The great regression: genitive variability in Late Modern English news texts

Benedikt Szmrecsanyi

to appear in: Börjars, Kersti, David Denison & Alan Scott (eds.) (in press), *Morphosyntactic categories and the expression of possession*. Amsterdam: Benjamins.

Abstract

Utilizing the variationist method, this contribution is concerned with the alternation between the *s*-genitive (*the president's speech*) and the *of*-genitive (*the speech of the president*) in Late Modern English news prose as sampled in ARCHER. A frequency analysis reveals that text frequencies of the *s*-genitive collapsed in the early 19th century, but recovered afterwards. Linear regression analysis indicates that slightly over half of this frequency variability is induced by "environmental" changes in the news genre habitat, such as varying input frequencies of human possessors. To investigate the remaining variability, we fit a logistic regression model and show that genitive choice grammars changed genuinely in regard to four language-internal conditioning factors: POSSESSOR ANIMACY, GENITIVE RELATION, POSSESSUM LENGTH, and POSSESSOR THEMATICITY. Applying customary grammaticalization diagnostics, we conclude that while the *s*-genitive was subject to grammaticalization in the 19th century, it actually degrammaticalized during the 20th century.

Keywords: genitive variability, historical linguistics, variationist paradigm, ARCHER, news prose, regression analysis, grammaticalization

Acknowledgments

I wish to thank Maren Holzkamp, Anja Kapust, and Angela Kurtz for help in coding the present study's genitive dataset, Katharina Ehret for help with formatting the manuscript, and Martin Hilpert, Lars Hinrichs, Anette Rosenbach, Christoph Wolk, and two referees for extremely helpful comments and suggestions. I am also grateful for the feedback to an earlier version of this paper presented at the 2009 Manchester Workshop on 'Morpho-syntactic categories and the expression of possession'. The usual disclaimers apply.

1. Introduction

This paper is an exercise in variationist model building (in the sense of, e.g., Labov 1969) based on historical corpus data. Tapping ARCHER (*A Representative Corpus of Historical English Registers*), we will seek to document the history of the grammatical alternation between the *s*-genitive, as in (1), and the *of*-genitive, as in (2), in the Late Modern English period.

(1) They daily expect here Plenipotentiaries from Holland, with a final answer upon **[the King]$_{por}$s [last Propositions]**$_{pum}$. <ARCHER 1672lon2.n2b>

(2) The Project for granting to the Government six-twelfths of the taxes for the service of the present year was agreed to, in the Chamber of Peers, on Thursday, by a majority of 108 to 1; and, having received **[the sanction]$_{pum}$ of [the KING]$_{por}$**, it is published officially as a law, in the Moniteur. <ARCHER 1819mor1.n5b>

While ARCHER samples many registers, the present study is concerned with genitive variability in the news genre specifically. News is known to be a particularly innovative and agile text type that is "unusually receptive to (and in a good many instances also productive of) innovations or changes 'from below.' " (Hundt and Mair 1999: 236). A frequency analysis will demonstrate that *s*-genitive frequencies in news texts collapsed rather abruptly around the first half of the 19th century, and gradually recovered afterwards. Why are we seeing these ups and downs? To what extent are "environmental" factors (i.e. genre-internal fads and fashions affecting the textual genitive habitat) to blame? To which degree are probabilistic changes in genitive choice grammars implicated? How do such probabilistic changes relate to the status of the *s*-genitive as an inflection or a clitic? Aside from such substantial questions, the paper contributes to the methodological state-of-the-art by marrying the interpretation of historical variability to modern probabilistic analysis techniques.

At this point a few comments are in order concerning the theoretical and analytical framework guiding this study. Our starting point is the understanding that many interesting aspects in language are probabilistic rather than categorical in nature (cf., for example, Bresnan et al. 2007 and the papers in Bod et al. 2003). The present study is specifically interested in syntactic change, a phenomenon which we take – in line with the probabilistic linguistics commitment – as involving not necessarily categorical changes (such as, say, the disappearance of verb-second word order in the history of English) but also more subtle changes affecting syntactic probabilities. Why is this subtlety interesting? Variationist sociolinguists have known for a long time that all change involves non-categorical synchronic variability and heterogeneity (Weinreich et al. 1968), with a sense (cf. Labov 1982: 75) that delicate changes affecting probabilistic factor weights are where the analyst can observe language change in action:

> In the process of change, linguistic forms gradually shift from one function to another. This trajectory can be viewed in the varying strength and distribution of independent linguistic features associated with one of the evolving grammatical morphemes. (Tagliamonte 2001: 747–748)

In exactly this spirit, we take the existence of competing syntactic variants (the *s*-genitive versus the *of*-genitive) as given and explore the degree to which syntactic change manifests in changing probabilistic weights, as a function of real time, of the conditioning factors (such as possessor animacy or the principle of end weight) that govern the genitive alternation. In the present study's parlance, any such subtle changes are evidence of 'changing genitive choice grammars'. The empirical challenge that will concern us throughout this paper is, of course, to keep apart genuine probabilistic changes from more trivial genre or topic induced frequency fluctuations.

The present study is structured in the following way. Section 2 gives a short history of the English genitive alternation. Section 3 presents the data source, ARCHER. Section 4 defines the variable context. Section 5 traces genitive frequencies in historical time. Section 6 discusses the eight language-internal conditioning factors on which the subsequent analyses are based. Section 7 is concerned with the extent to which changes in the news genre habitat are implicated in the overall frequency variability. Section 8 investigates genuine changes in genitive choice grammars. Section 9 is a discussion of the present study's major findings.

2. The history of genitive variation in English: an overview

As a backdrop for the subsequent empirical analysis, the following is a sweeping overview of genitive variability from Old English through the Present-Day English period. Needless to say, in historical terms the *of*-genitive is the incoming form, which appeared during the 9th century. Yet according to figures produced by Thomas (1931: 284) (cited in Mustanoja 1960: 75), the inflected genitive vastly outnumbered the periphrasis with *of* up until the 12th century. In the Middle English period, we begin to witness "a strong tendency to replace the inflectional genitive by periphrastic constructions, above all by periphrasis with the preposition *of*" (Mustanoja 1960: 70), to the extent that the inflected genitive came close to extinction (Jucker 1993: 121). The frequencies calculated by Thomas (1931) show that by the 14th century, the *of*-genitive had a market share of about 84% while the inflected genitive was increasingly confined to a functional niche coding animate possessors, possessive/subjective genitive relations, and topical possessors (Rosenbach 2002: 180-181). Somewhat surprisingly, then, Early Modern English sees a revival of the *s*-genitive, "against all odds" (Rosenbach 2002: 184). More specifically, during the Early Modern English period

> the *s*-genitive preferably expresses possessive and subjective function. Second, the s-genitive is almost exclusively restricted to genitive NPs with the semantic features [+animate] [+human] [+definite] [+referential]. Within these contexts the *s*-genitive stays in the English language and increases in occurrence. (Rosenbach and Vezzosi 2000: 301)

