A register perspective on grammar and discourse: variability in the form and use of English complement clauses
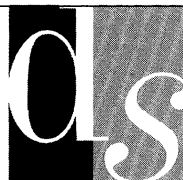
DOUGLAS BIBER
NORTHERN ARIZONA UNIVERSITY

ABSTRACT   This article explores the importance of register variation
for analyses of grammar and discourse. The general theme is
illustrated through consideration of variability in the form and use of
English complement clauses. First, the patterns of use for four related
grammatical constructions are considered: *that*-clauses and *to*-
clauses, headed by verbs and by nouns. The differing discourse
functions of each construction type are explored by considering their
lexico-grammatical associations (i.e. the verbs or nouns most
commonly occurring as the head of each type). However, it is shown
that the characteristic uses of each type are conditioned by register.
That is, each construction type has a different distribution across
spoken and written registers, with a different set of associated lexical
heads.A second study provides an even more striking illustration of
this interaction between grammar, discourse, and register: the
contextual factors conditioning the retention vs omission of the
complementizer *that*. In this case, it is shown that each register has
an overall norm, and that contextual factors are influential only
when they work in opposition to that register norm. These case
studies are presented to make the general point that analyses of
grammar and discourse are often inadequate and misleading when
they disregard register differences. Instead, a register perspective is
required to capture the range of variability associated with
grammatical patterns of use.

KEYWORDS: *complement clauses, corpus linguistics, discourse function,
English grammar, lexico-grammar, register variation*

1. Introduction

There have been numerous studies of grammar and discourse over the past two
decades, as researchers have come to realize that the description of grammatical

function is as important as structural analysis. In most cases, these studies focus on grammatical features that have two or more structural or semantic variants. By studying these features in naturally occurring discourse, researchers have been able to identify systematic differences in the functional use of each variant.

Research of this type became popular in the late 1970s and 1980s. For example, Prince (1978) compared the discourse functions of *wh*-clefts and *it*-clefts, considering a large number of examples from conversation and various written sources. Sandra Thompson and Deborah Schiffrin each carried out several studies of this type. For example, Thompson investigated word-order variation with detached participial clauses (1983), and adverbial purpose clauses (1985), as well as the discourse conditions associated with the omission of the complementizer *that* (Thompson and Mulac, 1991a, 1991b), and variation in the form and informational properties of relative clauses (Fox and Thompson, 1990). Schiffrin has studied the discourse factors influencing grammatical variation in verb tense (1981), causal sequences (1985a), and discourse markers (1985b, 1987). Other more recent studies of this type include Ward (1990) on verb phrase (VP) preposing, Collins (1995) on dative alternation, and Myhill (1995, 1997) on the discourse functions of modal verbs.

These studies are all empirical, in that they are based on analysis of grammatical features in actual texts. In addition, most of these studies have used both quantitative and qualitative analysis. That is, quantitative techniques are used to determine the distribution of grammatical variants across contexts, while detailed analyses of text extracts are used to interpret the distributional patterns in functional terms.

Despite these characteristics, there has often been relatively little concern with the generalizability of the database of texts used for analysis. Many of these studies have used a 'convenience' sample: a collection of texts that was readily available to the reseacher. The implicit assumption underlying this methodological decision seems to have been that any body of naturally occurring discourse will illustrate the same patterns of use. However, these text samples have often been small and, more importantly for the present purposes, there has often been no systematic control for register. Some studies are based on a single register; others are based on discourse examples with disregard to register; while others incorporate a comparison of use across registers.

More recently, researchers on discourse and grammar have begun to use the tools and techniques available from corpus linguistics, with its greater emphasis on the representativeness of the database, and its computational tools for investigating distributional patterns in large text collections (see Biber et al., 1998, for an introduction to this analytical approach).

There have been numerous research papers using corpus-based techniques to study English grammar and discourse. The edited volumes by Aarts and Meyer (1995), Aijmer and Altenberg (1991), and Johansson and Stenström (1991) provide good introductions to work of this type. There are also a number of book-length treatments reporting corpus-based investigations of grammar and dis-

course: for example, Tottie (1991) on negation, Collins (1991) on clefts, Granger (1983) on passives, Mair (1990) on infinitival complement clauses, Meyer (1992) on apposition, and several books on nominal structures (e.g. Varantola, 1984; De Haan, 1989; Geisler, 1995; Johansson, 1995).

In most cases, corpora are designed to represent some register differences, and thus many grammatical studies based on corpora have a register component. For example, Tottie (1991) and Geisler (1995) report differences for speech vs writing; Johansson (1995) distinguishes among Press, Fiction, and Academic prose for some analyses; and Granger (1983) distinguishes among several different spoken registers (including conversation, oration, commentary, interviews). At the same time, other corpus-based studies disregard register distinctions in their studies of grammar and discourse, focusing exclusively on a detailed analysis of contextual factors (e.g. De Haan, 1989; Mair, 1990; Sinclair, 1991).

Here I take a strong position on the importance of register for studies of discourse and grammar, arguing that most functional descriptions of a grammatical feature will *not* be valid for the language as a whole. Rather, characteristics of the textual environment interact with register differences, so that strong patterns of use in one register often represent only weak patterns in other registers. Thus, a complete functional analysis must consider the patterns of use in several registers.

