Variationist versus text-linguistic approaches to grammatical change in English: nominal modifiers of head nouns — Source link ⎘

Douglas Biber, Jesse Egbert, Bethany Gray, Rahel Oppliger ...+1 more authors

Institutions: Brigham Young University, Iowa State University, University of Zurich

Related papers:

- Register as a predictor of linguistic variation

- Propositional modifiers in early English medical prose: A study on their historical development in noun phrases

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

- Longman Grammar of Spoken and Written English

- The Development of English Aspectual Systems: Aspectualizers and Post-verbal Particles

Year: 2016

Variationist versus text-linguistic approaches to grammatical change in English: nominal modifiers of head nouns

Biber, Douglas ; Egbert, Jesse ; Gray, Bethany ; Oppliger, Rahel ; Szmrecsanyi, Benedikt

21

Variationist versus text-linguistic approaches to grammatical change in English: nominal modifiers of head nouns

Douglas Biber, with Jesse Egbert, Bethany Gray,
Rahel Oppliger, and Benedikt Szmrecsanyi

21.1   Introduction

21.1.1   Types of grammatical change

Numerous major grammatical developments have occurred in English in earlier historical periods, including the change to a relatively fixed SVO word order, the loss of most inflectional morphology (especially case suffixes), the increase in the range of function words (including prepositions, auxiliary verbs, infinitive marker *to*), and the introduction of the dummy auxiliary verb *DO* (see Rissanen 1999, van Gelderen 2006; see also Chapter 14 by Fischer in this volume). However, grammatical change over the last 300 years – the period of Late Modern and Present-day English – has been less dramatic, with no major structural innovations (see the surveys in Denison 1998, Mair 2006, Leech et al. 2009, Brinton and Bergs 2012). Instead, recent changes have been of two general types:

1   grammatical innovations that result in particular words being used for new grammatical functions;
2   shifts in the use (frequency and functions) of core grammatical features.

The first type of grammatical change has been studied mostly under the rubric of 'grammaticalization', which focuses on the way in which content words evolve over time to be used as grammatical function words. Examples include the use of *have to* and *got to* as semi-modals, *wanna* with modal

auxiliary functions, and *get* as an auxiliary verb in passive constructions. Other examples include the use of *GO*, *BE all*, and *BE like* as quotative verbs, *well* as a discourse marker, *pretty* as a hedge or intensifier, and sequences like *in spite of*, *with regard to*, and *because of* used as complex prepositions (see, e.g., Krug 2000; Hopper and Traugott 2003; Hoffmann 2004a; Lindquist and Mair 2004; Nevalainen 2004 Tagliamonte 2004; Buchstaller and van Alphen 2012).

In contrast, the second type of change involves the use of a grammatical feature: its overall frequency, changes in the (probabilistic) constraints on the choice among variants, changes in discourse function, and the co-occurrence of the grammatical feature with an increasing (or decreasing) set of words, associated with expanding (or shrinking) semantic domains. Examples include the increasing use of progressive verbs, multi-word verbs, analytical rather than synthetic comparison, and regular (versus irregular) verb inflections; and the decreasing use of modals, passive voice verbs, reflexives, and the relative pronoun *whom* (see, e.g., Hundt and Mair 1999; Mair 2006; Hundt 2007; Leech et al. 2009; Rohdenburg and Schlüter 2009).

Most grammatical changes over the past 300 years are of this second type. Even grammatical innovations, like the development of semi-modals and the *get*-passive, have also gradually continued to increase in frequency and functionality over this period (see, e.g., Leech et al. 2009). Thus, investigations into shifts in use have become increasingly important for the study of recent grammatical change:

> changes in the realm of syntax are often a function of quantity, rather than quality; that is, certain structures have expanded in number and frequency of occurrence during the PDE period. (Fennell 2001: 173)

> Since relatively few categorical losses or innovations have occurred in the last two centuries, syntactic change has more often been statistical in nature, with a given construction occurring throughout the period and either becoming more or less common generally or in particular registers. The overall, rather elusive effect can seem more a matter of stylistic than of syntactic change. (Denison 1998: 93)

In the present chapter, we undertake a historical exploration of variation and change in one grammatical characteristic that has exhibited major shifts in use over the past 300 years: the modification of English noun phrases. We focus especially on noun phrases that express genitive relationships, where one noun modifies another noun. In most previous studies, these constructions have been analysed in terms of two structural variants – the *s*-genitive and the *of*-genitive – which are often interchangeable. Traditionally, these constructions are associated with meanings of possession:

s-genitive: *the family's car*
of-genitive: *the car of the family*

However, in actual use, genitives express a wide array of meaning relations (see Biber et al. 1999: 303), for example:

Attribute: <u>*Martha's courage*</u> *failed her.*
Subjective: <u>*Chiang's recognition*</u> *of the priority of the spoken language*
                explained . . .
Partitive: <u>*This section*</u> *of the discussion concerns . . .*
Defining: *I live in* <u>*the city of Lahore*</u>*.*
Objective: <u>*The brutal murder of a child*</u> *causes . . .*

When this full set of meaning relations is considered, it becomes apparent that there is actually a third structural option that should be compared to traditional genitive constructions: nouns as premodifiers of a head noun (see also Rosenbach 2006, 2007). Thus consider the following example from a newspaper article:

. . . the Pope met Mr Gierek, the <u>*Communist Party chief*</u> *. . .*

The final noun phrase in this example illustrates a noun serving as premodifier of a head noun:

noun + head noun: *the Communist Party chief*

This noun phrase could also be paraphrased with the other two genitive variants, all expressing the same basic attributive meaning relationship:

noun-*'s* + head noun: *the Communist Party's chief*
head noun + *of*-phrase: *the chief of the Communist Party*

While several previous studies have investigated the choice between *s*-genitives and *of*-genitives (e.g. Rosenbach 2002; Kreyer 2003; Stefanowitsch 2003; Hinrichs and Szmrecsanyi 2007; Szmrecsanyi and Hinrichs 2008), only a few previous studies have focused on the choice between *s*-genitives and noun–noun constructions (e.g. Rosenbach 2006, 2007); and to our knowledge, no previous study has investigated the patterns of variation and change among all three variants.

   In the sections below, we explore this issue. These are historical changes of the second type: shifts in the overall frequencies and functions of structural variants. Although we present descriptive findings about the historical development of these noun phrase structures, our goals are also methodological: to carefully document the analytical procedures required for such an analysis, and to explore the consequences of different analytical decisions.

21.1.2 Corpus-based investigations of grammatical change

Corpus-based analysis is ideally suited to the study of historical change in the overall frequencies and functions of structural variants. As noted above, this type of grammatical change is both quantitative and qualitative, involving expansions (or decreases) in frequency, range of lexical co-occurrence

and functionality (including sensitivity to contextual factors), and changing sociolinguistic usage patterns. While it might be possible to notice some of these changes by reading texts from different periods, there is no way to reliably study this range of phenomena systematically without access to a large and representative collection of texts: a corpus.

Most recent investigations of historical change that focus on the use of grammatical features have employed corpus-based analyses. The first step for such analyses is to construct a corpus that represents the targeted language varieties and historical periods. For example, Mair (2006) and Leech et al. (2009) were interested in twentieth century grammatical change, comparing American English (AmE) to British English (BrE). The Brown family of corpora are very well suited to such research questions. These one-million-word corpora were designed to replicate the first large corpus of English: the Brown Corpus, which consists of 500 AmE written text samples published in 1961, taken from fifteen text categories (e.g., newspaper reportage, editorials, biographies, fiction, academic prose). Parallel corpora with this same design have been constructed for 1961 BrE (the LOB corpus), 1992 AmE (the Frown Corpus), 1991 BrE (the F-LOB Corpus), 1931 BrE (the BLOB-1931 Corpus), and 1901 BrE (the Lancaster BrE Corpus). Thus, by applying the same methods to this suite of corpora, it is possible to track quantitative patterns of grammatical change over the course of the twentieth century.