As for Present-Day English, empirical research has reported comparatively high frequencies of the *s*-genitive (cf., for instance, Hinrichs and Szmrecsanyi 2007; Szmrecsanyi and Hinrichs 2008; Rosenbach 2002). The consensus is that the *s*-genitive is spreading right now (for instance, Potter 1969; Dahl 1971; Raab-Fischer 1995; Rosenbach 2003; Hinrichs and Szmrecsanyi 2007; Szmrecsanyi 2009). There is other variability as well, though. For example, the more informal

the setting, the greater the preference for the *s*-genitive (Altenberg 1982: 284). The *s*-genitive is also more popular in American English than in British English (for instance, Rosenbach 2003; Hinrichs and Szmrecsanyi 2007; Szmrecsanyi and Hinrichs 2008), and is – interestingly – fairly widespread in presumably more conservative British English dialects (Szmrecsanyi 2006; Szmrecsanyi and Hinrichs 2008). With regard to the news genre specifically, Hinrichs and Szmrecsanyi (2007) diagnose a spread of the *s*-genitive in late 20th century press English which appears to be due to a process of "economization". Their argument is that tight information packaging is becoming increasingly more important, particularly in informational genres (cf. Biber 2003: 170). Because the *s*-genitive is the "characteristically more compact" (Biber et al. 1999: 300) coding option, the *s*-genitive is argued to have an edge over the *of*-genitive in news prose. The probabilistic analyses presented in Hinrichs and Szmrecsanyi (2007) and Szmrecsanyi and Hinrichs (2008) suggest that this edge manifests itself in tendencies such as the *s*-genitive being increasingly attracted to lexically dense genitive environments or thematically prominent possessors.

What is crucial in the context of the present study is that we know next to nothing about genitive variability in the Late Modern English period – a gap in the literature that the present study seeks to remedy, with an eye on the status of the *s*-genitive as an inflection or as a clitic.

3. Data

Our data source is ARCHER (*A Representative Corpus of Historical English Registers*), release 3.1 (cf. Biber et al. 1994). ARCHER covers the period between 1650 and 1990, spans 1.8 million words of running text, and samples eight different registers and two major varieties of English: British and American. The corpus design categorizes all texts into seven subperiods of 50 years, although the precise year of composition is also available for each text. In this study, we tap ARCHER's British English news texts section. This subcorpus comprises 70 texts (10 per 50-year period) à 2,000-2,500 words, totalling roughly 160,000 words of running text. The texts contain extracts taken from 27 different publications, which are listed in Table 1 and are typically up-market – at least in periods where the compilers had a choice between up-market and down-market publications. In this subcorpus, we are dealing with approximately 7,300 instances of the token *of*, 680 instances of *-'s*, and – in the first two periods – 5,500 tokens ending in *-s*. All of these instances were manually inspected and coded as genitival or not according to the guidelines set forth in the next section.

4. The variable context

Circumscribing the variable context and thus defining interchangeable genitive contexts is a tricky issue in historical linguistics. Variationist (socio)linguists working on modern data typically rely, implicitly or explicitly, on speaker intuitions to define the envelope of variation. In the historical realm, this procedure is of course not feasible as we do not have access to, say, an 18th-century news writer's intuitions on genitive interchangeability. Instead, the present study

1650-1699	Mercurius Politicus; Intelligencer, Published For Satisfaction and Information of the People; The London Gazette; The True Protestant Mercury or, Occurrences Foreign and Domestick; The Post Man and Historical Account
1700-1749	The Daily Courant; The Evening Post; The Daily Journal; Read's Weekly Journal, or British Gazetteer; The London Gazette
1750-1799	The London Evening Post; The Public Advertiser; The London Chronicle; The General Evening Post; The Star
1800-1849	Johnson's British Gazette and Sunday Monitor; The Morning Chronicle; Evans and Ruffy's Farmer's Journal and Advertiser; The Times; The Manchester Guardian
1850-1899	The People's Paper; Pall Mall Gazette; Glasgow Sentinel; The Times; The Manchester Guardian
1900-1949	Pall Mall Gazette; Daily Herald; The Times; Manchester Guardian
1949-1989	Manchester Guardian; The Sunday Times; Observer; Sunday Telegraph; The Times

Table 1: Publications sampled in ARCHER's British English news subcorpus.

adopts an exploratory approach to approximate contexts in which the *s*-genitive is interchangeable with the *of*-genitive in the time span subject to analysis (cf. Rosenbach 2002: 41 for a similar procedure).

We proceeded in the following way. Using `*'s`, `of`, and `*s` (the latter in the first two periods only, when a substantial number of *s*-genitives were spelt without an apostrophe) as search strings, we started out by manually extracting, in a strictly semasiological fashion, all occurrences matching the following patterns: (i) [full NP]'s [full NP], as in (3), (ii) [full NP]s [full NP], as in (4), and (iii) [full NP] *of* [full NP], as in (5).[1] Note that at this stage, we also hand-coded the boundaries of the possessor and possessum NP phrases (indicated by square brackets), as well as the possessor NP head noun (in italics).

(3) THE King and Queen are very well at present, **[her *Majesty*]$_{\text{por}}$'s [late Distempers]**$_{\text{pum}}$ having lasted but two days. <ARCHER 1697pos2.n2b>

(4) [. . .] and the Enemy not giving him any occasion to exercise his valour, his Excellence is returned according to **[his *Majestie*]$_{\text{por}}$s [order]$_{\text{pum}}$**, and within view of this Coast, to be revictualled, and enforced with a new Equipage. <ARCHER 1665int2.n2b>

(5) Meanwhile, **[a further tightening]$_{\text{pum}}$ of [baggage inspection *procedures*]$_{\text{por}}$** is likely to emerge from a review of Britain's airline and airport security. <ARCHER 1989tim2.n8b>

Crucially, we ignored cases where the possessor or possessum phrase is not nominal (i.e. does not have an actual noun as head) but pronominal (as in *some of them* or *her distempers*), since pronominal encoding of genitive relations is a different matter altogether (cf. Rosenbach 2002: 40). Moreover, in the case of the *of*-genitive specifically, we excluded *of*-tokens that are part

of titles (e.g. *the Duke of Normandy*), constructions where the possessor would be a nation or location (e.g. *the Borough of Southwark*), and the phrase *(by) the name of X*. However, heavily postmodified possessor/possessum phrases, as in (6), were included.

(6) Letters from Constantinople state that the only remaining point of difference between Turkey and Russia relate to **[the surrender]$_{\textbf{pum}}$ of [*Prince Suzzo* and other persons [who took refuge in Russia in the early part of the insurrection in Walachia and Meldavia]]$_{\textbf{por}}$**. <ARCHER 1822eva1.n5b>

The above guidelines left us with a total of $N = 4,507$ genitive instances in the dataset. In a next step, these instances were coded for (i) whether or not the possessum is explicitly marked for indefiniteness, as indicated by the presence of determiners such as *a* and *an*, and (ii) for the eight conditioning factors detailed in Section 6 below. This included coding genitive occurrences for the underlying genitive relation, distinguishing between six categories: ownership (e.g. *Tom's car*), kinship (e.g. *Tom's father*), body part (e.g. *Tom's knee*), part-whole (e.g. *the frame of the window*), partitive (e.g. *a number of beers*), and "other" (e.g. *a man of honor*). A subsequent frequency analysis showed that

1. as in Present-Day English, the *s*-genitive does not occur with indefinite possessums in the ARCHER data;
2. in partitive genitive relation contexts, the *s*-genitive has a negligible share of only .8%;
3. similarly, in part-whole genitive relations, the *s*-genitive has a share of only 1.1%.

For all intents and purposes, therefore, genitives with explicitly indefinite possessums as well as partitive and part-whole genitives do not constitute interchangeable contexts in our dataset. In order to not skew results, we therefore excluded such contexts, a step that left us with a pruned dataset spanning $N = 3,421$ genitives. The *s*-genitives and *of*-genitives in this dataset – which is fairly comparable to the modern datasets analyzed in Hinrichs and Szmrecsanyi (2007), Szmrecsanyi and Hinrichs (2008), and Szmrecsanyi (2010) – can be considered roughly interchangeable such that one variant may be substituted for the other with no semantic change (Labov 1966a,b).