In the following sections, I illustrate the interaction of grammar, discourse, and register with corpus-based analyses adapted from the *Longman Grammar of Spoken and Written English* (Biber et al., in press). The analyses are based on texts from four registers: conversation, fiction, newspaper language, and academic prose. Although these are general registers, they differ in important ways from one another (e.g. with respect to mode, interactiveness, production circumstances, purpose, and target audience). The analyses were carried out on the Longman Spoken and Written English Corpus, which contains around 40 million words of text, with around 4–5 million words from each of these four registers. All frequency counts reported here have been normalized to a common basis (a count per 1 million words of text), so that they are directly comparable across registers.

2. *Variation in the form and use of English complement clauses*

2.1. THAT-CLAUSES VS TO-CLAUSES

The two most common types of complement clause in English are *that*-clauses and *to*-clauses. These clauses can be controlled by verbs, adjectives, or nouns. In the following examples, the controlling element is given in brackets, and the complement clause is italicized:

> Controlled by a verb:
> I [hope] *that I can go.*
> I [hope] *to go.*

Controlled by an adjective:

I'm [happy] *that we're going to Sarah's house.*
I'm [happy] *to go to Sarah's house.*

Controlled by a noun:

He supported the [proposal] *that secure accommodation should be provided for juvenile offenders.*
He supported the [proposal] *to provide secure accommodation for juvenile offenders.*

As these examples illustrate, that-clauses and to-clauses can sometimes be used in similar grammatical contexts with similar meanings. However, empirical text analysis shows that the typical use of these structures is quite different. The following discussion focuses only on complement clauses controlled by verbs and nouns.

Figures 1 and 2 present 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 generally rare in conversation but common in formal written registers. However, of these four clause types, only *that*-clauses controlled by nouns provide a straightforward illustration of this distributional pattern, showing an increasing cline in frequency from conversation to academic prose. Interestingly, *that*-clauses controlled by verbs show exactly the opposite pattern: they are most common in conversation and notably rare in academic prose.
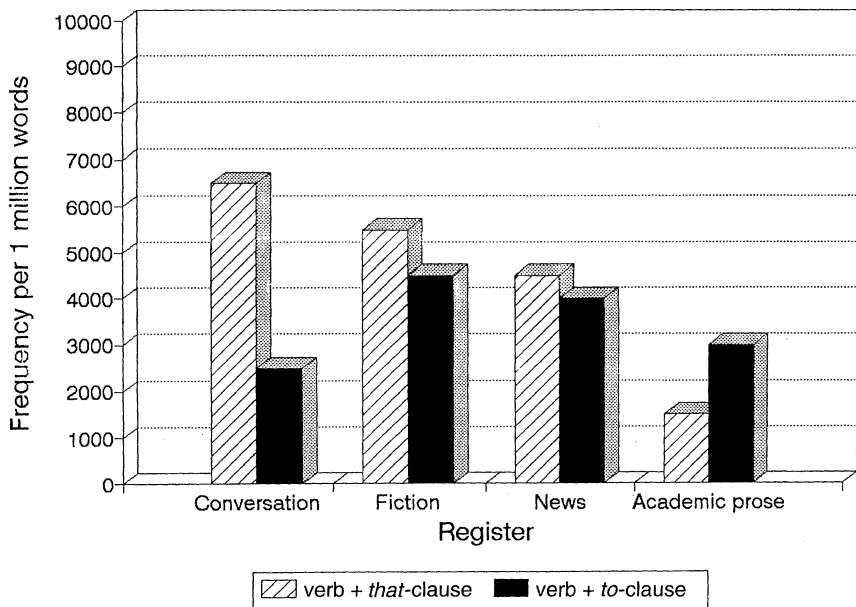

FIGURE 1. *Register distribution of verb +* that-*clause and verb +* to-*clause (based on Biber et al., in press, Chapter 9)*
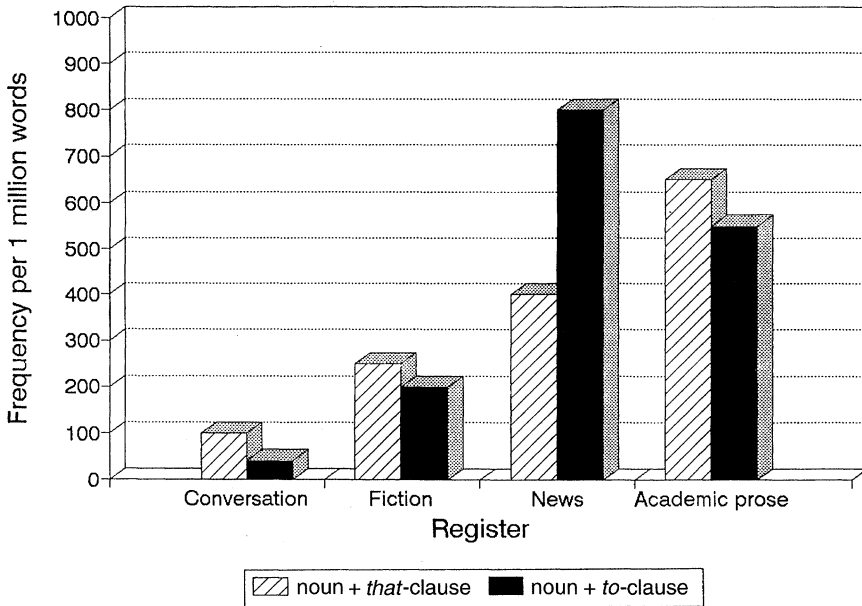
FIGURE 2.   *Register distribution of noun +* that-*clause and noun +* to-*clause (based on Biber et al., in press, Chapter 8)*

To-clauses are distributed in very different ways. *To*-clauses controlled by verbs have roughly the same frequency in conversation and academic prose, but they are considerably more common in fiction and news. In contrast, *to*-clauses controlled by nouns are extremely rare in conversation but very common in academic prose; however, these constructions are by far most common in news.

