3 Corpus-Based Comparisons of Registers

Douglas Biber and Susan Conrad

3.1 Introduction

Speech has many different characteristics from writing. The most obvious difference is the physical mode of production. In addition, speech is usually interactive and speakers usually do not plan their language ahead of time. In contrast, writing is usually not interactive. In fact, writers are usually addressing a large audience, rather than a single reader. However, a writer can plan and revise the text as much as he or she wants. The final written text includes only the revised and edited language.

Chapter 2 by Ronald Carter dealt with some general differences between speech and writing. However, it would be misleading to suggest that all spoken texts are the same. Rather, spoken texts can differ from one another with respect to some of these same characteristics. For example, although a formal lecture is a type of spoken language, it is very different from conversation: it is carefully planned ahead of time, it is addressed to a large audience, it will probably not be interactive, and the speaker will usually want to communicate information about the world rather than telling us a lot of personal information.

Similarly, written texts can differ with respect to these characteristics. For example, email messages are a type of writing, but they are very different from textbooks: emails might not be edited or revised carefully; they probably will be written to a single person, who will respond; and they might talk about the personal feelings and activities of the writer.

We use the term **register** to refer to these different kinds of texts. Different cultures recognize different registers. One way to figure out the registers in a culture is to list the text categories that have names, such as 'letters', 'textbooks', 'email messages', 'newspaper articles', 'biographies', 'shopping lists', 'term papers', and 'novels'. Each of these registers differ in their characteristics. They are planned to different extents, interactive to different extents, and addressed to different kinds of audiences. They also have different topics, and they are written with different goals or purposes.

Registers are identified by non-linguistic or situational characteristics, such as the setting, the audience, interactiveness, and extent of planning. Registers can be described at different levels of specificity. For example, academic prose is a very general register. Research articles and textbooks are two more specific registers

within academic prose. Chemistry textbooks for lower division undergraduates is an even more specific register within the general register of textbooks.

Although registers are defined by their situational characteristics, it turns out that there are important linguistic differences among registers. This is because linguistic features serve important communicative functions, and therefore they tend to occur in registers with certain situational characteristics. For example, the use of first and second person pronouns (*I* and *you*) is a linguistic characteristic. Linguistic analysis of conversation shows that first and second person pronouns are extremely common in that register. This linguistic characteristic is associated with the normal situational characteristics of conversation. Speakers in conversation talk a lot about themselves, and so they use the pronoun *I* a lot. They are also interactive, talking to another individual person, and so they use the pronoun *you* a lot.

In contrast, newspaper articles have very few first and second person pronouns. These linguistic features are not needed because of the situation associated with a newspaper article: the writer of a newspaper article does not normally talk about his or her personal details, and does not address a specific individual reader. As a result, there is little need to use *I* or *you*.

In the present chapter, we explore some other important linguistic differences among registers, and explain how those differences in the language are related to situational characteristics. We compare four general registers: conversation, fiction, newspaper writing, and academic prose. These four registers illustrate some of the major ways in which language varies in systematic ways across situations of use.

Linguistic descriptions of a register are based on analysis of a collection of texts: a corpus. In the following section, we briefly introduce the methodology of corpus-based analysis. Then, in Section 3.3, we introduce the corpus of texts used for the analyses in this chapter.

Section 3.4 presents the linguistic comparisons of registers. We present three case studies: (1) the register distribution of content or lexical word classes (nouns, verbs, adjectives, and adverbs); (2) the use of verb tense and aspect in different registers; and (3) the use of *that*-clauses and *to*-clauses. Finally, in Section 3.5 we describe how these case studies illustrate general patterns of register variation found across a wider range of registers and linguistic features.

3.2 Corpus-Based Analyses of Registers

The linguistic description of a register can be based on **intuitions** or on **corpus-based analyses**, which are empirical studies of texts from the register. Most speakers of a language have intuitions: a sense of how a grammatical feature is used. These intuitions often include register differences, so speakers have the sense that selected linguistic features are preferred in particular registers. However, it turns out that these intuitions are often incorrect. As a result, we need empirical investigations of actual texts to accurately describe the preferred linguistic characteristics of a register.

For example, one of the most widely held intuitions about language use among English language teachers is the belief that the progressive aspect (as in 'He *is reading* a book') is the normal choice in conversation. As a result, traditional textbooks have often introduced the progressive in the very first chapter. However, corpus analysis shows that this intuition is wrong: rather, simple aspect verbs (e.g. *read*) are much more common in conversation than progressive verbs. We discuss these patterns in Section 3.4.2 below. Our purpose here is simply to illustrate how intuitions can be wrong.

In contrast to intuition, corpus-based analysis provides empirical methods for determining what features are common or rare. A corpus is a large, systematic collection of texts or text samples stored on computer. Corpus-based analyses have the following characteristics (see Biber *et al.*, 1998: 4):

1 They are empirical, analysing the actual patterns of use in natural texts.
2 They utilize a large and principled collection of natural texts: a corpus.
3 They make extensive use of computers for analysis, using both automatic programs and interactive techniques in which the user interacts with the computer (as users do with a spellchecker).
4 They depend on both quantitative and qualitative analytical techniques – that is, they are concerned with both counts of frequencies (quantitative analysis) and the way that the features are used (qualitative analysis).

