

DOUGLAS BIBER/BETHANY GRAY

Is conversation more grammatically complex than academic writing?

Abstract

Conversation is usually considered to be grammatically simple, while academic writing is often claimed to be structurally complex, associated primarily with a greater use of dependent clauses. Our goal in the present paper is to challenge these stereotypes, based on the results of large-scale corpus investigations. We argue that both conversation and professional academic writing are grammatically complex but that their complexities are dramatically different. Surprisingly, the traditional view that complexity is realized through extensive clausal embedding leads to the conclusion that conversation is more complex than academic writing. In contrast, written academic discourse is actually much more ‘compressed’ than elaborated, and the complexities of academic writing are realized mostly as phrasal embedding rather than embedded clauses.

1. Introduction

Grammatical complexity is often linked with elaboration and clausal embedding in linguistic theory. A ‘simple’ clause has only a subject, verb, and object or complement. A ‘simple’ noun phrase has a determiner and head noun. Additions to these structures represent elaboration, resulting in ‘complex’ grammar. In particular, there is widespread agreement that embedded clauses are an important type of grammatical complexity (often contrasted with ‘simple’ clauses; see e.g., Huddleston 1984: 378, Willis 2003: 192, Purpura 2004: 91, Carter / McCarthy 2006: 489).

Conversation has long been described as grammatically simple in these terms. Conversational participants share time and place, and they normally also share extensive personal background knowledge. As a result, pronouns and vague expressions are common, and referring expressions generally do not need to be elaborated in conversation. Because of these factors, conversational grammar is assumed to be generally not complex, employing “simple and short clauses, with little elaborate embedding” (Hughes 1996: 33).

In contrast, academic writing is claimed to be structurally complex, shown by longer sentences, longer ‘t-units’ (a main clause plus all associated dependent clauses), “longer and more complex clauses with embedded phrases and clauses” (Hughes 1996: 34), and generally a greater use of subordinate clauses (see, e.g., O’Donnell / Griffin / Norris 1967, O’Donnell 1974, Kroll 1977, Chafe 1982, Brown / Yule 1983).

These stereotypical portrayals of conversation and academic writing reflect the most salient characteristics of both. For example, some of the most noticeable characteristics of conversation are the hesitations, false starts, and short non-clausal utterances, because none of these features are normally appropriate in formal writing. The following conversational excerpt illustrates these characteristics:

Text Excerpt 1: Conversation

Non-clausal utterances are marked in **bold**

- Barry: *I went to the Institute of Terror.*
- Wendy: *You went to where?*
- Barry: ***The Institute of Terror.***
- [...]
- Wendy: ***Oh.***
- Barry: *It's pretty cool. You want to go? I've got free tickets.*
- Wendy: *Is it – it's a – how long is it going to be open?*
- Barry: ***Until the thirty first.***
- Wendy: ***Cool.*** *It's, it's an, it's actually pretty scary and stuff?*
- Barry: *I wouldn't go so far as to say it's really scary*
- Wendy: *But it's cool.* [laugh]
- Barry: ***Yeah.*** [...] *I'll go with you. I wouldn't pay for you or anything*
- Wendy: [laugh]
- Barry: *But I'll go with you.*
- Wendy: *It's expensive, isn't it? It's like five bucks.*
- Barry: ***Yeah,*** *this one's six.*
- Wendy: ***The one down here?*** *And you have free tickets?*
- Barry: ***Well, yeah.*** [...]
- Wendy: ***Wow. Cool.***

This conversation additionally illustrates the reliance on short, simple clauses, such as *where's that?*, *it's pretty cool*, *I've got free tickets*, *I'll go with you*, and *It's like five bucks*. If most conversation included only the grammatical features illustrated in Text Excerpt 1, we would be justified in making the generalization that conversation was generally not grammatically complex (as measured by the traditional criteria).