Other studies have utilized corpora specifically designed to represent a range of registers and sub-registers over time. For example, the ARCHER corpus is a 1.8-million-word corpus of texts, organized in terms of eight speech-based and written registers sampled from 1650–1990 (see Biber et al. 1994a; Yáñez-Bouza 2011); the corpus of Early Modern English Medical Texts (EMEMT) contains sub-corpora for sub-registers of medical writing such as scientific journals, general treatises or textbooks, surgical and anatomical treatises, recipe collections, and health guides (see Taavitsainen and Pahta 2010); and the Corpus of English Dialogues 1560–1760 (CED), which includes text types to represent both authentic and constructed dialogue (see Culpeper and Kytö 2010).

More recently, researchers have been using much larger historical corpora, such as the Corpus of Historical American English (see Davies 2012b). This is a 400-million-word corpus of language from fiction (*c*.50 per cent of the total), magazines, newspapers, and other books, organized in decades from 1810–2010, with a target of *c*.20 million words sampled from each historical period (earlier periods have smaller samples, more recent periods are represented by larger samples). 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 per cent sample of texts from the magazine for the period 1923–2006 (see Davies 2013 and Millar 2009 for a study of modal verbs based on this corpus).

The present study is based on three registers from ARCHER (letters, newspaper reportage, and science research articles), which enables a detailed

exploration of genitive constructions using a relatively small corpus while still allowing inclusion of register as a factor in linguistic change. Extracting and coding genitive constructions is a relatively labour-intensive task (see section 21.2), which is why we restrict our attention to a manageable dataset. However, because genitive constructions are textually frequent, it is possible to study them quantitatively even on the basis of a relatively small corpus.

21.1.3   Methodological issues for empirical investigations of grammatical change

Methodological issues are always prominent in corpus-based studies of grammatical change. For example, researchers almost always describe the size and design of the corpus, addressing the extent to which the corpus sample represents the target discourse domain. Further, researchers are usually careful to document their corpus analysis procedures, describing the methods used to identify occurrences of the target linguistic features (e.g. through concordancing, hand analyses, or automatic tagging/parsing). In many cases, there is also careful discussion of the quantitative analyses, related to issues like norming and the application of appropriate statistical techniques.

Other methodological issues actually arise before the analysis begins, and these are less often addressed (or even noticed). In the present chapter, we focus on three of these issues: (1) the set of linguistic variants included in the analysis, (2) the set of registers included in the analysis, and (3) the research design employed for the analysis.

The first issue concerns the need to consider the full set of linguistic variants in order to have a complete understanding of historical shifts in use (see Labov's 1966: 49 'principle of accountability'). Most studies of grammatical variation have instead focused on the choice between only two variants. For example, most previous research on genitive noun phrases has focused on the binary choice between the *of*-genitive and the *s*-genitive (e.g. Gries 2002; Jankowski 2009; Grafmiller 2014; Shih et al. 2015). Rosenbach (2006, 2007) is exceptional in that she considers the use of premodifying nouns, but that study similarly focuses mostly on a binary opposition: between premodifying nouns and *s*-genitives.

This methodological restriction – which more often than not is a matter of convenience rather than conviction – has important implications for the conclusions drawn from a study. So, for example, studies on genitive constructions have generally concluded that the *of*-genitive is declining in use, being replaced by the *s*-genitive. Leech et al. (2009: 225) show that the *of*-genitive declined in use by 24 per cent from 1961 to 1991, while the *s*-genitive increased in use by 24 per cent; both trends are reported relative to the combined total of *of*-genitives and *s*-genitives. However, as we show in the following sections, inclusion of a third structural variant (nouns as noun premodifiers) in the same analysis leads to somewhat different conclusions

concerning these historical changes (because premodifying nouns have been increasing in use much more rapidly than *s*-genitives).[1]

The second issue has to do with the sample of texts considered in the analysis. Many previous studies of historical variation have been based on analysis of a general purpose corpus, or analysis of a single register. However, as we show below, there are important differences in historical change across registers: in the types of linguistic change, the magnitude of change, and even the direction of change. As a result, historical studies based on a general purpose corpus might fail to capture the actual patterns of change, while studies based on a single register will probably provide only a partial picture.

Finally, quantitative studies of language use can be undertaken with different research designs, which address different research questions. The choice of a particular research design and its influence on the type of quantitative analyses conducted are rarely explicitly discussed within the context of a particular study, aside from describing what sorts of procedures are carried out on the quantitative data. In reality, one of the first decisions a researcher makes involves determining the nature of that quantitative data, which in turn restricts the types of procedures and conclusions that can be drawn. The following section takes a closer look at the issue of research design, comparing how two approaches impact the most basic quantitative measure: how to measure the frequency of use of linguistic features.

21.1.4   Perspectives on 'frequency': variationist versus text-linguistic research designs

Empirical research on grammatical change in English has been carried out from two major perspectives: variationist and text-linguistic. These two perspectives approach the quantitative description of language use in fundamentally distinct ways. Simply put, variationist research studies investigate proportional preferences, while text-linguistic studies investigate the rates of occurrence in texts (see Biber 2012: 12–17).

The variationist approach was originally developed for sociolinguistic research (see Labov 1966, Cedergren and Sankoff 1974; see also Chapter 1 by Romaine in this volume) but has since been extended to other applications, like research in the Probabilistic Grammar framework (e.g. Bresnan and Hay 2008). The variationist method is based on analysis of the variants of a linguistic variable. To be included in the analysis, variants must be interchangeable (i.e. they are both grammatically possible and equivalent in meaning). Tokens of each variant are coded for a range of contextual factors, and then quantitative analysis (often statistical regression analysis) is used to determine the extent to which contextual and possibly language-external

---

[1] In the preceding section of their book, Leech et al. provide an extended discussion of nouns premodifying a head noun (2009: 211–22); however, they do not consider this structural option as an alternative to other genitive constructions.

constraints (in Labovian parlance, 'conditioning factors') favour or disfavour particular variants.

In contrast, in the text-linguistic approach, the quantitative analysis describes the rates of occurrence for linguistic features in texts, typically without extensive annotation of individual occurrences. In this approach, features are not necessarily contrasted with competing features (as they are in the variationist approach). When linguistic variants are contrasted in a text-linguistic design, each variant is treated as a separate linguistic feature.[2]

In terms of their research designs, the primary difference between these two analytical approaches is the unit of analysis (or the 'observations'):

- In variationist studies, the unit of analysis is each occurrence or non-occurrence of a linguistic feature ('variant'). Variationists are thus interested in individual linguistic choices, and their constraints.
- In text-linguistic studies, the unit of analysis is each individual text (or each sub-corpus – see below). Text linguists thus analyse linguistic use on a coarser level of granularity.

The units of analysis are the 'observations' that are described in a study. For the most part, each observation in a variationist study (i.e. a token of a linguistic feature) has categorical rather than continuous characteristics; the overall patterns can be quantified by counting the frequency of each category across the full set of observations. By contrast, each observation in a text-linguistic study (i.e. each text) is analysed in terms of quantitative characteristics. Variationist studies tell us the proportional preference for one variant over another, but they are typically agnostic about how often we will encounter a grammatical feature in a text. In contrast, text-linguistic studies are designed for this latter purpose.

In the present chapter, we illustrate this methodological difference through two related studies of noun phrases that express genitive relationships, where one noun modifies another noun. The first case study employs a variationist research design, while the second case study employs a text-linguistic research design. As we show in the following sections, these two approaches answer different research questions and lead to different conclusions. Taken together, they provide a more complete description of historical change than either taken on its own.