Certain aspects of these distributional patterns can be explained in terms of general register characteristics. In particular, many researchers since Wells (1960) have argued that 'nominal' styles are typical of formal writing, while 'verbal' styles are more colloquial. Reflecting this basic distinction, the analysis here shows that complement clauses controlled by nouns are generally common in the formal, written registers but rare in conversation. Similarly, in conversation complement clauses controlled by verbs are generally common, while complement clauses controlled by nouns are generally rare.

Previous research has also led us to expect writing generally to have a much higher degree of subordination than speech (e.g. O'Donnell, 1974; Kay, 1977; Kroll, 1977; Chafe, 1982; Brown and Yule, 1983), accounting for the increasing cline for *that*-clauses controlled by nouns (rare in conversation; very common in academic prose). However, this generalization does little to account for the patterns of variation found for the other three types of complement clause.

Some of these research questions have been addressed by MultiDimensional studies of register variation (Biber, 1988). Such studies have shown that dis-

course complexity is a multi-dimensional construct, that different types of structural elaboration reflect different discourse functions, and that different spoken and written registers are complex in different ways (see especially Biber, 1992). Further, it turns out that similar patterns are found cross-linguistically (see Biber, 1995, Chapter 7). For example, adverbial clauses tend to be found most commonly in conversation and other colloquial registers cross-linguistically, while nominal modifiers tend to be found most commonly in written, informational registers. Complement clauses are found in both spoken and written registers cross-linguistically, reflecting the patterns shown in Figures 1 and 2.

Although these previous studies help account for the overall distributional patterns of dependent clauses, we are left with several specific patterns in Figures 1 and 2 which run counter to many expectations and cannot be fully explained from previous research. For example, why should any type of dependent clause be more common in conversation than in written registers (as with *that*-clauses controlled by verbs)? What discourse functions are those clauses typically performing? Similarly, why should *to*-clauses controlled by nouns be so common in newspaper language, or what are the discourse functions of *that*-clauses controlled by nouns that make them so common in academic prose?

It is difficult to address such questions in a principled manner by a casual inspection of examples; such analyses do not provide an adequate basis for generalizations about the typical discourse functions of a feature in a register. However, an alternative approach is to analyze the differing lexical associations for each type of complement clause in each register, based on the assumption that the most common controlling verbs and nouns will provide an indication of the typical discourse functions of each type. This approach is used in the following subsections, to discuss the most common verbs controlling *that*-clauses and *to*-clauses (discussed in 2.1.1) and the most common nouns controlling *that*-clauses and *to*-clauses (discussed in 2.1.2). These analyses show that strikingly different lexico-grammatical patterns are 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.

2.1.1. Common verbs controlling that-*clauses versus* to-*clauses* Although a few verbs can control both *that*-clauses and *to*-clauses (e.g. *hope, decide,* and *wish*), most verbs can control only one or the other type of complement clause. For example, 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.

A complementary perspective is to consider the actual lexico-grammatical patterns of use, identifying the verbs that most commonly co-occur with each type of complement clause. 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 Figures 3 and 4 show, the most common verbs controlling a *that*-clause constitute a completely separate set from

	Conversation	Fiction	News	Academic Prose
think	********************	*********	***	*
say	************	********	*****************	**
know	*******	********	*	*
see	**	**	*	**
believe	*	**	***	*
find	*	**	*	**
feel	*	**	*	
show	*	*	**	***
suggest		*	*	**

FIGURE 3. *Most common verbs controlling a* that-*clause. Each * represents about 100 occurrences per million words (based on Biber et al., in press, Chapter 9)*

	Conversation	Fiction	News	Academic Prose
want	***********	*******	****	
try	**	****	***	*
like	**	**	*	
seem	*	****	*	***
begin		****	*	*
appear		*	*	**
continue		*	*	*
allow NP		*	*	*
expect NP		*	*	
fail			*	*
BE expected			**	
agree			*	
tend				**
attempt				*

FIGURE 4. *Most common verbs controlling a* to-*clause each * represents about 100 occurrences per million words (based on Biber et al., in press, Chapter 9)*

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 illustrate how investigations of use require a register perspective, since the most common controlling verbs vary considerably from one register to the next.

Some of these verbs (such as *want* and *try*) are grammatical 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 grammatical controlling both types of complement clause; however, these verbs have strong association patterns with only one clause type. Thus, although there is some overlap between the two types of complement clause 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 semantic domains –

mostly mental/perceptual verbs (e.g. *think, know, see, believe, feel*) or communication verbs (e.g. *say, suggest*). However, some of those verbs are extremely common controlling *that*-clauses, especially the verbs *think, say*, 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 is 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, learn*), communication verbs (e.g. *ask, promise*), verbs of desire (e.g. *want, like*), verbs of decision (e.g. *decide, intend*), verbs of effort or facilitation (e.g. *try, attempt, allow, enable, fail*), aspectual verbs (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,

(1)
I think he will. Actually, I think he's quite good, don't you?

(2)
I know it's sort of miserable.

(3)
Maureen said that Ryan was sick.

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

(4)
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 percent 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). Example 5 illustrates this use (as do 1 and 4):

(5)
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 appropri-

ate to report one's own personal desires, and this is most commonly done using the single verb *want* as a controlling element. For example,

(6)
I wanted to get rid of it.

(7)
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 is 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 discourse functions: *want* expressing personal desire; *try* expressing effort; *seem* as a marker of likelihood; and *begin* as an aspectual verb. For example:

(8)
She wanted to go to Mexico.