In contrast, one of the most noticeable characteristics of academic writing is that sentences tend to be long, and readers usually attribute that fact to the presence of numerous embedded clauses. Text Excerpt 2 illustrates this style of discourse:

Text Excerpt 2: Academic writing: Philosophy textbook

Embedded clauses marked with []

[Even if propositional attitude accounts succeeded in their own terms], they would not explain most of [what should be explained by a theory of emotion]. Propositional attitude theories are often presented [as if they were a simple consequence of the idea [that emotions involve the occurrence of mental states [which represent states of affairs in the world (states with "content")]]].

[...]

[What is distinctive about the propositional attitude theory] is the interpretation [it gives to the words thought and belief]. The mainstream philosophical tradition [in which Lyons is located] assumes [that our everyday understanding of these notions is adequate for a theory of emotion].

Here again, if most written academic texts incorporated this same dense use of embedded clauses, we would be justified in making the generalization that academic writing was highly complex as measured by that criterion.

However, consideration of a single text excerpt from a register does not provide an adequate basis for such conclusions. Rather, this is exactly the kind of research question that corpus-based research can contribute to (see, e.g., Biber/Conrad/Reppen 1998, McEnery/Tono/Xiao 2006). By basing analyses on large, representative collections of texts, it is possible to discover patterns of use that are generalizable to a register, rather than more specific patterns that characterize only particular texts. Further, corpus-based methods usually entail quantitative analysis, permitting description of the extent to which a

linguistic pattern is typical of a register. Both of these analytical characteristics are important here. First, corpus research shows that the general ‘complexity’ characteristics of conversation and academic writing are quite different from those that are especially salient in individual texts. And second, corpus research shows that both conversation and academic writing use grammatical complexity features to some extent; the major difference between them is in the quantitative extent to which they rely on different sets of features.

Our goal in the present paper is to challenge the stereotypes described above, based on the results of large-scale corpus investigations. We argue that both conversation and professional academic writing are grammatically complex – but their complexities are dramatically different. Surprisingly, if we adopt the traditional view that complexity is realized through extensive clausal embedding, the evidence presented below would lead us to conclude that conversation is more complex than academic writing. In contrast, written academic discourse is actually much more ‘compressed’ than elaborated, and the complexities of academic writing are realized mostly as phrasal embedding rather than embedded clauses.

The following sections present the results of large-scale corpus analyses that document these patterns of use. Section 2 introduces the corpora and linguistic features used for the analyses. Then, the analyses themselves are discussed in Section 3, which surveys the synchronic patterns of use for features associated with structural elaboration versus compression. In conclusion, we briefly discuss functional motivations for these patterns of use.

2. Corpus and grammatical features used for the analysis

We employ corpus-based analysis to describe the typical discourse styles of conversation and academic writing, investigating the extent to which both registers employ grammatical devices associated with structural elaboration. Previous corpus-based studies have documented the different complexities of spoken and written registers. For example, multi-dimensional studies of register variation (e.g., Biber 1988, 1992, 2006) have shown repeatedly that certain dependent clause types (e.g., *because*-clauses and *WH*-clauses) are more strongly associated with speech than writing. The *Longman Grammar of Spo-*

ken and Written English (Biber et al. 1999) provides more detailed descriptions of the grammatical features that are common in conversation versus those that are common in academic writing.

Building on this previous research, the present study focuses on the grammatical devices in English that are associated with structural elaboration. The descriptions below contrast the patterns of use in conversation to those in professional academic writing, based on analysis of a large corpus of texts for each of these two registers.

The conversation subcorpus is taken from the *Longman Spoken and Written Corpus* (see Biber et al. 1999: 24-35). The subcorpus includes 723 text files and c. 4.2 million words of American English conversation. These are conversations collected by participants who agreed to carry tape recorders for a two-week period. The corpus thus represents one of the largest collections of natural face-to-face conversations available.

We constructed a corpus of academic research articles (c. 3 million words), sampled from four general disciplines: science / medicine, education, social science (psychology), and humanities (history). We collected texts from three 20-year intervals (1965, 1985, 2005) to enable the description of short-term historical change. However, for the purposes of the present study, we consider these as a single group (429 texts, c. 2.9 million words), contrasted with conversation.