Taken together, these characteristics result in a scope and reliability of analysis not possible otherwise. Corpus analyses are used to study the preferences that speakers or writers have for one grammatical choice over another. These analyses can specify the frequency of alternative structures and the conditions that are associated with their use.

There are many introductory textbooks for more comprehensive information on corpus linguistics (e.g. McEnery and Wilson, 1996; Biber *et al.*, 1998; Kennedy, 1998; Hunston, 2002; Meyer, 2002). The following sections illustrate the application of corpus methods to the study of register variation.

3.3 Conversation, Fiction, Newspapers, Academic Prose: A Comparison of Non-Linguistic Characteristics

As explained in Section 3.1, registers are different varieties of language that are associated with different situations and purposes. Table 3.1 summarizes the situations and purposes of the four registers compared here.

The most general distinction is mode: conversation is spoken, while the other three registers are written. Conversation is also distinguished from the written registers by being interactive and produced in real time. With respect to communicative purpose, the registers fall into three major categories: (1) conversation focuses on personal communication; (2) fiction on pleasure reading; and (3) newspapers and academic prose share a more informational purpose. However, even between newspapers and academic prose there are differences. Much of academic prose is more concerned with building an argument than newspaper

Table 3.1 Situational characteristics of four language registers

	Conversation	Fiction	Newspapers	Academic prose
mode	spoken	written	written	written
interactiveness and real-time production	yes	restricted to fictional dialogue	no	no
shared situation?	yes	no	no	no
main communicative purpose/content	personal communication	pleasure reading	information/ evaluation	information/ argumentation/ explanation
audience	individual	wide-public	wide-public	specialist

Table 3.2 Composition of the LSWE Corpus used for the register descriptions

Register	Number of texts	Number of words
conversation	3436	3,929,500
fiction	139	4,980,000
newspapers	20,395	5,432,800
academic prose	408	5,331,800

writing is. Further, academic prose has a more specialized audience than a newspaper, which is written with a wide audience in mind.

As the following sections show, registers have grammatical differences as a reflection of their different communication circumstances. That is, the circumstances of a register have direct associations with language forms that are common in the register. The four registers discussed here do not provide a complete picture of register variation in English. But they are major varieties that illustrate the kinds of linguistic differences found among a wider range of registers.

The register descriptions presented in the following sections are taken from the *Longman Grammar of Spoken and Written English* (Biber *et al.*, 1999). Those descriptions are based on analysis of the Longman Spoken and Written English Corpus, summarized in Table 3.2.

3.4 Linguistic Differences Among Registers: Some Case Studies

3.4.1 Content word classes

One of the easiest linguistic characteristics for register comparisons is the use of the basic part-of-speech categories: nouns, verbs, adjectives, and adverbs. Figure 3.1 plots the frequency of these word classes, showing important linguistic differences

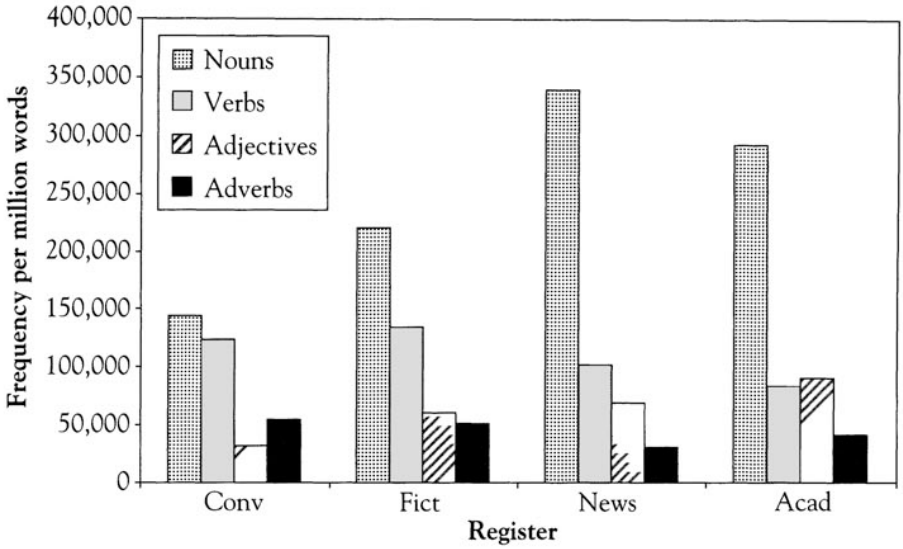

Figure 3.1 Content word classes across registers

Source: Based on Biber *et al.* (1999, Figure 2.2)

across the four registers in our study. The most obvious difference is for nouns: extremely common in newspapers and academic prose, but much less common in conversation. Verbs are less frequent overall, but they show the opposite distribution: more common in conversation than in newspapers or academic prose. Fiction is intermediate in the use of both nouns and verbs.

Adjectives and adverbs mirror the distribution of nouns and verbs: adjectives are most common in academic prose and newspapers, while adverbs are most common in conversation and fiction.