(9)
Before she went, Margotte wanted to kiss the old man.

(10)
He probably tried to save it.

(11)
She was trying to divert his attention.

(12)
It seemed to be a lot wilder than anything I remembered.

(13)
Toby seemed to be gone for a long time.

(14)
Then I began to laugh a bit.

(15)
Then I felt the post begin to slide upwards through my hands.

In sum, this section has briefly discussed the most common verbs taking each type of complement clause. Each register has a different pattern of lexical associations, which is in turn associated with the typical discourse functions of the

clause type in that register. The following section shows how a similar approach can be used to investigate the discourse functions of noun-complement clauses.

2.1.2. Common nouns controlling that-*clauses vs* to-*clauses*   Similar to controlling verbs, a few nouns can control both *that*-clauses and *to*-clauses (e.g. *proposal* and *claim*). However, most nouns can control only one or the other type of complement clause. For example, the nouns *hope, possibility,* and *fact* can control only a *that*-clause, while the nouns *attempt, opportunity,* and *effort* can control only a *to*-clause. Unlike complement clauses controlled by verbs, both *that*-clauses and *to*-clauses controlled by nouns are primarily features of written language, being especially common in the informational, non-fiction registers. However, Figure 2 shows a surprising difference in the register distribution of the two types of complement clauses controlled by nouns: *that*-clauses are much more common in academic prose than in news, while *to*-clauses show the opposite distribution.

Consideration of the most common nouns controlling each type of complement clause helps to explain this distributional difference. Figure 5 presents the register distribution of the most common nouns controlling a *that*-clause, while Figure 6 presents the most common nouns controlling a *to*-clause. Figure 5 shows that academic prose has the widest range of common nouns controlling *that*-clauses, including some nouns that are especially common (*possibility, fact,* and *assumption*). Figure 6 shows the opposite pattern, with news having the widest range of common nouns controlling *to*-clauses, including several nouns that are particularly common: *chance, attempt, opportunity, effort, ability, decision, right, plan,* and *bid.*

The nouns controlling *that*-clauses and *to*-clauses are for the most part from quite different semantic domains. Most of the common nouns controlling *that*-clauses refer to cognitive constructs (e.g. *idea, impression, knowledge, sense, opinion, belief*) or logical constructs (e.g. *possibility, conclusion, fact, hypothesis, proposition*). In academic prose, these noun + *that*-clause combinations function as one of the primary devices used to mark stance. In these constructions, the *that*-clause reports a proposition, while the head noun reports the author's stance towards that proposition. Two primary kinds of stance information are given in academic prose by these constructions (although many head nouns can express both types of information):

(A) An assessment of the certainty of the proposition in the *that*-clause, as with *fact, possibility, hypothesis*; for example:

(1)
We are not here concerned with the elusive though connected *fact* that causal circumstances in a different sense explain their effects.

(2)
There is a *possibility* that some sediment could get into milk which could lead to a prosecution or rejection by the buyer.

	Conversation	Fiction	News	Academic
idea		**	*	**
hope		*	**	*
possibility		*	*	***
impression		*	*	*
knowledge		*	*	*
news		*	**	
conclusion		*		*
sense		*		*
opinion		*		*
fear		*		
thought		*		
doubt			**	**
suggestion			**	**
belief			**	**
fact			*	****
view			*	**
indication			*	*
claim			*	*
ground(s)			*	*
report			**	
sign			*	
rumor			*	
warning			*	
assumption				****
observation				**
notion				*
hypothesis				*
assertion				*
conviction				*
proposition				*

FIGURE 5. *Most common nouns controlling a* that-*clause. Each* * *represents about 10 occurrences per 1 million words (based on Biber et al., in press, Chapter 8)*

(3)
This recently reported structure thus provides direct evidence supporting our *hypothesis* that the release of torsion-angle strain takes place during phosphorylation.

(B) An indication of the source of the knowledge expressed in the *that*-clause. Three primary sources can be distinguished as:

• linguistic communication, as with *claim, report, suggestion, proposal, remark*
• cognitive reasoning, as with *assumption, idea, notion, observation*
• personal belief, as with *belief, doubt, hope, opinion*

For example:

(4)
That is, the Papago case turns out to support Chomsky and Halle's *claim* that place alone never distinguishes anterior coronals.

	Conversation	Fiction	News	Academic
chance	**	***	*******	*
attempt		**	******	*******
opportunity		**	****	****
effort		**	****	***
desire		**	*	**
ability		*	***	*********
power		*	*	***
inability		*	*	*
duty		*	*	*
permission		*	*	
decision			******	*
right			***	***
failure			**	***
capacity			*	**
commitment			*	*
intention			*	*
willingness			*	*
plan			*******	
bid			******	
battle			**	
proposal			**	
agreement			*	
deal			*	
determination			*	
freedom			*	
refusal			*	
scheme			*	
tendency				***
responsibility				*

FIGURE 6. *Most common nouns controlling a to-clause. Each * represents about 10 occurrences per 1 million words (based on Biber et al., in press, Chapter 8)*

(5)
Implicitly or explicitly, there is an *assumption* that formal care systems have in some way been deficient.

(6)
Their frustrations were the product of their *belief* that the leadership was not responding adequately to the party's 'crisis'.