The corpora were grammatically annotated ('tagged') using software developed for the *Longman Grammar of Spoken and Written English* and earlier corpus studies of register variation (e.g., Biber 1995). Then, more specialized computer programs were developed for detailed linguistic analyses of specific types of structural elaboration.

Table 1 lists the types of dependent clauses that we considered for our analysis of structural elaboration. These dependent clauses can serve three major syntactic functions: complement clauses, which usually function as the direct object of a verb; adverbial clauses, which modify the main verb; and post-nominal relative clauses, which modify a head noun. In addition, dependent clauses can be finite (with tense overtly marked) or non-finite.

Grammatical feature	Examples
Finite complement clauses	<i>I wonder how he is today.</i> <i>I thought that was just too funny.</i>
Non-finite complement clauses	<i>We'd love to come.</i> <i>They talk about building more.</i>
Finite adverbial clauses	<i>She won't narc on me, because she prides herself on being a gangster.</i> <i>You can have it if you want.</i>
Finite relative clauses	<i>A method that would satisfy the above conditions...</i> <i>a repressor substance which prevents the initiation...</i>
Non-finite relative clauses	<i>the assumptions given above ...</i> <i>initiatives involving local authorities ...</i>

Table 1: Selected grammatical features associated with structural elaboration

We also considered grammatical devices that result in a ‘compressed’ rather than ‘elaborated’ discourse style, illustrated in Table 2. These are all phrases rather than dependent clauses, used to modify a head noun. Attributive adjectives and pre-modifying nouns occur before the head noun (‘pre-modifiers’), while prepositional phrases occur after the head noun (‘post-modifiers’).

Grammatical feature	Examples
Attributive adjective (adjective as noun pre-modifier)	<i>a large number, unusual circumstances</i>
Noun as noun pre-modifier	<i>human actions, membrane structure</i>
Prepositional phrase as noun post-modifier	<i>the scores for male and female target students in the class</i> <i>the mechanism for penetration of protein through the ovariole wall</i>

Table 2: Selected grammatical features associated with structural compression

Most of these features could be identified accurately using automatic computer programs. However, prepositional phrases required hand coding to determine when the phrase was functioning as a noun modifier versus adverbial. This analysis was based on a sub-sample of tokens (every fourth occurrence) from a sub-sample of the corpus (48 conversations and 41 academic research articles). The counts for all linguistic features were converted to a 'normed' rate of occurrence (per 1 000 words) for each text (see Biber / Conrad / Reppen 1998: 263-264).

3. Structural elaboration and compression in conversation versus academic writing

As noted above, researchers have usually focused on dependent clauses (or subordinate clauses) as the primary measure of grammatical complexity or structural elaboration. What they have less often noticed is that there is extensive clausal embedding in conversation. In particular, complement clauses (also called 'nominal clauses') are very common, especially *that*-clauses and *WH*-clauses. Complement clauses normally fill a direct object slot, making it possible for a relatively short utterance to have multiple levels of embedding. For example, the following short utterance has two embedded complement clauses:

You know [you could get [what you wanted]]

Unlike adverbial clauses and relative clauses, complement clauses are not optional structures; rather, they take the place of a required noun phrase. In conversation, the complement clause usually occurs with a transitive verb (e.g., *think*, *know*, or *want*): the complement clause substitutes for the noun phrase as the direct object of the verb. As a result, these structures can contain multiple levels of structural embedding. For example, the following relatively short sentence from conversation has four embedded complement clauses, each occurring as the object of the preceding main verb:

But I don't think [we would want [to have it [sound like [it's coming from us]]]]].