Text samples 1 and 2 illustrate these characteristics in conversation and academic prose:

Text Sample 1. Conversation (At a dinner party)
Verbs are underlined italics; **nouns are bold underlined**

B: Hmm, *wait* until you *taste* this.
A: What *is* it?
C: Yum, **quiche**.
B: *Wait* until you *taste* this.
C: *Let* me *see*, am I gonna *sit* here or over there?
B: I don't *know*. Whichever **way** you *want* to *do* it.
A: This *looks* like nice **salad**. What *is* this, some **eggs**?
B: Hmm. Okay. No *taste* it.
A: What? *Taste* what? I'*m* not sure what that *is*. It'*s* **egg**. I *think* that'*s* what this *is*.
B: Well, you have to *taste* it.

A: Ooh, this *is* good. Mm, hmm. You didn't *put* any **dressing** on.
B: I *know*. It's right there in that little **thing**. I *'m* curious to *know* what you *think* of the **quiche**.
A: Hmm. Well I haven't *tasted* it yet.

Text Sample 2. Academic prose sample
Verbs are underlined italics; **nouns are bold underlined**; ***adjectives are bold italics***.

Table 7-5 *measures* in several ***different*** **ways** the **significance** of the ***various*** **industry classifications**. **Column** 2 *indicates* the ***numerical*** and **percentage distribution** of the **business population** among the ***various*** **industries**. **Column** 3 *shows* in both ***absolute*** and ***relative*** **terms** the **portion** of the ***national*** **income** *originating* in the ***various*** **industries**. Several **points** in Table 7-5 *are* ***noteworthy:*** A ***large*** **number** of **firms** *is engaged* in **agriculture**, but **agriculture** *is* relatively ***insignificant*** as a **provider** of **incomes** and **jobs**.

The core grammatical characteristics of these two text samples are strikingly different. The academic textbook sample has only seven lexical verbs. Verbs like *indicates*, *shows*, and *are/is* do not communicate much information; rather, their main job is to connect long and complex noun phrases, which convey most of the new information in the passage.

Newspaper prose has similar characteristics to academic writing. In fact, newspaper stories often employ even more nouns than academic prose, especially when companies, people, and their positions are described:

Text Sample 3. Newspaper sample
Verbs are underlined italics; **nouns are bold underlined**

Ariana's **board** *includes* **Bernard Giroud**, who *started* **Intel France** in 1971 and *left* the **company** last **year** a corporate **vice president** to *become* a **venture capitalist**, **Gerard Yon**, formerly **sales** and **marketing manager** of **Chorus Systemes**, the **microkernel house**, and now **president** of **VST**, a French **start-up** in electronic **document management**, and **Pascal LeVasseur**, technical **director** of **Dell France**.

In contrast, the conversation excerpt relies heavily on verbs and short clauses, a total of 30 main verbs in this short interaction. These verbs communicate much of the essential information: actions and events (*wait*, *taste*) and the speaker's mental states and attitudes (*know*, *want*, *think*). In contrast, only eight nouns occur in this conversation. Further, some of these nouns have almost no specific meaning (*way*, *thing*).

Instead of nouns, the conversation excerpt uses pronouns. Text Sample 1 is repeated below, with these pronouns marked:

Text Sample 1 repeated from above (conversation), with **pronouns marked as bold underlined**
B: Hmm, *wait* until **you** *taste* **this**.
A: What *is* **it**?

C: Yum, quiche.
B: *Wait* until **you** *taste* **this**.
C: *Let* **me** *see*, am **I** gonna *sit* here or over there?
B: **I** don't *know*. Whichever way **you** *want* to *do* **it**.
A: **This** *looks* like nice salad. What *is* **this**, some eggs?
B: Hmm. Okay. No *taste* **it**.
A: What? *Taste* what? **I** *'m* not sure what **that** *is*. **It** *'s* egg. **I** *think* **that** *'s* what **this** *is*.
B: Well, **you** have to *taste* **it**.
A: Ooh, **this** *is* good. Mm, hmm. **You** didn't *put* any dressing on.
B: **I** *know*. **It** *'s* right there in that little thing. **I** *'m* curious to *know* what **you** *think* of the quiche.
A: Hmm. Well **I** haven't *tasted* **it** yet.

Figure 3.2 shows that this dense use of pronouns is typical of conversation. In fact, there are slightly more pronouns than nouns in this register. This pattern is very different from newspapers and academic prose, where nouns are many times more common than pronouns (as Samples 2 and 3 illustrate).

These basic differences in the use of word classes are related to basic situational differences among the registers. Conversation focuses on personal information and so uses the pronoun *I* and verbs that express feelings and mental states. It is directly interactive and so it also uses the pronoun *you*. There is also a shared situation, so speakers can refer to that situation with third person pronouns without being explicit about the intended meaning. In contrast, newspapers and academic