5. Genitive frequencies over time: an overview

Reading the literature reviewed in Section 2, one might expect – what with the comeback of the *s*-genitive "against all odds" (Rosenbach 2002: 184) in the Early Modern English period – to see a further gradual linear increase in *s*-genitive frequencies between the Early Modern English period (during which the *s*-genitive reclaimed some of the market share it had lost to the *of*-genitive during the Middle English period) and the Present-Day English period, for which previous research has reported comparatively high text frequencies of the *s*-genitive. The frequency overview in Table 2, however, puts paid to this expectation. What we see is not a gradual linear expansion of the *s*-genitive but a V-shaped pattern: The *s*-genitive started out with a share of 19% in the 1650-1699 period. Frequencies then started to decline in the 1750-1799 period,

	s-genitive		*of*-genitive		Total	
1650-1699	89	(19%)	370	(81%)	459	(100%)
1700-1749	116	(24%)	363	(76%)	479	(100%)
1750-1799	78	(17%)	388	(83%)	466	(100%)
1800-1849	30	(5%)	539	(95%)	569	(100%)
1850-1899	85	(17%)	411	(83%)	496	(100%)
1900-1949	84	(17%)	410	(83%)	494	(100%)
1950-1989	127	(28%)	331	(72%)	458	(100%)
Total	609	(18%)	2,812	(82%)	3,421	(100%)

Table 2: Interchangeable genitive frequencies in ARCHER news texts.

reaching their low point in the 1800-1849 period (mean share of the *s*-genitive: 5%), and recovering subsequently. In point of fact, the 1950-1989 period even surpasses the first ARCHER period, exhibiting a 28% *s*-genitive share. In this connection, observe that the V-shaped pattern manifests in relative genitive frequencies (i.e. percentages) *and* absolute genitive frequencies (i.e. token frequencies), and that the *s*-genitive slump is unlikely to be a sampling issue, as the total number of observations in the ARCHER's middle periods is not any lower than, e.g. in the starting period. As an aside, we also note that the V-shaped frequency pattern survives inclusion of all genitive contexts such as partitive and part-whole relations, not just interchangeable ones.

Figure 1 is a frequency graph that puts the numbers in Table 2 into a wider historical context, combining ARCHER frequencies with frequencies published in the literature. While the measurements underlying the graph are not maximally homogeneous – the data points prior to ARCHER are not based on news texts, and Thomas' (1931) figures do not refer to interchangeable genitive contexts only – the overall picture is clear: Seen against the backdrop of long-term diachronic variability, the collapse of *s*-genitive frequencies in news texts around the middle of the Late Modern English period is quite marked, and the subsequent increase fairly steep. In short, Figure 1 looks like an *s*-curve in trouble. The task before us, then, is to explain the up and down of *s*-genitive frequencies in the Late Modern English period: Why did the *s*-genitive fall out of fashion in the period between 1650 and 1850? Why did it become popular again after the middle of the 19$^{\text{th}}$ century?

6. Conditioning factors

To learn about the determinants of historical genitive variability, we need to investigate the factors that condition genitive choice. Every one of the $3,421$ genitive occurrences in the dataset is annotated for three language-external variables: the CORPUS FILE where the genitive occurs (this can be used to approximate author idiosyncracies), ARCHER TIME SLICE (in which of ARCHER's 50-year periods was the news piece composed), and the exact YEAR in which the news piece was composed. In addition, the genitive occurrences were coded for eight major language-internal conditioning factors, which are the usual suspects according to the literature: two factors relat-

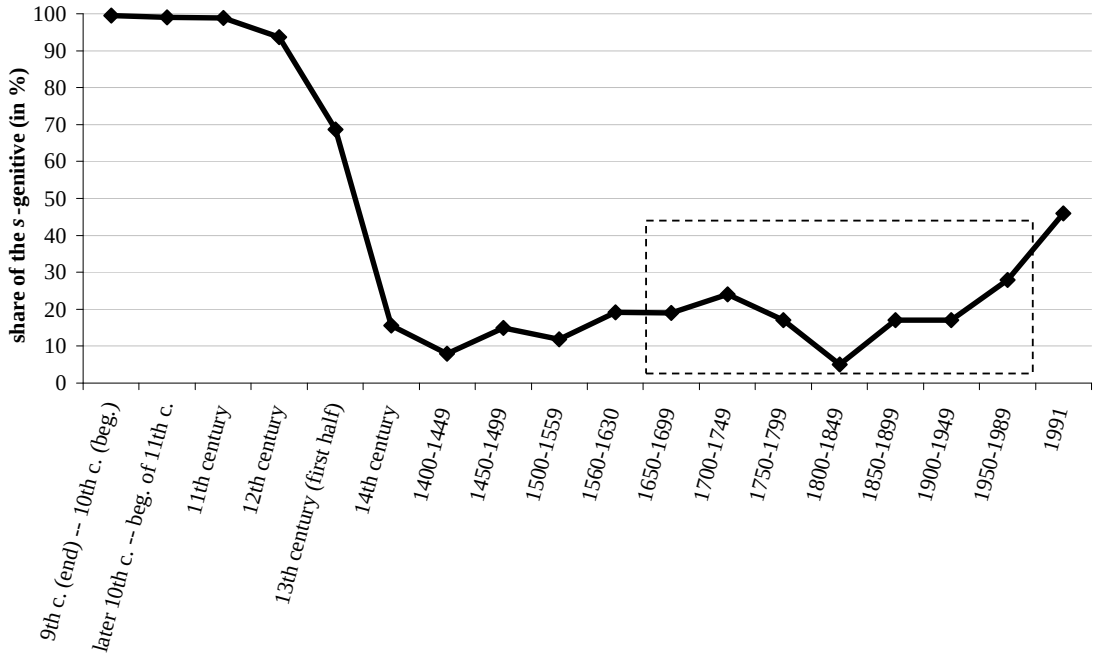

Figure 1: Long-term relative frequencies of the *s*-genitive (or "inflected genitive"). Dotted box: ARCHER news texts, 1650 through 1989. Other sources: (i) 9th century through 14th century according to Thomas (1931), cited in Mustanoja (1960: 75); (ii) 1400–1449 through 1560–1630 according to Rosenbach and Vezzosi (2000: Table 1); (iii) 1991 according to Hinrichs and Szmrecsanyi (2007) on the basis of data from news texts in F-LOB.

ing to semantics (GENITIVE RELATION and POSSESSOR ANIMACY), two factors pertaining to the domain of processing (POSSESSOR LENGTH and POSSESSUM LENGTH), a phonological factor (POSSESSOR PHONOLOGY), an information-structure factor (POSSESSOR GIVENNESS), and two text linguistic factors (POSSESSOR THEMATICITY and LEXICAL DENSITY). This section discusses the technicalities of the annotation process, and reports univariate distributions.

