Table of Contents

1 Looking at language in use—some preliminaries

1.1 Introduction

1.1.1 Thinking about *goalless*, *shall* and cars

Let's start by having a look at the following three questions:

a) What is the meaning of *goalless*?
b) How is the word *shall* used in Present-day British English? Suggest one or two typical examples to illustrate your description.
c) Who talks more about cars, British men or British women?

Question a) concerns the meaning of a single word—this type of question could, for example, be asked by a learner of English as a foreign language who has come across *goalless* without sufficient context to fully understand its meaning. In contrast, the second question goes beyond lexical meaning; *shall* is a modal verb (like *will*, *must* and *can*) and is therefore normally used together with other verb forms (like *run*, *sing* and *be*). In other words, rather than simply asking a question about the meaning of a certain word, question b) is about how this word can be combined with other elements of the English language to express a particular grammatical relationship or function. This question might for example be asked by an English teacher who is preparing a lesson on modal verbs. Question c), finally, broadly deals with the relationship between language and society. It is admittedly a bit of an odd question—calling up common clichés and stereotypes about the difference between the two sexes—and you are probably more likely to meet questions of this form during a dinner table conversation than as part of a linguistic enquiry. But there's a deeper reason for asking this question here, which will become apparent when we discuss possible answers, so let's just for the time being assume that this is a perfectly sensible thing to ask.

Task:
Spend a few moments thinking about possible answers to the questions above. Then ask some fellow students or friends the same questions and compare their answers to yours. Do you all agree on what the correct answers are? If not, think about the reasons why these differences may have occurred.

If you are a native speaker of English (or a highly proficient speaker of English as a second or foreign language), you may feel that your intuitions about the language will be fully sufficient to provide answers to all three of them. However, and this may have been confirmed if you did the above task as a group of people, even native speakers quite often disagree about certain aspects of language and its use, and these three questions may be no exception. For example, when answering question a), many people immediately think of *goalless* as meaning 'aimless, purposeless; having no destination'. Interestingly, typically only few people think of a second meaning of the word, namely that which is used in football to refer to 'a game in which no goals were scored on either side'.

Moving on to question b), your intuition may have told you that *shall* is quite old-fashioned and slowly dying out, while speakers nowadays prefer *will* and other future time expressions such as *going to* or *gonna*. You may also have worked out that the modal auxiliary *shall* is followed by the infinitive without *to*, and perhaps even that *shall* is used most frequently when the subject is a first person pronoun (that is, *I* or *we*). As a result, the typical example you gave might have looked something like this:

(1) I *shall* ring you up as soon as I arrive.

Alternatively, you might also have thought of a use of *shall* in offers, suggestions, requests for instructions, and requests for advice. This use takes the form of a question, i.e. the subject (e.g. *I*) follows the modal *shall*. A typical sentence is shown in (2).

(2) *Shall* I carry your bag?

When asked about the level of formality of this second type of use, people are usually quite undecided. However, the majority have the impression that this is a particularly polite—and therefore formal—usage. Furthermore, when asked about which of the two structures is more frequent, people often don't feel confident in providing a clear answer.

As for question c), most people would answer this by stating that men talk more about cars than women.

This quick summary clearly shows that the intuition-based approach can result in a considerable range of possible answers, and it is not clear how close to the "truth"—or perhaps better, how close to actual usage—they really are. In order to determine this, you may therefore want to find independent confirmation. Let us consider some ways in which this could be done. For example, dictionaries will easily help you with question a). Indeed, the *Oxford English Dictionary* (OED) lists both of the meanings of *goalless* that were mentioned above. Yet

you may also want to know which of the two senses is more common in Present-day English: unfortunately, the OED does not give you any help there.[1]

For the second question, grammar books are an obvious source of additional information. However, in this context it is important to ask what authority the author of a particular grammar book has for writing up his or her description. If its contents are heavily based on the author's intuitions about the English language, they may in fact also not fully reflect actual usage, even considering that an author of a grammar book is likely to be very knowledgeable about such matters.[2]

