

## 1.1 Introduction

Every time we write or speak, we are faced with a large array of choices: not only choices of what to say but of how to say it. The vocabulary and grammar that we use to communicate are influenced by a number of factors, such as the reason for the communication, the setting, the people we are addressing, and whether we are speaking or writing. Taken together, these choices give rise to systematic patterns of choice in the use of English grammar.

Traditionally, such patterns have not been included as part of grammar. Most grammars have focused on structure, describing the form and (sometimes) meaning of grammatical constructions out of context. They have not described how forms and meanings are actually used in spoken and written discourse. But for someone learning about the English language for the purposes of communication, it is the real use of the language that is important. It is not enough to study just the grammatical forms, structures, and classes. These tell us what choices are available in the grammar, but we also need to understand how these choices are used to create discourse in different situations.

The year 1999 saw the publication of a large-scale grammar of English with the aim of meeting the above needs: the *Longman Grammar of Spoken and Written English (LGSWE)*. This was the outcome of an international research project which lasted seven years. Now, we have written a revised, simplified, and shortened version of *LGSWE* for use by advanced students and their teachers.

*LGSWE* made important innovations in the method of grammatical study. It was based on a large, balanced corpus of spoken and written texts. These texts were electronically stored and analyzed with the aid of computers. The analysis produced information about the frequency of grammatical features in different kinds of language. (We use the term 'feature' broadly in this book, to refer to any grammatical form, structure, class, or rule.) The results of the analysis were then studied by the team of grammarians. The goal was to explain not just what is *possible* in English grammar, but what is more or less *probable* in different situations.

This book, the *Student Grammar of Spoken and Written English (SGSWE)*, presents the insights and discoveries of *LGSWE* to advanced students of English and their teachers. *SGSWE* is designed to be used with the accompanying workbook as a textbook (for use in class or for self-study), or alternatively as a reference grammar:

- **SGSWE is a pedagogical textbook on grammar.** For students systematically studying English grammar, *SGSWE* begins with the 'basics' in the opening two chapters, and moves progressively into more advanced territory. It ends with a chapter devoted to the special characteristics of conversational grammar—an important topic which has generally been neglected in grammars up to now.
- **SGSWE can also be used as a reference grammar.** It covers all major features, structures, and classes of English grammar, together with their meaning and use. With the help of the glossary of grammatical terms and the index, students and teachers can use this book as a reference guide. For more detailed information about a feature or its use, you can consult corresponding sections in the larger *LGSWE*.

The greatest innovation of *SGSWE* is that it is a corpus-based grammar. The new methodology of large-scale corpus study developed in *LGSWE* produces results which are very useful for students. In general, advanced students of English want to understand not only the structural rules of English, but how the language is used for communication. For this purpose, new insights, explanations, and information from corpus-based studies of English are an important advantage.

## 1.2 A corpus-based grammar

A corpus is a large, systematic collection of texts stored on computer. The corpus used for *LGSWE* and *SGSWE*—the Longman Spoken and Written English Corpus (the *LSWE Corpus*)—contains approximately 40 million words of text, providing a sound basis for the analysis of grammatical patterns. Because the *SGSWE* is based on analysis of this corpus, it offers a number of advantages over traditional grammars:

- **Real examples:** The book contains over three thousand examples of English in use. These are authentic examples from the corpus, showing how real people use real language. Invented examples that sound artificial—a familiar feature of many other grammars—are entirely absent. At the same time, we have taken care to avoid corpus examples that are overly difficult because they require understanding a complicated context.
- **Coverage of language variation:** The core corpus we have used for this book represents four major registers of the language: conversation, fiction writing, news writing, and academic prose. In the past, grammars have usually presented a single view of the language, as if the grammar of English were one fixed and unchanging system. This clearly is not true. Although there is an underlying system of grammar, speakers and writers exploit that system very differently to meet their communication needs in different circumstances. The corpus-based information included in *SGSWE* describes differences between a spoken register (conversation) and three written registers (fiction, newspapers, and academic prose). In addition, our corpus contains data from both American and British conversations and newspapers. Where they occur, important differences between these geographical dialects are pointed out. (See 1.3 for more about registers and dialects.)
- **Coverage of preference and frequency:** This grammar gives information about the preferences speakers or writers have for one grammatical choice over another. Specifically, we discuss the frequency of alternative structures and the conditions that are associated with them. This is clearly a major matter of interest to advanced learners of the language and future teachers of the language. For example, it is not enough simply to describe the structural differences between active and passive constructions, because students and teachers need to know how users choose between these two options. For example, passive verbs are ten times more frequent in academic prose than they are in conversation. And there are good reasons for this, which we discuss in 6.6.1.
- **Interpretations of frequency: context and discourse:** Information about frequency needs to be explained by human interpreters. That is, a corpus

grammar needs to present the evidence of the corpus as a means of exploring *why* users of English make one choice in one situation, and a different choice in another. Usually several factors are relevant, including register, expressing personal attitudes, giving informational emphasis, or other more specific contextual conditions. The *SGSWE* gives attention to the conditions under which grammatical choices are made, so the grammar has a discourse orientation.