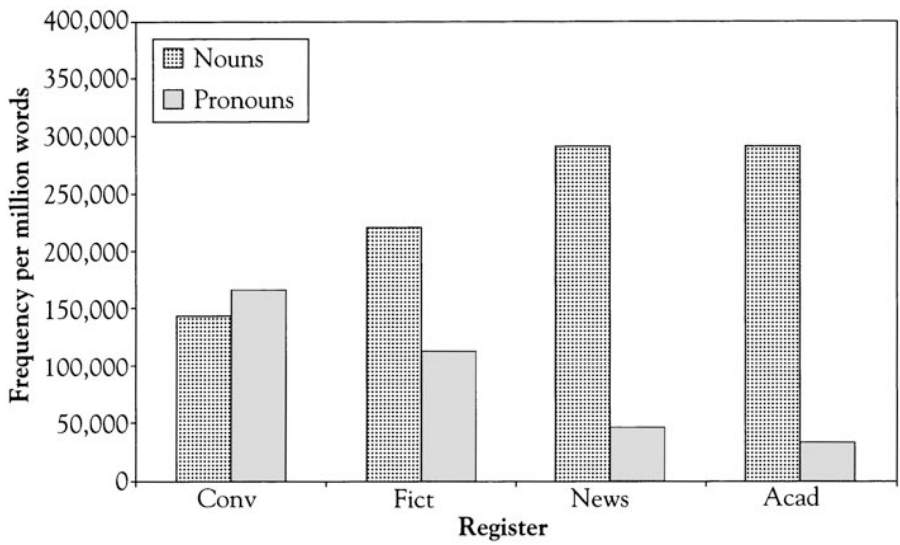

Figure 3.2 Nouns vs. pronouns across registers

Source: Based on Biber *et al.* (1999, Figure 4.1)

prose tend to use full nouns to avoid possible ambiguity. (Thus note the repetition of *agriculture* in the last line of Sample 2. If the pronoun *it* had been used here instead, readers might mistakenly assume that it referred to the noun *number*.)

3.4.2 Verb tense and aspect

Another basic linguistic feature that shows important register differences is verb tense. Figure 3.3 compares the use of present vs. past tense verbs across registers. In contrast to the use of nouns vs. verbs (Section 3.4.1 above), conversation and academic prose are surprisingly similar in their strong preference for present tense over past tense. Text Samples 1 and 2 in the last section illustrate this pattern: only one verb phrase in the conversation is past tense (*didn't put*) and none of the verbs in the academic excerpt are past tense. However, the preference for present tense reflects slightly different uses in these two registers. In both registers, present tense verbs are often used to describe states that exist at the present time. Conversation uses copular verbs like *be* and *look* for this purpose, while academic prose uses verbs like *become*, *change*, *develop*, *occur*, *include*, and *involve*. Conversation, though, has an additional common use for present tense verbs: to express personal attitudes or states with mental verbs like *think*, *want*, *know*, *see*, and *mean*.

In contrast, fiction shows a strong preference for past tense verbs. This is not especially surprising, since fiction focuses on the narration of past actions and events. Text Sample 4 illustrates the use of past tense verbs in fiction.

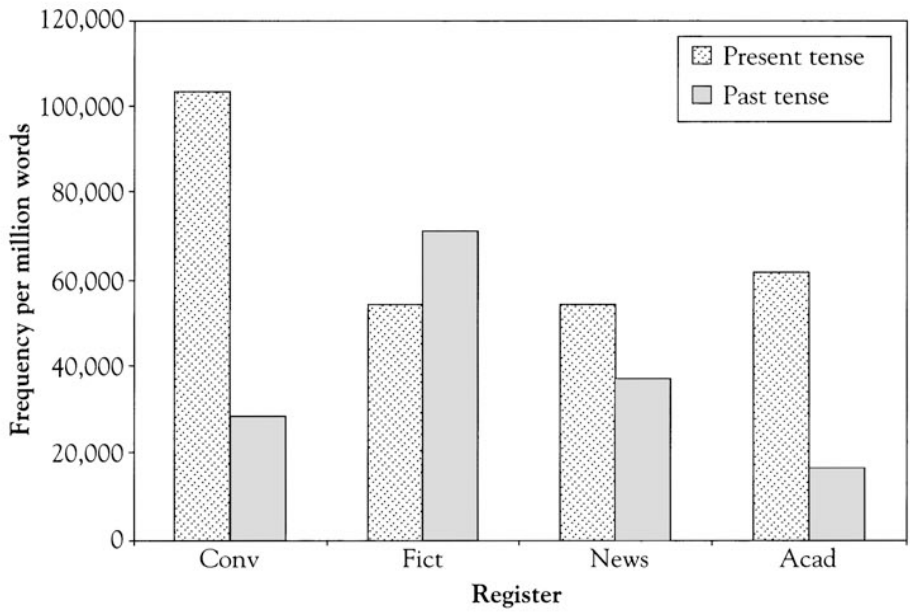

Figure 3.3 Present vs. past tense verbs across registers

Source: Based on Biber *et al*. (1999, Figure 6.1)

Text Sample 4. Fiction text sample
Past tense verbs are bold underlined.

Then one day Leon **announced** he **was leaving** home, at the age of sixty-two. Roman **had** dreams at night in which he **rounded** up pigs who **were** loose in the corn, and they **kept** Leon awake. So he **went** over the hill and **made** him a sod house in the meadow.