Many of the common nouns taking *that*-clauses in academic prose are nominalized equivalents of verbs that can control *that*-clauses, including:

hope, doubt, suggestion, belief, conclusion, claim, fear, knowledge, sense, report, assumption, thought, hypothesis, and *observation*

Examples 7 and 9 illustrate the use of *that*-clauses with controlling nouns in academic prose, while examples 8 and 10 illustrate the use of corresponding verbs controlling *that*-clauses in conversation:

(7)
Lagrange seemed to cherish *hopes* that his work would show the way to the solution of the general quintic.

(8)
I just *hope* that I've plugged it in properly. (Conversation)

(9)
Affirmative assessment starts from a basic *belief* that human beings are purposeful beings who have intentions which guide their behavior.

(10)
I just can't *believe* I did that.

As illustrated by these examples, the expression of stance is backgrounded and turned into an abstraction in noun-complement clauses when compared to verb-complement clauses. With verb-complement clauses, the subject of the controlling verb is often a human agent or experiencer – usually referring to the speaker (*I*) in conversation – so that the stance reported by the verb can be attributed directly to that person. In contrast, the stance conveyed by controlling head nouns in academic prose is not attributed to anyone, so that readers must infer that the noun reports the stance of the writer.

The opposite distributions of *that*-clauses controlled by verbs (preferred in conversation) and *that*-clauses controlled by nouns (preferred in academic writing) can thus be attributed to two factors. First, conversation has an overall preference for verbal rather than nominal structures, while academic prose tends to integrate information in noun phrases. In addition, the differing primary purposes and conventions of conversation and academic prose are important here: conversational participants are very interested in each others' personal feelings and attitudes, and thus stance is expressed prominently and directly attributed to participants. In contrast, academic readers and writers are generally much more interested in the information being conveyed than personal attitudes; thus, when stance is expressed, it tends to be backgrounded and not directly attributed to the author.

The common use of definite head noun phrases with *that* complement clauses further backgrounds the author's stance, since it carries the implication that the stance expressed by the controlling noun is generally accepted or known information (e.g. *the fact, the possibility, the assumption*). In most cases, readers will not have already adopted the stance expressed by this controlling noun. However, the use of the definite article with the controlling noun suggests that the expressed stance is generally accepted, further backgrounding the fact that the noun actually presents the personal stance of the author.

Unlike *that*-clauses, the head nouns most commonly taking *to*-clauses do not typically present a personal stance towards the proposition in the complement clause. Instead, the common head nouns taking *to*-clauses present human goals, opportunities, or actions; for example, *chance, attempt, effort, ability, opportunity, decision, plan, bid*. These meanings fit the typical purposes of news reportage, with a focus on human goals and actions. For example,

(11)
We need to give decent people a *chance* to elect a sensible council.

(12)
Mr Golding believed the parents had suffered enough and any *attempt* to prosecute the doctor should be dropped.

(13)
Meetings on Friday morning also ended without a *decision* to accept the deal which provides for Palestinian self-rule in the Gaza Strip.

(14)
A loyalist politician has admitted talking to Sinn Fein about his *plans* to travel into the heart of republican areas of Belfast.

(15)
Now it's back to stroke play as Gordon Fairweather makes another *bid* to put the family name back on the trophy.

(16)
The leader's gunshot wounds are taking their toll, complicating *efforts* to persuade him to surrender.

Thus, from a grammatical perspective, *that*-clauses and *to*-clauses have complementary sets of controlling nouns, and these lexical associations correspond to different typical discourse functions for the two complement clause types: providing an assessment of the status of information (certainty or source) in the case of the nouns controlling *that*-clauses, and describing human goals, opportunities, or actions in the case of the nouns controlling *to*-clauses. However, a register perspective is needed to uncover these associations: *that*-clauses controlled by nouns are used especially in academic prose; *to*-clauses controlled by nouns are found primarily in news; and neither clause type is common in conversation or fiction.

3.   *Retention vs omission of the complementizer* that

In most *that*-clauses, the complementizer can be freely omitted with no substantial change in meaning. For example, compare:

(a) I hope *I'm not embarrassing you.*
(b) I hope *THAT Paul tells him off.*

There are several characteristics of the textual environment that influence the retention vs omission of *that*, and these textual factors interact in important ways with register differences. First, as Figure 7 shows, different registers have different overall norms for *that* retention vs omission. In conversation, *that*-omission is the typical case, with the complementizer being omitted in about 85 percent of all occurrences. At the other extreme, academic prose almost always retains the complementizer *that*.

These overall distributional patterns correspond to the differing production circumstances, purposes, and levels of formality found across registers.
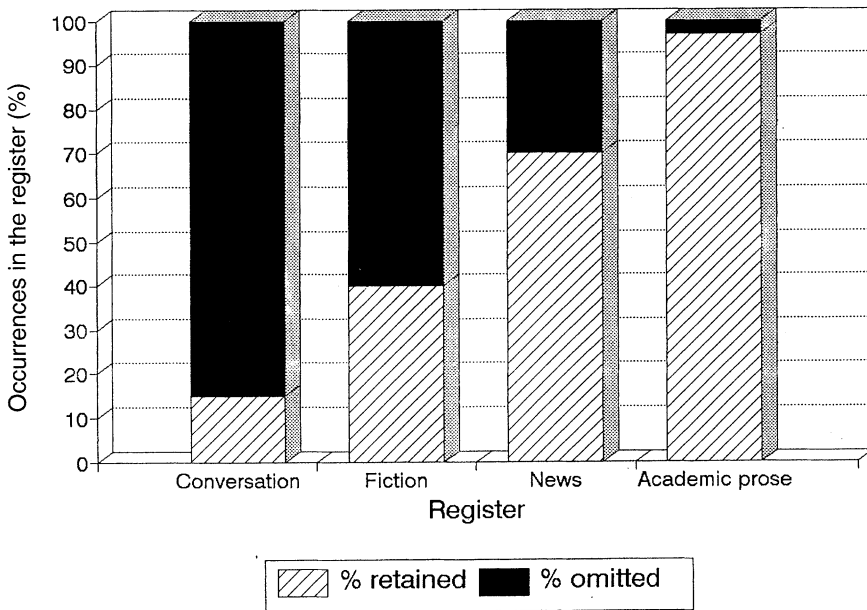
FIGURE 7. *Proportional retention vs omission of* that, *by register (based on Biber et al., in press, Chapter 9).*