- **Lexico-grammatical patterns:** Another distinctive feature of the *SGSWE* is that it brings together the study of grammar and vocabulary. Traditionally, both in theory and pedagogical practice, grammar has been separate from vocabulary, as if they were two totally independent aspects of language and language learning. This separation is artificial, as becomes evident to anyone who uses a large corpus for studying grammar. What becomes clear is that, when they use a language, people bring together their knowledge of word behavior (lexis) with their knowledge of grammatical patterns. These two aspects of language interact in lexico-grammatical patterns. For example, there is one set of verbs that commonly occur with a *that*-clause (e.g. *think, say, know*) and a different set of verbs that commonly occur with a *to*-clause (e.g. *want, like, seem*). In addition, each register prefers different verbs with these clause types. These patterns help to explain the typical meanings and uses of each clause type in each register.

### 1.3 More about language variation

We have mentioned the coverage of language variation as one of the strengths of this grammar. It is important to recognize that there are two major types of language varieties: registers and dialects.

#### 1.3.1 Registers

Registers are varieties of language that are associated with different circumstances and purposes. For example, Table 1.1 compares the circumstances and purposes of the four registers compared in *SGSWE*.

Comparisons between registers can be made on many different levels. For example, the most general distinction can be made on the basis of mode: conversation is spoken, while the other three registers are written. If we consider a more specific characteristic—the main communicative purpose—the registers fall into three categories. Conversation focuses on personal communication,

Table 1.1 Circumstances of the four main registers in *SGSWE*

	CONV	FICT	NEWS	ACAD
mode	spoken	written	written	written
interactiveness and real-time production	yes	restricted to fictional dialog	no	no
shared situation	yes	no	no	no
main communicative purpose/content	personal communication	pleasure reading	information/evaluation	argumentation/explanation
audience	individual	wide-public	wide-public	specialist

fiction on pleasure reading, and newspapers and academic prose share a more informational purpose. (We sometimes call these last two expository registers.) However, even between newspapers and academic prose there are differences. Academic prose is more concerned with building an argument than newspaper writing is. What is more, academic prose has a more specialized audience than a newspaper, which is written with a wide audience in mind.

Register variation is the main aspect of language variation we consider in this book. We choose this focus because registers differ greatly in their grammar usage, reflecting their different communication circumstances. In other words, the circumstances of a register have a direct impact on which grammar habits are common in the register. To take an easy example, consider how pronouns and nouns are used in conversation and news. Conversation is interactive as a form of personal communication. It is not surprising, then, that conversation shows a frequent use of the first-person pronouns *I* and *we*, and the second-person pronoun *you*. In contrast, newspaper writing is not directly interactive: it is not addressed to an individual reader, and it often does not have a stated author. But it has a function of conveying general information of current interest. Not surprisingly, in newspaper texts, first- and second-person pronouns are relatively rare. Instead, proper nouns, referring to people, places, and institutions, are particularly common.

The four registers that we concentrate on are, of course, far from a complete picture of register variation in English. However, they have the advantage of (a) being major registers, likely to be frequently encountered by any advanced student of the language, and (b) being sufficiently different from one another to show important differences as well as important similarities.

Within each register there are also sub-registers. For example, fiction can be broken down into different sub-categories—detective fiction, fantasy fiction, romance fiction, etc. The compilers of the *LSWE* Corpus took pains to represent all the categories and to obtain as good a balance as possible between different sub-varieties within each category. (More details are given in 1.4 below.) In order to limit the size of this book, though, we make only a few references to sub-registers, such as commenting on a feature common in, say, sports reporting or detective fiction.

#### 1.3.2 Dialects

Dialect variation interacts with register variation. Dialects are varieties according to the identity of speaker(s) or writer(s)—their geographic area, gender, socio-economic class, and so on. Dialect is less important for grammatical purposes than register. From the grammatical point of view, dialect differences are arbitrary, while register differences are functional, reflecting the way that grammar varies according to communicative purpose.