The use of tense in newspapers is more surprising. Because the 'news' is usually events that have already occurred, we might expect that newspapers would be like fiction in having a strong preference for past tense verbs. It is the case that past tense verbs are much more common in newspapers than in academic prose. However, Figure 3.3 shows that present tense verbs are more common than past tense verbs in newspapers. Sample 5 illustrates this pattern of use:

Text Sample 5. Newspaper text sample
Past tense verb phrases are bold underlined; *present tense verb phrases are italics underlined* (non-finite verb phrases are not marked)

A bill signed by Gov. Pete Wilson yesterday making it easier for the San Fernando Valley to split from Los Angeles may spark secession movements in other cities around the state. The governor's office **received** calls urging him to sign the bill from some residents of La Jolla, who *have complained* in the past about services provided by the city of San Diego. 'This bill *sends* a wake-up call to the entire state – city governments must become more accountable to the people they *serve*,' **said** Wilson. The measure, which *becomes* law Jan. 1, no longer *allows* a city council to unilaterally veto a secession proposal. A co-author of the bill, Assemblyman Robert Hertzberg, D-Sherman Oaks, **said** at a bill-signing ceremony that the San Fernando Valley's drive to form a new city *has* already *sparked* interest in three other Los Angeles communities – Wilmington, San Pedro and Hollywood. 'It really *deals* with this underlying issue of people wanting a closeness to their government, not necessarily less government,' **said** Hertzberg. After signing the bill, Wilson lightheartedly **thanked** Hertzberg and the other co-author, Tom McClintock, R-Simi Valley, for not mentioning that the state's last split **was** Coronado from San Diego.

Past tense verbs are used in newspaper stories for the same purposes as in fiction: to report past events. Many of these are speech act verbs, reporting what somebody *said*. However, the reports of these events are interspersed with statements giving commentary on the significance of those past events, and present tense verbs are common in those statements (e.g. verbs like *becomes*, *allows*, *deals*). In addition, newspaper stories include quotes of direct speech, which are similar to conversation in their preference for present tense. Finally, newspaper stories also use perfect aspect verb phrases with present tense to express past actions with continuing consequences (e.g. *have complained*, *has sparked*).

Similar to the register differences for verb tense, verb aspect also shows important differences across register. There are two marked verb aspects in English: progressive aspect and perfect aspect. For example:

	Present tense	Past tense
Progressive aspect	*is eating*	*was eating*
Perfect aspect	*has eaten*	*had eaten*

We have space here for only a brief consideration of their use, focusing on progressive aspect.

As noted above, English language teachers and materials writers often believe that the progressive aspect is the normal choice in conversation. This belief is sometimes reflected in the frequent use of progressive verbs in made-up dialogues. For example, consider the following conversation from *As I was Saying: Conversation Tactics* (Richards and Hull, 1987):

> Text Sample 6. Made-up dialogue from an English language coursebook
> **Progressive aspect verb phrases are bold underlined**
>
> Doctor: Hello Mrs. Thomas. What can I do for you?
> Patient: Well, I've **been having** bad stomach pains lately, doctor.
> Doctor: Oh I'm sorry to hear that. How long have you **been having** them?
> Patient: Just in the last few weeks. I get a very sharp pain about an hour after I've eaten.
>
> ...
>
> Doctor: Well, I don't think it's anything serious. Maybe you eat too quickly. You don't give yourself time to digest your food.
> Patient: My husband **is always telling** me that.

As Figure 3.4 shows, it is correct that progressive aspect verb phrases are more common in conversation than in other registers. The contrast with academic prose is especially noteworthy: progressive aspect is rare in academic prose but common in conversation. However, as Figure 3.5 shows, it is not at all correct to conclude that progressive aspect is the normal choice in conversation. Rather, simple aspect is clearly the normal choice. In fact, simple aspect verb phrases are more than 20 times as common as progressive verb phrases in conversation. Text Sample 1, in Section 3.4.1 above, illustrates the normal pattern – all 20 of the verb phrases in this interaction are simple aspect. In cases like this, we see how corpus research can correct mistaken perceptions that are widespread among language professionals.

3.4.3 *That*-clauses and *to*-clauses: the interaction of words, syntax, and register

That-clauses and *to*-clauses are the two most common types of complement clause in English. They are called **complement clauses** because they complete the meaning

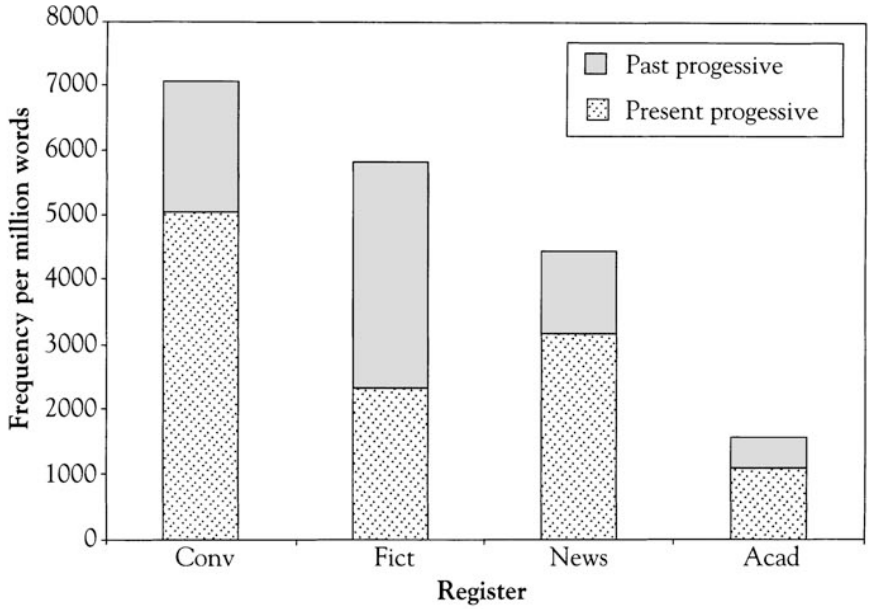

Figure 3.4 Progressive aspect verb phrases across registers

Source: Based on Biber *et al*. (1999, Figure 6.4)