Adverbial clauses are optional rather than obligatory clause elements. However, these clause types are also commonly found in conversation, as in:

*She married him [because Clinton's father died before Clinton was born]
[If anybody wakes me up early] they die*

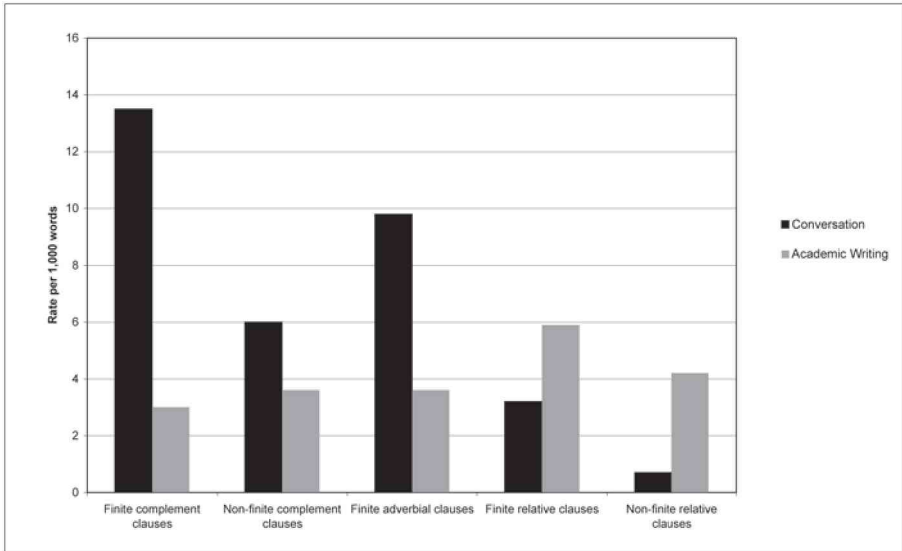


Figure 1: Common dependent clause types in conversation vs. academic writing

Our corpus investigations show that the structures illustrated above represent strong general differences between academic writing and conversation: that they are much more frequent in conversation than in academic writing. Thus, Figure 1 shows that both complement clauses and adverbial clauses are much more frequent in conversation than in academic writing. These differences are strongest for finite clauses (e.g., *that*-clauses and *WH*-clauses functioning as complement clauses; *because*-clauses and *if*-clauses functioning as adverbial clauses). However, the same general pattern holds for non-finite complement clauses (*to*-clauses and *ing*-clauses). In contrast, relative clauses are more frequent in academic writing than in conversation (especially non-finite relative clauses, such as *the concept of society proposed here*).

Text Excerpt 3 illustrates the pervasive use of embedded clauses in conversation. Unlike Text Excerpt 1 above, this conversational excerpt shows how certain kinds of dependent clauses can occur with extreme density in normal conversational interactions. For the most part, these kinds of structural elaboration do not feel complex, and they certainly do not inhibit normal communication. However, they are clearly ‘complex’ according to the definition of embedded clauses added on to simple clauses. Overall, Figure 1 shows that there are around twice as many dependent clauses in conversation as in academic writing. Thus, if we limited our comparison to these features, we would be forced to conclude that conversation is more complex than academic writing.

Text Excerpt 3: Conversation

Dependent clauses are marked in **bold**

Gayle: *And Dorothy said **Bob's getting terrible with, with the smoking.** Uh, he's really getting defiant about it **because there are so many restaurants where you can't smoke** and he just gets really mad and won't go to them.*

[...]

Peter: *Well they, they had a party. I forget **what it was.** They had it at a friend's house. I can't remember **why it wasn't at their house** any way. And they had bought a bottle of Bailey's **because they knew I liked Bailey's.***

[...]

Gayle: *I can't remember **who it was.** One of us kids.*

[...]

Peter: *Oh. I'll tell you **I think the biggest change in me is since I had my heart surgery.***

Gayle: *Really? Yeah I guess my, I mean **I know my surgery was a good thing but***

Peter: *<?> It makes you **think.** You realize **it can happen to you.***

The obvious question at this point is to ask why academic research writing seems grammatically complex. That is, given that dependent clauses are generally more frequent in conversation than in writing, we need to account for the perception that academic texts are hard to process. Part of this perception is caused by difficult subject matter and complex vocabulary. However, there are also grammatical features that make a major contribution to this complexity. In particular, the structural elaboration of academic writing is realized mostly as phrases without verbs. For example, consider the following sentence from a Biology research article:

The knowledge of tissue distribution of each novel molecular species is the first step toward the understanding of its possible function.

