Journal of English Linguistics

The online version of this article can be found at:

Published by:

$SAGE

Additional services and information for *Journal of English Linguistics* can be found at:

>> Version of Record - May 20, 2013

What is This?

Being Specific about Historical Change: The Influence of Sub-Register

Douglas Biber[1] and Bethany Gray[2]

Abstract

This article argues that historical linguistic change is mediated by register differences at a highly specific level. As a result, seemingly minor differences in register can correspond to meaningful and systematic differences in the patterns of linguistic change. Two specific case studies of twentieth-century historical change are presented. The first explores variation among sub-registers of news reportage, comparing the patterns of change in magazine articles from *Time* magazine to those found in newspaper articles from the *New York Times*. This case study shows that the differing readerships and purposes of magazines versus newspapers result in different historical-linguistic patterns of use. The second case study then explores variation among sub-registers of academic research writing. This study shows how differences associated with academic discipline (science vs. social science vs. humanities) correspond to systematically different trends in historical change. Even more surprising, this study shows that science articles published in journals aimed at multidisciplinary audiences differ from articles published in journals targeted toward specialized audiences. In the conclusion, we briefly consider the theoretical issue of whether these case studies illustrate historical change within a register or change to new registers. However, the primary goal of the article is methodological, to argue for more attention to register differences in corpus-based historical research.

[1]Northern Arizona University, Flagstaff, AZ, USA
[2]Iowa State University, Ames, IA, USA

Corresponding Author:
Douglas Biber, Northern Arizona University, Box 6032, Flagstaff, AZ 86011-6032, USA.
Email: douglas.biber@nau.edu

The Role of Register in Historical Research

Over the past decade, numerous studies of linguistic variation and change in English have been carried out using corpus-based analyses. In most cases, the corpora used for such studies have been constructed reflecting two major design considerations: size and coverage of registers. General purpose corpora, like the Brown Corpus, the British National Corpus, and the Corpus of Contemporary American English, differ in size, but they are all similar in their inclusion of a range of registers.

The coverage of registers is typically given even more weight in the design of historical corpora (see http://www.helsinki.fi/varieng/CoRD/corpora/index.html). For example, the Brown family of corpora are constructed to have the exact same coverage of written registers (e.g., press reportage, press editorials, skills and hobbies, popular lore, learned, general fiction, adventure, and western fiction) across samples from different years in the twentieth century (1931, 1961, 1991). Similarly, the ARCHER corpus is constructed to represent a range of written and speech-based corpora (e.g., medical prose, science prose, newspaper reportage, fiction, personal letters, diaries, drama) in fifty-year historical periods from 1650 to 1990. The much larger Corpus of Historical American English (COHA) is structured around four major register categories (fiction, magazines, newspapers, and nonfiction books) with a continuous sample across years from 1810 to 2009 (Davies 2010–).

Many other historical corpora are designed to represent a particular register, such as the Corpus of English Dialogues (1560–1760), the Corpus of Early English Medical Writing (1375–1800), and the Corpus of English Religious Prose (1500–1699). These corpora give special attention to register, categorizing texts for specific "sub-registers."[1] So, for example, the texts in the Corpus of Early Modern English Medical Writing (EMEMT) are categorized for sub-registers such as scientific journals, general treatises or textbooks, surgical and anatomical treatises, recipe collections, and health guides (Taavitsainen & Pahta 2010). An alternative approach was used for the construction of the *TIME* Magazine Corpus, which focuses on a single register/genre but includes nearly a 100 percent sample of texts from the magazine for the period 1923–2006 (Davies 2007).

Not surprisingly, the researchers who have given the most attention to register differences in corpus construction are also the researchers who have systematically considered those differences in their linguistic analyses. So, for example, Culpeper and Kytö (2010) compare the registers of trial proceedings, witness depositions, play texts, and dialogue in prose fiction as part of their more general analysis of historical change in "spoken" English dialogues. Similarly, the chapters in Taavitsainen and Pahta (2010, 2011) document the linguistic characteristics of specific sub-registers like surgical treatises versus recipe collections in their description of historical change in English medical writing.

In many other cases, though, researchers have been interested in general historical change in English, attempting to chart developments for the language as a whole. These researchers have disregarded register differences even when they are encoded

in the corpora used for analysis. In such analyses, the inclusion of multiple registers in a corpus is considered to be a necessary requirement to achieve representation of the language in general, with the assumption that a corpus with this design enables analysis of overall historical developments in English.

A recent exchange of articles between Geoffrey Leech and Neil Millar is informative in this regard. Leech (2003), based on analysis of the Brown family of corpora (1961 and 1991), concludes that modal verbs are declining in English. Millar (2009), based on analysis of the much larger *TIME* Corpus, argues that Leech is wrong and that at least some modals are increasing in use in English. Leech (2011), based on analysis of COHA, argues that Millar is wrong and that the modal verbs really are declining in English.

The interesting consideration for our purposes here is that both Leech and Millar want to make generalizations about language change for English as a whole. They both assume that their corpora represent English, providing the basis for general conclusions about historical change in the language. Thus, Leech (2003) argues that the Brown family of corpora are "balanced" and have matched designs (equivalent samples for the same register categories) across historical periods, and thus they can be used to track general historical change in English. Millar (2009) criticizes the 2003 study, noting that the Brown family of corpora are relatively small and sampled from only two years (1961 and 1991); in contrast, the *TIME* Corpus is much larger, and sampled continuously across years from 1923 to 2006. Millar finds an increasing use of modal verbs in the *TIME* Corpus and concludes that this provides a more accurate picture of the overall trend for English. Finally, Leech (2011) counters that Millar's study is based on only a single register—*Time* magazine—and therefore the increasing use documented for that corpus does not represent the actual (declining) trend for the language as a whole. Leech addresses the possible importance of register differences, noting that Millar's analysis has limited generalizability because it represents change in only a single specific register (Leech 2011:549-550). But he does not otherwise consider register differences, arguing instead that balanced multiregister corpora can be used to represent overall change in the language as a whole.

Our point here is not to disagree with Leech's claim that modals are decreasing in use. In fact, Biber (2004:118) shows this same trend for both written and speech-based registers in the ARCHER corpus. Rather, our main goal here is to challenge the assumption that historical change should be documented for the language as a whole. We argue instead that register is crucially important as a mediating factor for historical developments, that change should be studied relative to particular registers, rather than attempting a kind of average for English.

We support this claim through consideration of historical change within the register domains of news reportage and academic prose, focusing mostly on the increasing use of "literate" linguistic features associated with structural compression. Several previous studies have focused on the opposite trend: the increasing use of colloquial linguistic forms in English written registers over the past two centuries. That is, written registers have shown an increasing use of lexical and grammatical features associated

with conversation, such as first person pronouns, contractions, and semimodals (e.g., *be going to, have to*). This trend, which has accelerated in the twentieth century, has been referred to as the "drift" of written registers toward more "oral" styles (Biber & Finegan 1989), "informalization" (Fairclough 1992), and "colloquialization" (Hundt & Mair 1999; Mair 2006; Leech et al. 2009).

Earlier investigations of these patterns suggested that they represent a general historical trend for written English as a whole. However, subsequent investigations have shown that there are important differences among written registers, in that some written registers have not participated in this historical drift. For example, Biber and Finegan (1997/2001) show that written registers like fictional novels and personal letters have been strongly influenced by the shift to more colloquial linguistic styles, but written academic registers have not participated in these changes. Hundt and Mair (1999) also note this difference, distinguishing between "agile" written registers (e.g., newspaper prose) that are receptive to these changes, and "uptight" written registers (e.g., academic prose) that resist such changes.

More recent studies show that the patterns of historical change are actually more complicated. In particular, it is not accurate to describe academic prose as conservative and resistant to change. In fact, academic research writing has been the locus of other historical linguistic changes that are at least as noteworthy. Rather than increasing in the use of "colloquial" linguistic features, academic prose has changed historically to become more "compressed," relying on noun phrase structures rather than clauses, and relying heavily on phrasal (rather than clausal) modification (see, e.g., Halliday & Martin 1993/1996; Mair 2006; Biber & Clark 2002; Biber 2009; Biber & Conrad 2009; Leech et al. 2009; Biber & Gray 2010, 2011).

Following Hundt and Mair (1999), we regard news writing and academic prose as especially interesting registers with respect to the historical trends of colloquialization and compression. News writing is interesting because it has been influenced by both historical trends, becoming both more colloquial and more compressed (Biber & Gray 2012; see further details below). Academic prose is interesting because the historical changes toward grammatical compression are more pronounced in this register than probably any other register in English (see Biber & Gray 2010, 2011). For both registers, these have been rapid historical developments, occurring especially over the past one hundred years.

These findings support the claim that register differences are crucially important for understanding grammatical change. In the present article, we argue that such differences are even more important than we had previously recognized—that even specific sub-registers differ in pervasive and systematic ways in their historical development.

## The Dilemma for Corpus Construction: Size versus Careful Text Selection

One methodological challenge for tracking grammatical change is the limited availability of historical corpora that are both large and sampled carefully with detailed

register information. For example, the ARCHER corpus (see, e.g., Biber, Finegan & Atkinson 1994; Biber & Finegan 1997/2001) was designed to carefully represent different register categories (e.g., drama, personal letters, diaries, newspaper reportage, medical research writing). However, ARCHER is limited with respect to size: the corpus is structured in terms of fifty-year periods, and each register is represented by only ten text samples per historical period. The corpus has been expanded considerably in recent years (see Yáñez-Bouza 2011), but it is still mostly designed for the analysis of long-term grammatical change across registers, rather than detailed lexical descriptions of real-time change. The Brown family of corpora has been designed with similar priorities: careful attention to the selection of individual texts representing particular registers, but relatively small size (one million words and five hundred text samples in each corpus; see Leech et al. 2009).

In contrast, more recently constructed corpora like COHA (Davies 2010–) and American Google Books (Davies 2011–) are extremely large (COHA comprises 400 million words; American Google Books comprises 155 billion words). Such corpora permit analysis of historical change on a scale not imaginable with earlier corpus designs, allowing for a nearly continuous tracking of change across time, and the analysis of historical patterns for individual words as well as grammatical features (see Davies in press-a, in press-b). However, it is not feasible to verify the specific register classification of individual texts in a corpus of this size, making these corpora somewhat less suitable for detailed investigations of registers and register variation. For example, while these large corpora specify general register categories for these texts (e.g., newspaper articles or academic writing), less attention is provided concerning the specific sub-registers of texts within these relatively broad categories.