6.1. Semantics: GENITIVE RELATION

The majority view in the literature is that in Present-Day English, possessive relations have a privileged status in the semantics of the *s*-genitive (for instance, Taylor 1989). Conversely, the *of*-genitive is favored with, e.g., partitive and objective genitive relations (Biber et al. 1999: 303). In the spirit of these claims, we initially utilized a coding scheme with six categories to hand-code genitive occurrences in our dataset, differentiating between (i) ownership relations, both prototypical (i.e. alienable legal) as in *Tom's car* and non-prototypical as in *God's justice*, (ii) kinship relations, (iii) body part relations, (iv) part-whole relations, (v) partitive relations,

	% *s*-genitive: ownership relation	% *s*-genitive: other relations	% all genitives: ownership relation
1650-1699	61	6	24
1700-1749	69	8	26
1750-1799	58	5	22
1800-1849	28	1	16
1850-1899	50	5	26
1900-1949	55	5	24
1950-1989	79	11	25

Table 3: The factor GENITIVE RELATION: relative frequencies by ARCHER time slice.

and (vi) "other" relations, such as subjective genitive relations (e.g. *my wife's love*). Next, we removed part-whole and partitive genitives from the dataset due to insufficient variability (cf. Section 4), and conflated kinship and body part genitives with the "other" category because of low cell counts (for example, there are no kinship genitives in the 1900-1949 period). In the subsequent analyses, we will thus be dealing with a binary distinction between ownership genitive relations, as in (7) , and "other" genitive relations, as in (8) (which is actually an 'objective' genitive relation, in the parlance of Quirk et al. 1985: 321–322).

(7) Yesterday Morning the Lady of the Right hon. the Earl of Ashburnham, was safely delivered of a Daughter, at **[his Lordship]$_{\text{por}}$'s [house]$_{\text{pum}}$** in Dover-street. <ARCHER 1762pub1.n4b>

(8) The Mareschal de Villeroy continues still at Lessines, and the Mareschal de Catinat before Aeth, to see **[the Fortifications]$_{\text{pum}}$ of [that place]$_{\text{por}}$** repaired, and Monsieur de Bousslers near Roeux to observe our Army which is still in the same Post as in the our former. <ARCHER 1697pos2.n2b>

Table 3 provides a frequency overview. We observe that, as predicted by the literature, the *s*-genitive is typically more popular (i.e. has a share of more than 50%, as can be seen in the leftmost column) in ownership relation contexts than the *of*-genitive. The exception is the 1800-1849 period, in which the *s*-genitive only codes 28% of all ownership relations. With non-ownership relations, the *s*-genitive is comparatively marginal (middle column). In all, the diachronic development of the share of the *s*-genitive according to ownership relations mirrors the overall V-shaped frequency pattern (cf. Section 5). At the same time, the rightmost column in Table 3 suggests that ownership relation contexts have overall dropped in the 1750-1800 period and, especially, the 1800-1849 period.

	% *s*-genitive: human p'ors	% *s*-genitive: non-human p'ors	% all genitives: human p'or
1650-1699	47	4	35
1700-1749	53	4	42
1750-1799	47	3	32
1800-1849	20	1	23
1850-1899	36	8	33
1900-1949	42	10	23
1950-1989	60	14	30

Table 4: The factor POSSESSOR ANIMACY: relative frequencies by ARCHER time slice.

6.2. Semantics: POSSESSOR ANIMACY

Animacy of the possessor NP is often seen as one of the most important determinants of genitive choice. The more human and animate a possessor, and the more it conveys the idea of animate things and human activity, the more likely it is to take the *s*-genitive (for instance, Altenberg 1982; Biber et al. 1999; Dahl 1971; Jucker 1993; Quirk et al. 1985; Taylor 1989; Rosenbach 2005, 2008). All possessors in the dataset were thus hand-coded for possessor animacy, drawing on a comparatively simple classification that distinguishes between human possessors (category HUMAN in Zaenen et al. 2004) and non-human possessors. So, in (9), we find a possessor NP whose head noun (*murderer*) was coded as "human"; in (10), the possessor NP head noun (*justice*) is inanimate and thus classified as "non-human".

(9) [. . .] one of the Children hung about one of **[the *murderer*]$_{\textbf{por}}$'s [legs]$_{\textbf{pum}}$**, yet was pull'd off and thrown after the rest. <ARCHER 1653merc.n2>

(10) Thus you see **[the footsteps]$_{\textbf{pum}}$ of [God's *justice*]$_{\textbf{por}}$**. <ARCHER 1653merc.n2>

The frequency overview in Table 4 demonstrates that the *s*-genitive is and always has been more frequent when the possessor is human instead of non-human. Having said that, there is substantial diachronic volatility: As for human possessors, we again find a V-shaped pattern such that the *s*-genitive is relatively frequent with human possessors in the early and late ARCHER periods, but comparatively infrequent in the 19th century. With respect to non-human possessors, we observe that the *s*-genitive becomes markedly more frequent in this domain starting in the middle of the 19th century. There is also volatility in terms of overall input frequencies, however: Human possessors were comparatively frequent in the 1700-1749 period and comparatively infrequent in the 1800-1849 and 1900-1949 periods.

	mean values *s*-genitive	mean values *of*-genitive	mean values Total
1650-1699	12	20	18
1700-1749	13	22	19
1750-1799	13	18	17
1800-1849	12	22	21
1850-1899	12	24	22
1900-1949	13	25	23
1950-1989	13	26	22

Table 5: The factor POSSESSOR LENGTH: mean values by ARCHER time slice.

	mean values *s*-genitive	mean values *of*-genitive	mean values Total
1650-1699	12	11	11
1700-1749	13	11	11
1750-1799	12	12	12
1800-1849	14	12	12
1850-1899	16	13	14
1900-1949	19	12	13
1950-1989	21	12	15

Table 6: The factor POSSESSUM LENGTH: mean values by ARCHER time slice.

6.3. Processing: POSSESSOR LENGTH and POSSESSUM LENGTH

The principle of "end-weight" (for example, Behaghel 1909/1910; Wasow 2002) postulates that speakers and writers tend to place "heavier" elements after shorter ones, probably due to properties of the human parser (cf. Hawkins 1994). Hence, if the possessor is heavy, there should be a general preference for the *of*-genitive because it places the possessor last. If the possessum is heavy, a general preference for the *s*-genitive is expected. There is a literature on how to approximate "weight" (for an overview and evaluation, cf. Wasow 1997; Szmrecsanyi 2004). In the present study, "weight" is operationalized as a constituent's length in graphemic characters. Consider (11):

(11) A 'strictly confidential' report by the council's internal audit branch puts much of the blame for **[the construction branch]$_{\textbf{por}}$'s [losses]$_{\textbf{pum}}$** on Mr Dover and his immediate subordinate. <ARCHER 1979obs2.n8b>

In (11), POSSESSOR LENGTH is 21 graphemic characters, and POSSESSUM LENGTH is six graphemic characters. We relied on Perl scripts that automatically annotated every genitive observation in the dataset for possessor and possessum NP length.

	% *s*-genitive: final sibilant in p'or present	% *s*-genitive: final sibilant in p'or absent	% all genitives: final sibilant in p'or present
1650-1699	2	25	25
1700-1749	11	29	28
1750-1799	2	22	25
1800-1849	0	8	31
1850-1899	1	23	28
1900-1949	1	23	27
1950-1989	6	34	23

Table 7: The factor POSSESSOR PHONOLOGY: relative frequencies by ARCHER time slice.