This sentence consists of only a single main clause, with the main verb *is*. There are no dependent clauses in this sentence. The sentence is relatively long because there are multiple prepositional phrases:

*of tissue distribution
of each novel molecular species
toward the understanding
of its possible function*

In addition, many of the noun phrases include extra nouns or adjectives as pre-modifiers before the head noun:

tissue distribution
novel molecular species
possible function

In their main clause syntax, sentences from academic writing tend to be very simple. Thus, consider the following sentence from a Psychology research article:

*This may indeed be **part** [**of** the reason [**for** the statistical link [**between** schizophrenia and membership [**in** the lower socioeconomic classes]]]]].*

Similar to the example from biology above, the clausal syntactic structure of this sentence is extremely simple, with only one main verb phrase:

X may be Y (*This may be part*)

All of the elaboration here results from prepositional phrases added on to noun phrases. Thus, unlike conversation, academic writing does **not** frequently employ dependent clauses for structural elaboration. Rather, we find a more 'compressed' style, employing embedded phrases rather than fuller dependent clauses.

As Figure 2 shows, academic writing relies heavily on non-clausal phrases instead of dependent clauses to add information. Most of these phrases occur embedded in noun phrases. Many of these structures are adjectives modifying a head noun (e.g., *theoretical orientation*) or nouns pre-modifying a head noun (e.g., *system perspective*). But the most striking difference from conversation is for the use of prepositional phrases as noun post-modifiers. Many of these are *of*-phrases (e.g., *an interpretation **of the general form of mitochondria***), but other prepositions are also commonly used for this function (e.g., *the complex relations **between three components**; understanding rational approach **to politics***). Prepositional phrases used as adverbials (e.g., ***From the systems perspective**, these stages are marked by...*) are also more common in academic writing than in conversation, but the difference is much less strong.

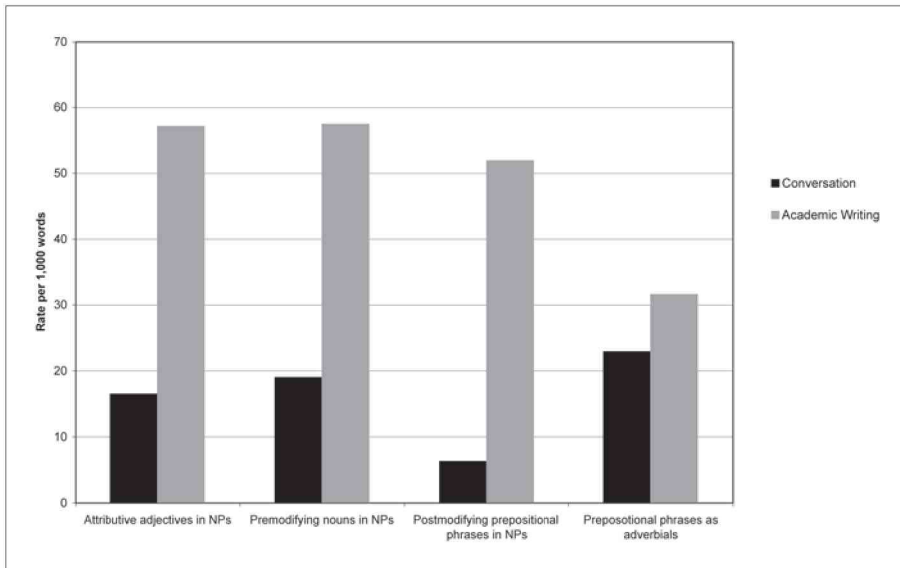


Figure 2: Common dependent phrasal types in conversation vs academic writing

It is not the case that there are no dependent clauses in academic writing. Rather, as Figure 1 above shows, dependent clauses are relatively frequent in academic writing, especially noun modifiers and non-finite clauses. Text Excerpt 4 illustrates these patterns.