Conversations are spoken and produced on-line; they typically have involved, interpersonal purposes; and they are casual and informal in tone. These characteristics are associated with omission rather than retention of *that* as the norm. Academic prose has the opposite characteristics: careful production circumstances; an expository, informational purpose; and a formal tone. Correspondingly, *that* retention is the norm in academic prose.

Textual factors are also important in influencing the choice between omission and retention of *that* (Thompson and Mulac, 1991a, 1991b). These factors can be divided into two groups:

(A) Textual factors favoring the omission of *that*:
The omission of *that* is favored when the grammatical characteristics of the surrounding discourse conform to the most common uses of *that*-clauses. To the extent that a construction conforms to the characteristics typically used with *that*-clauses, listeners and readers can anticipate the presence of a *that*-clause without the explicit marking provided by the *that* complementizer.
Two of the most important typical characteristics are:
• The use of *think* or *say* as the main clause verb (these are by far the two most common verbs taking a *that*-clause; see Figure 3);
• The occurrence of co-referential subjects in the main clause and the *that*-clause (which is more common than non-co-referential subjects).

(B) Textual factors favoring the retention of *that*:
The retention of *that* is favored with grammatical characteristics that are not typical
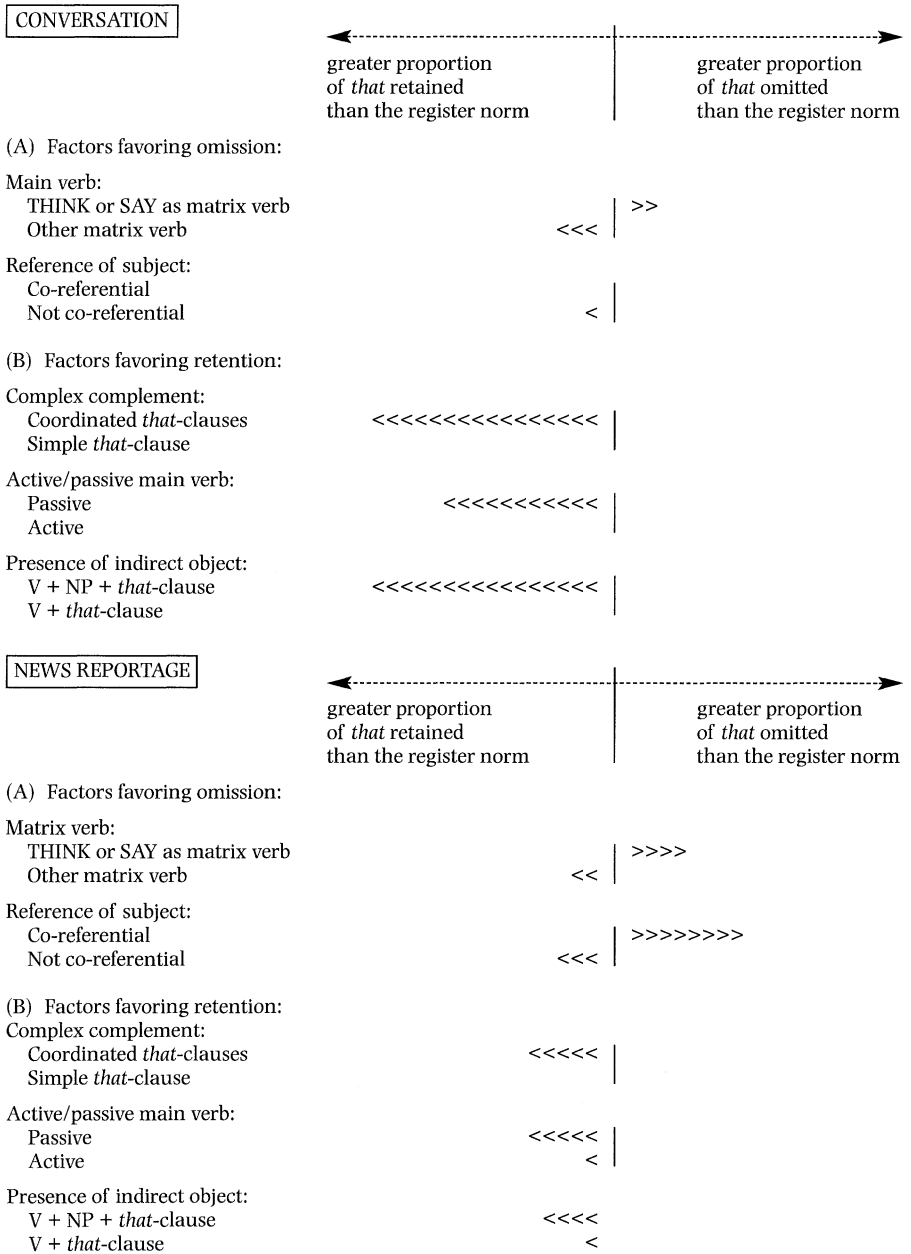
| CONVERSATION | | |