In the variationist research design, each occurrence of a genitive noun phrase is treated as an observation; the analysis is restricted to those occurrences of genitive noun phrases that are interchangeable with other structural variants (see discussion in sections 21.2 and 21.3). Each of these noun phrases is coded for several linguistic factors, such as the animacy of the modifying noun, the thematic status of the head noun and modifying noun, the length of the head noun phrase and modifying noun phrase. Then,

---

[2] Biber et al. (1998: 269–74), Biber and Jones (2009), and Biber (2012) provide detailed discussions of these different research designs in synchronic studies of linguistic variation and use.
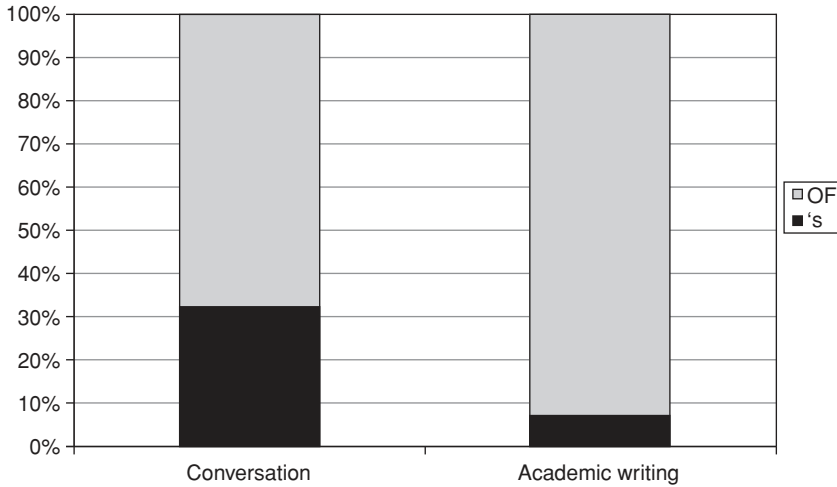
Figure 21.1 *Proportional use of s-genitives versus of-genitives modifying a head noun, in late twentieth-century conversation vs. academic writing (based on Biber et al. 1999: 302)*

considering the full set of all interchangeable genitive noun phrases, it is possible to determine the factors that favour one linguistic variant over another. For example, we might find that 75 per cent of all interchangeable noun phrases with an *s*-genitive have an animate modifying noun (e.g., *the president's* book), while only 20 per cent of the noun phrases with an *of*-genitive have an animate modifying noun (e.g. *the main goal of the president*). In this case, we could conclude that an animate modifying noun favours the *s*-genitive over the *of*-genitive. (It is also possible to analyse the complete set of interacting predicting factors through a logistic regression; see below.)

In contrast, each text is an observation in the text-linguistic design. In this case, the rate of occurrence is determined for each grammatical feature in each text, and it is subsequently possible to compute means and standard deviations for those rates in different registers. For example, from this perspective *of*-genitives occur with a mean of 30.2 per 1,000 words in a corpus of 2,005 science research articles (standard deviation = 6.0), and 36.6 per 1,000 words in a corpus of 2,005 history research articles (standard deviation = 8.1) (see Biber and Gray 2013: 122).

These two types of research design can lead to opposite conclusions regarding which linguistic form is more 'common'. For example, Figure 21.1 presents corpus findings regarding the use of *s*-genitives versus *of*-genitives from a variationist perspective (based on Biber et al. 1999: 302).[3] At first sight, these findings might lead one to conclude that *s*-genitives are more

---

[3] These findings are based on analysis of all *s*-genitives and all *of*-phrases modifying a head noun, with no consideration of interchangeability.
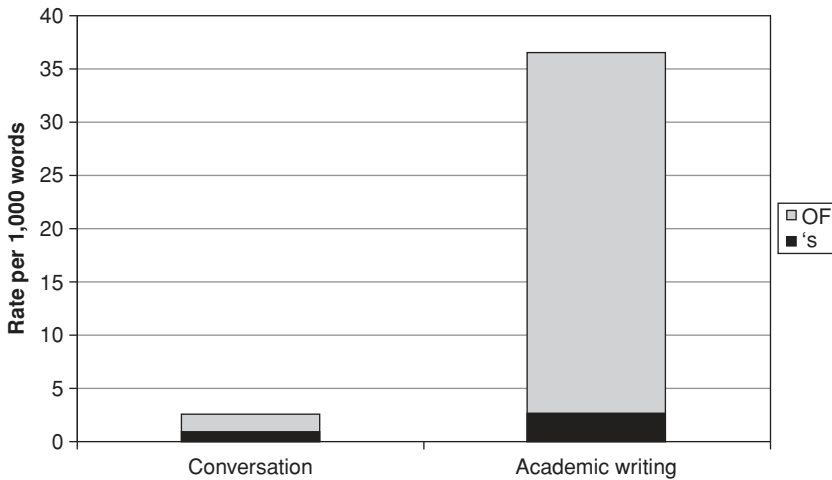
Figure 21.2 *Rate of occurrence for genitive constructions in late twentieth-century conversation vs. academic writing (based on Biber et al. 1999: 302)*

common in conversation than in academic writing: *c*.30 per cent of all genitives in conversation are *s*-genitives, while only *c*.5 per cent of all genitives in academic writing are *s*-genitives. However, because these findings are presented from a variationist perspective, they report proportional preference; they do <u>not</u> actually tell us how often a listener/reader will encounter these structures in texts.

In contrast, a text-linguistic design can be used to investigate the rates of occurrence for these different grammatical features. In this case, as Figure 21.2 shows, we would come to exactly the opposite conclusion: *s*-genitives have a higher rate of occurrence in academic writing (*c*.2.5 occurrences per 1,000 words) than in conversation (*c*.0.8 times per 1,000 words).

The apparent contradiction between the two approaches arises because the overall use of genitive constructions (combining all *s*- and *of*-genitives) is much higher in academic writing than in conversation: only *c*.2.5 total genitives per 1,000 words in conversation versus *c*.34 total genitives per 1,000 words in academic writing. As a result, the proportion of *s*-genitives is higher in conversation (0.8/2.5 = *c*.30 per cent; see Figure 21.1), while the actual rate of occurrence for *s*-genitives is higher in academic writing (see Figure 21.2).

In addition to the nature of the observations (or units of analysis), a second major difference between the two research designs has to do with the population of linguistic instances included in the analysis. Variationist analyses are restricted to a sample of linguistic tokens that are *interchangeable* variants of the same linguistic variable (Labov's 1966: 49 'principle of accountability'). The theoretical motivation is that analyses should be restricted only to those linguistic tokens where speakers are genuinely making a choice. Identifying interchangeable tokens of a linguistic variable is a major step in variationist

analyses, which often results in a greatly reduced sample of linguistic tokens (see section 21.2). For example, in Hinrichs and Szmrecsanyi (2007), *c*.64 per cent of all genitives were considered to be interchangeable; the other 36 per cent were excluded from the analysis.

A related consideration is that the variants of a linguistic variable can differ dramatically in the extent to which they are interchangeable. For example, in the variationist comparison of genitives by Hinrichs and Szmrecsanyi (2007), *c*.80 per cent of all *s*-genitives in their sample were interchangeable and therefore included in the analysis, but only *c*.56 per cent of all *of*-phrases (modifying a head noun) were coded as interchangeable.

These differences in inclusion criteria can have major implications for subsequent conclusions. For example, relying on a variationist sample of interchangeable tokens for genitives, Leech et al. (2009: 225; figure 10.5) show that 58 per cent of these structures were *s*-genitives in 1991; and based on that finding, they conclude that 'by 1991, the *s*-genitive had overtaken the *of*-genitive in frequency'. However, it is crucially important to be aware of the methodological basis of such claims: this finding is based on the set of *interchangeable* tokens, which is very different from the total set of occurrences for these features. Thus, Figure 21.2 – based on *all* occurrences of *s*-genitives and *of*-phrases modifying a noun – shows a very different pattern, with *of*-phrases being much more common than *s*-genitives (especially in academic writing).[4]

In the following sections, we further discuss and illustrate these methodological considerations through a case study of historical change in the use of genitive constructions. We define 'genitives' broadly to include any constructions that involve a noun phrase serving as modifier of a head noun. In particular, we investigate the use of three structural variants: *s*-genitives, *of*-genitives, and premodifying nouns. This three-way choice can be studied from a variationist perspective, and all three linguistic features can be investigated from a text-linguistic perspective. The following descriptions compare and contrast the kinds of historical patterns that can be discovered through each approach.