Like register, dialect distinctions can be made at different levels of specificity. 'American English' or 'British English' is a very general level; 'the speech of female teenagers in the South Bronx area of New York City' would be a more specific dialect. In the *SGSWE* we focus only on the high-level distinction between American English and British English (although the spoken corpus was developed following sampling methods for geographical regions, socio-economic class, gender, and age, so more detailed dialect studies are possible).

The corpus used for this book also includes some representation of Australian, Canadian, Caribbean, Irish, and West African English in the fiction subcorpus (see details in 1.4). For full coverage of dialect variation, we would need to have coverage of many more world varieties of English. The contrasts between British and American English in this book serve as just one example of regional dialect variation. We have chosen these two dialects because they serve as a target for many learners and teachers of English.

### 1.3.3 Standard and non-standard (vernacular) English

There is no official academy that regulates usage for the English language, but there is still a prevailing world-wide view that there is a 'standard English': the language variety that has been codified in dictionaries, grammars, and usage handbooks. However, in the corpus, especially in conversation, usage regarded as non-standard (also called the **vernacular**) is also found. For example, consider the vernacular use of *what* as a relative pronoun in 1, and of *ain't* and multiple negative constructions in 2 and 3:

- 1 *They were by the pub what we stayed in.* (CONV)
- 2 *I ain't done nothing.* (CONV)
- 3 *'There ain't nothing we can do.'* (FICT)

As 3 shows, vernacular forms also occur in fiction texts, particularly in representing the speech of fictional characters. However, in general, vernacular forms are rare in the written corpus.

When vernacular forms are discussed or illustrated in the rest of the book, they are generally noted as 'non-standard'. In general, our description of grammar is limited to standard English. However, it is worth bearing in mind that conversation contains many vernacular features of language, and so such features can be expected in the conversation part of our corpus, whereas they are strictly avoided in written language (except where written language deliberately mimics speech, as in fictional dialog). In a similar way, conversational transcriptions often contain informal or non-standard spellings like *gonna* and *cos* (see 1.6.2 below).

#### A Variation in standard English

The term 'standard English' can be misleading. 'Standard' in some contexts means 'uniform, unchanging', and so it is assumed that only one form of a grammatical feature is accepted in standard English. But this is clearly false. For example, both of the following conform to standard English, even though one begins a relative clause with *that* and the other with *which*:

- 1 *I could give you figures that would shock you.* (FICT)
- 2 *This chapter is devoted to a discussion of various flow processes which occur in open systems.* (ACAD)

The relative pronouns *that* and *which* could exchange places in these sentences, although there are a number of factors that favor one over the other (see 9.7, 9.8.1).

Little of the variability within standard English is due to dialect differences between American English and British English, which actually show very few

grammatical differences in their standard dialects. We point out such differences where important in SGSWE. For example, one well-known difference is that American English has two past participles for the verb *get* (*got* and *gotten*), whereas British English has only one:

*Angie, I think we've got a leak.* (BrE CONV)

*They've got money.* (AmE CONV)

*He must have got to the door just as the bomb landed.* (BrE CONV†)

*And we still haven't gotten a damn pumpkin.* (AmE CONV)

Note: Another aspect of vernacular English is illustrated by the last example: the word *damn* (or *damned*), which is a mild 'swearword' or taboo term. Swearwords can cause offence or be considered impolite, especially where used in the wrong context, and so we point them out when they occur in examples.

#### B Prescriptive v. descriptive grammars

Most cases of variation within standard English (e.g. *that* and *which* in 1 and 2 above) do not attract attention from ordinary language users. However, speakers do tend to be aware of some aspects of disputed usage and sometimes have strong opinions about what forms are 'correct'. Thus while the use or the omission of *that* is rarely noticed or commented on, the choice between *who* and *whom* can rouse strong feelings (see 9.8.2). Prescriptive grammars dictate how people 'should' use the language. For example, a prescriptive grammarian would insist that only *whom* should be used when the pronoun refers to a human and functions as an object or prepositional complement. In contrast, speakers in conversation regularly prefer *who* in actual usage:

*There's a girl who I work with who's pregnant.* (CONV)

In fact, many speakers would find the use of *whom* unusual in any informal, conversational situation.

In this grammar, we do not argue that any one alternative is correct in cases like these. Rather than a prescriptive grammar, the SGSWE is a descriptive grammar. We focus on describing the actual patterns of use and the possible reasons for those patterns. However, we do refer from time to time to some cases of disputed usage. Although these may not be so important from the viewpoint of communication, they often play a significant role in people's judgments of what is 'good grammar'. These judgments, in turn, may have an influence on actual patterns of use.