Another way of trying to find answers to at least the first two questions is by asking a wide range of informants who are native speakers of English. This is best done by giving them apparently unrelated questions whose context will trigger the use of the feature in question (e.g. *shall* vs. *will*). This method of "informant testing" is often more accurate than a direct appeal to native speaker intuitions, as the information provided is less likely to be influenced by factors such as self-censorship or accommodation. For example, when asked directly, an informant may opt to use *I will* or *I'll*—instead of *I shall*—because he or she does not want to give the impression of being old-fashioned. However, the same informant may not have any problems with using *I shall* in situations where they are not aware of the fact that the questions or tasks are designed to extract information about their use of *will* vs. *shall*. Although this informant-based method is clearly more informative than relying purely on the intuitions of a single speaker, it is obviously also much more difficult and time-consuming to carry out.

Finally, you could simply decide to observe what's happening around you and draw your conclusions on the basis of the data you collect. Every time someone talks about a car, you take note of the speaker's sex. Every time someone uses *shall*, you look at the type of construction in which it is used. And every time you read or hear *goalless*, you use the context to find out more about the meaning of this word. Once you have noted down a sufficient number of instances, you will have a reliable basis for a description of what is really going on with *goalless*, *shall* and talk about cars in today's English. However, there are two major problems with this method. First, with fairly infrequent words and expressions (e.g. *goalless*), you will have to wait a very long time before you have enough data to make any general claims. Secondly, and more importantly, your language experience may differ dramatically from that of other people who also use English. If, for example, you are a student at a British university, a large

1 However, some learner dictionaries (e.g. the *Collins COBUILD Advanced Learner's English Dictionary* 2006) do indicate whether certain senses are particularly common or rare.

2 It has to be pointed out, however, that many modern descriptions of English are no longer purely intuition-based. Instead, grammar books nowadays are often based on exactly the kind of data and methodology that we will describe in this book.

part of your language use will take place in interactions with other students and a considerable part of what you read will be academic texts (like the one you are reading right now). This is very different from the language experience of an average coal miner, lawyer, or jazz musician. And maybe the experience of these other types of language users will be particularly different from yours just in the context of the three questions you are trying to answer.

This book is about a method—and a tool—that will allow you to eliminate these two major problems to a very large extent. Suppose you had access to a huge collection of texts and conversations produced by a cross-section of today's population in Britain—i.e. by students, lawyers, jazz musicians, coal miners and a whole range of other types of language users. Further suppose that you would have access in such a way that it is possible to easily search the complete collection in a matter of seconds, and that you would also be able to get further information about the search results that are retrieved (e.g. about the type of speaker or writer, the kind of context in which it was produced, etc.). This is exactly what the British National Corpus (BNC) and *BNCweb* will give you.

1.1.2 Clues from a corpus—the BNC

The BNC is a 100 million word collection of samples of written and spoken language from a wide range of sources. It was put together to represent a wide cross-section of current British English, and contains a large number of language samples from different kinds of texts, produced by different kinds of language users and made available in different ways. A more detailed description of the corpus—including an account of how it was compiled, what type of texts it contains and what additional information is available about these texts—will be given in Chapter 3. *BNCweb* is a user-friendly web-based interface that was created to search (or as we say, to query) the data contained in the BNC. It gives you easy access to a wide range of functions that allow you to linguistically analyze the results of your queries. Originally developed at the University of Zurich by Hans Martin Lehmann, Sebastian Hoffmann and Peter Schneider (see Lehmann et al. 2000), *BNCweb* is nowadays maintained and further extended by Sebastian Hoffmann and Stefan Evert. The functionality of *BNCweb* is described in detail in the remaining chapters of the book.