21.2   Methods

The study is based on an analysis of a sub-corpus of ARCHER (see Biber et al. 1994a), including all BrE texts from the registers of personal letters, newspaper reportage, and science articles. In total, the corpus used in this study comprises 327 texts and nearly 390,000 words. These three registers were chosen because they differ with respect to their primary communicative purposes, their interpersonal focus, and their intended audience. Taken

---

[4] The Leech et al. findings are based on analysis of a general corpus of written registers, while Figures 21.1 and 21.2 present results for specific registers.

together, inclusion of these registers allowed us to investigate the ways in which patterns of linguistic variation are mediated by register differences (see also Biber 2012).

We coded all texts to identify occurrences of *s*-genitives, *of*-genitives, and premodifying nouns, and determine their interchangeability with the other two variants. For *s*-genitives and *of*-genitives, we followed the methods used in Szmrecsanyi and Hinrichs (2008) and Wolk et al. (2013), and we then developed a similar set of methods for coding premodifying nouns.

The first step was to automatically identify potential cases of each of the three variants. For the genitives, we searched for *of* and final **'s/*s'* (as well as final **s* in seventeenth- and eighteenth-century texts, because apostrophes were often omitted from *s*-genitives during that period). For the premodifying nouns, we used the Biber Tagger[5] (see Biber et al. 1999: 35–6) to automatically identify nouns and search for instances of two adjacent nouns. We then manually coded each occurrence, to eliminate cases that were not genitives (e.g. *of*-phrases as part of prepositional verbs, such as *think of*, *speak of*, *be composed of*), and to then mark the boundaries of the two noun phrases in the remaining cases. In the following discussion, we refer to the two parts of these constructions as the 'possessor' and the 'possessum', even though most instances of genitives do not actually express the meaning of possession.

For the purposes of the variationist study, we analysed each genitive construction by hand to determine if it was 'interchangeable' with one or both of the other two variants. This step was based on our intuitions, deciding whether the structure was functionally equivalent and could be rephrased with another variant to express roughly the same meaning. (There has been considerable debate over the years of the extent to which grammatical variants are truly equivalent or interchangeable; see, e.g., Lavandera 1978; Dines 1980; Weiner and Labov 1983. Stefanowitsch (2003) includes a critical discussion of similar issues with respect to genitive constructions.) In general, we required that the rephrasing use the same words (e.g. *the county justices* versus *the county's justices*). There were two main modifications to this rule, where we additionally allowed:

1 the optional addition or deletion of a determiner to the possessum for *of*-genitives (e.g. *the government's policy* ↔ *the policy of the government*);
2 the optional pluralization or singularization of the possessor for premodifying nouns (e.g. *home prices* ↔ *prices of homes*).

The following special cases were coded as not interchangeable:

1 phrases that have been conventionalized (e.g. *Murphy's law*, *post office*);

---

[5] The Biber tagger has both probabilistic and rule-based components, uses multiple large-scale dictionaries, and runs under Windows. The tagger has been used for many previous large-scale corpus investigations, including multi-dimensional studies of register variation (e.g. Biber 1988) and the *Longman Grammar of Spoken and Written English* (Biber et al. 1999).

2   constructions in which an *s*-genitive is not followed by an explicit posses-sum phrase (e.g. *an associate of John's*);

3   titles of books, films, etc. that are premodified with an *s*-genitive (e.g. *Van Gogh's Starry Night*);

4   measures expressed as *of*-genitives (e.g. *three gallons of milk*);

5   *of*-genitives where the possessor noun phrase has a post-modifier (e.g. *the girlfriend of the man that I met*);

6   noun premodifiers that are not definite (since the possessum in *s*-genitives is always definite; e.g. *a London college*).

Beyond the guidelines enumerated here, coders were instructed to rely on their best judgement to determine interchangeability. After several rounds of trial coding and subsequent revisions to the coding scheme, two coders rated several texts in order to measure inter-coder reliability. Reliability was calculated for each of the nominal modifiers using simple percent agreement and Cohen's k.[6] The *s*-genitive ($N = 84$) coding achieved a simple agreement rate of 95 per cent and a 'very good' Cohen's k of 0.91. Reliability for *of*-genitives ($N = 112$) achieved a simple agreement of 90 per cent and a 'very good' Cohen's k of 0.80. Finally, the reliability analysis for premodifying nouns ($N = 91$) yielded a lower, yet still acceptable simple percent agreement of 85 per cent, with a 'good' Cohen's k of 0.69. Computer programs were developed to automatically count each of the features of interest. The text-linguistic analysis was carried out based on the normalized (per 1,000 words) rates of occurrence for each of the texts in the corpus. The variationist analysis, on the other hand, was based on calculating proportions of the raw counts.

21.3   The variationist analysis of genitive constructions

As described in preceding sections, the first step in the variationist analysis was to consider each linguistic token, to determine if it was interchangeable with one or both of the other variants. One methodological disadvantage of considering three variants is immediately apparent in such an analysis: there are many more alternatives to consider than in a study of a simple dichoto-mous choice. For example, *s*-genitives are analysed to determine whether they are interchangeable with *of*-genitives, interchangeable with nouns as nominal premodifiers, interchangeable with both *of*-genitives and nouns as nominal premodifiers, or not interchangeable at all. Similarly, *of*-genitives and nouns as nominal premodifiers are all coded to identify instances that are interchangeable with one, both, or neither of the other two variants.

---

[6] Reliability was calculated for each of the nominal modifiers using simple per cent agreement and Cohen's kappa (*k*). Cohen's *k* was chosen because it is a more robust measure of agreement than simple per cent of agreement in that it accounts for agreement that occurs by chance. If raters agree completely, $k = 1$, and $k = 0$ if agreement among raters is at or below the level expected by chance.

Table 21.1 *Interchangeable occurrences of* s-*genitives*

Register	Period	Total s-genitives	Interchangeable with *OF*	Interchangeable with N–N	Interchangeable with *OF* and N–N
Letters	18th c.	118	100 (85%)	7 (6%)	6 (5%)
	19th c.	96	81 (84%)	2 (2%)	0 (0%)
	20th c.	67	52 (78%)	3 (4%)	1 (1%)
News	18th c.	304	260 (86%)	23 (8%)	21 (7%)
	19th c.	131	109 (83%)	3 (2%)	1 (1%)
	20th c.	297	271 (91%)	33 (11%)	31 (10%)
Science	18th c.	90	84 (93%)	22 (24%)	22 (24%)
	19th c.	75	67 (89%)	18 (24%)	18 (24%)
	20th c.	37	35 (95%)	6 (16%)	5 (14%)

Tables 21.1–21.3 present the results of this coding. Table 21.1 shows the extent to which *s*-genitives are interchangeable with the other two variants. Most *s*-genitives are interchangeable with *of*-genitives, across registers and across historical periods. In contrast, few *s*-genitives are interchangeable with premodifying nouns: as low as 5–10 per cent in letters and newspaper articles, and *c*.20 per cent in science prose. Nearly all *s*-genitives that are interchangeable with premodifying nouns are also interchangeable with *of*-genitives (as shown by the last column of Table 21.1).