The challenge for corpus-based historical analyses is to balance both considerations. But in the past, the difficulty in obtaining suitable historical corpora for detailed analyses has sometimes resulted in historical comparisons that are not well controlled for register. Our primary goal in the present article is to show how this practice can lead to inaccurate conclusions; in fact, even minor register differences can sometimes confound historical comparisons.

The problem can be illustrated with an unpublished historical study that we carried out several years ago. The goal of that study was to track historical change in written news reportage, with a particular focus on the patterns of change in the twentieth century. We began our investigation with the newspaper sub-corpus from ARCHER, which included ten news reportage text samples from British English newspapers published in the first half of each century and twenty text samples published in the second half of each century (ten British English and ten American English). In previous studies, the news sub-corpus from ARCHER had been used to track major grammatical developments in news reportage over the past three centuries (see, e.g., Biber & Clark 2002; Biber 2003), but we concluded that this sampling was not adequate to provide the basis for detailed investigations of change in the twentieth century. For this reason, we supplemented the twentieth-century sub-corpus from ARCHER with a relatively large sample of twentieth-century news reportage articles from *Time* magazine. The
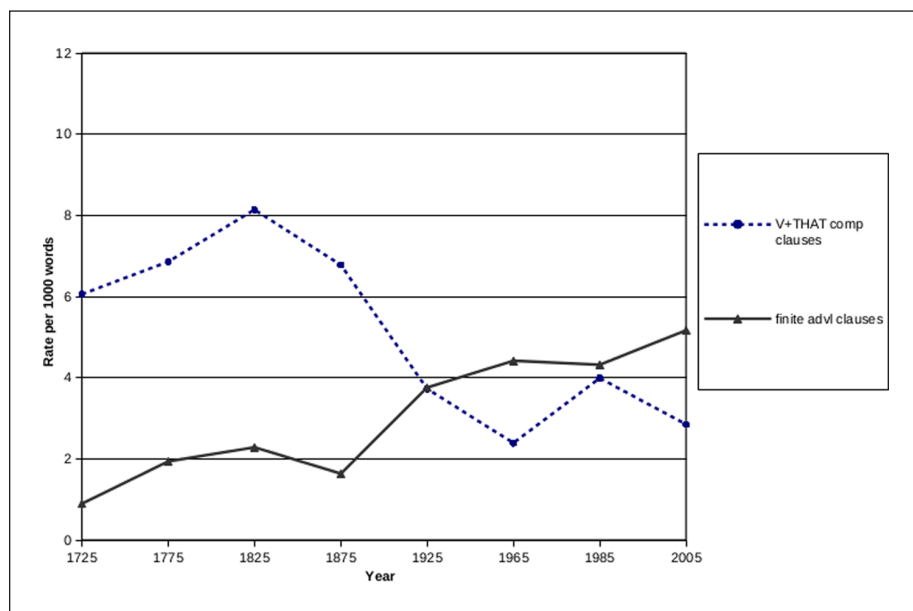
Figure 1. Finite dependent clauses in written news reportage (ARCHER and *TIME* combined).

combined corpus included two hundred articles published in five specific years: 1925, 1945, 1965, 1985, and 2005 (i.e., one thousand articles total). We checked these articles by hand to include only news reportage texts (i.e., excluding editorials, essays, etc.). Our expectation was that we could combine the newspaper texts from ARCHER with the news reports from *Time* to provide a more detailed account of historical change in the twentieth-century news reportage (while also using the eighteenth- and nineteenth-century samples from ARCHER to see the longer-term trends in newspaper reportage).

Figure 1 shows the kinds of findings resulting from this analysis, tracking the use of *that* complement clauses controlled by verbs (as in 1) and finite adverbial clauses (as in 2):

(1) Admiral Alexeieff warned the Tsar that the Japanese would forestall the Russians

(2) such an occupation would be very advantageous to Russia because it would convict Japan of a violation of all international law).

("A Chapter of War History: Alexeieff and the Tsar Interesting Exchange of Messages," *Pall Mall Gazette*, June 1, 1905)
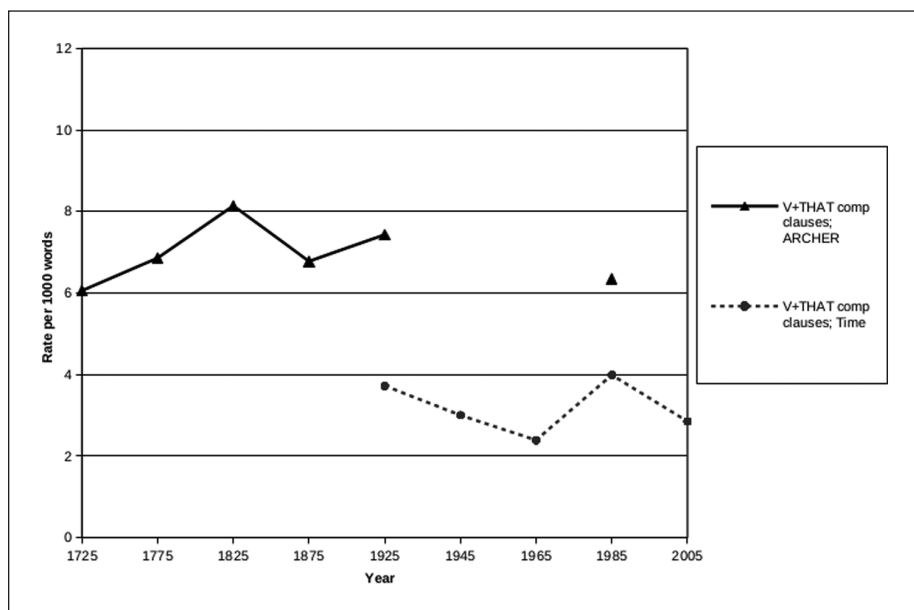
Figure 2. Finite *that* complement clauses in ARCHER (newspapers) versus *Time* magazine.

The most notable changes shown in Figure 1 occurred between 1875 and 1925, with a fairly dramatic decrease in *that*-clauses and a relatively large increase in adverbial clauses.

However, it turns out that this apparent historical change is confounded by the sub-register difference between news reports in newspapers versus news reports in *Time* magazine. Figure 2 displays the same data for *that*-clauses as Figure 1, except the ARCHER newspaper texts are plotted separately from the *Time* magazine articles. Figure 2 shows that there is a large difference even in 1985 between the relatively frequent use of *that*-clauses in newspaper articles versus the considerably less frequent use of this feature in *Time* magazine. The apparent historical decrease in the use of this feature that was seen in Figure 1 is actually due to the addition of the *Time* magazine sample, rather than representing an actual historical change in newspaper writing. That is, when the data for ARCHER and *Time* are considered separately for 1985 in Figure 2, it appears that the use of *that*-clause constructions has remained relatively constant over time. Similarly, Figure 3 shows that there has been a small and gradual increase in the use of finite adverbial clauses in both newspaper articles and *Time* magazine articles. However, the apparent large increase in use between 1875 and 1925 shown in Figure 1 actually reflects the addition of the *Time* magazine sample (which exhibits consistently higher frequencies of adverbial clauses than ARCHER newspapers), rather than a genuine historical change in news reportage generally.
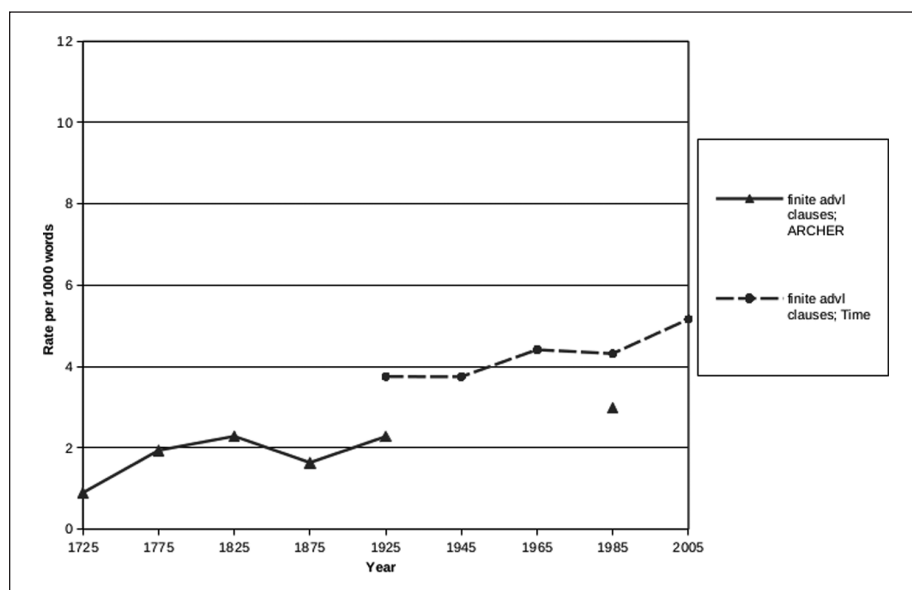
Figure 3. Finite adverbial clauses in ARCHER (newspapers) versus *Time* magazine.

We later stumbled across a similar discrepancy in our analyses of academic writing. Similar to our investigations of historical change in newspaper writing, we initiated our historical studies of academic writing with the ARCHER corpus, and then supplemented that with additional texts. When the ARCHER corpus was constructed, the sampling for scientific academic writing was considered to be one of the least problematic registers because there is a continuous record of publication for the *Philosophical Transactions of the Royal Society of London* (*PT*) from 1665 to the present day (see Atkinson 1999). Thus, ARCHER relies solely on *PT* articles to represent science writing. In some more recent studies, we supplemented the ARCHER corpus with a larger sample of twentieth-century scientific research articles from specialist journals (see, e.g., Biber & Gray 2010). Although it was not the intended goal of our investigation, we began to notice consistent differences in the linguistic styles of specialist science research articles (i.e., articles published in journals aimed at highly specific and specialized disciplines and/or subdisciplines) versus the articles published in *PT*.