To whet your appetite, let us quickly return to our three questions and see what clues we can find with the help of the BNC and *BNCweb*. A quick search for *goalless* shows that there are only 86 instances in the whole corpus. So on average, the word occurs less than once in every million words. Figure 1.1 displays how *BNCweb* will present the results of the search—or query—to you. This kind of output is generally referred to as a concordance.

Your query "goalless" returned 86 hits in 39 different texts (98,313,429 words [4,048 texts]; frequency: 0.87 instances per million words)

|< << >> >| Show Page: 1 Show Sentence View Show in random order New Query Go!

No	Filename	Hits 1 to 20 Page 1 / 5		
1	A1N 339	and drew a large attendance. The only redeeming feature of a	goalless	, worthless event at Stamford B
2	A1N 409	were created by Gillian Coultard in midfield, and the game finished	goalless	Photograph: Peter Jay Football
3	A22 50	clear mild night, emerged as worthy winners on aggregate after a	goalless	draw. Rangers, precariously ah
4	A2E 548	had managed just two goals in their first nine games. A	goalless	first leg at Hillsborough had lef
5	A2E 568	Sansom, their former England left-back, was left behind after the	goalless	draw against Stockport County
6	A3L 237	but less interesting and inventive. As one observer put it after	goalless	draws with Castellon and Mall
7	A40 589	's final game in the Lada Classic at Luton yesterday - a	goalless	draw with the world champion
8	A4B 401	has said he noted faults in Terry Butcher's game during the	goalless	draw in Sweden last month. Bu
9	A52 40	BARCLAY in Chorzow Poland. .0 England. .0 A SECOND successive	goalless	draw saw England through to t
10	A5C 5	team made sure of a place in Italy next summer with a	goalless	draw, their third in six qualifyi
11	A5U 147	leaders, Ealing, visit Leicester. Leicester could only manage a	goalless	draw midweek with Sutton Co
12	A80 377	different locations. The rush for tickets was sparked by Sunday's	goalless	draw between the US and El S
13	A8C 364	win at Hillsborough was important for morale, but last week's	goalless	draw with QPR at Plough Lan
14	A99 281	at least with the Swedes you can be reasonably sure of a	goalless	draw. All the top seeds will wa
15	A9H 175	quantity now, their most crucial result in the qualifiers was the	goalless	draw with the Soviet Union in

Figure 1.1: The first 15 hits of a search for *goalless* in the BNC (cropped view)

Looking at this concordance, it is immediately obvious that football appears to be the predominant context in which British English speakers make use of the word *goalless*. In fact, if you were to look at all 86 instances in more detail, you would find that every single one is from the field of sports. Now, this does not of course mean that the other meaning of *goalless*—i.e. 'aimless'—does not exist at all in Present-Day English. After all, although the BNC contains nearly 100 million words, it is actually quite tiny in comparison with the totality of language use in Britain, and it is entirely possible that some very infrequent features are not represented at all in the corpus. However, you can now safely say that the 'aimless' meaning of *goalless* is very marginal indeed. The other obvious point to note from this list of results is that *goalless* often co-occurs with *draw*, referring to a game during which no goals are scored.[3] Of the total of 86 instances, 51 (59 per cent) co-occur with *draw*. If you are a learner of English as a foreign language, this is useful information because it will not only allow you to understand the most common meaning of the word but it will also give you the opportunity to notice how it is used idiomatically by native speakers.

What can the BNC tell us about the second question, i.e. how *shall* is used in Present-day English? A simple lexical search of *shall* gives you many more hits than you will want to look at: there are 19,505 instances of *shall* in the whole

3 At least this is the case in British English. Speakers of other varieties of English may prefer the expression *goalless tie* instead.

BNC. However, we could restrict our investigation by looking at the spoken part of the corpus only. A good reason for doing this is that we suspect that *shall* is becoming less common nowadays: it is widely assumed in linguistics that when something changes in a language, that change generally starts in the spoken rather than the written variety.