◄--------------------------------|-------------------------------►

greater proportion | greater proportion
of *that* retained | of *that* omitted
than the register norm | than the register norm

(A) Factors favoring omission:

Main verb:
   THINK or SAY as matrix verb   >>
   Other matrix verb   <<<

Reference of subject:
   Co-referential
   Not co-referential   <

(B) Factors favoring retention:

Complex complement:
   Coordinated *that*-clauses   <<<<<<<<<<<<<<<
   Simple *that*-clause

Active/passive main verb:
   Passive   <<<<<<<<<<<
   Active

Presence of indirect object:
   V + NP + *that*-clause   <<<<<<<<<<<<<<<
   V + *that*-clause

| NEWS REPORTAGE | | |

◄--------------------------------|-------------------------------►

greater proportion | greater proportion
of *that* retained | of *that* omitted
than the register norm | than the register norm

(A) Factors favoring omission:

Matrix verb:
   THINK or SAY as matrix verb   >>>>
   Other matrix verb   <<

Reference of subject:
   Co-referential   >>>>>>>>
   Not co-referential   <<<

(B) Factors favoring retention:
Complex complement:
   Coordinated *that*-clauses   <<<<<
   Simple *that*-clause

Active/passive main verb:
   Passive   <<<<<
   Active   <

Presence of indirect object:
   V + NP + *that*-clause   <<<<
   V + *that*-clause   <

FIGURE 8. *Departure from the register norms for retention vs omission of the complementizer* that, *depending on textual factors (based on Biber et al., in press, Chapter 9). Each '<' or '>' represents a 5% departure from the register norm, for all occurrences of* that-clauses *in that register with the stated textual factor: '<' marks proportionally greater use of* that RETENTION *than the register norm; '>' marks proportionally greater use of* that OMISSION *than the register norm.*

of *that*-clauses making these structures difficult to process if *that* were omitted. Three of the most important such factors are:

- The use of coordinated *that*-clauses;

- The use of a passive voice verb in the main clause;

- The presence of an intervening noun phrase between the main clause verb and the *that*-clause.

For the present discussion, the most interesting aspect of these discourse factors is that they are mediated by register considerations. That is, textual factors are most influential when they operate *counter* to the overall register norm. Figure 8 presents these patterns for conversation and news reportage.

Because conversation has a strong register norm favoring the omission of *that*, the discourse factors favoring omission have little influence in that register. In contrast, the discourse factors favoring *that* retention are very powerful in conversation (resulting in departures from the overall register norm of 55%–80%):

- The use of coordinated *that*-clauses:
(*That* is 80% more likely to be retained than the overall register norm)

For example:
(3) Cos every time they use it, she reminds them *that it's her television* <and> *that she could have sold it.*
(4) I'm sure *they think I'm crazy* <and> *that I'm in love with him or something.*

- A passive voice verb in the matrix clause:
(*That* is 55% more likely to be retained than the overall register norm)

For example:
(5) I <was told> *that Pete was pissed.*
(6) About two weeks after that it <was diagnosed> *that she had cancer of the ovary.*

- The presence of an intervening noun phrase between the matrix clause verb and the *that*-clause:
(*That* is 80% more likely to be retained than the overall register norm)

For example:
(7) Then I <told> him *that I'm not doing it anymore.*
(8) I was busy trying to <convince> him *that he had to go to the doctor.*
(9) I <promised> her *that I wouldn't play it.*

News reportage shows the opposite tendencies: the overall register norm favors *that* retention and thus the factors favoring retention have comparatively little influence. In contrast, the factors favoring *that* omission are relatively influential in news (resulting in departures from the overall register norm of 20–40%). The following sentences from news reportage illustrate the most common main verbs, together with co-referential subjects, co-occurring with *that*-omission:

(10) After a month she said (0) *she couldn't cope with it.*
(11) He thought (0) *he was being attacked.*

The patterns of use described in this section show that register interacts in important ways with other discourse factors. In fact, these findings suggest that register is actually the more basic factor. That is, for at least some linguistic features, register sets the overall norm of use, and discourse factors are relatively uninfluential except when they run counter to the overall register overall norm.

4. Summary and conclusion

This article has illustrated several ways in which a register perspective is important for analyses of grammar and discourse. First, grammatical features are used to differing extents in different registers, depending on the extent to which the typical discourse functions of the feature fit the typical communicative characteristics of the register. However, there are also much more complex patterns of association, with textual factors interacting with register patterns in intricate ways. Although patterns such as those described here must be interpreted much more fully, the present article has illustrated the systematicity and importance of register patterns in describing the use of related grammatical features.

In retrospect, this interaction between register, grammar, and discourse is not surprising. Functional grammar is based on the premise that linguistic variability is communicatively functional, so that studying the use of variants in actual discourse contexts allows us to understand why particular variants are chosen at particular points in a text. Research on register variation has shown that register differences are mediated by similar functional considerations operating at a more global level. Given that communicative function is central to both kinds of linguistic variability (functional grammar and register variation), it obviously makes sense to combine the two perspectives.

Methodological difficulties are one of the reasons that these two approaches have not been combined more often in previous research. It can be difficult to assemble a principled text database that allows investigation of discourse factors across registers. Further, it is often difficult and time consuming to analyze complex patterns of use across multiple registers. The analyses presented here show how a corpus-based approach can be used for such investigations, providing an important complementary perspective to detailed analyses of linguistic features in individual texts.