Although we both specialize in the study of register variation, we were quite surprised by these findings. In the case of news, we had carefully selected news reportage articles from *Time* magazine, and our expectation was that those texts represented the same register as news reportage articles from newspapers. However, the research findings displayed in Figures 2 and 3 strongly suggest that this expectation was wrong—that instead, minor differences in audience, purpose, and perhaps editorial policy have resulted in systematic differences in the use of linguistic features. Similarly, we had simply assumed that academic science writing in *PT* was the same

register as academic science writing in specialist research journals. And here again, those assumptions were challenged by our empirical linguistic comparisons (likewise associated with differences in audience and purpose).

The present study explores such differences in more detail, investigating the extent to which historical change is mediated by specific sub-register patterns. We base our analyses on relatively large corpora of sub-registers within news reportage and within academic prose (described in detail below). Our goal was to construct strong tests of the hypothesis that sub-register plays a fundamentally important role for understanding historical grammatical change, by comparing sub-registers that are minimally different in their situational characteristics. As the following sections show, this hypothesis is clearly supported by the linguistic development of a range of grammatical features over the past two centuries.

Historical Development in News Reportage

Corpus and Method

Table 1. Corpus Composition for News Reportage.

Register	Historical period (number of texts)					Total Texts	Total Words
	1850	1900	1925	1985	2005		
New York Times	50	50	50	50	50	250	ca. 242,000
Time	—	—	200	200	200	600	ca. 427,000

To analyze historical change in news reportage, we constructed a corpus representing two sub-registers: newspaper articles and news magazine articles (see Table 1). Newspaper articles are taken from the *New York Times* (*NYT*). Five periods are included in this sub-corpus: 1850, 1900, 1925, 1985, and 2005. Magazine articles that report on current news are taken from *Time* magazine, which was first published in 1923. Thus, our *Time* sub-corpus includes articles from three historical periods over the past eighty years—1925, 1985, and 2005—allowing for direct comparison with *NYT* in those periods.[2]

We followed the same analytical methods for both the case study of news reportage as well as the subsequent case study of academic writing. First, all texts were "tagged" using the Biber automatic grammatical tagger (see Biber 1988; Biber et al. 1999). Additional computer programs were written to calculate rates of occurrence for linguistic features associated with colloquialization and compression (see below). Then rates of occurrence were calculated for each text in the corpus by "norming" the raw counts for each linguistic feature to a standard rate per one thousand words (see Biber, Conrad, & Reppen 1998:263-264). Both case studies employ a "Type B" design, in which each text is treated as an observation (see Biber & Jones 2008:1298-1300; Biber 2012). This design allowed us to compute mean rates of occurrence and standard deviations of linguistic features for each register in each historical period and to also use standard inferential statistics (ANOVA) to test for significant differences among those mean scores.

Table 2. Descriptive Statistics for Four Grammatical Features in *New York Times* and *Time* (rates of occurrence per 1,000 words).

Register	Year	N	Communication verb + *that*-clause		Passive verbs		Noun + *of*-phrase		Noun + noun	
			M	SD	M	SD	M	SD	M	SD
New York Times	1925	50	2.9	3.8	18.8	11.8	30.0	11.4	28.4	15.0
	1985	50	3.2	2.9	13.4	5.6	25.9	7.9	39.8	15.2
	2005	50	4.0	2.6	9.6	3.2	19.7	5.4	39.0	13.5
Time	1925	200	1.9	2.2	20.5	6.9	34.5	10.0	21.7	10.4
	1985	200	2.7	2.5	16.9	6.9	22.7	6.7	34.0	11.2
	2005	200	2.8	2.4	12.9	5.9	19.0	6.6	34.2	11.8

Table 3. Summary of the ANOVA Factorial Models for News Reportage Sub-registers.

	Model F Score	Model significance	Model R^2	Year	Sub-register	Year × sub-register
Communication verb + *that*-clause	6.85	< .0001	.044	< .0001	< .0001	*ns*
Passive verbs	37.9	< .0001	.203	< .0001	< .0001	*ns*
Noun + *of*-phrase	88.7	< .0001	.369	*ns*	< .0001	< .0001
Noun + noun	40.5	< .0001	.211	< .0001	< .0001	*ns*

Twentieth-Century Patterns of Change in News Reportage

In the present section, we consider four linguistic features that have been undergoing rapid change in use over the course of the twentieth century: direct and indirect quoted language, passive voice verbs, noun + *of*-phrase, and noun–noun sequences. The analyses presented below show that there are systematic differences between news sub-registers in the development of these features. In most cases, news reports in *NYT* are more innovative than news reports in *Time*. That is, *NYT* shifts to an innovative use of these linguistic features earlier than *Time*, and in most cases, *NYT* persists in using these features to a greater extent in later historical periods.

Table 2 gives descriptive statistics (mean and standard deviation) for each feature in each register and time period. Table 3 provides the results from a factorial ANOVA, testing the statistical significance of the differences in mean scores between sub-registers and across time periods.

Speech (and written communication) is reported through the use of two structural devices: direct quotes and indirect reports with *that*-clauses controlled by communication verbs (e.g., *said that* . . .). As Figure 4 shows, both structures have increased in use over the course of the twentieth century. The increase for *that*-clauses has been gradual, but in general, *NYT* has been more innovative than *Time* (in any given year, *NYT* uses *that*-clauses to a greater extent than *Time*). In contrast, *Time* magazine in its earliest years was more innovative in its use of direct quotes. *NYT* followed the lead in this
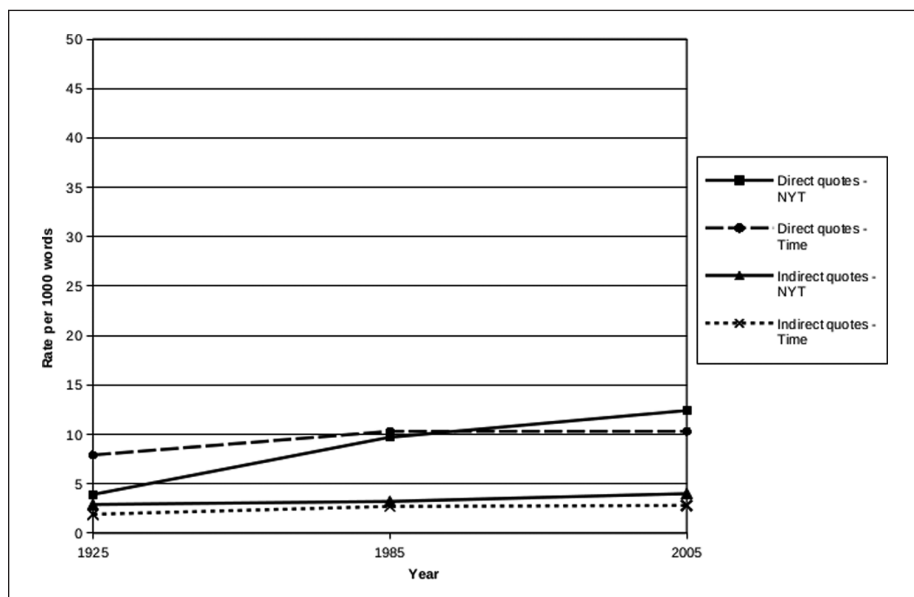
Figure 4. Direct and indirect quoted speech in *New York Times* versus *Time*.

case, matching the style of *Time* by the 1980s, and then employing direct quotes to a greater extent than *Time* by 2005.

Text Sample 1 illustrates the frequent use of both indirect reported language and direct quotes in present-day *NYT* reportage:

Text Sample 1

Eight former executives of the accounting firm KPMG and one outside lawyer have been indicted on charges stemming from their role in designing and selling questionable tax shelters. Yet so far, no court has decided that the shelters were improper.

One tax shelter promoter has asked a federal judge in San Francisco to tackle that very question: the legality of the shelters. And federal prosecutors in Manhattan appear to be worried that the judge might do that. Lawyers for the government have twice argued that the judge should wait for the criminal case to conclude.

. . . The defendants in New York are awaiting the government's response—due next week in the federal court in San Francisco—to the summary judgment motion filed in the civil case brought by the shelter promoter arguing that the shelter at issue, known as Blips (for bond-linked issue premium structure), was proper. The Internal Revenue Service has never considered Blips valid. . . .

Lawyers for the government <u>contend that</u> the structure as described does not reflect what happened in the transaction, <u>that</u> as a result there is no abstract issue to be decided and <u>that</u> the proceedings could interfere with the criminal case.

"The court cannot decide the narrow legal issue presented in the delicately crafted motion for summary judgment, until the parties can take discovery about, and present a fully developed record on, the factual dispute," government lawyers wrote last month.

Lawyers for Presidio responded: "The government makes no argument that ruling on ripe legal issues will 'interfere' with a parallel case. It seems that the only possible 'interference' to the government is that the ruling might go against the government's position." . . .

The indictment charges the nine men with conspiracy to defraud the government, and prosecutors have <u>said</u> [Ø] they intend to add charges and defendants in the next few weeks.

(Jonathan D. Glater, "Tax Shelters and KPMG: Sorting Out Legal Issues," *New York Times*, October 7, 2005)

These linguistic features are used to support and personalize the narrative line of a news story, associating particular perspectives and personal reactions with the people and institutions who expressed them. This style of reportage has increased over the course of the twentieth century, with *NYT* at present using both features to a greater extent than *Time*.