With *BNCweb*, it is easy to restrict searches to sub-parts of the corpus, e.g. spoken texts only. This part of the BNC contains about 10 million words, but *shall* still occurs 2,735 times. This suggests that *shall* is still in common use in Present-day English—compare this to the 86 instances of *goalless* in the whole corpus—and that it is still a long way from vanishing from the language altogether. Figure 1.2 shows a screenshot of the first five hits that are returned by *BNCweb*.

As you can see, both types of uses mentioned above are found in these first few sentences, e.g. ***shall*** *we listen to you* (no. 1, where the personal pronoun follows *shall*) and *I* ***shall*** *be contacting him* (no. 4, where the personal pronoun is placed first). But which of the two patterns is more frequent, and can we find out more about preferences among particular (types of) speakers?

Your query "shall" in spoken texts returned 2735 hits in 482 different texts (10,409,858 words [908 texts]; frequency: 262.73 instances per million words) [0.202 seconds - retrieved from cache]

|< << >> >| Show Page: 1 Show KWIC View Show in random order New Query Go!

No	Filename	Hits 1 to 50 Page 1 / 55
1	D91 721	or **shall** we listen to you?
2	D95 182	Or **shall**, what do you think of haven't it petitioned?
3	D95 226	Well, what I think I **shall** do now is I think I should take this a little further about this union business, I think I should get in touch with Dave [gap:name] , the Editor of the T U C to find out what the exact position of these, this so called union is because it doesn't sound like a union to me, it sounds, it sounds like an .
4	D95 278	I did do that along with Ron [gap:name] and er they were speaking in terms of er a conjurer at under a pound a time and thing of that nature which should then come to a the pensioner's category at Poole, so I took it back to Stuart [gap:name] and he said oh see what I can do Norman, and at the present moment it rests there because I haven't been able to contact Stuart at the moment owed to the holiday, but I **shall** be contacting him and hopefully we will also be doing two days, which is the Tuesday and the Thursday, also what they, er, he's, he's promised to do is to come half way with the cost of the jazz band, which is a great help.
5	D95 279	Er, so you can say that er Mr [gap:name] is a friend of pensioner's, he said, he said he would be prepared to, what I, I, I approached him and said er, what about Harlow Caring Council, are they prepared to assist the pensioners in any way or do they wish to join in on this, oh yes he said, of course Norman he said, how much are you paying er Ron [gap:name] I said well his asking forty pound for the, for the morning, oh he said I'll go half way with that, then he came out and said to me, pull me up afterwards and ask me to go to leisure services about the Tuesday, and so I'm still following that up and hopefully we will have two days on pensioner's week, because you want to have as much impact as possible and in a few moments, when I nearly finished here, I **shall** be reading you something where you'll see that it is important that we make an impact on the people of Harlow.

Figure 1.2: Result of a search for *shall* in the spoken component of the BNC

One way of proceeding from here would now be to look at every single one of the 2,735 instances of *shall* returned by the search, always noting down information about the speaker (if available) and the grammatical pattern in which it is used. However, this would be very tedious and time-consuming. Fortunately there are quicker and more convenient ways of seeing patterns in the way *shall* is used. Let's for example consider the age of speakers who use *shall*. Our intuition might tell us that older speakers are typically more conservative and might therefore more likely use an old-fashioned form. If this were true we might then expect the use of *shall* to be more frequent among older speakers than among younger ones. *BNCweb* allows you to test this hypothesis in just a few simple clicks (using the so-called DISTRIBUTION feature).

Age:				
Category	**No. of words**	**No. of hits**	**Dispersion (over speakers)**	**Frequency per million words**
0-14	385,234	189	66/258	490.61
25-34	1,120,516	368	90/351	328.42
15-24	594,400	185	81/302	311.24
60+	1,137,433	311	89/318	273.42
35-44	1,075,749	287	81/335	266.79
45-59	1,638,364	400	133/436	244.15
total	**5,951,696**	**1,740**	**540/2,000**	**292.35**

Figure 1.3: Distribution of *shall* over the category "Age of speaker" in the spoken component of the BNC

As you can see in Figure 1.3, the data is not conclusive: older speakers do not use *shall* more frequently than younger ones; in fact, it is the youngest group that can be seen to use this modal most often, while the oldest age group is found somewhere in the middle of the table. Clearly, this finding does not support the view that *shall* is archaic and in the process of dying out.