Of-genitives are more consistently interchangeable with both of the two other variants, but they show greater differences across registers and periods (see Table 21.2). In letters, *c*.30–35 per cent of the occurrences of *of*-genitives are interchangeable with *s*-genitives, and *c*.25 per cent are interchangeable with premodifying nouns. Those patterns hold across periods. A higher proportion of *of*-genitives are interchangeable with *s*-genitives in newspaper articles: *c*.50 per cent across periods. But newspaper articles are similar to letters in that only *c*.25 per cent of *of*-genitives are interchangeable with premodifying nouns. Science writing is interesting in that it shows an apparent historical increase in the proportion of *of*-genitives that are interchangeable: from *c*.30 per cent in the eighteenth century to *c*.50 per cent in the twentieth century. This same pattern holds for both interchangeability with *s*-genitives and with premodifying nouns. Finally, the last column in Table 21.2 indicates that interchangeability with *s*-genitives versus interchangeability with premodifying nouns are relatively independent, since the figures for

Table 21.2 *Interchangeable occurrences of* of-*genitives*

Register	Period	Total of-genitives	Interchangeable with 'S	Interchangeable with N-N	Interchangeable with 'S and N-N
Letters	18th c.	767	275 (36%)	176 (23%)	54 (7%)
	19th c.	420	144 (34%)	94 (22%)	34 (8%)
	20th c.	331	101 (31%)	87 (26%)	19 (6%)
News	18th c.	1778	910 (51%)	487 (27%)	168 (9%)
	19th c.	1558	796 (51%)	346 (22%)	128 (8%)
	20th c.	1240	617 (50%)	336 (27%)	160 (13%)
Science	18th c.	1589	484 (30%)	485 (31%)	465 (29%)
	19th c.	1595	678 (43%)	682 (43%)	663 (42%)
	20th c.	1597	753 (47%)	817 (51%)	651 (41%)

Table 21.3 *Interchangeable occurrences of nouns as noun modifiers*

Register	Period	Total nouns as noun modifier	Interchangeable with 'S	Interchangeable with OF	Interchangeable with 'S and OF
Letters	18th c.	72	21 (29%)	43 (60%)	18 (25%)
	19th c.	69	10 (14%)	34 (49%)	8 (12%)
	20th c.	159	28 (18%)	74 (47%)	26 (16%)
News	18th c.	299	23 (8%)	85 (28%)	19 (6%)
	19th c.	379	52 (14%)	156 (41%)	42 (11%)
	20th c.	984	109 (11%)	330 (34%)	88 (9%)
Science	18th c.	212	43 (20%)	120 (57%)	42 (20%)
	19th c.	263	22 (8%)	120 (46%)	13 (5%)
	20th c.	1122	176 (16%)	577 (51%)	135 (12%)

three-way interchangeability are considerably lower than either of the other two columns.

The patterns of interchangeability for nouns as premodifiers (Table 21.3) are similar to those for *s*-genitives: relatively few nouns as premodifiers are interchangeable with *s*-genitives (across registers and periods), but *c*.50 per cent of nouns as premodifiers are interchangeable with *of*-genitives (with

lower proportions in newspaper articles). And here again, we see that if a premodifying noun is interchangeable with an *s*-genitive, that token will usually also be interchangeable with an *of*-genitive (as shown by the last column of Table 21.3).

Two general patterns are noteworthy here as background to the interpretation of variationist findings:

1   In general, many tokens of genitive constructions are <u>not</u> interchangeable. In fact, only 10–50 per cent of occurrences are interchangeable for many of the comparisons. The one exception here is for *s*-genitives, which are usually interchangeable with *of*-genitives (Table 21.1). But otherwise, fewer than 50 per cent of the occurrences of these constructions are interchangeable with other variants.
2   The extent of interchangeability varies considerably across constructions, across registers, and to some extent, across periods.

While identifying interchangeable tokens is one of the first steps in a variationist perspective, there is usually little consideration of the extent of interchangeability. That is, the analysis is focused on the linguistic variable, operationally defined as the set of interchangeable occurrences. As a result, the extent to which that set of variants represents the total pool of linguistic occurrences has generally been disregarded as theoretically irrelevant. However, a complete historical description of a structural domain must also account for the patterns of variation and change for the non-interchangeable occurrences. We briefly return to this point below and then again in the conclusion.

In the remainder of the present section, though, we adopt the variationist perspective, considering the patterns of variation within the set of interchangeable occurrences of genitive constructions. Figure 21.3 presents our findings for the alternation that has been the focus for most previous work on genitives: the choice between *s*-genitives versus *of*-genitives in constructions where the two are interchangeable.[7] The patterns shown in this figure provide some support for earlier claims that the *s*-genitive has been increasing historically at the expense of the *of*-genitive (see, e.g., Potter 1969; Leech et al. 2009). However, this figure also shows that this historical trend is mediated by register differences. Thus, in personal letters, the *s*-genitive increased proportionally in use during the nineteenth century, but that pattern then remained relatively stable over the course of the twentieth century. In newspaper reportage, *of*-genitives became even more strongly preferred in the nineteenth century, followed by a strong shift towards *s*-genitives during the twentieth century. As a result, newspaper writing and personal letters are relatively similar in showing *c.*30–35 per cent proportional use of *s*-genitives in the latter part of the twentieth century. However, the historical

---

[7]   Figure 21.3 is based on all tokens of *s*-genitives and *of*-genitives that can be interchangeable with one another, including tokens that could also be interchangeable with nouns as nominal premodifiers.
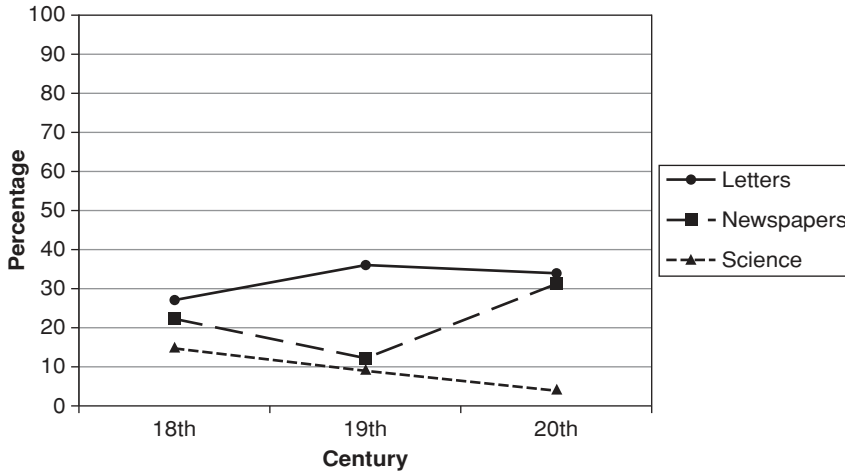
Figure 21.3 *Historical change in the proportional use of the* s-*variant (vs. the* of-*variant)*
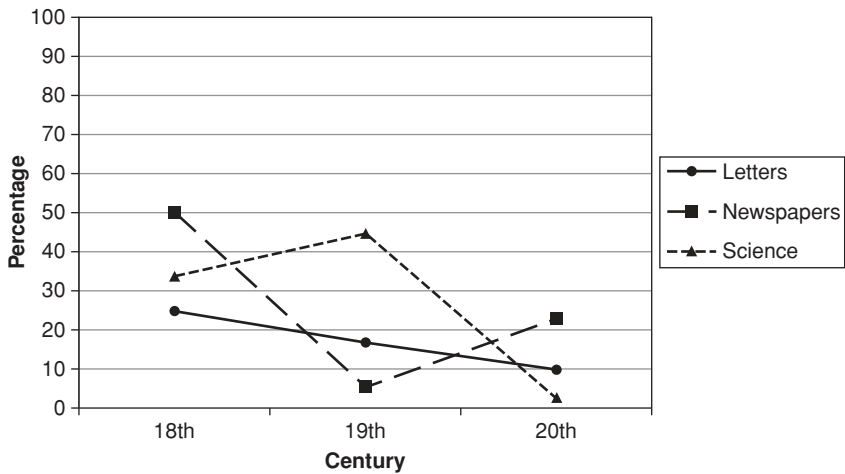

Figure 21.4 *Historical change in the proportional use of the* s-*variant (vs. the premodifying noun variant)*

trend in science prose contrasts with both letters and newspaper reportage: a small reliance on *s*-genitives in the eighteenth century, followed by a steady decline in proportional use over the following two centuries. As a result, only *c.*5 per cent of interchangeable genitive constructions are realized as *s*-genitives in twentieth-century science prose.

