also say *Montmorency bought a Porsche*, but not conversely.
ix You can't insert *every* at random points in the sentence *A man's got to do what a man's got to do* and get a grammatical result every time.
x You can't sincerely say *I believe Sally slapped the boss* if you think that *The boss was slapped by Sally* is a lie.

4. You may have heard people talk about a noun functioning as an adjective or a verb functioning as a noun. Under the framework we've described here, what's wrong with such a description? (See §§1.3.2 and 1.3.3.)

5. Consider the sentence *Jupiter and Saturn are planets*. Here, the subject is *Jupiter and Saturn*. Based on this, we form the following hypothesis: When the subject has two constituents joined by *and*, if the choice is between *are* and *is*, *are* is always correct. To check that this is true, should you search for examples with the form 'NP *and* NP *are*', or for examples with the form 'NP *and* NP *is*'? Explain why.

6. List three or four grammar 'rules' you recall hearing about (e.g., Don't start a sentence with *and*). For each rule, say whether you believe that violating that rule would (a) result in an ungrammatical sentence, (b) produce a grammatical sentence in a non-standard dialect but not in Standard English, (c) produce a markedly informal Standard English sentence, or (d) produce Standard English that some people don't like for no obvious reason.

7. [Supplementary exercise] Choose one or two language references (of any language; grammars, style books, usage guides, or dictionaries) from those accessible to you. Say whether you think they are focused on describing the language or advising the language user, and explain why.

8. [Supplementary exercise] We have explained that Standard English is 'just a name', but it still offends some people (because of the association between 'not standard' and 'substandard'). You will see throughout this book that we believe finding the right term for a concept is important. Can you think of an alternative name for Standard English that would be better? Explain the pros and cons of one or two candidate terms.

2 Overview of the Book

This short chapter aims to provide a few of the tools needed for grammatical description, to introduce some key concepts and technical terms employed in subsequent chapters, and to give a sense of how the rest of the book is organized.

We start by drawing an absolutely crucial distinction between words as counted by a word processor and words as listed in a dictionary (§2.1). In §2.2 we explain how phrases and clauses are built around words. And then in §2.3 to §2.16 we give brief indications of the content of Chapters 3 to 16, so to some extent this chapter can be used as a rough guide to the structure of the book.

Our plan for organizing our coverage of the grammar of English involves defining a particularly simple kind of unit called the **canonical clause**. We concentrate entirely on the structure of canonical clauses right up to the end of Chapter 8. Then from Chapter 9 to Chapter 16 we cover the details of how other kinds of clause depart from the pattern of the canonical ones. The initial sketch of canonical clauses is in this chapter (§2.4), and the fuller account is in Chapter 4. At the end of this chapter there is an appendix summarizing our notational conventions.

2.1 Word Forms and Lexemes

There are two completely different entities that people call 'words'. To see that the difference is real, ask yourself how many words there are in this sentence:

[1] *We love our cat, and she loves her cats.*

Most people will say it has nine words. Any word processor will agree. And there's nothing wrong with that answer, which takes words to be unbroken sequences of letters.

But it's not compatible with the idea that a dictionary LISTS THE WORDS IN A LANGUAGE and explains their meanings. Some of the words in [1] are missing from typical dictionaries. You'll find a dictionary entry for *cat*, but none for *cats*; and *love* will get an entry but *loves* won't. Why? The answer is that the dictionary is implicitly treating *cats* as a special form of *cat* that is required when reference is to any number of cats other than one, and it treats *loves* as the form of *love* that you have to use after *he* or *she* (but not after *I* or *we*). It is in effect assuming that your

knowledge of the grammar of English will suffice to tell you such things, which means the dictionary can ignore them.

You'll probably also find if you look up *our* that although there's an entry for it, that entry says little more than that it is the 'possessive form of the pronoun *we*', or something similar. (We'd use the term **genitive** rather than 'possessive', but for now you can treat the two terms as equivalent.) And the entry for *her* will say it's a form of the pronoun *she*.

So if we want to stick with the obviously true statement that [1] is nine words long, we need a term other than 'word' for the units that get full entries in dictionaries. Most linguists call them **lexemes**.

When we want to emphasize that we intend the other sense of 'word', we may refer to letter strings as **word forms**. Sometimes, though, the distinction won't matter: when talking about *and* we can just say it's a word, because the lexeme it belongs to has only one word form.

This means we can avoid the apparent contradiction: [1] is correctly described as containing nine words (i.e., word forms): it contains *and*, *cat*, *cats*, *her*, *love*, *our*, *she*, and *we*. But what a dictionary needs to list is a different matter. [1] contains word forms of just five lexemes. We can give them names by putting their most basic word form in bold italics: the five lexemes are ***and***, ***cat***, ***love***, ***she***, and ***we***.

A related point is that some lexemes happen to be written with an internal space. For example, most grammarians would treat ***no one*** as a single lexeme, just like its synonym ***nobody***, since absolutely no word can separate the two parts without changing the meaning. A small number of other lexemes (not many) are written with a space: note the underlined items in sentences like *It sort of misled me*; *Take this in case you get hungry*; *It is no doubt apocryphal*; *Santa Cruz is north of Monterey*; *He stepped back so as not to be seen*.

The distinction between word forms and lexemes is important in many practical ways. A literary scholar assessing Shakespeare's word use, or a developmental psychologist studying the size of a child's vocabulary, should surely count lexemes. Learning ***give*** and ***take*** should count as adding two words to a child's vocabulary – not ten (*give*, *take*, *gives*, *takes*, *gave*, *took*, *given*, *taken*, *giving*, *taking*).

2.1.1 Lexeme Categories

Lexemes belong to clusters that have strikingly different grammatical properties. These clusters of grammatically similar lexemes are known as **lexical categories**. (Traditional grammars call them 'parts of speech', an odd term that we will not use.) We'll employ nine lexical categories (some having important subcategories), listed in [2].