But let's dig a little deeper. Another thing you can do with *BNCweb* is to find out which words occur particularly often before or after *shall*. In this way, you could confirm your hunch—if this is what you came up with in response to question b)—that the first person pronoun subjects *I* and *we* are very frequent both before and immediately after *shall*. It turns out that nine out of every ten instances of *shall* occur together with *I* or *we*. The interesting question now is whether there are any differences among the various age groups with respect to the two possible sentence types, i.e. *I/we shall* vs. *shall I/we*. Again, *BNCweb* gives you this type of information very quickly—the results are shown in Figures 1.4a and 1.4b.

Age:				
Category	**No. of words**	**No. of hits**	**Dispersion (over speakers)**	**Frequency per million words**
60+	1,137,433	183	60/318	160.89
45-59	1,638,364	192	72/436	117.19
35-44	1,075,749	118	48/335	109.69
25-34	1,120,516	106	49/351	94.6
15-24	594,400	52	30/302	87.48
0-14	385,234	11	9/258	28.55
total	**5,951,696**	**662**	**268/2,000**	**111.23**

Figure 1.4a: Distribution of *I/we shall* in the spoken component of the BNC

Age:				
Category	**No. of words**	**No. of hits**	**Dispersion (over speakers)**	**Frequency per million words**
0-14	385,234	175	62/258	454.27
25-34	1,120,516	244	69/351	217.76
15-24	594,400	126	63/302	211.98
35-44	1,075,749	149	55/335	138.51
45-59	1,638,364	197	92/436	120.24
60+	1,137,433	103	51/318	90.55
total	**5,951,696**	**994**	**392/2,000**	**167.01**

Figure 1.4b: Distribution of *shall I/we* in the spoken component of the BNC

As you can see, the two patterns show an opposite trend: *I/we shall* is most often used by older speakers (182 instances, on average 160 times per million words), but the same group of speakers use *shall I/we* the least (103 instances—about 91 instances per million words). The reverse is true for the youngest speakers, who use *shall I/we* most often (175 instances, 454 instances per million words) but hardly use *I/we shall* at all (only 11 instances).

Now that you have obtained these findings—or **DESCRIPTIVE STATISTICS**—you have quite a good foundation for answering the second of the three questions at the start of this chapter. First of all, you can say that *shall* is still quite frequent in Present-day English—although of course you haven't yet checked how much more frequent *will* is. Secondly, you can say that one of the two uses, i.e. *I shall* or *we shall* is predominantly used by older speakers, suggesting that the declarative form may indeed be old-fashioned. Furthermore, you can say that the other type of use, which includes offers, suggestions and requests for instructions expressed by *shall I?* or *shall we?*, is mainly used by younger speakers. Finally—and most crucially—you could look at this age distribution as a snapshot of a change in the English language that is still ongoing, and from this predict what the future of this use might be. Think about it: what will happen

when the young speakers represented in the BNC will be sixty or older? Will they have started using *I/we shall* more frequently by then because that's simply what older speakers do? Probably not. A much more likely interpretation of the data is that the declarative use is slightly dated and indeed slowly leaving the language—it is dying out. The use of *shall* for offers and suggestions, on the other hand, is probably going to increase even further. If this is true, perhaps it would make sense for teachers of English as a second or foreign language to introduce this type of use first, and only later go on to present the more marginal and archaic uses.

Even though we have extracted all sorts of information from the corpus, we have of course not yet answered the question whether the use of *shall* in offers and suggestions is particularly polite or not. Unfortunately, the tables we have compiled so effortlessly do not help us find this answer. Instead, we will have to look more closely at a sufficient number of instances of this particular use of *shall* in context. Descriptive statistics are almost always only one side of the coin, and a comprehensive description of a linguistic phenomenon will often require both a quantitative and a qualitative analysis of the data.