The historical trends are less consistent in Figure 21.4, which plots the proportional use of interchangeable *s*-genitives versus nouns as premodifiers. This is due in part to the fact that these two construction types are in general not interchangeable, and thus the proportions shown in Figure 21.4 are based on very small samples. For example, the nineteenth-century
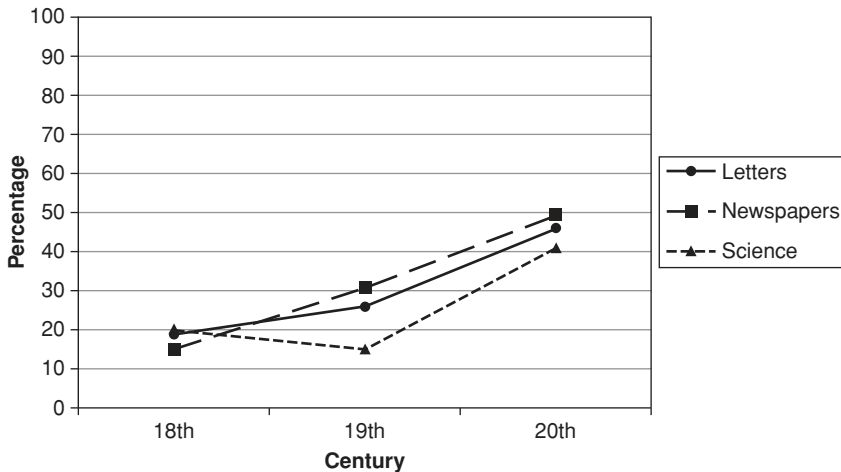
Figure 21.5 *Historical change in the proportional use of the premodifying noun variant (vs. the* of*-variant)*

proportion of 45 per cent *s*-genitives in science prose is based on a sample of only forty tokens (eighteen *s*-genitives that are interchangeable with noun-premodifiers – see Table 21.1, and twenty-two noun-premodifiers that are interchangeable with *s*-genitives – see Table 21.3). Despite the fluctuations, the overall historical trends are consistent across registers, with a notable increase in the proportional use of noun-premodifiers (and decline in the proportional use of *s*-genitives) across the centuries. Science prose shows the strongest increase, with noun-premodifiers being used over 95 per cent of the time in interchangeable constructions from the twentieth century.

The sample of interchangeable occurrences for the *of*-genitive versus noun-premodifier alternation is much larger (see Tables 21.2 and 21.3), and correspondingly, the historical trends shown in Figure 21.5 are much more consistent across centuries. For all three registers, there is a strong increase in the proportional use of noun-premodifiers at the expense of *of*-genitives. Letters and newspaper reportage take the lead in this change during the nineteenth century, while science prose shifted strongly towards increased noun-premodifier variants during the twentieth century (see also Biber and Gray 2011a, 2013; Berlage 2014).

Finally, Figure 21.6 plots the proportional use of variants for those occurrences of genitives that can take all three variants. Similar to Figure 21.4, Figure 21.6 is based on small sample sizes for many data points (see the right columns in Tables 21.1–21.3), but one trend stands out from Figure 21.6: the historical increase in the preference for noun-premodifiers in cases where all three variants are possible. This increase is most pronounced in letters and newspaper reportage, but the same trend occurs to a lesser extent in science prose.
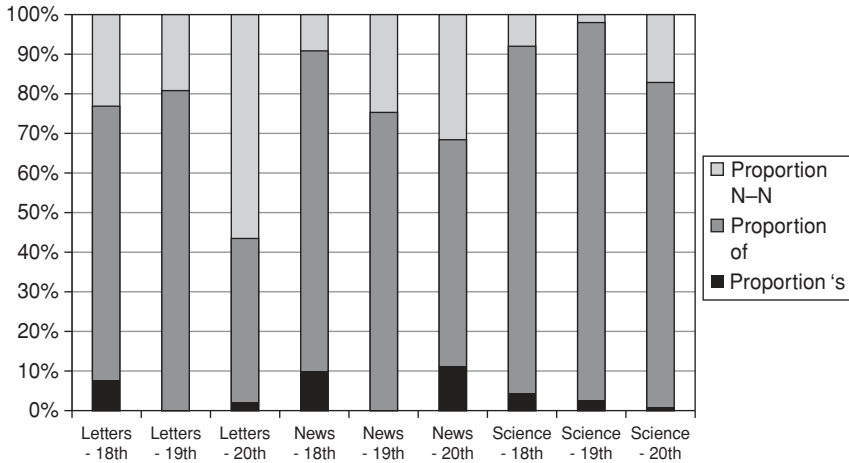
Figure 21.6 *Proportional use of* s-*genitives,* of-*genitives, and nouns as nominal premodifiers – for noun phrases that can take all three variants*

In summary, the variationist findings show that

1   In some registers (letters and newspapers), there has been an increase in the proportional use of s-genitives at the expense of of-genitives.
2   In other registers (science prose), s-genitives have actually decreased proportionally in comparison to of-genitives.
3   There has been a strong increase, across registers, in the proportional use of noun-premodifiers, at the expense of both s-genitives and of-genitives.

21.4   The text-linguistic analysis of genitive constructions

There are actually two different research designs that can be used for text-linguistic analyses of a grammatical feature. The simplest design is to treat each sub-corpus as an observation, computing an overall rate of occurrence for each sub-corpus. For example, Figure 21.7 plots historical change in the rate of occurrence for *of*-phrases (regardless of syntactic function), based on analysis of COHA (see Davies 2012b). In this case, the sub-corpus for each decade is treated as a single observation, and so we computed a single rate for all the combined texts within a decade. The advantages of this approach are that it is efficient, and in the case of corpora like COHA, it permits consideration of very large samples. The major disadvantage of this approach is that it is not possible to compute a statistical measure of dispersion, so it is difficult to determine the extent to which the use of a feature varies across texts within a sub-corpus.

An alternative research design used for text-linguistic analyses is to treat each individual text as an observation. That is, we can compute a rate of occurrence for the grammatical feature in each text, making it possible to then compute a mean score for all the texts in a category (e.g. a register
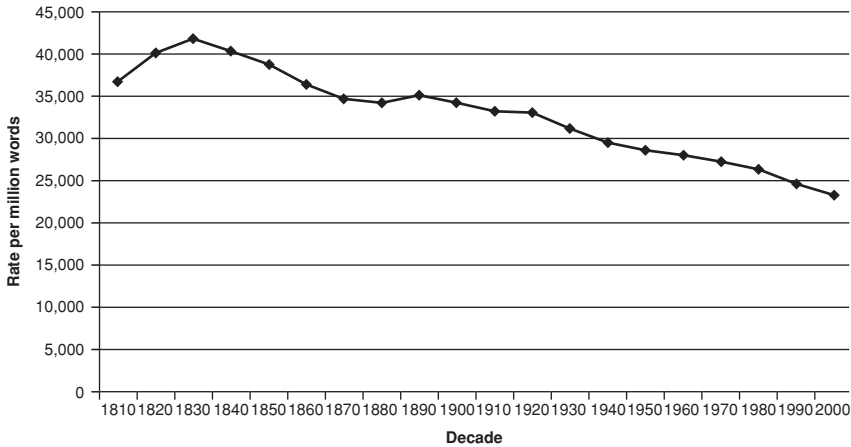
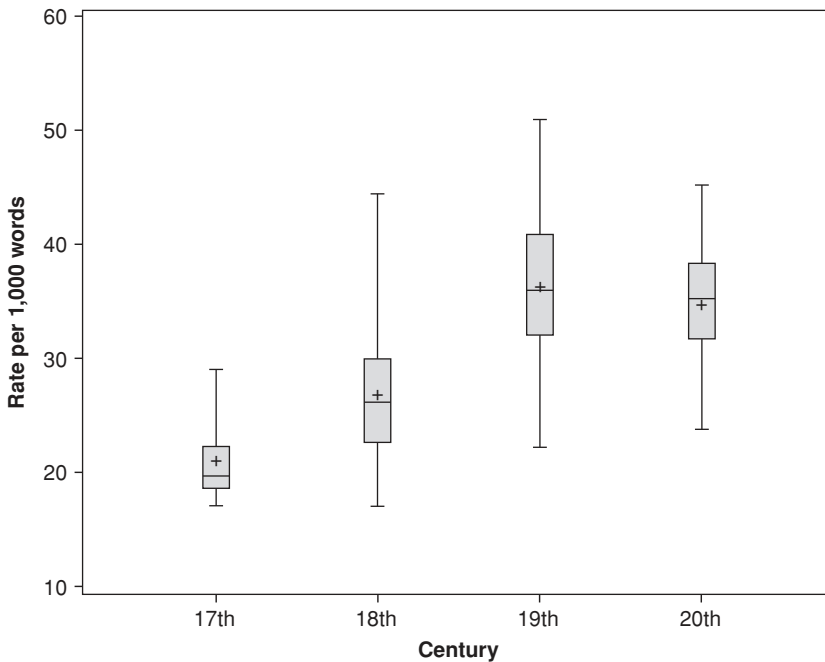Figure 21.7 *Historical change in the rate of occurrence for* of*-phrases (in COHA)*