Accompanying the increased use of direct and indirect quotes, we find a decreased use of stereotypically "literate" features. Two of these features are especially noteworthy: passive voice verbs and the *of*-genitive (see also Leech et al. 2009:148-153, 224-225). The analysis of our news reportage corpus confirms the decrease in use of passive voice verbs and *of*-genitives documented by Leech et al. (2009). Figures 5 and 6 display the use of the passive voice and noun + *of*-phrases, respectively, in *Time* and *NYT* from 1850 to 2005. These two figures illustrate the frequent use of both features in the last half of the nineteenth century and beginning of the twentieth century. Thus, almost any article from 1850 *NYT* reportage illustrates the dense use of both passive voice verbs (<u>underlined</u>) and the *of*-genitive (*ITALIC CAPS*); for example:

Text Sample 2

At four o'clock Monday morning, the arrest *OF* representatives *OF* the people <u>was commenced</u>. CHANGARNIER <u>was taken</u> at his house where a number *OF* his fellow members *OF* the Assembly <u>were</u> collected. Among other *OF* the prominent members seized that morning, are LA- MORICIEKE, CAVAIGNAC, Gen. HEDEAU, who <u>is reported</u> to have made resistance and to have <u>been</u>
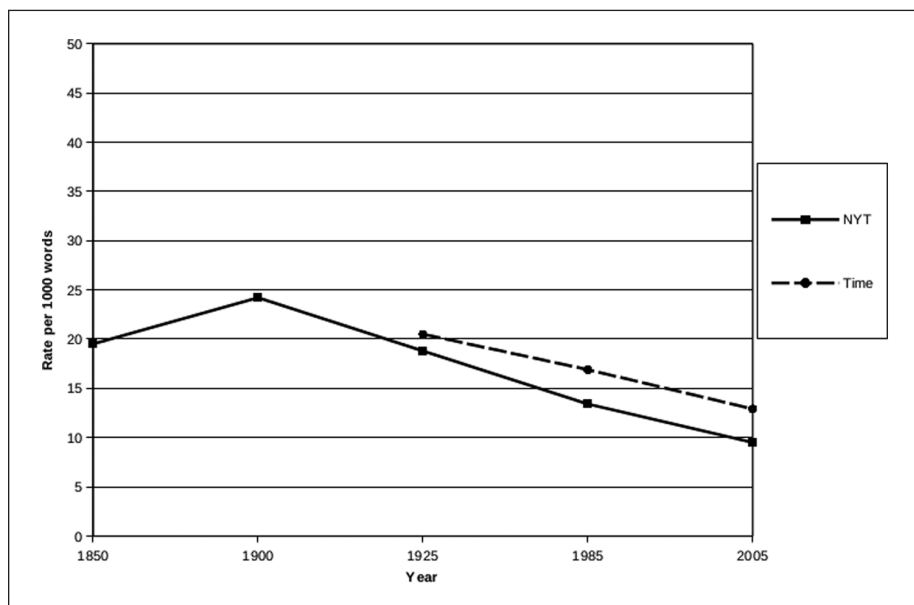
**Figure 5.** Historical use of passive voice in *New York Times* versus *Time*.

wounded in the struggle, in which one *OF* his domestics lost his life . . . and others *OF* lesser note from all parties have been apprehended.

("The Revolution at Paris," *New York Times*, December 4, 1851)

In contrast, the longer Text Sample 1 above (from *NYT* 2005) shows only one example of a finite passive voice verb (*have been indicted*) and only two instances of the *of*-genitive (*executives of the accounting firm, legality of the shelters*).

Figure 5 shows that passives were still increasing from 1850 to 1900, but that they then declined strongly in use over the twentieth century. The decline is found in both *NYT* and *Time*, but *NYT* takes the lead. Thus, in any given year, *NYT* uses fewer passive verbs than *Time*. Similarly, Figure 6 shows a strong and consistent decline in use for *of*-genitive constructions over the entire period of 1850 to 2005. In this case, *NYT* is more innovative than *Time* in the earliest period that allows a comparison (1925), but *Time* quickly adopts this change, so that the two registers are nearly identical in their infrequent use of *of*-genitive constructions by 2005.

The statistical tests reported in Table 3 confirm these two different patterns. Differences in the mean scores for passive verbs are statistically significant across historical periods and between the two sub-registers. The interaction effect is not significant, indicating that the difference between the two sub-registers remains relatively constant across historical periods. In contrast, the mean scores for *of*-genitives are
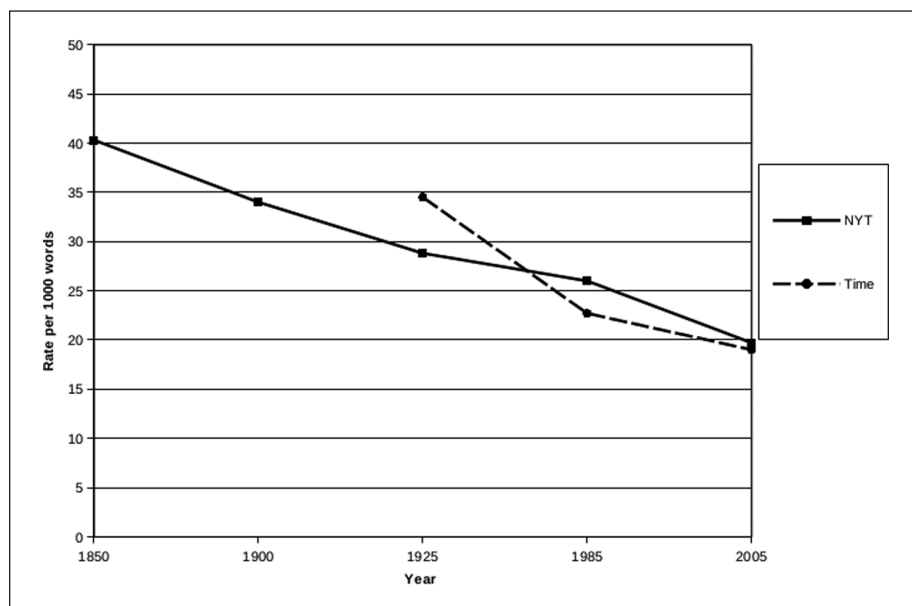
Figure 6. Historical use of noun + *of*-phrase in *New York Times* versus *Time*.

significantly different only between the two sub-registers, but the interaction effect is also significant, indicating that there is significant change over time, but the difference between the two sub-registers also changes over time (i.e., there is not a significant difference between sub-registers in the most recent period).

The three grammatical changes discussed to this point can be related to the colloquialization of news reportage during the twentieth century, resulting in an increased use of conversational features (direct and indirect quoted language) and a decreased use of stereotypically literate features (passive verbs and *of*-genitives). We have shown that sub-register distinctions are important for understanding these developments. In the use of direct quotes, *Time* magazine was innovative in its earliest periods, but otherwise *NYT* has generally taken the lead historically. Thus, although the trend toward greater colloquial styles (and decreased "literate" styles) has influenced both registers, *Time* has generally lagged behind *NYT* in the implementation of these changes.

However, we noted in the introduction that news reportage has also been influenced by a complementary historical trend in the twentieth century: the development of more "compressed" linguistic styles. Perhaps the most noteworthy grammatical change of this type is the dramatic increase in the use of nouns modifying a head noun, as in sequences like *government action, business community, school board, health insurance*, and *security measures*. The dense use of this device can be observed in the 2005 *NYT* Text Sample 1 above, including examples such as *tax shelters, summary judgment motion, bond-linked issue premium structure*, and *tax shelter promoter* (this is not a comprehensive list of all noun + noun sequences in Text Sample 1).

Figure 7. Historical use of noun + noun in *New York Times* versus *Time*.

Figure 7 shows the dramatic increase in the use of noun–noun sequences over the past 150 years. Noun–noun sequences were relatively rare in mid-nineteenth-century newspaper reportage, but they increased sharply in use at the end of that century, and that strong increase continued throughout most of the twentieth century. Table 3 shows that those differences are statistically significant for both historical period and for sub-register. Figure 7 indicates, however, that this historical development appears to have reached a saturation point at the end of the twentieth century, so that there has been little change in use over the past thirty years.

The interesting pattern for our purposes here is that once again, *NYT* has adopted these changes to a greater extent than *Time* magazine. That is, in any given year, *NYT* makes greater use of noun–noun sequences than *Time*. In summary, the analyses that we have presented here have shown that with respect to several major grammatical features, *NYT* leads *Time* in the implementation of historical developments. The following section shows that patterns of change in academic writing are likewise mediated by seemingly minor differences in sub-register.

Historical Development in Academic Research Writing

Corpus

The case study on academic research writing compares multiple sub-registers, and as a result, the academic corpus is more complicated than the news reportage corpus. As

Table 4. Corpus Composition for Academic Research Writing.

Register	Historical period (number of texts)							
	1850	1875	1900	1965	1985	2005	Total texts	Total words
Nineteenth-century research articles								
Nineteenth-century science/medicine	55	46	50	—	—	—	151	ca. 1.8 million
Nineteenth-century humanities (history books)	19	18	—	—	—	—	37	ca. 5 million
Twentieth-century research articles								
Modern specialist science	—	—	—	81	64	52	197	ca. 1.1 million
Modern specialist social science	—	—	—	58	58	57	173	ca. 2.3 million
Modern multidisciplinary science (*PT* and *Science*)	—	—	—	—	40	40	80	ca. .5 million
Modern humanities (research articles and books)	—	—	—	30	30	60	120	ca. 2.5 million

a baseline for our study, we include two sub-corpora of academic texts from the nineteenth century: science research articles and humanities books (history). However, the main focus of the analysis is on four sub-corpora of disciplinary research writing from the past fifty years, with samples from 1965, 1985, and 2005: specialist science research articles, specialist social science research articles, multidisciplinary science research articles, and history research articles (and books). A fuller description of the six sub-corpora is provided in Table 4.

The nineteenth century was a period of transition in which academic sub-disciplines and specializations were beginning to form. However, the major disciplinary distinction of the time was between humanities academic writing (e.g., in philosophy, history, and literary discussions) and science/medical prose. Thus, to provide a baseline for our analyses of sub-registers in the twentieth century, we collected two sub-corpora from nineteenth-century academic writing: one for science/medical writing and one for humanities (history) writing. We sampled science articles from three periods: 1850, 1875, and 1900. These sub-corpora came from the science/medical research articles in ARCHER, astronomy research articles from the Corpus of English Texts on Astronomy (CETA; see Crespo García & Moskowich-Spiegel Fandiño 2010), *PT*, and the journal *Science* (for 1875 and 1900).