Finally, let's have a quick look at the third question—but how do we do this? How can we really answer the question whether men or women talk more about cars? A very basic approach would be simply to look for the word *car* and to have *BNCweb* calculate the same kind of distributional statistics as for *shall* above, just this time for the sex of speakers rather than age. Figure 1.5 displays the result of this calculation. Interestingly, women seem to use the word *car* more often than men. Notice, by the way, that the number of actual hits is higher for men (1,789 male vs. 1,597 female uses), but we need to take into account that there are more words in this corpus uttered by men than by women. This is why measuring the frequency across the same amount of text—as occurrences per million words, for example—is important: 485 instances per million words (pmw) for women vs. 361 pmw for men. We will—or we shall?—return to this issue again in later chapters.

Sex:				
Category	**No. of words**	**No. of hits**	**Dispersion (over speakers)**	**Frequency per million words**
Female	3,290,569	1,597	333/1,360	485.33
Male	4,949,938	1,789	438/2,448	361.42
total	**8,240,507**	**3,386**	**771/3,808**	**410.9**

Figure 1.5: Distribution of the word *car* over male and female speakers in the spoken component of the BNC

But what have we actually answered by looking at Figure 1.5? If you think about it, not all that much. First of all, we have forgotten an important part of the use of the word *car*: the plural form *cars*. Secondly, and much more importantly, what does it actually mean to "talk about cars"? Do you always need the lexical item *car* to do so? If someone says *I bought a Merc yesterday*, clearly this is also talking about a car. Conversely, what about mentioning a *car boot sale*? The word *car* is used here, too, but is the speaker really talking about cars? You can probably see that finding a reliable answer to the third question involves much more than a simple search and a few clicks in *BNCweb*—and this is a valuable insight. Some research questions are much easier to answer with the help of corpora than others, and it is important to know both the opportunities and the limitations that the use of corpora involves.

1.2 Why read this book?

This book is mainly about the practical steps involved in answering relevant linguistic research questions with the help of the BNC and *BNCweb*. As you will quickly realize, *BNCweb* is a very user-friendly tool: it is easy to perform a simple search of the corpus, and a few mouse-clicks are usually sufficient to give you lots of further information about your query. You might therefore wonder: is it really necessary to read a detailed manual? Our answer to this is: first, this book is not just a software manual—it was written by linguists interested in language study, and goes beyond a description of what the software can do. It is focused on what linguistic questions you can answer using the software and how you can go about interpreting the data generated by it in a meaningful way. The ease of use of *BNCweb* makes corpus-based language study appear simpler and more straightforward than it really is, and masks some considerations that should be part of every enquiry.

First and foremost, it is necessary to know more about the corpus: What is actually in the BNC? How did the compilers of the BNC choose the texts? How much do we know about the speakers and writers of the texts and the conditions of their production?

Second, it is necessary to learn the theoretical bases and methodological steps in corpus-based research: *How do I interpret the results presented by* BNCweb*? What do they tell me about British English as a whole or the text varieties that I chose to examine? What do they* ***not*** *tell me? How do I compare results from different searches? How can I be sure the results are reliable? How do I know that my searches really are relevant to answering my research questions?* This book will help you answer these important questions, and you will learn about theory and methods as you work your way through the chapters. It will help you avoid the potential problems and pitfalls that could turn the first

steps of a novice corpus user into a potentially frustrating or misguided experience.

In this book, methodological points are addressed and illustrated in the context of actual investigations of language use. It is this combination of theory with extensive hands-on practice that makes the book different from others in the field of corpus linguistics. The functionality of the various features of *BNCweb* are explained through "real-life" examples of linguistic issues, combining "how-to" with a discussion of theoretical and methodological considerations.