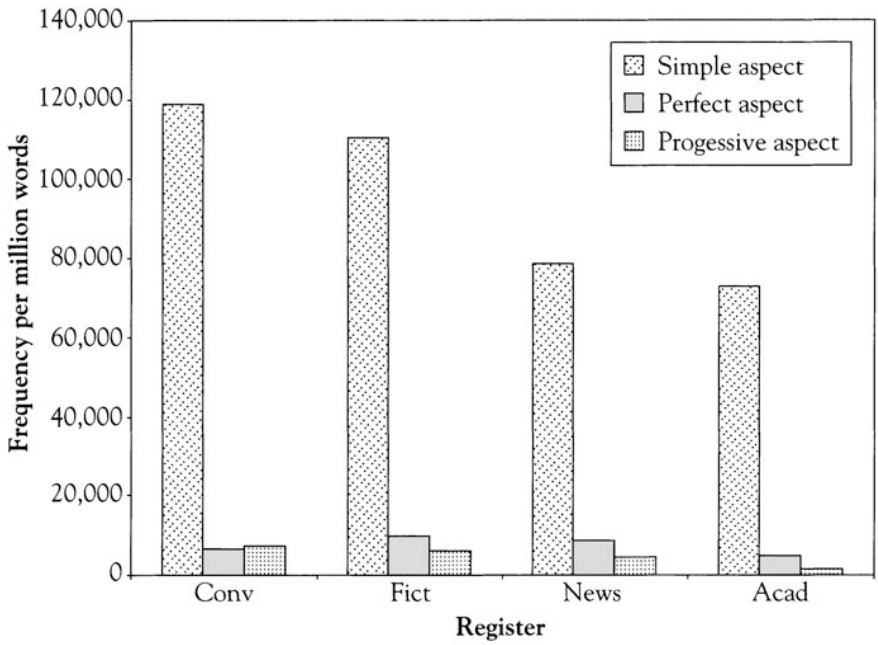

Figure 3.5 Simple, perfect, and progressive aspect verb phrases across registers

Source: Based on Biber *et al*. (1999, Figure 6.2)

of another element in the clause. Often, they complete the meaning of the verb, for example (complement clauses are underlined):

I hope that I can go.
I hope to go.

In these examples, the complement clauses complete the meaning of the verb *hope*. Alternatively, the complement clauses are said to be controlled by the verb. They serve as direct objects of the verb, telling what it is that I hope.

As these examples illustrate, *that*-clauses and *to*-clauses can sometimes be used in similar grammatical contexts with similar meanings. However, corpus analysis shows that these clause types are usually not overlapping in their use: they tend to occur with different sets of controlling verbs, in different registers, serving distinct communicative functions.

Figure 3.6 presents the overall distribution of each clause type across registers. Even at this general level of analysis, we are confronted with findings that show the importance of register – and that run counter to popular expectations. In particular, there is a widespread perception that dependent clauses are rare in conversation but common in formal written registers. However, *that*-clauses controlled by verbs show exactly the opposite pattern: they are most common in conversation and notably rare in academic prose. In contrast, *to*-clauses controlled by verbs have roughly the same frequency in conversation and academic prose, but proportionally they are the