The twentieth century witnessed incredible disciplinary diversification within the general domain of science research, including the development of disciplines in the social sciences. Most science articles from the nineteenth century still presented the results of observational research in general disciplines, such as biology, chemistry, and astronomy; these articles were read by relatively wide audiences of scientists (and often educated lay readers) from multiple disciplines. In contrast, research articles in the twentieth century usually report the results of experimental studies, written by specialists in narrow sub-disciplines, to be read by a small group of specialists from the same sub-disciplines. This diversification and subsequent shift in content, readership, and purposes have led us to question whether the development of sub-registers within academic science writing is associated with systematic differences in historical linguistic development. Thus, the twentieth-century science sub-corpora are designed to allow us to test this hypothesis.

The primary focus of the present study is register diversification in the late twentieth century (using the nineteenth-century sub-corpora as a baseline for comparison). Over the course of the twentieth century, numerous academic sub-disciplines emerged out of the more general disciplines that existed in earlier centuries, including the "social science" disciplines and highly specialized disciplines within "science." While it would be possible to analyze the patterns of variation among specific sub-disciplines (see, e.g., Gray 2011), we include only a two-way distinction here: science versus social science. For science, we collected research articles from specialist journals in biology (*Journal of Cell Biology, Biometrics, Journal of Animal Ecology*), medicine (*American Journal of Medicine*), and physiology (*Journal of Physiology*). For social science, we collected research articles from specialist journals in education (*American Educational Research Journal, Journal of Educational Measurement*) and psychology (*American Journal of Psychology, Developmental Psychology*).

Interestingly, some of the science research journals that were influential in the nineteenth century have continued to publish research articles up to the present day. We include sub-corpora for two of these journals: *PT* and the journal *Science*, for the periods 1985 and 2005.

Our initial inclination was to track the history of science writing in English by restricting the analysis to these influential research journals that have had a long continuous history. After all, *PT* has been published since 1665, and *Science* since 1880. In many ways, these journals have remained relatively constant in purpose and readership, intended as outlets for the most important science research findings from across the full range of scientific disciplines, and written for an audience from that same breadth of disciplines. By analyzing the same research journals, we (and other researchers who have previously used these journals as the basis for diachronic studies) should be able to isolate the influence of historical time.

However, that approach would only partially capture the historical evolution of science writing. In the nineteenth century, most science writing was published in multidisciplinary venues and read by a wide multidisciplinary audience. Thus, journals like *PT* and *Science* provide a fairly good representation of the universe of science writing from that period. In contrast, articles of that type represent only a very small proportion of all science writing at present. Instead, most science writing today is published in highly specialized journals associated with specific sub-disciplines, and read mostly by specialists from those same sub-disciplines. Thus, we consider multidisciplinary science writing (*PT* and *Science*) as a separate sub-register from specialist science and social science research journals.

Finally, we also include a sample of academic research writing in history (representing humanities research): research articles in history (from the *Journal of Contemporary History* and the *Journal of the History of Ideas*), published in 1965, 1985, and 2005 (supplemented by a sample of books from 2005). Humanities serves as an interesting comparison point to science and social science writing because it has been much less influenced by the trend toward experimental designs and quantitative methods.

Our central goal here was to investigate the extent to which these relatively subtle sub-register differences are influential as determinants of linguistic change. Thus, we
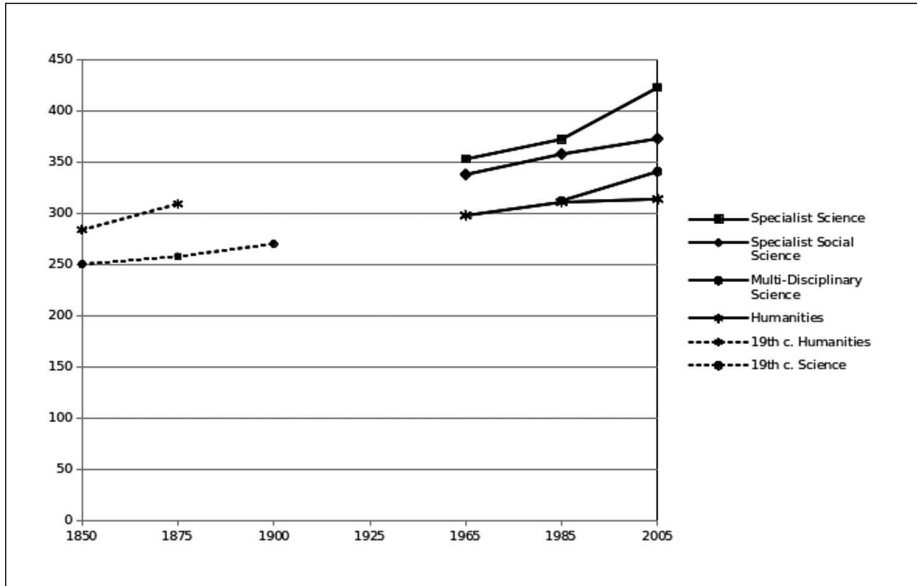
Figure 8. Nouns in academic registers.

compare the four major sub-registers described above: specialist science research articles, multidisciplinary science research articles, specialist social-science research articles, and humanities research writing (history).

Twentieth-Century Patterns of Change in Academic Research Writing

As we note earlier, academic research writing is an especially interesting register in English, because the historical developments resulting in increased grammatical compression are more pronounced in this register than in most other registers of English (see Biber & Gray 2010, 2011). These have been rapid historical developments, occurring mostly in the twentieth century. The research question that we explore in the present section is the extent to which these developments have applied generally to all sub-registers of academic writing.

The most obvious characteristic of modern academic writing is its heavy reliance on nouns. Thus, as early as 1960, Wells documented the "nominal style" of academic writing, in contrast to the "verbal style" of other varieties. Figure 8 shows that the density of common nouns has been increasing steadily in academic writing over the past 150 years. Interestingly, Figure 8 also shows that the nominal style noted by Wells in 1960 represents a density of use that is considerably less extreme than in present-day academic research writing.

Table 5 presents the actual mean scores and standard deviations for the academic sub-registers in the recent historical periods (1965, 1985, 2005). Taken together, Table 5

Table 5. Descriptive Statistics for Five Grammatical Features in Academic Sub-registers from Three Recent (1965, 1985, 2005) Historical Periods (rates of occurrence per 1,000 words).

Register	Year	N	Nouns		Nominalizations		Relative clauses	
			M	SD	M	SD	M	SD
Specialist science	1965	81	353.1	37.6	53.5	13.9	4.4	2.3
	1985	64	372.5	49.1	56.2	13.7	4.2	2.5
	2005	52	423.0	26.4	61.0	16.8	4.5	2.2
Specialist social science	1965	58	337.8	32.8	64.6	13.4	6.0	2.6
	1985	58	358.1	31.9	67.6	24.9	5.5	1.9
	2005	57	372.8	36.7	70.5	13.5	6.1	2.3
Multidisciplinary science	1985	40	312.4	30.7	61.8	20.4	7.4	2.5
	2005	40	340.9	32.5	72.1	13.0	6.6	2.1
Humanities (history)	1965	30	298.4	25.2	57.1	10.7	9.8	2.7
	1985	30	310.9	26.2	57.5	11.8	8.7	2.2
	2005	60	314.2	32.1	62.8	14.9	9.3	2.8

Register	Year	N	Noun + *of*-phrase		Noun + noun	
			M	SD	M	SD
Specialist science	1965	81	35.7	8.6	50.3	15.1
	1985	64	33.1	7.0	60.1	19.4
	2005	52	30.2	6.0	76.6	15.1
Specialist social science	1965	58	34.8	8.3	46.2	17.0
	1985	58	34.4	7.2	58.4	14.0
	2005	57	29.8	5.0	66.3	18.6
Multidisciplinary science	1985	40	41.4	8.5	43.2	15.1
	2005	40	35.5	5.2	56.7	13.5
Humanities (history)	1965	30	39.4	7.3	16.5	7.1
	1985	30	37.5	6.3	23.3	11.8
	2005	60	36.6	8.1	24.3	9.2

and Figure 8 show that the increase in nouns has not occurred uniformly in all sub-registers. For example, history research writing has changed little over this period: it had a relatively dense use of nouns in the nineteenth century, and it has maintained roughly that same density of nouns up to the present time. In contrast, the use of nouns has increased dramatically in science research writing, and it appears that the increase is still in progress.

The trend is by far most pronounced in specialist science articles, where nouns have increased by over 10 percent in just the past 20 years. Specialist social science articles follow the same increasing trend as the specialist science articles, but to a less extreme extent. Finally, the trend toward increased use of nouns has also affected multidisciplinary science articles, but it is considerably less pronounced than in either of the specialist sub-corpora.

Table 6. Summary of the ANOVA Factorial Models for Four Academic Sub-registers (specialist science, specialist social science, multidisciplinary science, humanities) in Three Recent Historical Periods (1965, 1985, 2005).

	Model *F* score	Model significance	Model R^2	Year	Sub-register	Year × sub-register
Nouns	51.02	< .0001	.477	< .0001	< .0001	< .0001
Nominalizations	7.90	< .0001	.124	< .0001	< .0001	*ns*
Relative clauses	33.6	< .0001	.376	*ns*	< .0001	*ns*
Noun + *of*-phrase	10.5	< .0001	.158	< .0001	< .0001	*ns*
Noun + noun	71.1	< .0001	.560	< .0001	< .0001	< .001

Table 6 presents the results of a factorial ANOVA, testing the statistical significance of the mean differences across the three recent historical periods and across the four academic sub-registers. Both main effects show significant differences. In addition, there is a significant interaction effect here, reflecting the fact that humanities research writing has changed little in the use of nouns, while nouns have increased dramatically in use in the science sub-registers.

There are numerous other grammatical features that can be associated with the "nominal style" of modern academic writing. Most of these have increased in use over the course of the twentieth century, although the specific patterns of sub-register development are sometimes surprising. Perhaps the most studied of these characteristics is nominalizations, which have been described as the most important manifestation of the "grammatical metaphor" that is prevalent in modern science writing (see, e.g., Halliday 1988/2004; Banks 2008). Figure 9 shows that the use of nominalizations has increased in all types of academic writing in our study, including history research writing as well as all sub-registers of science research writing.[3] Surprisingly, though, the pattern of sub-register diversification is the opposite to that found for common nouns: nominalizations have increased the most in specialist social science articles and in multidisciplinary science articles, while the increase is considerably less pronounced in specialist science articles. Table 6 shows that these mean differences are statistically significant for both historical period and sub-register.