## 1.4 More detail on the LSWE Corpus

For a corpus-based grammar, the design of the corpus is an important concern. Detailed discussion of the LSWE Corpus can be found in Chapter 1 of LGSWE, especially pp. 24–35. Here we give only a brief overview of the corpus.

The entire corpus contains approximately 40 million words. Most of the analyses comparing the four registers used a subcorpus of approximately 20 million words. Additional texts for the dialect comparisons and occasional comparisons with supplementary registers account for a further 20 million words. Table 1.2 provides an overview of the overall composition of the corpus.

Table 1.2 Overall composition of the LSWE Corpus

	number of texts	number of words
<b>core registers</b>		
conversation (BrE)	3,436	3,929,500
fiction (AmE & BrE)	139	4,980,000
news (BrE)	20,395	5,432,800
academic prose (AmE & BrE)	408	5,331,800
<b>AmE texts for dialect comparisons</b>		
conversation (AmE)	329	2,480,800
news (AmE)	11,602	5,246,500
<b>supplementary registers</b>		
non-conversational speech (BrE)	751	5,719,500
general prose (AmE & BrE)	184	6,904,800
<b>total Corpus</b>	<b>37,244</b>	<b>40,025,700</b>

The strength of the LSWE Corpus does not just lie in its size—although size can be important, especially for the study of rare grammatical features. More important qualities are the diversity and balance of the corpus. The LSWE Corpus represents a comparatively wide range of register and dialect variation within the language, and each category of texts is represented by a wide range of writers/speakers and 'sub-registers'. In 1.3.1 and 1.3.2 we mentioned some of the sub-categories within the registers and dialects of the corpus. Tables 1.3–6 provide further details.

Like the written registers, conversation is also a diverse register, but no effort was made to identify sub-registers or list all the topics of conversation. Most of the conversations in the LSWE Corpus are private (often domestic) talk. However, occasionally other kinds of talk are included, like service encounters in a store, or one side of a telephone call. Planned speech, such as lectures, speeches, and sermons, are in a separate register of 'non-conversational speech'.

Table 1.3 Approximate numbers of speakers in the BrE and AmE conversation subcorpora by gender

gender	BrE	AmE
female	270	292
male	225	199

Table 1.4 Distribution of fiction texts across national varieties

national variety	number of texts	number of words
AmE	41	1,095,200
BrE	79	3,347,100
other	19	537,700

Table 1.5 Breakdown of the British and American news subcorpora by topic

topic	BrE	AmE
arts/entertainment	418,400	325,000
business/economics	542,800	1,545,000
domestic/local/city news	1,233,900	995,000
foreign/world news	1,156,100	680,000
sports	1,218,700	260,000
all other topics (including editorials, law, social news, science/medicine/technology, etc.)	862,500	1,485,000

Table 1.6 Breakdown of the academic prose subcorpus

major categories	number of texts	number of words
academic book extracts	75	2,655,000
academic research articles	333	2,676,800

Note: Subjects included are agriculture, anthropology/archaeology, biology/ecology, chemistry/physics, computing, education, engineering/technology, geology/geography, law/history/politics, linguistics/literature, mathematics, medicine, nursing, psychology, sociology.

## 1.5 Overview of the book

We have organized our discussion of grammar into several major sections:

- **Chapters 2–3 Key concepts and categories in English grammar**  
These chapters present a basic introduction to English grammar, providing the foundation for our discussion of particular areas in later chapters. They introduce the basic terms for structures, rules, and classes in English grammar, illustrating them throughout with real corpus examples. It may be useful to refer back to these chapters (as well as the glossary) if you come across a puzzling term later in the grammar. Because they review all of grammar in a simplified way, these chapters have less room for the information about discourse choice found more plentifully in later chapters.
- **Chapters 4–7 A close look at the major phrase types**  
Chapters 4–7 cover the major classes of 'content words' or lexical words (noun, verbs, adjectives, and adverbs) together with the related phrases (noun phrases, verb phrases, adjective phrases, and adverb phrases):  
4 Nouns, pronouns, and the simple noun phrase  
5 Verbs  
6 Variation in the verb phrase: tense, aspect, voice, modality  
7 Adjectives and adverbs
- **Chapter 8 Clause grammar**  
This chapter introduces the structure and function of independent and dependent clauses. It is a pivotal chapter roughly half way through the book. It looks back to the phrase chapters, Chapters 4–7, showing how those phrase types are used in clauses. At the same time, it looks forward to the later clause

## 10 Chapter 1: Introduction: a discourse perspective on grammar

chapters, Chapters 9–12, which describe the use of clauses in more complex constructions.