1.3 Organization of the book

The organization of the chapters is as follows: Chapter 2 introduces some of the fundamental concepts of corpus-linguistic methodology. This is followed by a detailed description of the British National Corpus in Chapter 3. In Chapter 4, we then illustrate the basic search functionality of *BNCweb* and show how a query result—in the form of concordance lines—can be investigated to gain insights into the use of a particular word or phrase. This is followed by a second methodology chapter—Chapter 5—which covers a number of important issues relating to the comparability and reliability of findings made through *BNCweb*. We focus on why normalized frequencies are important (and how they are calculated), introduce the concepts of "precision" and "recall", and testing for statistical significance. In Chapter 6, we offer a detailed description of *BNCweb*'s "Simple Query Syntax" and show how it can be used to perform highly sophisticated searches of the corpus.

The next three chapters are then devoted to various ways of further manipulating and analyzing your query result. Chapters 7 (DISTRIBUTION and SORT) and 8 (COLLOCATIONS) cover ways of exploring your query results automatically, i.e. without the need to look at concordance lines individually (or, as it is often called, manually). In Chapter 9, we then turn to the manual annotation of concordance lines and guide you through the process of adding your own classifications to a query result (either within *BNCweb* itself or with the help of third-party programs such as Microsoft *Excel*).

For many research questions, it will be necessary to restrict searches to a subsection of the whole BNC—a so-called "subcorpus". Chapter 10 illustrates the various ways in which subcorpora can be defined. Furthermore, we will show how user-defined subcorpora can be employed to make repeated searches of (sub-parts of) the BNC more efficient. *BNCweb* also offers two additional functions—the FREQUENCY LIST and KEYWORD features—that can be used to explore the corpus data from a more "whole-text" or macro perspective (i.e. without starting from a concordance); these will be covered in Chapter 11.

In addition to the Simple Query Syntax introduced in Chapter 6, *BNCweb* also accepts queries in something called "CQP Query Syntax", whose advanced features allow users to perform even more powerful and flexible searches of the corpus. Given the much less intuitive nature of this query syntax, however, the description offered in Chapter 12 is likely to appeal predominantly to more advanced users. Chapter 13, finally, concerns practical issues in the running of *BNCweb*. It covers such aspects as the difference between standard users and users with administrator rights, and it also describes some internal aspects of the workings of the software that have been designed to optimize access by whole groups of users. The chapter concludes by outlining some issues relating to the installation and maintenance of *BNCweb*.

1.4 How to use this book

This book is probably best read while sitting in front of a computer with access to *BNCweb*. This will make it possible for readers to gain hands-on experience in using the tool by following the step-by-step descriptions of the many sample analyses. Each chapter also contains a number of tasks and exercises that will offer further opportunities for enhancing and broadening the practical skills of readers. However, the book has been written in such a way as to make independent reading of its contents a worthwhile experience.

Several of the chapters contain a considerable amount of information—in fact, it may be too much to fully "digest" everything in one sitting. This especially applies to the two chapters which introduce the Simple Query Syntax and the CQP Query Syntax (Chapters 6 and 12), as their descriptions are designed to be useful as a comprehensive reference to the query language. Although it may be informative to read these chapters in one go, you will probably find yourself returning to their contents at some stage in the future, as your need to make more complex searches arises.

A similar comment applies to the chapter describing the BNC (Chapter 3) and to the methodologically oriented Chapters 2 and 5. While we recommend that you consult these chapters thoroughly before you conduct any serious studies on the BNC, we would like to encourage you to explore the different features and options of *BNCweb* at your own pace, so don't worry if you don't fully understand everything the first time around. As you become more experienced and more familiar with the output provided by *BNCweb*, you will likely get a better grasp of the more theoretical aspects of corpus linguistic methods that we discuss in these chapters. They are therefore well worth revisiting. In sum, we are confident that this book will give you a thorough grounding in corpus linguistic theory and methods, as you learn by doing—as we guide you through this powerful yet user-friendly program.