Figure 21.8 *Distribution of* of*-genitives in science articles across centuries*

or a historical period). In this case, it is also possible to compute measures of dispersion, showing the extent of variability among the texts within a category. For example, Figure 21.8 displays box plots for the use of *of*-genitives in science articles, providing information about the central tendency and the range of variation in each century. (For example, the '+' on Figure 21.8 shows the mean score, and the boxes show the range of the first and third quartiles.) Similarly, Figure 21.9 displays a scatter plot for newspaper texts, showing the year of each individual text correlating with the rate of occurrence
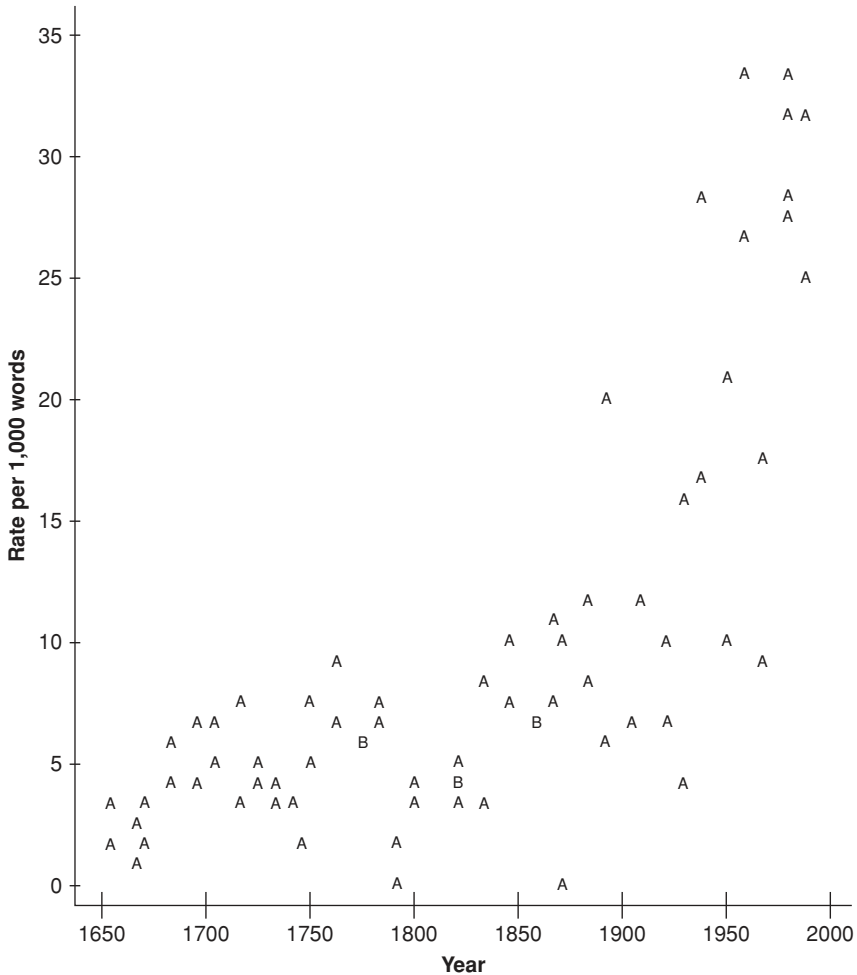
Figure 21.9 *Distribution of nouns as premodifiers in newspapers, across years.*
Legend: A = 1 observation, B = 2 observations

for noun premodifiers in that text. Similar to Figure 21.7, these graphs capture historical trends in the use of grammatical features. However, they additionally show the variability among texts within historical periods.

Because they allow measures of dispersion (and also generally include a large number of observations – the texts), text-linguistic designs based on analysis of each text (rather than each sub-corpus) also allow us to compute various statistics that test for significant differences among categories, and measure the strength of relationships. For example, Table 21.4 presents Pearson correlations for the use of the three types of noun phrase modifiers correlated with time (i.e. the year of the text). Pearson correlations measure the strength of the relationship between two numeric variables. Correlation coefficients have a scale of −1 to +1: a value near −1 represents a strong

Table 21.4 *Historical change in the use of general linguistic features, shown by Pearson correlation coefficients (r) for the rate of occurrence correlated with date (1650–1990)*
Key:
.60 to .99 = +++
.30 to .59 = ++
.20 to .29 = +
−.20 to −.29 = −
−.30 to −.59 = −−

	Letters N = 187 texts		Newspapers N = 70 texts		Science prose N = 70 texts	
	trend	*r*	trend	*r*	trend	*r*
s-genitives		−.09	+	.26		−.10
of-genitives	−	−.27		.12	+++	.61
premodifying nouns	++	.41	+++	.74	+++	.73

decrease in use over time; a value near +1 represents a strong increase in use over time; and a value near 0.0 indicates that there has not been any consistent pattern of change over time. These correlations measure linear historical trends, regardless of the overall extent to which a feature is used. For example, *of*-genitives in newspaper reportage have a small correlation of only .12 with year. This correlation tells us that the rate of occurrence for *of*-genitives in newspaper reportage has changed little over time; but, it does not tell us whether *of*-genitives have been frequent or rare overall.

Table 21.4 shows that there have been important historical changes in the use of these features. *S*-genitives have increased slightly in newspaper reportage but their use has remained essentially unchanged in the other two registers. *Of*-genitives have increased strongly in use in science articles, but otherwise have changed little in use in the other registers (and actually declined slightly in personal letters). In contrast, premodifying nouns have increased notably in all registers, and very strongly in newspapers and science articles (see also Biber and Gray 2011a, 2013; Berlage 2014).