Table 5 provides mean possessor NP lengths by ARCHER time slice. In accordance with the principle of end-weight, *of*-genitives do have longer possessor NPs than *s*-genitives. While the weight of *s*-genitive possessor NPs (leftmost column) has remained fairly stable over time, we observe a modest increase in *of*-genitive possessor NP weights (middle column). At the same time, there is also a general increase in possessor lengths (rightmost column). As for possessum NP lengths (Table 6), notice that – again in accordance with the principle of end-weight – *s*-genitive possessum NPs are typically weightier than *of*-genitive possessum NPs, though the difference is clearly less marked than with possessor NPs. In total, *of*-genitive possessum NP length has remained rather stable over time while *s*-genitive possessum NP length has consistently and rather clearly increased. Lastly, as with possessor NPs, possessum NPs have generally become heavier in real time.

6.4. POSSESSOR PHONOLOGY

A final sibilant in the possessor NP is claimed to encourage usage of the *of*-genitive due to a haplology or *horror aequi* effect (for instance, Altenberg 1982; Zwicky 1987). Perl scripts automatically annotated all genitive occurrences in the dataset as to whether the possessor phrase ended in <s>, <ce>, <sh>, or <tch>, as in (12).[2]

(12) The delay which has necessarily taken place in **[the preparation]$_{\text{pum}}$ of [this despatch]**$_{\text{por}}$ has made it possible to give you the returns from two of the other chief towns of the province. <ARCHER 1833tim2.n5b>

Table 7 makes amply clear that the *s*-genitive is indeed dispreferred when the possessor NP ends in a sibilant. From the diachronic perspective, the *s*-genitive-specific figures in Table 7 once again follow the familiar V-shaped pattern: In the 1800-1849 period, *s*-genitives are rare regardless of whether or not the possessor NP ends in a final sibilant. It is interesting to note, however, that possessors with final sibilants were significantly ($p = .043$) more frequent in the 1800-1849 period than in ARCHER news on the whole.

	% *s*-genitive: p'or is given	% *s*-genitive: p'or is not given	% all genitives: p'or is given
1650-1699	23	19	9
1700-1749	33	23	10
1750-1799	12	17	11
1800-1849	9	5	12
1850-1899	14	18	13
1900-1949	20	16	16
1900-1949	33	27	17

Table 8: The factor POSSESSOR GIVENNESS: relative frequencies by ARCHER time slice.

6.5. Information status: POSSESSOR GIVENNESS

To enhance discourse flow, discourse-new possessors prefer the *of*-genitive because it places the possessor last, thus maintaining old-before-new order. Conversely, if the possessum is discourse-new, the *s*-genitive is the preferred option, all other things being equal (Biber et al. 1999: 305; Quirk et al. 1985: 1282). We operationalized this factor by having Perl scripts establish for every genitive observation in the dataset whether the possessor NP head noun had been mentioned in a discourse context of 50 words prior to a given genitive observation. To illustrate, consider (13), where the possessor NP head noun in the *s*-genitive construction (*the vessel's arrival at Boston*) is, in fact, mentioned in the preceding sentence (*the said vessel*).

(13) The owners have published the following account, viz. That they gave express orders to their correspondents in London, that none of the East-India Company's tea should on any terms be shipped on board the said vessel, however advantageous the offer, or great the loss on the voyage. That on **[the vessel]$_{\text{por}}$'s [arrival at Boston]**$_{\text{pum}}$, they, the owners, freely and publickly declared their willingness to send her back [Ĕ] <ARCHER 1774lon1.n4b>

We hasten to acknowledge that this measure – which boils down to checking for previous mention – is a rather blunt instrument, but one that yields consistent measurements across texts.

The picture emerging from Table 8 is not clear-cut: In five of the seven ARCHER time slices, the *s*-genitive is slightly more frequent when the possessor NP head noun is given, but in two periods (1750-1799 and 1850-1899) it is not. What is clear-cut, however, is the fact that given possessors have become consistently more frequent over time (almost doubling their share, from 9% in the first ARCHER time slice to 17% in the last ARCHER time slice).

	mean values *s*-genitive	mean values *of*-genitive	mean values Total
1650-1699	3	3	3
1700-1749	3	2	3
1750-1799	3	3	3
1800-1849	2	2	2
1850-1899	2	2	2
1900-1949	4	3	3
1950-1989	5	3	4

Table 9: The factor POSSESSOR THEMATICITY: mean values by ARCHER time slice.

6.6. Text linguistics: POSSESSOR THEMATICITY

Osselton (1988) has claimed that the general topic of a text co-determines which nouns in that text can take the *s*-genitive. In this sprit, we assume that increased text frequency of a possessor NP, and thus increased thematicity, will make the *s*-genitive more likely (cf. Hinrichs and Szmrecsanyi 2007; Szmrecsanyi and Hinrichs 2008). So, Perl scripts determined the text frequency of the possessor NP head noun in the corpus text where the genitive occurred. Consider the genitive phrase (*the Pope's arrival*) in (14): The possessor NP head noun is *Pope*, and in ARCHER text 1979stm2.n8b (which is actually a piece about the Pope's 1979 visit to Poland), the token *Pope* has a rather high local text frequency of 21 occurrences.

(14) The turnout for **[the Pope]$_{\mathbf{por}}$'s [arrival]$_{\mathbf{pum}}$** was a stunning and heartening sight for the Church in Poland [...] <ARCHER 1979stm2.n8b>

Table 9 reports mean possessor NP head noun frequencies over time. Succinctly put, possessor thematicity did not make a distributional difference up until the 20$^{\text{th}}$ century, when *s*-genitive possessors started to attract more frequent possessors. There is no substantial longitudinal variability concerning the overall thematicity of genitive possessors.

6.7. Text linguistics: LEXICAL DENSITY

There is evidence (Hinrichs and Szmrecsanyi 2007; Szmrecsanyi and Hinrichs 2008) that the *s*-genitive is preferred in contexts where informational density is high. Therefore, when news writers feel a need to economically code more information in a given textual passage, the *s*-genitive has an edge due to being the more compact and economic coding option (Biber et al. 1999: 99). We used Perl scripts to establish the number of different word types of the textual passages – 50 words before and 50 words after a given genitive construction – where the genitive occurrence in question is embedded. Higher type-token ratios indicate increased lexical density. In (15), for instance, we find an *s*-genitive phrase (*the Emperor's Gallery*) embedded in a passage with a fairly high type-token ratio (82 different types per 1000 word tokens).

	mean values *s*-genitive	mean values *of*-genitive	mean values Total
1650-1699	72	72	72
1700-1749	72	70	71
1750-1799	72	71	71
1800-1849	72	71	71
1850-1899	72	71	71
1900-1949	72	71	71
1950-1989	75	73	74

Table 10: The factor LEXICALDENSITY: mean values by ARCHER time slice.

(15) The Banditti who have taken the Part of the Genoese, and who occupy the Fort of Aleria, have made several Attempts to penetrate into the Country, but without Success; and we hope even to dislodge them from the Fort. Florence, May 10. M. Louis Siries, Director of the Works in **[the Emperor]$_{\text{por}}$'s [Gallery]$_{\text{pum}}$**, died here the 6th in an advanced Age. His great Excellency as a Jeweller and Goldsmith, the Fruit of his own untaught Genius, is well known. He first made himself famous, by giving such Hardness to Gold, that it would bear an Edge like Steel. <ARCHER 1762publ.n4b>

An inspection of the mean values in Table 10 reveals that there is a good deal of diachronic stability, except for the 20$^{\text{th}}$ century, where we observe a marked increase in overall lexical densities (from 71 to 74; $p < .001$). However, it is *s*-genitive contexts rather than *of*-genitive contexts that absorb most of this increase.