- **Chapters 9–12 Building on the clause**

Building on Chapter 8, these chapters explore additional clause-based or clause-derived structures. For example, complex noun phrases (Chapter 9) may not seem to involve clauses, yet in fact they do, because noun phrases can contain relative clauses and other clause-level constructions. These constructions cannot be explained properly until clauses have been covered.

- 9 Complex noun phrases

- 10 Verb and adjective complement clauses

- 11 Adverbials

- 12 Word order choices

- **Chapter 13 The grammar of conversation**

Chapter 13 places the spotlight firmly on spoken language, and shows how the grammar of conversation is adapted to the particular demands of spontaneous spoken interaction. It also highlights some special features of conversational grammar: for example, dysfluencies, discourse markers such as *well* and *okay*, ellipsis, interjections, and taboo expressions ('swearwords').

- **Reference section**

The book ends with a glossary of grammatical terms and an index. The reference section of the book is designed for all users, but especially to help teachers and students using it as a reference grammar.

## 1.6 Conventions used in the book

The list of Abbreviations, Symbols and Conventions on p. iv covers many of the conventions used in this book. But some aspects of the transcription of conversation and the use of tables and figures deserve a more detailed explanation here.

### 1.6.1 Transcription of speech

Spoken language must be transcribed before it can be studied. That is, the transcriber must listen to a tape recording and write down exactly what was said. For the LSWE Corpus, the transcribers produced an orthographic transcription. This transcription uses the ordinary symbols of written texts, including the conventional spellings of words (in most cases). Conventional punctuation symbols—particularly hyphens, periods, commas, and question marks—are used to reflect typical intonation associated with those symbols. Thus, a period reflects falling intonation and a question mark reflects rising intonation.

### 1.6.2 Spelling variations, reduced pronunciation, and limitations of the transcription

You may sometimes notice variants which are irrelevant to the study of grammar, such as variant spellings like *OK* and *okay*, or American spellings (such as *center*) in some texts and British spellings (such as *centre*) in others. These differences are not significant, and can be ignored from the grammatical point of

view. More relevant linguistically are spellings of reduced pronunciations, such as *gonna*, *gotta*, and *wanna* instead of *going to*, *got to*, and *want to*, and *cos* or *cause* instead of *because*. These are semi-conventional spellings, capturing the reduced pronunciations which are very common in casual conversation. They help give an impression of what the speech was like. You will also notice 'words' like *um* and *er* signaling filled pauses (a type of dysfluency—see 13.2.5) and exclamatory words like *hm* and *ooh* which it would be rare to find in ordinary written texts. However, a strict, phonetically accurate transcription was not the goal of the corpus, nor would a phonetically detailed transcription be feasible with a corpus of this size.

Since the corpus was not transcribed phonetically, some features of speech, such as stress and intonation, are not available. In the vast majority of cases, however, the transcription provides plenty of detail for grammatical analyses.

### 1.6.3 Visible frequency: the use of tables and figures

We have said that frequency is important for understanding how the grammar of English is actually used in different registers, dialects, and situations. At the same time, we recognize that tables of statistics are often not useful to the average grammar student. To overcome this problem, we have used two main ways of representing frequency, both of them avoiding the use of lists of numbers:

- The first method of indicating frequency is to use ordinary words such as *often*, *rarely*, *common*, *uncommon*. While not precise, these terms are useful in giving a general idea of frequency differences. These generalizations are based on corpus analysis, which is often reported in *LGSWE* with tables and figures.
- The second method is to use figures (bar graphs). These figures enable you to compare frequencies in an immediately visible way, by looking at the length of the bars being compared. In these figures, the registers are always presented in the order conversation, fiction, news writing, academic writing. For an example, see Figure 2.1 on p. 23.

To make comparisons easier in the figures, frequencies are normed to the standard measure of 'occurrences per million words'. Thus, although the sizes of the registers of the corpus are somewhat different, the comparisons are based on a standard measurement of *relative* frequency. It may be difficult to envision a million words, so (although there is great diversity in print size and rate of speech) handy comparative measures are as follows:

- Books average about 350 words/page, so one million words = about 3,000 pages.
- Speakers average about 120 words/minute, so one million words = about 140 hours.