Text Excerpt 4: Academic research article

Main Verbs are underlined; main verbs in dependent clauses marked in **bold**

*A number of important themes have emerged from previous research **exploring** the links among gender, interaction, and collaborative learning. First, in mixed-gender interactions boys tend to **dominate** apparatus, teacher attention, and peer discussion within the classroom (for a review, see Howe 1997). Second, the comparative context and, in particular, task are important in **determining** how children engage in collaborative interaction (Holmes-Lonergan 2003). Third, a child's gender and that of his or her conversation partner appear to **affect** the dynamics of conversation but not necessarily the answer that children agree on (Leman 2002): Specifically, boys tend to **show** greater resistance to girls' arguments, but ultimately all children tend to **opt** for the developmentally more advanced answer after interaction, regardless of whether a girl or a boy has **put** these arguments forward.*

A further consideration for studies of gender and interaction is the influence of both speakers and partners gender on conversation. Leaper (1991) examined both these types of gender effect on communication between pairs of children in two different age groups (5 and 7 years) who were engaged in play with a puppet.

However, this same text excerpt also illustrates the more important grammatical pattern that sharply distinguishes between academic writing and conversational discourse: the heavy reliance on phrasal rather than clausal elaboration. Thus, Text Excerpt 4 is repeated as Text Excerpt 5 below, highlighting the complex noun phrases with phrasal rather than clausal modifiers. A quick glance at this excerpt shows that the majority of this text is composed of such structures.

Text Excerpt 5: Academic research article [repeated from Text 4]

Complex noun phrases with no clausal embedding are marked in **bold**

*A number of important themes have emerged from **previous research** exploring the links among gender, interaction, and collaborative learning. First, in **mixed-gender interactions** boys tend to dominate **apparatus, teacher attention, and peer discussion within the classroom** (for a review, see Howe, 1997). Second, **the comparative context and, in particular, task** are important in determining how children engage in **collaborative interaction** (Holmes-Lonergan, 2003). Third, **a child's gender and that of his or her conversation partner** appear to affect **the dynamics of conversation but not necessarily the answer** that children agree on (Leman, 2002): Specifically, boys tend to show **greater resistance to girls' arguments**, but ultimately all children tend to opt for **the developmentally more advanced answer after interaction**, regardless of whether a girl or a boy has put these arguments forward.*

A further consideration for studies of gender and interaction is the influence of both speakers and partners gender on conversation. Leaper (1991) examined both these types of gender effect on communication between pairs of children in two different age groups (5 and 7 years) who were engaged in play with a puppet.

In fact, relatively few noun phrases in Text Excerpt 5 are simple noun phrases. The majority of noun phrases contain some manner of phrasal complexity. Furthermore, noun phrases in academic prose often have multiple levels of phrasal embedding within a single noun phrase. For example, the following complex noun phrase contains two prepositional phrases functioning as noun postmodifiers (head noun of phrase in bold, prepositional postmodifiers bracketed):

the influence [of both speakers and partners gender] [on conversation]

Thus, despite stereotypical beliefs about the complexity of academic writing stemming from subordinate clauses, it appears that one of the more distinctive complex structures of academic prose have been largely overlooked: phrasal modification.

4. Conclusion

In summary, the stereotype that writing is more elaborated than speech is not supported by corpus evidence. In fact, using traditional measures of elaboration – considering the use of dependent clauses – we would conclude that the opposite was the case: that conversation is more complex and elaborated than academic writing. However, that conclusion would also be an over-simplification, because it does not fully capture the characteristics of either conversation or academic writing.

However, the elaboration of conversation is very restricted in nature. As noted above, most of the dependent clauses in conversation are integrated into the clause structure: complement clauses normally fill an object slot controlled by a transitive verb. As such, these dependent clauses are not ‘elaborating’ in the same way that adverbial clauses and relative clauses are. In addition, the structural patterns in conversation are very restricted lexically. For example, although there are over 200 different verbs that can control a *that* complement clause (e.g., *assume, ensure, feel, hear, imply, indicate, propose, realize, suggest*), only three verbs account for c. 70% of all occurrences of this clause type in conversation: *think* (35%), *say* (20%), *know* (13%) (see Biber et al. 1999: 667-670). The lexical restriction is even stronger with *to* complement clauses, where c. 50% of all occurrences are controlled by the verb *want* (see *ibid.*: 710-714). Thus, the overall frequency of dependent clauses in conversation is largely due to a few high frequency lexico-grammatical patterns.