6.8. Language-internal conditioning factors: interim summary

The series of univariate frequency analyses in this section was not conclusive. We have seen that the V-shaped overall frequency picture is echoed in V-shaped genitive distributions according to factors such as GENITIVE RELATION, POSSESSOR ANIMACY, and POSSESSOR PHONOLOGY. Because in each case the *s*-genitive dip happened in all conditions (ownership relations and other relations, human and non-human possessors, final sibilants present and absent), the aforementioned factors do, at first view (and subject to the limits of univariate analysis), not appear to be terribly explanatory. On the other hand, it would seem that we are dealing with more linear distribution changes with regard to factors such as POSSESSOR and POSSESSUM LENGTH, POSSESSOR THEMATICITY, and LEXICAL DENSITY. The problem with all of this is that the habitat – read: the input frequencies in the news genre – has also been subject to diachronic variability: For instance, ownership relation contexts in general were rare at the beginning of the 19$^{\text{th}}$ century (discouraging *s*-genitive usage), while possessors with final sibilants were markedly widespread during the same period (encouraging *of*-genitive usage). At this point, we thus cannot rule out that in the time span studied here, nothing actually happened to the grammar of genitive choice *per se*, and that changes in input frequencies are solely to blame for the overall V-shaped fluctuation in genitive frequencies. This is the hypothesis that we shall test in the next

section.

7. Environmental factors: on the impact of changing input frequencies

This section explores whether and to what extent overall genitive frequency changes over time are merely a function of variable input frequencies – of, say, variable proportions of human possessors among all possessors. To investigate this issue, we will aggregate genitive observations and fit a *linear regression* model (cf. Baayen 2008: 84-97) to predict a continuous dependent variable (in our case, *s*-genitive proportions) from of several independent variables, such as the share of human possessors among all possessors.[3] We will operate on the assumption that input frequency changes cause changes in genitive frequencies.

7.1. Model fitting and model simplification

First, we aggregated the $3,421$ genitive observations in our dataset to the level of the CORPUS FILE, which yields an aggregate dataset with 70 observations (one per ARCHER corpus file). The dependent variable is the share of the *s*-genitive among all genitives in the corpus file (for instance, in file 1959gua1.n8b, the *s*-genitive is used in 30.2% of all cases, hence the value of the dependent variable is 30.2). The independent variables are – in the case of originally categorical variables – corresponding per cent figures (for instance, in file 1959gua1.n8b, 25.6% of all genitive relations are ownership relations) or – in the case of originally continuous variables – mean values (for instance, in file 1959gua1.n8b, mean possessum length is 11.7 graphemic characters). Next, we specified the maximal model, regressing the share of the *s*-genitive per corpus file as the dependent variable against the eight aggregated conditioning factors discussed in the previous section as independent variables. Next, the model was simplified via a stepwise removal of non-significant independent variables, moving from less significant to more significant independent variables. The final, minimal adequate model – reported in Table 11 – contains only three independent variables: PERCENTAGE OWNERSHIP RELATIONS, PERCENTAGE HUMAN POSSESSORS, and MEAN POSSESSUM LENGTH.

7.2. Model evaluation and model discussion

The regression coefficients in Table 11 reveal that for every per cent point increase in the share of ownership relations, the share of *s*-genitives per corpus text will typically rise by .46 per cent points. The corresponding effect size in regard to the share of human possessors is .40. Also, for every one-unit increase in mean possessum length (for example, 12 graphemic characters instead of just 11), the share of the *s*-genitive per corpus text will increase by 1.08 per cent points.

The summary statistics in Table 11 indicate that there is a highly significant multiple correlation between the three independent variables considered in the model and the dependent variable, *s*-genitive shares ($p < .001$). The adjusted R^2 value is .53, which means that the model, comparatively simple as it is and actually completely ignorant about diachrony, accounts for 53% of

	b	p-value	
model intercept	-17.61	.009	**
PERCENTAGE OWNERSHIP RELATIONS	.46	.002	**
PERCENTAGE HUMAN POSSESSORS	.40	.001	**
MEAN POSSESSUM LENGTH	1.08	.018	*
N			70
residual standard error			7.454, 66 df
F-statistic			26.94 on 3/66 df, $p < .001$
adjusted R^2			.530

* significant at $p < .05$, ** $p < .01$.

Table 11: Predicting *s*-genitive rates per corpus file from input frequencies: the minimal adequate linear regression model.

the variance in the dependent variable. In Figure 2 we find a graph that plots observed corpus frequencies (black dots) against frequencies as predicted by the linear regression model in Table 11 (white dots); data points are arranged according to year of composition. The solid trend line estimates the overall relationship between observed *s*-genitive frequencies and real time, thus adding more granularity to the frequency overview in Section 5. The dotted trend line depicts the frequency predictions by the linear regression model, calculated according to input frequencies. The point is, then, that the linear regression model *does* predict a decrease of *s*-genitive frequencies during the first half of the 19th century, and a subsequent recovery of *s*-genitive frequencies. The model underestimates, however, the extent to which frequencies slumped in the 19th century, and does not predict the phenomenal rise of *s*-genitive frequencies in the 20th century.

7.3. Environmental factors: interim summary

Any baby – even an otherwise happy one – will cry when it is getting cold. This does not mean that the baby has changed. What has changed is the baby's environment. By the same token, the analysis in this section has demonstrated that environmental factors, i.e. fluctuating input frequencies of (i) ownership relations, (ii) human possessors, and (iii) possessum lengths, explain slightly more than half of the overall volatility in genitive frequencies over time. A discussion of the reasons why those input frequencies fluctuate is beyond the scope of the present study, although it seems safe to assume that we are dealing here with genre-internal fads and fashions. What is important is that environmental factors fail to fully account for the extreme unpopularity of the *s*-genitive in the first half of the 19th century, nor do they predict its rising popularity in the 20th century. So, something else must have changed – something that affects how news writers choose between the *s*-genitive and the *of*-genitive.

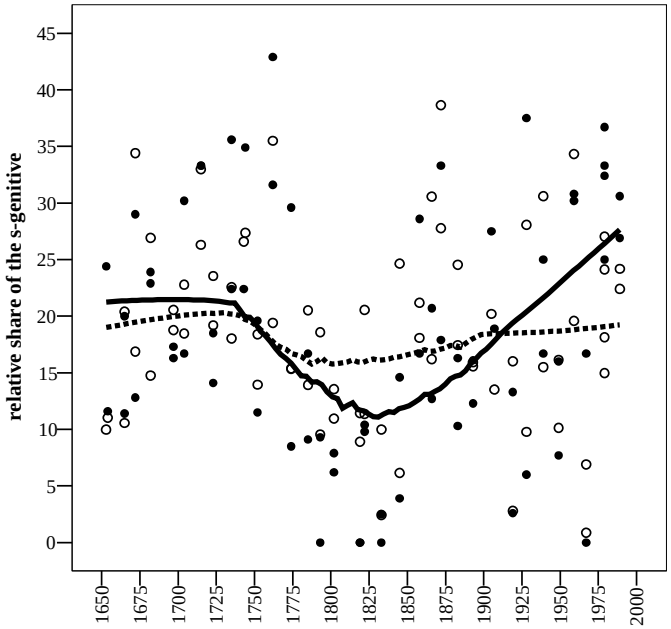

Figure 2: Predicting *s*-genitive shares: scatterplot, observed relative frequencies (black dots) versus predicted relative frequencies in linear regression (white dots). Each data point represents one corpus file. Horizontal axis: year of composition. Solid line: LOESS curve estimating the relationship between real time and observed relative frequencies. Dotted line: LOESS curve estimating the relationship between real time and predicted relative frequencies.