Figures 21.10–21.12 summarize these historical developments graphically, and further compare the actual magnitude of use for each of the three features. Figure 21.10 plots the patterns of change in personal letters: very little change in the use of *s*-genitives; a moderate decline in the use of *of*-genitives; and a moderate increase in the use of premodifying nouns. Both *s*-genitives and premodifying nouns are considerably less common than *of*-genitives in this register. Newspaper reportage (Figure 21.11) shows somewhat different historical patterns: little change in the use of *s*-genitives, but with a slight increase in the twentieth century; a nineteenth-century increase in the use of *of*-genitives, followed by a twentieth-century decline; and a strong
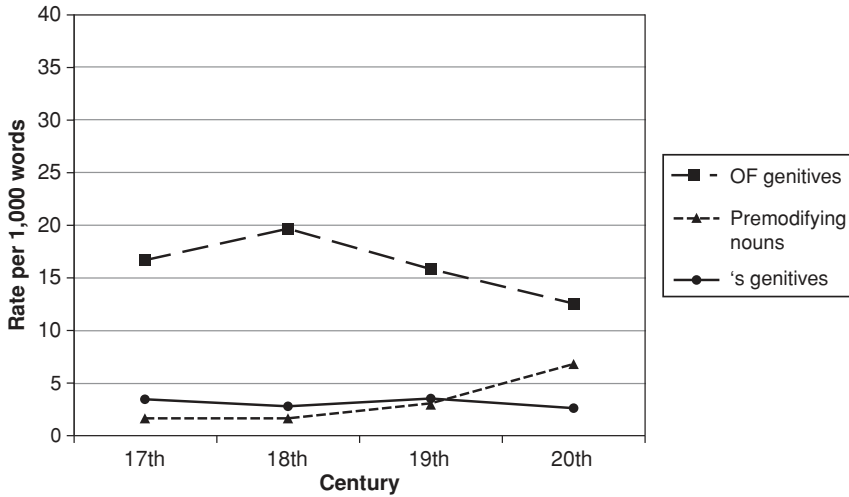
Figure 21.10 *Historical change in the use of genitive features in personal letters*
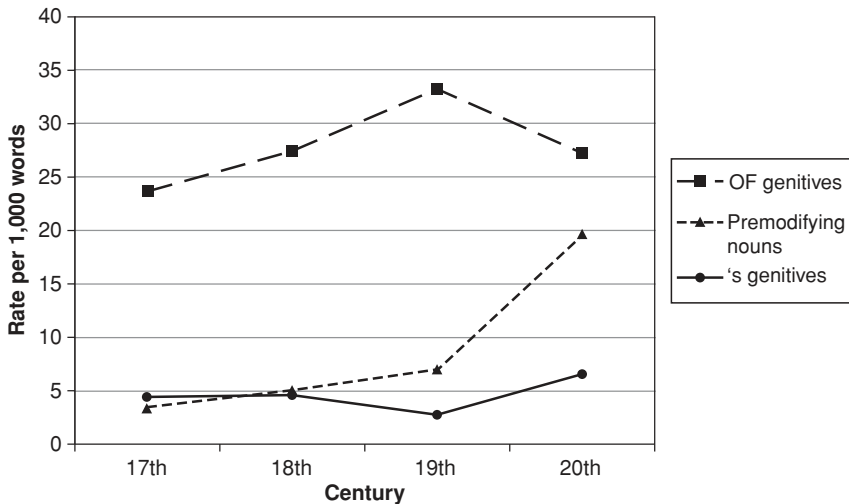

Figure 21.11 *Historical change in the use of genitive features in newspaper reportage*

twentieth-century increase in the use of premodifying nouns. In science articles (Figure 21.12), *s*-genitives have always been rare, and they have become even less common in the twentieth century. In contrast, *of*-genitives have always been relatively common; they increased strongly in use during the nineteenth century; and they have decreased only slightly in the twentieth century. However, the most notable historical change in science articles is the strong twentieth-century increase in use for premodifying nouns.
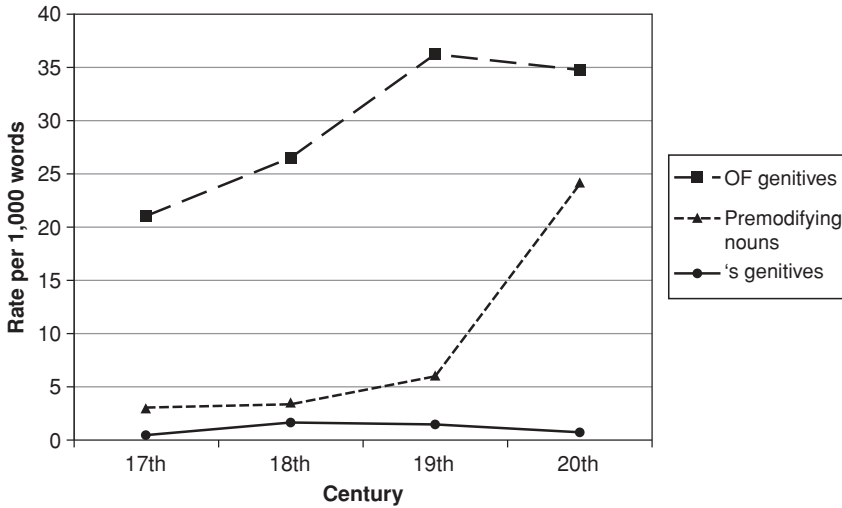
Figure 21.12 *Historical change in the use of genitive features in science articles*

In summary, these findings show several general patterns:[8]

1 There are important differences across registers, in the extent of historical change, the direction of change, and the particular features affected by change. Thus, consideration of only a single register, or analysis of a general corpus with no consideration of register differences, will obscure these more systematic patterns of change within registers.

2 S-genitives are generally rare in Modern English in comparison to these other options for noun modification. (S-genitives have increased slightly in newspaper reportage, but they are still rare in comparison to of-genitives and premodifying nouns.)

3 The of-genitive was especially important in informational prose in the nineteenth century, when it increased in use in both newspaper

---

[8] Text-linguistic research designs based on analysis of each text also allow the application of inferential statistical techniques to test for significant differences. For example, the following table summarizes the results of a factorial ANOVA, testing the statistical significance of the mean differences across historical periods and across the three registers. Both main effects show significant differences (except for *s*-genitives across historical periods). In addition, there are significant interaction effects for all three linguistic features, reflecting the different directions and extents of change across registers.

Table 21.1n *Summary of the ANOVA factorial models for three registers (letters, newspapers, science prose) across centuries*

	Model F-Score	Model signif.	Model R^2	Period	Register	Period* Register
s-genitives	6.84	<.0001	0.193	ns	<.0001	<.001
of-genitives	34.83	<.0001	0.549	<.0001	<.0001	<.0001
premodifying nouns	38.68	<.0001	0.575	<.0001	<.0001	<.0001

reportage and science articles. This structure has declined in use in the twentieth century, especially in newspaper reportage.

4   Premodifying nouns are increasing in use in all three written registers. This increase has occurred primarily in the twentieth century, and it has been strongest in the informational written registers (especially science articles).

21.5   Putting it all together

The analyses presented in sections 21.3 and 21.4 have shown how the variationist and text-linguistic approaches yield distinct, yet complementary, descriptions of grammatical change in the use of genitive constructions. At the same time, we hope to have demonstrated the importance of three methodological practices for historical analysis:

1   the need to include the full set of linguistic variants that are potentially relevant in a structural shift;
2   the need to include a range of register variation;
3   the need to consider both variationist and text-linguistic research designs.

As the descriptions in sections 21.3 and 21.4 show, incomplete – and possibly misleading – conclusions would result from more restricted analyses. For example, consideration of only s-genitives versus of-genitives would fail to capture the important shift to the use of premodifying nouns, apparently becoming the preferred choice at the expense of both of the other two options. Consideration of only newspaper reportage would suggest an increase in the use of s-genitives, and a decrease in the use of of-genitives – patterns that are opposite to those found in science articles. And consideration of only variationist or text-linguistic designs, which approach quantitative data in distinct ways, would result in very different conclusions about the magnitude and direction of these historical changes.

   In future research, we hope to explore these patterns in much more detail. For example, we plan to use regression analyses to identify the contextual factors that are most influential in predicting these linguistic choices. We also plan to further explore the reasons for non-interchangeability, including consideration of why some variants are more likely to be interchangeable than others.

   Our goals here, however, have been more methodological, arguing that the study of grammatical change requires carefully crafted empirical research designs. First, grammar is not (necessarily) a set of binary grammatical alternations, so analysts should consider the full set of variants. Second, when choosing data sources it is crucial to keep in mind that register variation may interact with historical variation, and vice versa. Third, the choice

between variationist and text-linguistic research designs has important consequences for subsequent conclusions: the former method explores the factors that influence the linguistic choices that language users make, while the latter approach explores the frequency with which language users use particular linguistic forms in texts. The choice of method also has practical ramifications: variationist designs require potentially laborious coding for interchangeability, while the frequency measurements that underpin the text-linguistic approach are typically more straightforward. Thus, our main goal here has been to lay the foundation for an integrated approach that reconciles the two research designs.