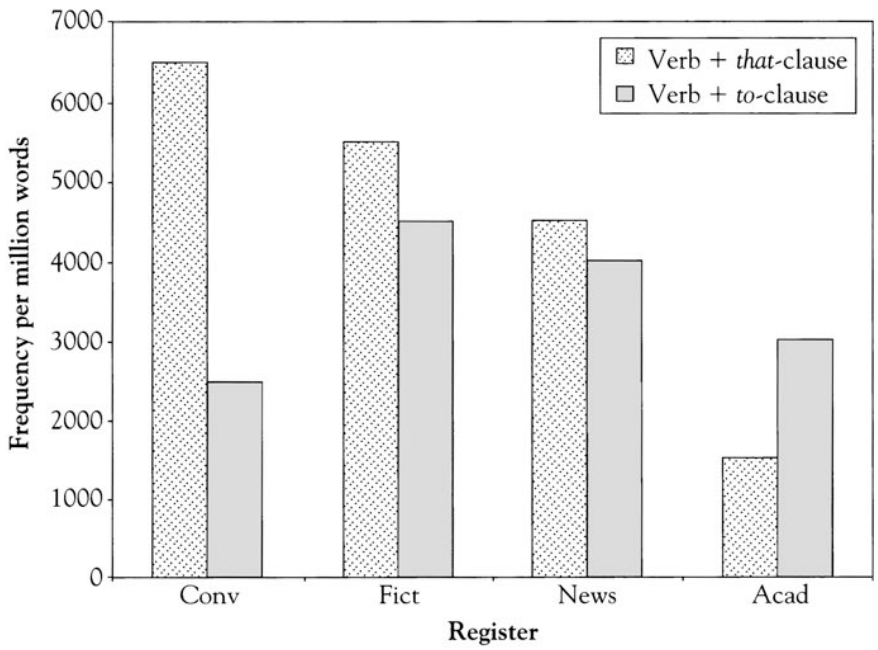

Figure 3.6 *that*-clauses and *to*-clauses across registers

Source: Based on Biber *et al.* (1999, Figures 9.6 and 9.11)

preferred choice in academic prose. *To*-clauses are considerably more frequent in fiction and news, where they are almost as common as *that*-clauses.

The patterns of use shown in Figure 3.6 raise several interesting questions. For example, why should some dependent clauses be more common in conversation than in written registers? What communicative functions are those clauses typically performing? Conversely, why should *that*-clauses controlled by verbs be so rare in academic prose, while *to*-clauses are much more common?

It is difficult to address such questions by a casual inspection of examples. An alternative approach, however, is to analyse the lexical associations for each type of complement clause in each register. That is, a corpus-based investigation can determine which verbs occur most often with each type of clause. These are called **lexico-grammatical patterns** because they show how lexical items (words – in this case, the verbs) are associated with the grammatical structures (the different types of complement clauses). It is unlikely that the verbs would occur randomly, and so analyses proceed based on the assumption that the most common controlling verbs will provide an indication of the typical communicative functions of each type of clause. In fact, it turns out that there are strikingly different lexico-grammatical patterns associated with each complement clause type and with each register, and that those associations can be explained in terms of the typical topics and communicative purposes of each register.

Many grammar books will identify *possible* lexico-grammatical patterns. They will note that a few verbs can control both *that*-clauses and *to*-clauses (e.g. *hope*, *decide*, and *wish*), but most verbs can control only one or the other type of complement clause (e.g. the verbs *imagine*, *mention*, *suggest*, *conclude*, *guess*, and *argue* can control a *that*-clause but not a *to*-clause; the verbs *begin*, *start*, *like*, *love*, *try*, and *want* can control a *to*-clause but not a *that*-clause). However, a corpus-based investigation is different because it focuses not on what is *possible*, but on the actual lexico-grammatical patterns of use. This perspective allows us to investigate the typical uses of each clause type (as opposed to the patterns that are grammatical in theory but might occur rarely in practice).

As Tables 3.3 and 3.4 show, the most common verbs controlling a *that*-clause constitute a completely separate set from the most common verbs controlling a *to*-clause, even though some of these verbs are grammatical with both types of complement clause. Further, these tables show important register differences, since the most common controlling verbs vary considerably from one register to the next.

Some of these verbs (such as *want* and *try*) are grammatically controlling only one type of complement clause, and they have strong lexical associations with that structural type. Other verbs – such as *think*, *say*, and *know* – are grammatically controlling both types of complement clause; however, these verbs have strong association patterns with only one clause type. Thus, although there is some overlap in the controlling verbs that are grammatical, corpus-based analysis shows that there is in fact very little overlap in the commonly occurring lexical associations.

Further, *that*-clauses and *to*-clauses are productive in different ways. *That*-clauses combine with relatively few verbs, from only a few categories of meaning (also called **semantic domains**) – mostly mental/perceptual verbs (e.g. *think*, *know*, *see*, *believe*, *feel*) or communication verbs (e.g. *say*, *suggest*). However, a few of those verbs are extremely common controlling *that*-clauses, especially the verbs *think*, *say*,

Table 3.3 Most common verbs controlling a *that*-clause

	Conversation	Fiction	Newspapers	Academic prose
think	**********	*********	***	*
say	**********	********	**********	**
know	*******	********	*	*
see	**	**	*	**
believe	*	**	***	*
find	*	**	*	**
feel	*	**	*	
show	*	*	**	***
suggest		*	*	**

Note: Each * represents c. 100 occurrences per million words.
Source: Based on Biber *et al.* (1999, Figures 9.2–9.5 and Table 9.2).

Table 3.4 Most common verbs controlling a *to*-clause

	Conversation	Fiction	Newspapers	Academic prose
want	**********	*******	****	*
try	**	****	***	*
like	**	**	*	
seem	*	****	*	***
begin		****	*	*
appear		*	*	**
continue		*	*	*
allow NP[1]		*	*	*
ask		*	*	
find NP		*		
expect NP			*	*
fail			*	*
agree			*	
BE expected[2]				**
tend				**
require NP				*
attempt				*
BE found				*
BE required				*

Notes: 1 NP indicates that a noun phrase occurs between the verb and the *to*-clause, as in *I would allow* ***the Press Council*** *to play its proper role.*
2 BE indicates any form of the verb *be* (i.e., making passive voice), as in *Heavy fighting was expected to break out soon.*
Each * represents c. 100 occurrences per million words.
Source: Based on Biber *et al.* (1999, Figures 9.13–9.16 and Table 9.8).

and *know* in conversation (and to a lesser extent fiction). The verb *say* controlling a *that*-clause is also extremely common in news.