Previous research has shown that the increased noun phrase complexity of the twentieth century is of a particular type, associated with a preference for phrasal as opposed to clausal elaboration (see, e.g., Biber & Gray 2010, 2011; Biber, Gray & Poonpon 2011). In fact, it turns out that some features of clausal elaboration have actually decreased in use. The most notable of these is finite relative clauses, which have declined in use over the course of the twentieth century. However, as Figure 10 shows, there are distinct patterns of change among sub-registers: the decrease in use is by far strongest in specialist science articles, where relative clauses have declined by 50 percent from 1900 to 1965. Specialist social-science articles also show a strong decline in use, while the multidisciplinary science articles show a lesser decline in use. And in
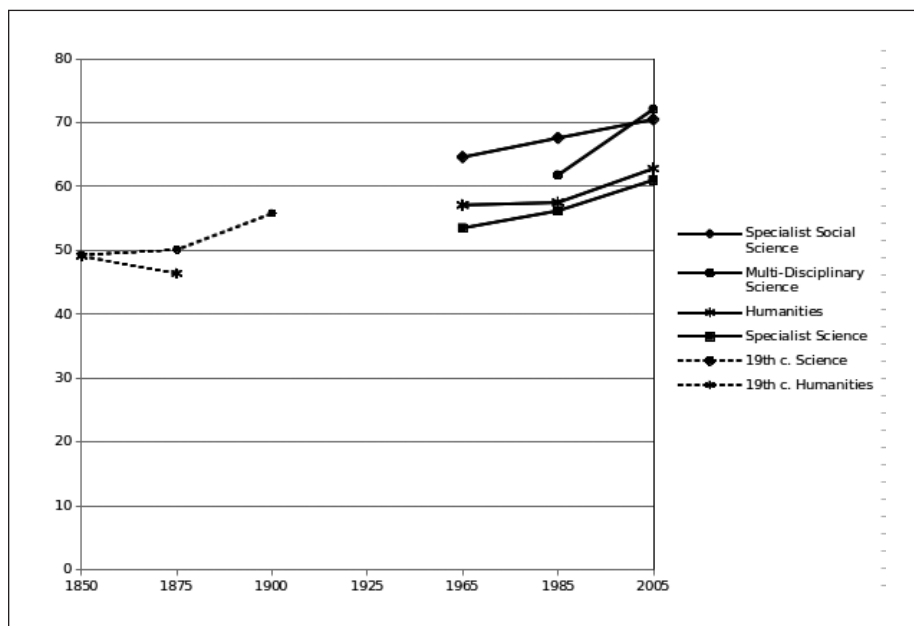
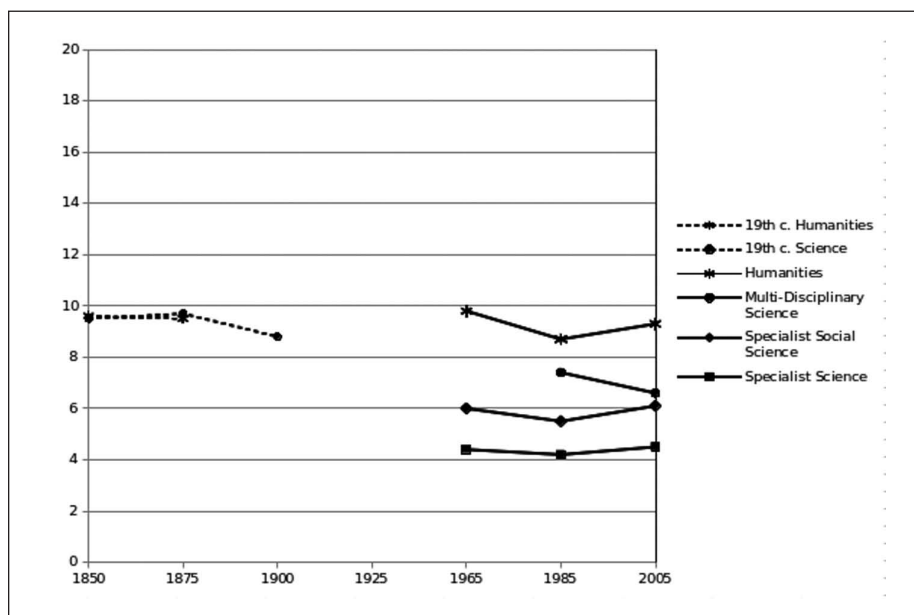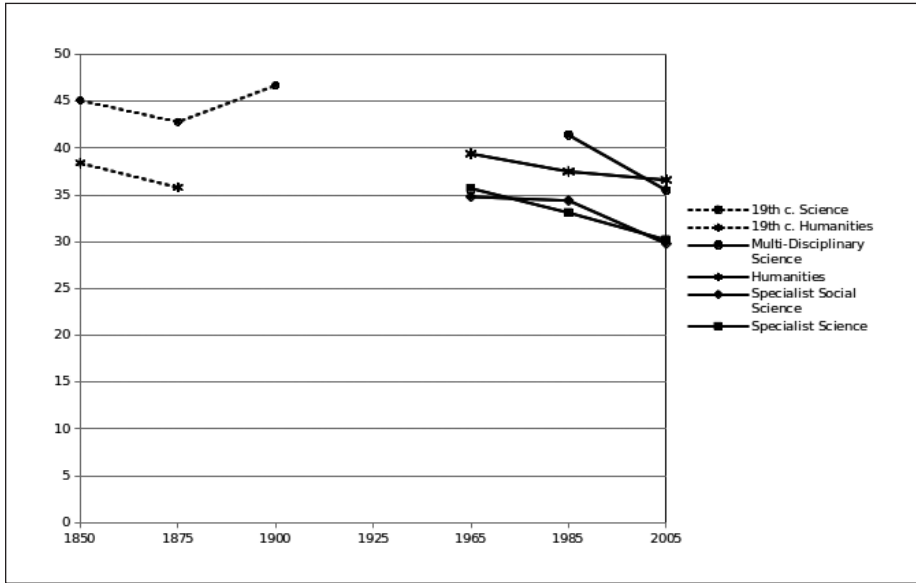**Figure 9.** Nominalizations in academic registers.

**Figure 10.** Finite relative clauses in academic registers.

Figure 11. Noun + *of*-phrase in academic registers.

contrast to all science sub-registers, history research writing has remained essentially unchanged over the past century in its heavy reliance on relative clauses for nominal elaboration. The ANOVA results reported in Table 6 are based on historical change for the three recent historical periods—1965, 1985, and 2005. Thus, the results for relative clauses show a significant difference for sub-register, but no significant difference for historical period. That is, the large historical decrease in the use of relative clauses in science writing occurred in the early twentieth century, while these registers have remained relatively unchanged in the use of relative clauses over the past fifty years.

Interestingly, postnominal *of*-phrases have followed a similar historical path (Figure 11): strong declines in specialist science and social-science articles, a lesser decline in multidisciplinary science articles, and essentially no change in history research writing. Those trends are in marked contrast to historical change in the use of other postnominal prepositional phrases (headed by, e.g., *in, on, with*), which have increased dramatically in use over the past century (e.g., Biber & Gray 2010, 2011).

As noted above, these findings are in agreement with our earlier research studies, which show a general twentieth-century shift in academic writing away from clausal styles of elaboration and toward phrasal, "compressed" styles. The most important grammatical device used to modify noun phrases in this compressed style is a noun used as a premodifier of a head noun (e.g., *patient history, case study*). As Figure 12 shows, premodifying nouns were only moderately common in the nineteenth century, but they have increased in use by as much as 400 percent in the twentieth century. This increase has been strongest in the specialist science sub-register, and also very strong
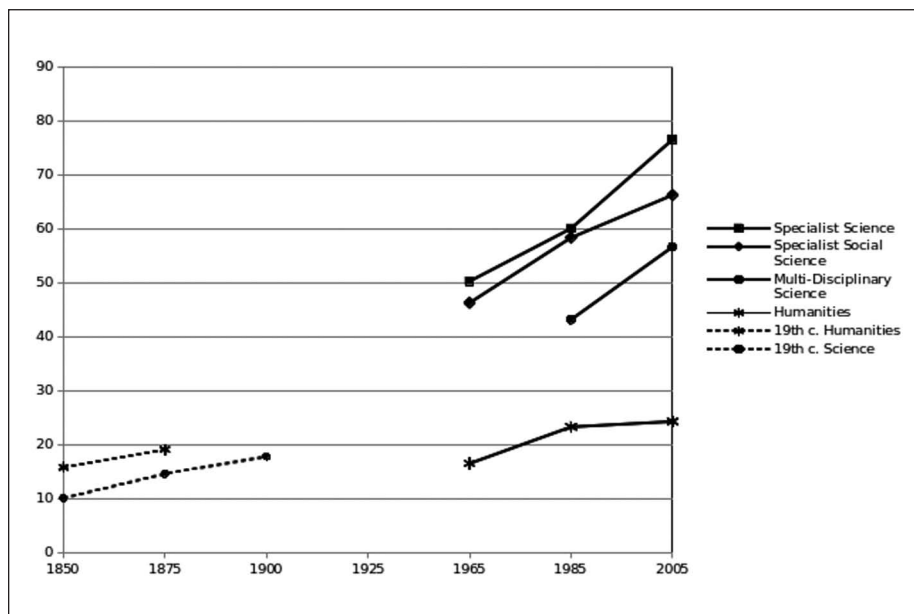
Figure 12. Nouns as nominal premodifiers in academic registers.

in the specialist social-science register. Multidisciplinary science writing is progressing in the same direction, but it lags behind the specialist sub-registers in the extent to which this device is employed. In contrast, history research writing has increased only slightly in the use of this feature. These developments are significant across historical period and across sub-registers (see Table 6). In addition, there is a significant interaction effect for this feature, reflecting the fact that humanities writing has changed little, in contrast to the dramatic changes over time observed for science writing.