On the other hand, the lack of elaboration in academic writing is in part an artifact of inadequate measures, rather than an accurate characterization of academic writing. That is, elaboration has normally been analyzed by considering the extent to which dependent clauses are used in a text. By that measure, we would conclude that academic writing is actually less elaborated than conversation. However, that measure misses the most important structural characteristic

of academic written discourse: the reliance on phrasal rather than clausal elaboration. Most sentences in academic prose are elaborated in the sense that they have optional *phrasal* modifiers, especially nominal pre-modifiers (adjectives or nouns) and nominal postmodifiers (e.g., prepositional phrases).

These phrasal modifiers are elaborating because they are optional, providing extra information. At the same time, though, these structures are condensed or compressed: the opposite of elaborated. That is, phrasal modifiers are alternatives to fuller, elaborated expressions that use clausal modifiers (e.g., *the effect of gender* can be paraphrased with a relative clause, as in *the effect which is caused by gender*).

There are good reasons why compressed, phrasal expressions are preferred over elaborated clausal expressions in academic writing: they are more economical; they allow for faster, more efficient reading; and they are equally comprehensible to the expert reader despite the fact that some explicit meaning is lost when fuller clauses are reduced to phrasal structures. In contrast, conversation relies on a relatively small set of very productive verbs controlling complement clauses to convey information, with much less reliance on complex noun phrases.

Thus, academic writing is dramatically different from speech but not in the ways that conform to the stereotypes of complexity created through the use of embedded dependent clauses. Rather, academic writing has developed a unique style, characterized especially by the reliance on nominal/ phrasal rather than clausal structures. Consequently, perhaps the question should not be which register is more or less complex, but instead, in what respects are conversation and academic writing each complex in their own distinctive ways?

References

- Biber, Douglas (1988): *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, Douglas (1992): On the complexity of discourse complexity: A multidimensional analysis. In: *Discourse Processes* 15: 133-163.
- Biber, Douglas (1995): *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Biber, Douglas (2006): *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, Douglas / Conrad, Susan / Reppen, Randi (1998): *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

- Biber, Douglas / Johansson, Stig / Leech, Geoffrey / Conrad, Susan / Finegan, Edward (1999): Longman grammar of spoken and written English. London: Longman.
- Brown, Gillian / Yule, George (1983): Discourse analysis. Cambridge: Cambridge University Press.
- Carter, Ronald / McCarthy, Michael (2006): Cambridge Grammar of English. Cambridge: Cambridge University Press.
- Chafe, Wallace (1982): Integration and involvement in speaking, writing, and oral literature. In: Spoken and written language: Exploring orality and literacy. Ed. by Deborah Tannen. Norwood, NJ: Ablex, 35-54.
- Huddleston, Rodney (1984): Introduction to the grammar of English. Cambridge: Cambridge University Press.
- Hughes, Rebecca (1996): English in speech and writing: Investigating language and literature. London: Routledge.
- Kroll, Barbara (1977): Ways communicators encode propositions in spoken and written English: A look at subordination and coordination. In: Discourse across time and space (SCOPII no. 5, eds. E.O. Keenan and T. Bennett). Los Angeles: University of Southern California, 69-108.
- McEnery, Tony / Xiao, Richard / Tono, Yukio (2006): Corpus-based language studies: An advanced resource book. London: Routledge.
- O'Donnell, Roy C. (1974): Syntactic differences between speech and writing. In: American Speech 49: 102-110.
- O'Donnell, Roy C. / Griffin, William J. / Norris, Raymond (1967): A transformational analysis of oral and written grammatical structures in the language of children in grades three, five, and seven. In: Journal of Educational Research 61: 36-39.
- Purpura, James (2004): Assessing Grammar. Cambridge: Cambridge University Press.
- Willis, David (2003): Rules, Patterns and Words: Grammar and Lexis in English Language Teaching. Cambridge: Cambridge University Press.