In contrast, apart from the verb *want* in conversation, no individual verb is extremely common controlling *to*-clauses. However, there are a large number of different verbs that can control a *to*-clause, and those verbs come from many different semantic domains: mental verbs (e.g. *expect*, *find*), communication verbs (e.g. *ask*, *promise*), verbs of desire (e.g. *want*, *like*), verbs of intention or decision (e.g. *agree*, *decide*, *intend*), verbs of effort or facilitation (e.g. *try*, *attempt*, *allow*, *enable*, *fail*, *require*), aspectual verbs (having to do with the status of an action as beginning, in progress, or completed – e.g. *begin*, *continue*), and likelihood verbs (e.g. *seem*, *appear*, *tend*).

These differing patterns of lexical association help to account for the overall differences in register distribution between *that*-clauses and *to*-clauses. Conversational partners tend to use a relatively restricted range of vocabulary, but it is almost always appropriate to report one's own thoughts (*I think that ...*, *I know that ...*) or the speech of others (*he/she said that ...*) with a *that*-clause. (Note that the complementizer *that* is usually omitted in conversation.) For example (*that*-clauses are underlined):

That-clauses in conversation:

I think he will. Actually, I think he's quite good, don't you?
I know it's sort of miserable.
Maureen said that Ryan was sick.

Multiple occurrences of these verb + *that*-clause combinations are often used in close proximity, as in:

He said it was so difficult for him. I think it was a real shock for him.

The verb *think* is especially common as a controlling element in conversation, accounting for about 30 per cent of all *that*-clauses in that register. In most cases, this verb is used as a hedging device to mark a proposition that the speaker is not entirely certain about (rather than reporting the actual 'thoughts' of the speaker). For example:

A: Is this plastic, or is it, perhaps, you know, resin?
B: I think it's plastic.

Because of the extremely heavy reliance on a few high frequency verbs as controlling elements – especially *think*, *say*, and *know* – *that*-clauses are very common in conversation.

Turning to the use of *to*-clauses in conversation, it is almost always appropriate to report one's own personal desires, and this is most commonly done using the single verb *want* as a controlling element. For example:

To-clauses in conversation:

I wanted to get rid of it.
And then he said, 'I don't mean to put pressure on, but I just want to get to know you, we've got so much in common, and, uh, I want to take you out for dinner ...'.

However, other uses of *to*-clauses are much less common in conversation, accounting for the generally lower frequency of this complement clause type in that register.

The three written registers show a very different pattern of use with the verbs controlling *to*-clauses: although no single verb is extremely common (except for *want* in fiction), there are a large number of verbs from different semantic domains that occur relatively frequently. *To*-clauses controlled by verbs are most common in fiction because it relies on a few high frequency verbs – especially *want*, *try*, *seem*, and *begin* – but also makes frequent use of a wide range of different verbs.

It is interesting to note that even the high frequency verbs controlling *to*-clauses in fiction are from four different semantic domains and thus represent different communicative functions: (1) *want* expressing personal desire; (2) *try* expressing effort; (3) *seem* as a marker of likelihood; and (4) *begin* as an aspectual verb. For example (controlling verbs are underlined):

To-clauses in fiction:

She wanted to go to Mexico.
Before she went, Margotte wanted to kiss the old man.

He probably tried to save it.
She was trying to divert his attention.

It seemed to be a lot wilder than anything I remembered.
Toby seemed to be gone a long time.

Then I began to laugh a bit.
Then I felt the post begin to slide upwards through my hands.

In sum, then, though many verb + complement clause combinations are possible, the corpus analysis reveals strong association patterns among the clause types, the verbs, and the registers. These patterns of use are tied to the function that each verb + clause combination serves in fulfilling a communicative need of the register, as well as to different levels of diversity in the communicative functions of the registers.

3.5 Conclusion

This chapter has provided a brief introduction to the use of corpus analyses for making comparisons among registers. In the past, register variation has been described with intuition or anecdotal evidence – both of which have been misleading at times. Corpus studies, in contrast, provide empirical evidence of how large numbers of speakers and writers adapt to the different communicative situations of different registers. Large-scale patterns are apparent on a variety of grammatical levels. Word classes, grammatical systems such as tense and aspect, and lexico-grammatical associations – such as between complement clauses and verbs – all give evidence of the ways that language choices can be adapted to meet the needs of different situations of use to our everyday lives.

Virtually all language users must adapt their language for different situations of use. Few people go through even a single day experiencing just one register – only

having casual conversations, or only reading a newspaper, or only writing an academic paper. Understanding register variation, therefore, is not a supplement to our understanding of grammar; it is central.

References

Biber, D., Conrad, S. and Reppen, R. (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. London: Longman.

Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Kennedy, G. (1998) *An Introduction to Corpus Linguistics*. Harlow: Addison-Wesley Longman.

McEnery, T. and Wilson, A. (1996) *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Meyer, C. (2002) *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.

Richards, J.C. and Hull, J.C. (1987) *As I Was Saying: Conversation Tactics*. Tokyo: Addison Wesley.