The following text samples illustrate these grammatical developments. Sample 3 is from multidisciplinary science writing in the mid-nineteenth century, illustrating the types of clausal elaboration that were common in earlier centuries (including finite adverbial clauses, complement clauses, and relative clauses). Finite relative clauses are underlined in this sample, and genitive-*of* is shown in *ITALIC CAPS*. There are several nominalizations (*division, operation, tenacity, explanation, alterations, comparison*), but no instances of noun–noun sequences in this text excerpt.

Text Sample 3

That division *OF* these nerves produces some serious lesion is proved by the death *OF* the animal, which generally takes place a few days after the operation. Considering the well-known tenacity *OF* life possessed by these animals this was quite an unexpected result, for which I am unable at present to afford any

satisfactory <u>reason</u>. . . . Whatever may be the true explanation, it is impossible not to regard this result with surprise, when we consider the serious lesions <u>which this animal is capable of undergoing at other points</u> OF <u>the frame without loss</u> OF <u>life</u>. The usual time <u>which it survives</u> is variable, and depends greatly on the season *OF* the year. . . . For the purpose *OF* avoiding this loss *OF* life, I adopt the plan *OF* dividing the glossopharyngeal on one side only *OF* the tongue, and I find that it has the desired effect *OF* preserving the life *OF* the animal, while we can observe the same alterations on the corresponding side, as well as when both nerves are divided. Another advantage found in the division *OF* a single nerve is, that on the uninjured side we have constantly at hand a means *OF* comparison <u>by which we can judge with certainty</u> respecting any alterations <u>that may be produced in the divided nerve</u>.

(Augustus Waller, "Experiments of the Section of the Glossopharyngeal and Hypoglossal Nerves of the Frog, and Observations of the Alterations Produced Thereby in the Structure of Their Primitive Fibres," *Philosophical Transactions of the Royal Society of London* 140 (1850): 423-429)

Present-day multidisciplinary science writing can sometimes be similar in discourse style to nineteenth-century articles. Thus, Text Sample 4 illustrates a modern *PT* text with extensive clausal elaboration (finite relative clauses are <u>underlined</u>) and use of the genitive-*of* (shown in *ITALIC CAPS*). However, this text also illustrates the recent shift toward increasing use of phrasal modification; thus, noun–noun sequences are shown in bold italics.

Text Sample 4

From Darwin derives not only the explicit assumption *OF animal-human continuity*, but also the implicit assumption that *behaviour patterns*, <u>from which intelligence must be inferred</u>, are proper candidates for evolutionary selection, a tenet <u>that became the</u> *foundation stone OF* <u>ethology and was also seminal for many branches</u> OF <u>psychobiology</u>. In one *OF* his notebooks Darwin wrote that 'he <u>who understands baboon</u> would do more towards metaphysics than Locke' (Notebook M, p. 84, 16 August 1838; cf. Gruber (1974) pp. 317-318). Darwin himself, however, is relatively silent, even impotent, when it comes to those behavioural descriptions <u>that might lead to a deep or systematic understanding</u> OF <u>the cognitive capacities</u> OF <u>animals, let alone metaphysics</u>, and his most concentrated treatment *OF* reasoning, for example, in the Descent *OF* man is little more than scattered anecdotes *OF* doubtful value.

(L. Weiskrantz, "Introduction: Categorization, Cleverness and Consciousness," *Philosophical Transactions of the Royal Society of London (Series B)* 308, no. 1135 (1985): 3-19)

The more typical style of recent multidisciplinary science writing is to employ much less clausal elaboration, coupled with a dense use of phrasal modifiers, as in Text Sample 5:

Text Sample 5

Tide gauge measurements suggest that global average *sea-levels* rose by between 1 and 2 mmyr-1 during the twentieth century (Church et al. 2001). Between 1993 and 2000, *satellite altimetry* indicated that the rate *OF* rise was approximately 2.5 mmyr-1 (Cabanes et al. 2001), which may indicate a recent acceleration *OF* the *long-term trend* or it could be the result *OF* interdecadal variability. In the future, global average *sea-level* is expected to increase more rapidly as a result *OF* anthropogenic climate change. The Third Assessment Report *OF* the Intergovernmental Panel on *Climate Change* concluded that between 1990 and 2100, the global average *sea-level* rise is likely to be between 9 and 88 cm (Church et al. 2001). While this *century-scale* rise in sea-level will threaten many low-lying unprotected coastal areas, it is the extremes *OF sea-level* associated with *storm surges* that will cause much OF the damage (for instance McCarthy et al. 2001). Extreme *water levels* will also increase as *time-average sea-levels* rise. In addition, changes in the number, path and strength *OF* atmospheric cyclonic storms may alter the formation and evolution *OF storm surges*. However, because individual *surge events* depend on the driving meteorology, we cannot predict individual surges more than, at most, a few days ahead.

(Jason A. Lowe and Jonathan M. Gregory, "The Effects of Climate Change on Storm Surges Around the United Kingdom," *Philosophical Transactions of the Royal Society of London (Series A)* 363 (2005): 1313-1328)

This style is similar to that found in recent specialist science writing, except that specialist science writing exhibits an even more extreme reliance on phrasal modification of nouns, coupled with even less use of dependent clauses. For example,

Text Sample 6

Population growth rate is a particularly powerful index for evaluating *harvest effects* because it measures the ability *OF* a population to increase when subjected to any specified level *OF* exploitation. . . . Selectivity *OF* the harvest on *Putauhinu Island* translates into large differences in *harvest rates* among *weight* classes. . . . The *population effects OF* removing an individual depends on quality (i.e. future contributions to reproduction) and on the contribution *OF* its stage to demography. The effect of *quality differences* among *chick classes* in our model is small because chicks move among classes during the nanao and rama (Table 1). The contribution *OF chick survival* to *population growth* is small, and regardless of

initial weight all chicks become identical pre-breeders at the end *OF* the first year (Fig. 1). . . . There is evidence for such links between characteristics *OF* young individuals and life history traits *OF* adults in many taxa. . . .

> (Christine Hunter and Hal Caswell, "Selective Harvest of Sooty Shearwater Chicks: Effects of Population Dynamics and Sustainability," *Journal of Animal Ecology* 75 (2005): 589-600.)

One interesting characteristic of the noun–noun sequences found in modern academic writing is that they are typically not technical terms. Of course, we can find counterexamples which fit the stereotype of jargon-heavy scientific academese, as in,

(3) The *Peutz-Jegher syndrome tumor-suppressor gene* encodes a *protein-threonine kinase*.

> (Reuben J. Shaw, Katja A. Lamia, Debbie Vasquez, Seung-Hoi Koo, Nabeel Bardeesy, Ronald A. DePinho, Marc Montminy & Lewis C. Cantley, "The Kinase LKB1 Mediates Glucose Homeostasis in Liver and Therapeutic Effects of Metformin," *Science* 310 (2005): 1642-1646)

However, the more common pattern is to modify a nontechnical head noun with another nontechnical premodifying noun. Specialist background knowledge is required to understand the meaning relationship between the two nouns, but the nouns themselves typically have transparent, everyday meanings. Many of these could be paraphrased with an *of*-genitive as nominal postmodifier:

(4) *sea level*   level *OF* the sea
    water level   level *OF* the water
    life history   history *OF* a life
    life history traits   traits *OF* a history *OF* a life

Other examples could be paraphrased with other prepositional phrases as nominal postmodifier:

(5) *quality differences*   differences in quality
    century-scale rise   a rise on the scale of a century
    population effects   the effects on the population

Even in these paraphrased examples, specialist knowledge is required to understand the meaning relationship between the head noun and the prepositional phrase. However, other examples are much more difficult to paraphrase; for example,

(6)   *tide gauge measurements*   measurements made by a gauge that calculates
                                    the level of the tide

long-term trend	a trend that lasts for the long term
time-average sea-levels	the level of the sea averaged across different time periods
storm surge	a surge in the sea level, associated with a storm
surge events	events that involve a surge in the sea level, associated with a storm
harvest rates	differences in the magnitude of a harvest
harvest effects	the extent to which the magnitude of a harvest varies
chick survival	the extent to which chicks survive

In sum, the individual nouns employed in noun–noun sequences are often not technical, but the meaning relationships underlying those sequences are technical and require specialist background knowledge for understanding. As a result, this grammatical device is most strongly preferred in specialist science writing, with a less dense use in science writing for multidisciplinary audiences.

Summary and Conclusion

Our goals in the present article have been both methodological and historical/descriptive. Methodologically, we hope to have demonstrated the importance of not only registers but also of sub-registers for studies of historical language use, arguing that register distinctions at all levels are important for the study of linguistic variation (see also Biber 2012). We undertook two strong tests of this claim: comparing two sub-registers of written news reportage and comparing four sub-registers of academic research writing. Both case studies show that even at a very specific level of register differentiation, there are systematic differences in the patterns of historical development. These differences are all statistically significant, and in most cases they are also statistically strong. Historical analyses that disregard these differences would confound the description of linguistic change with patterns that in fact reflect register differences. We believe that these case studies convincingly demonstrate that register differences matter, for both synchronic and diachronic descriptions of language use—and that they are important at a more specific level of analysis than has typically been suspected.

From a descriptive perspective, we have documented some surprising patterns of register variation and change. First, we have shown how American newspaper reportage in *NYT* has been consistently more innovative in historical changes than written news reports in *Time* magazine. It is conceivable that these linguistic differences are simply the result of different editorial policies, with *Time* magazine being more conservative and resistant to change than *NYT*. If this turns out to be the primary factor, it is noteworthy that the same policies have been maintained in both publications for approximately eighty years, and that those policies have applied to the use of both colloquial features as well as compression features. Interestingly, the grammatical patterns reported here are at odds with the findings of previous research, which has described *Time* magazine as especially innovative in its creation of new vocabulary (see Firebaugh 1940; Yates 1981).

Other possible factors include differences in the intended audiences of the two publications and differences in the typical communicative purposes of their news articles. For example, news reportage articles in *NYT* have generally included more personal narratives than *Time* articles. These differences are becoming less noticeable in recent periods, as it is also becoming more difficult to clearly identify news reports as distinct from other newspaper/magazine registers. However, it is likely that the historical linguistic differences documented in the present study reflect a range of influences, rather than any single factor.