8. Changing genitive grammars

To isolate changes in genitive choice grammars over time, this section will be concerned with whether or not the eight language-internal conditioning factors discussed so far exhibit stability in regard to their effect on genitive choice. For example, all other things being equal, did human possessors favor *s*-genitives more strongly in the 17th century than in the 20th century? To investigate issues like these, we utilize *mixed-effects binary logistic regression* analysis (see Baayen 2008: chapter 7).[4] We will predict binary outcomes (news writers producing either an *s*-genitive or an *of*-genitive) given several predictor variables, such as POSSESSOR ANIMACY and real time. We thus abandon the aggregate level of genitive *frequencies* which has concerned us in the precious section (dependent variable: *s* genitive rates per corpus text) and turn to the level of *individual linguistic decisions* (dependent variable: individual binary genitive outcomes – asking, for example, *why is the second genitive occurrence in corpus text 1762pub1.n4b an* s*-genitive and not an* of*-genitive*?), which is the level where genitive choice grammars can be observed in action. To the extent that we will see significant interactions between real time and the effect of language-internal conditioning factors, we will have unearthed evidence for changes in

news writers' genitive choice grammars.

8.1. Model fitting and model simplification

Mixed effects logistic regression modeling is the closest a corpus linguist can come to conducting a controlled experiment: The procedure models the combined contribution of all conditioning factors considered, systematically testing each factor while holding the other factors in the model constant. The random effects in the model take care of author idiosyncracies and differential propensities of different noun lemmas to co-occur with either genitive type. On a technical note, we observed the customary steps (cf. Crawley 2005; Baayen 2008) to obtain a minimal adequate regression model. We began by fitting the maximal model including, as main effects, all eight language-internal conditioning factors as well as the language-external factor ARCHER TIME SLICE, a 7-way categorical variable where each factor level represents one of ARCHER's seven 50-year time slices.[5] Crucial to this section's research question, the maximal model also contained eight interaction terms between the language-internal factors and ARCHER TIME SLICE. Standard diagnostics indicated that there is no serious collinearity between the predictor variables entered into the maximal model. As for random effects (i.e. adjustments to the intercept to accommodate non-repeatable effects), the model includes CORPUS FILE as a by-subject random effect approximating author idiosyncracies and POSSESSOR HEAD NOUN as a by-item random effect. Subsequently, the model was simplified by removing factors and interaction terms lacking significant explanatory power. We started the pruning process with the least significant higher-order interaction, moving towards main effects. Explanatory power of the interaction terms and main-effect categorical factors with more than two levels was assessed via likelihood-ratio tests. The pruning process resulted in the deletion of the following main effects and interaction terms from the model: POSSESSOR GIVENNESS, ARCHER TIME SLICE $\times$ POSSESSOR LENGTH, ARCHER TIME SLICE $\times$ POSSESSOR PHONOLOGY, and ARCHER TIME SLICE $\times$ LEXICAL DENSITY. Finally, we tested the justification of including the two random effects by means of likelihood ratio tests. Both random effects were retained, as models including only one of the two random effects turned out to fare significantly worse than a model with both random effects. The final, minimal adequate model is reported in Table 12.

8.2. Model evaluation and model discussion

The quality of the minimal adequate model in Table 12 is excellent. As for predictive accuracy, the model correctly predicts 96% of all genitive outcomes in the dataset (a "dumb" model categorically predicting the *of*-genitive would only be correct in 82% of all cases). Somers' Dxy, a rank correlation coefficient between predicted outcome probabilities and observed binary outcomes, is .98, which is likewise a very good value indicating that the model discriminates near-perfectly between genitive types. As for goodness of fit, the R^2 value for the correlation between predicted probabilities and observed genitive proportions in the dataset is a similarly near-perfect .98.

Inspection of the variances and standard deviations associated with the two random effects in

	odds ratio	b	p-value
model intercept	.00	-9.254	.000 ***
GENITIVE RELATION (default: 'other')			
ownership	242.88	5.493	.000 ***
POSSESSOR ANIMACY (default: non-human p'or)			
human possessor	11.75	2.464	.004 **
POSSESSOR LENGTH (1 unit $\widehat{=}$ 1 graphemic character)	.85	-.163	.000 ***
POSSESSUM LENGTH (1 unit $\widehat{=}$ 1 graphemic character)	1.21	.187	.000 ***
POSSESSOR PHONOLOGY (default: final sibilant absent)			
final sibilant present	.01	-4.828	.000 ***
POSSESSOR THEMATICITY (1 unit $\widehat{=}$ 1 frequency point)	1.05	.049	.560
LEXICAL DENSITY (1 unit $\widehat{=}$ 10 TTR points)	1.89	.638	.005 **
ARCHER TIME SLICE (default: 1950-1989)			
1650-1699	1.88	.631	.585
1700-1749	1.56	.448	.728
1750-1799	.53	-.642	.624
1800-1849	.10	-2.268	.146
1850-1899	2.05	.719	.509
1900-1949	2.19	.782	.495
interaction terms:			
GENITIVE RELATION × ARCHER TIME SLICE	(see Table 13)		.020 * [a]
POSSESSOR ANIMACY × ARCHER TIME SLICE	(see Table 13)		.000 *** [a]
POSSESSUM LENGTH × ARCHER TIME SLICE	(see Table 13)		.000 *** [a]
POSSESSOR THEMATICITY × ARCHER TIME SLICE	(see Table 13)		.095 + [a]
random effects:			
POSSESSOR HEAD NOUN			
(intercept, $N = 1,698$, variance: 10.472, $\sigma = 3.236$)			.000 *** [a]
CORPUS FILE			
(intercept, $N = 70$, variance: .906, $\sigma = .952$)			.000 *** [a]
summary statistics:			
N			3,421
% correct predictions (% baseline)			96 (82)
Somers' Dxy			.98
R^2 predicted probabilities/observed proportions (10 bins)			.98

[a] significance assessed by means of likelihood ratio tests

+ marginally significant at $p < .1$, * significant at $p < .05$, ** $p < .01$, *** $p < .001$.

Table 12: Predicting genitive choice in ARCHER news texts: the minimal adequate logistic regression model. Predicted odds are for the *s*-genitive.

the model reveals that POSSESSOR HEAD NOUN idiosyncracies are vastly more pronounced than CORPUS FILE (i.e. author) idiosyncracies. This is another way of saying that in the dataset, by-item effects dwarf by-subject effects. Among the head nouns receiving the highest positive adjustments to the intercept – thus favoring the *s*-genitive – we find lemmas such as *father* and *grace* (as in *What else may relate to his Grace's Obsequies, cannot yet be learn'd* <1723dai2.n3b>). The corpus text that receives the highest positive adjustment to the intercept is ARCHER text 1774lon1.n4b, a piece published in *The London Chronicle* whose author apparently had a soft spot for the *s*-genitive.