REFERENCES

Aarts, B. and Meyer, C., eds (1995) *The Verb in Contemporary English: Theory and Description.* Cambridge: Cambridge University Press.

Aijmer, K. and Altenberg, B., eds (1991) *English Corpus Linguistics.* London: Longman.

Biber, D. (1988) *Variation Across Speech and Writing.* Cambridge: Cambridge University Press.

Biber, D. (1992) 'On the Complexity of Discourse Complexity: A Multidimensional Analysis', *Discourse Processes* 15: 133–63.

Biber, D. (1995) *Dimensions of Register Variation: A Cross-Linguistic Perspective.* Cambridge: Cambridge University Press.

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. (in press) *Longman Grammar of Spoken and Written English.* London: Longman.

Brown, G. and Yule, G. (1983) *Discourse Analysis.* Cambridge: Cambridge University Press.

Chafe, W.L. (1982) 'Integration and Involvement in Speaking, Writing, and Oral Literature', in D. Tannen (ed.) *Spoken and Written Language: Exploring Orality and Literacy.* Norwood, NJ: Ablex.

Collins, P. (1991) *Cleft and Pseudo-Cleft Constructions in English.* London: Routledge.

Collins, P. (1995) 'The Indirect Object Construction in English: An Informational Approach', *Linguistics* 33: 35–49.

De Haan, P. (1989) *Postmodifying Clauses in the English Noun Phrase: A Corpus-Based Study.* Amsterdam: Rodopi.

Fox, B.A. and Thompson, S.A. (1990) 'A Discourse Explanation of the Grammar of Relative Clauses in English Conversation', *Language* 66: 297–316.

Geisler, C. (1995) *Relative Infinitives in English.* Uppsala: Uppsala University.

Granger, S. (1983) *The Be+Past Participle Construction in Spoken English with Special Emphasis on the Passive.* Amsterdam: Elsevier Science.

Johansson, C. (1995) *The Relativizers Whose and Of Which in Present-Day English: Description and Theory.* Uppsala: Uppsala University.

Johansson, S. and Stenström, A.-B. eds (1991) *English Computer Corpora: Selected Papers and Research Guide.* Berlin: Mouton.

Kay, P. (1977) 'Language Evolution and Speech Style', in B.G. Blount and M. Sanches (eds) *Sociocultural Dimensions of Language Change.* New York: Academic Press.

Kroll, B. (1977) 'Ways Communicators Encode Propositions in Spoken and Written English: A Look at Subordination and Coordination', in E. Ochs Keenan and T. Bennett (eds) *Discourse Across Time and Space (SCOPIL #5).* Los Angeles: University of Southern California Press.

Mair, C. (1990) *Infinitival Complement Clauses in English.* New York: Cambridge University Press.

Meyer, C. (1992) *Apposition in Contemporary English.* Cambridge: Cambridge University Press.

Myhill, J. (1995) 'Change and Continuity in the Functions of the American English Modals', *Linguistics* 33: 157–211.

Myhill, J. (1997) '*Should* and *Ought*: The Rise of Individually Oriented Modality in American English', *English Language and Linguistics* 1: 3–23.

O'Donnell, R.C. (1974) 'Syntactic Differences between Speech and Writing', *American Speech* 49: 102–10.

Prince, E.F. (1978) 'A Comparison of *Wh*-clefts and *It*-clefts in Discourse', *Language* 54: 883–906.

Schiffrin, D. (1981) 'Tense Variation in Narrative', *Language* 57: 45–62.

Schiffrin, D. (1985a) 'Multiple Constraints on Discourse Options: A Quantitative Analysis of Causal Sequences', *Discourse Processes* 8: 281–303.

Schiffrin, D. (1985b) 'Conversational Coherence: The Role of *Well*', *Language* 61: 640–67.

Schiffrin, D. (1987) *Discourse Markers*. Cambridge: Cambridge University Press.

Sinclair, J. (1991) *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.

Thompson, S. A. (1983) 'Grammar and Discourse: The English Detached Participial Clause', in F. Klein-Andreu (ed.) *Discourse Perspectives on Syntax*. New York: Academic Press.

Thompson, S.A. (1985) 'Grammar and Written Discourse: Initial vs. Final Purpose Clauses in English', *Text* 5: 55–84.

Thompson, S.A. and Mulac, A. (1991a) 'The Discourse Conditions for the Use of the Complementizer *That* in Conversational English', *Journal of Pragmatics* 15: 237–51.

Thompson, S.A. and Mulac, A. (1991b) 'A Quantitative Perspective on the Grammaticization of Epistemic Parentheticals in English', in E. C. Traugott and B. Heine (eds) *Approaches to Grammaticalization: Volume II*. Amsterdam: John Benjamins.

Tottie, G. (1991) *Negation in English Speech and Writing: A Study in Variation*. San Diego, CA: Academic Press.

Varantola, K. (1984) *On Noun Phrase Structures in Engineering English*. Turku: University of Turku Press.

Ward, G.L. (1990) 'The Discourse Functions of VP preposing', *Language* 66: 742–63.

Wells, R. (1960) 'Nominal and Verbal Style', in T. A. Sebeok (ed.) *Style in Language*. Cambridge, MA: MIT Press.

DOUGLAS BIBER holds the rank of Professor in the Applied Linguistics Program, English Department, at Northern Arizona University. His research interests include register variation (in English and cross-linguistically), English grammar and discourse, and the application of computational and corpus-based techniques to issues in linguistics. ADDRESS: Applied Linguistics Program, Department of English, Northern Arizona University, Flagstaff, AZ 86011-6032 USA. [email: Douglas.Biber@nau.edu]