With regard to academic research writing, we have shown how science (and social science) writing has been much more innovative than research writing in the humanities. However, our analyses further identify systematic differences at more specific levels, with science research writing being more innovative than writing in the social sciences, and specialist research articles being more innovative than articles written for a multidisciplinary audience. These linguistic changes have been of a particular type, marked by an overall decline in dependent clauses together with a strong increase in the use of phrasal modifiers. These grammatical shifts can be linked to the "information explosion" and the need to express more information in fewer words. At the same time, these grammatical developments result in prose that is less explicit in the expression of the meaning relationships among constituents. These two factors working together help to explain why these historical changes have been restricted mostly to informational writing, and why they have been most extreme in the sub-registers intended for specialist readerships.

In sum, the present study has shown how twentieth-century grammatical changes in news reportage and academic writing are mediated by specific sub-registers. In future research, we hope to explore in more detail the specific lexico-grammatical correlates of these changes, as well as their underlying social/situational motivating factors.

Declaration of Conflicting Interests

Funding

Notes

1. Following Biber and Conrad (2009:8-11), we use the term "register" to refer to situationally defined varieties. Registers can be identified and described at any level of specificity (Biber & Conrad 2009:32-33). The term "sub-register" is used in the present article to refer to specific registers within a more general register domain. Our point here is not to argue for a particular level of specificity, but rather to raise awareness that even highly specific register differences correspond to systematic differences in historical development.

2. Ideally, our corpus would have also included samples for 1945 and 1965, so that the analy-sis would have been based on five evenly spaced data points over the past eighty years. The actual corpus that we used allows a less complete description, but it is still more than adequate to demonstrate our central point concerning the importance of sub-register.

3. Nominalizations in the present study include only nouns formed through derivational morphol-ogy, such as the suffixes *-tion, -ance*, and *-ment*. These forms are identified automatically by the Biber Tagger (Biber 1988; Biber et al. 1999). It is worth noting that the notion of "nominal-ization" here differs from that adopted by Halliday (1988/2004) and Banks (2008), who also include cases of verb-to-noun conversion (e.g., words like *aim, form*, and *list* used as nouns).

4. Although it is not considered in the present study, Leech et al. (2009:222-225) and Hinrichs and Szmrecsanyi (2007) show that the decline in use for the *of*-genitive is in part offset by a dramatic historical increase in use for the *s*-genitive.

References

Atkinson, Dwight. 1999. *Scientific discourse in sociohistorical context: The philosophical transactions of the Royal Society of London, 1675–1975*. Hillsdale, NJ: Lawrence Erlbaum.

Banks, David. 2008. *The development of scientific writing: Linguistic features and historical context*. London: Equinox.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge, UK: Cambridge University Press.

Biber, Douglas. 2003. Compressed noun phrase structures in newspaper discourse: The compet-ing demands of popularization vs. economy. In Jean Aitchison & Diana Lewis (eds.), *New media discourse*, 169-181. London: Routledge.

Biber, Douglas. 2004. Historical patterns for the grammatical marking of stance: A cross-regis-ter comparison. *Journal of Historical Pragmatics* 5(1). 107-135.

Biber, Douglas. 2009. Are there linguistic consequences of literacy? Comparing the potentials of language use in speech and writing. In David Olson & Nancy Torrance (eds.), *Cambridge handbook of literacy*, 75-91. Cambridge, UK: Cambridge University Press.

Biber, Douglas. 2012. Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory* 8. 9-37.

Biber, Douglas & Victoria Clark. 2002. Historical shifts in modification patterns with complex noun phrase structures: How long can you go without a verb? In Teresa Fanego, María José López-Couso & Javier Pérez-Guerra (eds.), *English historical syntax and morphology*, 43-66. Amsterdam: John Benjamins.

Biber, Douglas & Susan Conrad. 2009. *Register, genre, and style*. Cambridge, UK: Cambridge University Press.

Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus linguistics: Investigating lan-guage structure and use*. Cambridge, UK: Cambridge University Press.

Biber, Douglas & Edward Finegan. 1989. Drift and the evolution of English style: A history of three genres. *Language* 65. 487-517.

Biber, Douglas & Edward Finegan. 1997/2001. Diachronic relations among speech-based and written registers in English. In Terttu Nevalainen & Leena Kahlas-Tarkka (eds.), *To explain the present: Studies in the changing English language in honour of Matti Rissanen*, 253-275.

Helsinki: Société Néophilologique. Reprinted in Susan Conrad & Douglas Biber (eds.), *Variation in English: Multi-dimensional studies*, 66-83. New York: Longman.

Biber, Douglas, Edward Finegan & Dwight Atkinson. 1994. ARCHER and its challenges: Compiling and exploring a representative corpus of historical English registers. In Udo Fries, Gunnel Tottie & Peter Schneider (eds.), *Creating and using English language corpora*, 1-14. Amsterdam: Rodopi.

Biber, Douglas & Bethany Gray. 2010. Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes* 9. 2-20.

Biber, Douglas & Bethany Gray. 2011. Grammar emerging in the noun phrase: The influence of written language use. *English Language and Linguistics* 15. 223-250.

Biber, Douglas & Bethany Gray. 2012. The competing demands of popularization vs. economy: Written language in the age of mass literacy. In Terttu Nevalainen & Elizabeth Traugott (eds.), *Rethinking the history of English*. Oxford, UK: Oxford University Press.

Biber, Douglas, Bethany Gray & Kornwipa Poonpon. 2011. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly* 45. 5-35.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *The Longman grammar of spoken and written English*. London: Longman.

Biber, Douglas & James Jones. 2008. Quantitative methods in corpus linguistics. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, 1286-1304. Berlin: Walter de Gruyter.

Crespo García, Begoña & Isabel Moskowich-Spiegel Fandiño. 2010. CETA in the context of the *Coruña Corpus*. *Literary and Linguistic Computing* 25(2). 153-164.

Culpeper, Jonathan & Merja Kytö. 2010. *Early Modern English dialogues: Spoken interaction as writing*. Cambridge, UK: Cambridge University Press.

Davies, Mark. 2007-. *TIME Magazine Corpus: 100 million words, 1920s-2000s*. accessed January 9, 2013, http://corpus.byu.edu/time/.

Davies, Mark. 2010–. *The Corpus of Historical American English: 400 million words, 1810–2009*. http://corpus.byu.edu/coha/.

Davies, Mark. 2011–. *Google Books Corpus*. (Based on Google Books n-grams). http://googlebooks.byu.edu/.

Davies, Mark. in press-a. Expanding horizons in historical linguistics with the 400 million word Corpus of Historical American English. *Corpora*.

Davies, Mark. in press-b. Recent shifts with three nonfinite verbal complements in English: Data from the 100 million word TIME Corpus (1920s–2000s). In Bas Aarts, Jo Close, Sean Wallis & Geoff Leech (eds.), *Current change in the English verb phrase*. Cambridge, UK: Cambridge University Press.

Fairclough, Norman. 1992. *Discourse and social change*. Cambridge, UK: Polity.

Firebaugh, Joseph J. 1940. The vocabulary of "Time" magazine. *American Speech* 15. 232-242.

Gray, Bethany. 2011. *Exploring academic writing through corpus linguistics: When discipline tells only part of the story*. Flagstaff: Northern Arizona University dissertation.

Halliday, M. A. K. 1988/2004. On the language of physical science. In M. A. K. Halliday, *The language of science*, 140-158. London: Continuum.

Halliday, M. A. K. & J. R. Martin. 1993/1996. *Writing science: Literacy and discursive power*. London: Falmer Press.

Hinrichs, Lars & Benedikt Szmrecsanyi. 2007. Recent changes in the function and frequency of standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language and Linguistics* 11. 437-474.

Hundt, Marianne & Christian Mair. 1999. "Agile" and "uptight" genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4. 221-242.

Leech, Geoffrey. 2003. Modal on the move: The English modal auxiliaries 1961–1992. In Roberta Facchinetti, Frank Palmer & Manfred Krug (eds.), *Modality in contemporary English*, 223-240. Berlin: Mouton de Gruyter.

Leech, Geoffrey. 2011. The modals ARE declining. Reply to Neil Millar's "Modal verbs in TIME: The English modal auxiliaries 1961–1991." *International Journal of Corpus Linguistics* 16(4). 547-564.

Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge, UK: Cambridge University Press.

Mair, Christian. 2006. *Twentieth-century English: History, variation and standardization*. Cambridge, UK: Cambridge University Press.

Millar, Neil. 2009. Modal verbs in TIME: Frequency changes 1923–2006. *International Journal of Corpus Linguistics* 14(2). 191-220.

Taavitsainen, Irma & Päivi Pahta. 2010. *Early modern English medical writing: Corpus description and studies*. Amsterdam: John Benjamins.

Taavitsainen, Irma & Päivi Pahta. 2011. *Medical writing in early modern English*. Cambridge, UK: Cambridge University Press.

Wells, Rulon. 1960. Nominal and verbal style. In Thomas A. Sebeok (ed.), *Style in language*, 213-220. Cambridge, MA: MIT Press.

Yáñez-Bouza, Nuria. 2011. ARCHER past and present (1990–2010). *ICAME Journal* 35. 205-236.

Yates, Norris. 1981. The vocabulary of *Time* Magazine revisited. *American Speech* 56(1). 53-63.

Author Biographies

Douglas Biber is Regents' Professor of English (Applied Linguistics) at Northern Arizona University. His research efforts have focused on corpus linguistics, English grammar, and register variation (in English and cross-linguistic; synchronic and diachronic). He has written over 180 research articles and 14 authored books and monographs, including a textbook on *Register, Genre, and Style* (Cambridge, 2009), the co-authored *Longman Grammar of Spoken and Written English* (1999), and other academic books published with Cambridge University Press (1988, 1995, 1998) and John Benjamins (2006, 2007).

Bethany Gray is Assistant Professor of English (TESL / Applied Linguistics program) at Iowa State University. Her research has focused on the application of corpus linguistics to the study of synchronic and diachronic register variation, with a particular focus on the grammar of academic writing, cross-disciplinary variation, and sub-registers within academic writing. Her research has resulted in articles published in *English for Specific Purposes, Journal of English for Academic Purposes, and TESOL Quarterly,* as well as forthcoming articles in *International Journal of Corpus Linguistics and Corpora*.