The main effects in the model are all in line with the previous literature – with the possible exception of POSSESSOR GIVENNESS, which turned out to be insignificant and is thus not included in the minimal adequate model (the explanatory failure of the predictor replicates similar findings by, e.g., Gries 2002 and Hinrichs and Szmrecsanyi 2007 in regard to Present-Day English genitive variation). Consider the signs of the regression coefficients (*b*) in Table 12: Ownership relations, human possessors, increased thematicity, and increased lexical density all favor the *s*-genitive. Conversely, longer possessors and final sibilants in the possessor favor the *of*-genitive. Observe also that for every one-unit increase in POSSESSOR LENGTH, the odds for the *s*-genitive decrease by 15% while every one-unit increase in POSSESSUM LENGTH, the odds for the *s*-genitive increase by 21%. A look at the odds ratios reported in Table 12 shows that among the categorical conditioning factors, ownership relations have the largest effect size: If the genitive relationship is one of ownership, the odds that an *s*-genitive will be chosen increase by a factor of approximately 243. Another strong effect can be observed for POSSESSOR PHONOLOGY: If the possessor ends in a final sibilant, the odds for an *s*-genitive decrease by about 99%. Notice that POSSESSOR THEMATICITY and ARCHER TIME SLICE are not statistically significant as main effects; the reason why they are in the model nonetheless is that they partake in significant higher-order interactions, as we shall see presently. Thus, the lacking statistical significance of ARCHER TIME SLICE as a main effect highlights the fact that it is not real time *per se* that has a bearing on genitive choice.

The discussion in the previous paragraph has centered on *the effect sizes* that individual conditioning factors have. Figure 3 visualizes the *overall importance* of the factors in the model by plotting the decrease in the model's Akaike Information Criterion (AIC; cf. Sakamoto and Akaike 1978) if a factor (including higher-order interactions in which it partakes) is removed from the minimal model (cf. Jaeger 2006: 84-85). More sizable decreases correspond to increased overall importance. GENITIVE RELATION, then, is the most important factor to predict the $N = 3,421$ genitive choices in the dataset, closely followed by POSSESSOR LENGTH, then POSSESSOR ANIMACY, and POSSESSOR PHONOLOGY. The real-time variable ARCHER TIME SLICE is the fifth most important predictor variable, virtually on par with POSSESSUM LENGTH which ranks sixth. LEXICAL DENSITY and POSSESSOR THEMATICITY are overall rather marginal factors. In all, the point is that real time is an important predictor of genitive variation in the dataset, but clearly not the most important one.

Having said that, how exactly does real time impact genitive variability, given that – as we have seen – ARCHER TIME SLICE is not a significant main effect? The answer is that the influence of diachrony is mediated through interactions between ARCHER TIME SLICE and the

	odds ratio	b	p-value
GENITIVE RELATION × ARCHER TIME SLICE			
ownership:1650-1699	.08	-2.501	.025 *
ownership:1700-1749	.08	-2.552	.009 **
ownership:1750-1799	.03	-3.362	.002 **
ownership:1800-1849	.08	-2.481	.077 +
ownership:1850-1899	.14	-1.935	.063 +
ownership:1900-1949	1.28	.244	.818
POSSESSOR ANIMACY × ARCHER TIME SLICE			
human possessor:1650-1699	8.69	2.162	.068 +
human possessor:1700-1749	27.42	3.311	.003 **
human possessor:1750-1799	35.55	3.571	.003 **
human possessor:1800-1849	9.97	2.300	.110
human possessor:1850-1899	2.20	.787	.459
human possessor:1900-1949	.29	-1.238	.254
POSSESSUM LENGTH × ARCHER TIME SLICE			
possessum length:1650-1699	.83	-0.188	.000 ***
possessum length:1700-1749	.89	-0.120	.023 *
possessum length:1750-1799	.91	-0.094	.074 +
possessum length:1800-1849	.91	-0.098	.053 +
possessum length:1850-1899	.84	-0.168	.000 ***
possessum length:1900-1949	.87	-0.134	.005 **
POSSESSOR THEMATICITY × ARCHER TIME SLICE			
possessor thematicity:1650-1699	1.02	.019	.892
possessor thematicity:1700-1749	.72	-0.330	.014 *
possessor thematicity:1750-1799	.87	-0.138	.340
possessor thematicity:1800-1849	.72	-0.333	.271
possessor thematicity:1850-1899	1.04	.040	.787
possessor thematicity:1900-1949	.99	-0.010	.942

+ marginally significant at $p < .1$, * significant at $p < .05$, ** $p < .01$, *** $p < .001$.

Table 13: Predicting genitive choice in ARCHER news texts: interaction terms in logistic regression. Default period: 1950-1989. Predicted odds are for the *s*-genitive.

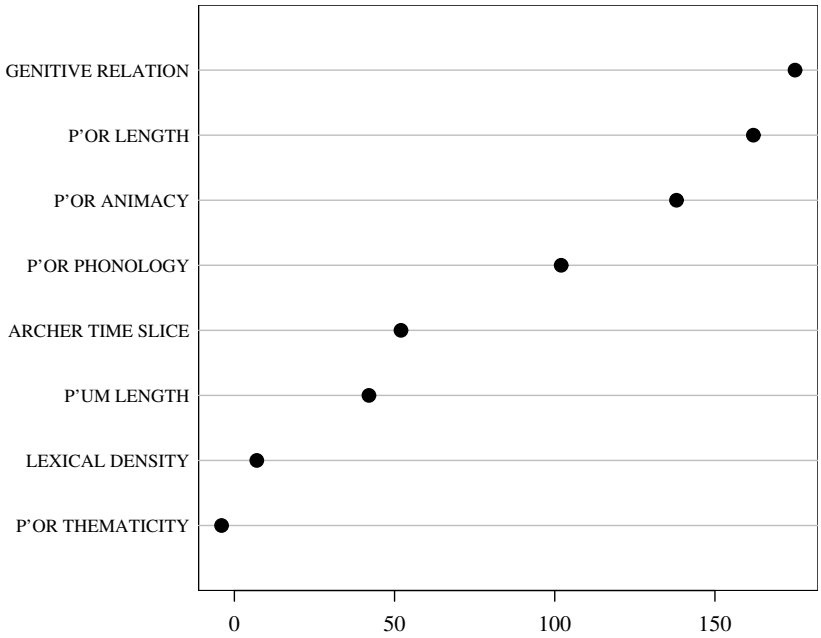

Figure 3: Importance of factors in model: decrease in Akaike Information Criterion (AIC) if factor removed.

effect size of four language-internal conditioning factors: GENITIVE RELATION, POSSESSOR ANIMACY, POSSESSUM LENGTH, and (marginally significantly) POSSESSOR THEMATICITY. The corresponding regression weights for different levels of the moderator variable ARCHER TIME SLICE are displayed in Table 13. In this table, odds ratios < 1, or coefficients (b) with a negative sign, indicate that during the ARCHER period in question, the interacting conditioning factor was less favorable to the *s*-genitive and/or more favorable to the *of*-genitive compared to ARCHER's 1950-1989 period, which serves as the baseline which effect size fluctuations are measured against. The interaction terms detailed in Table 13 may be summarized as follows:

- The exceedingly strong attraction of *s*-genitives to the GENITIVE RELATION of ownership is a 20th-century phenomenon. In the first ARCHER period, for instance, if the genitive relation to be encoded was an ownership relation (as opposed to another kind of genitive relation), the odds that the *s*-genitive would be chosen increased by a factor of roughly 19.[6] In the last ARCHER period, the corresponding value is, as we have seen, 243.
- The factor POSSESSOR ANIMACY had its heyday in the 18th century. In the 1750-1799 period, for example, if the possessor was human instead of non-human, the odds for the *s*-genitive increased by a factor of approximately 418. This effect is near-categorical. The factor's effect size then started to decline substantially in the first half of the 19th century, with a slight recovery during the second half of the 20th century. Hence, in Present-Day English, if the possessor is human instead of non-human, the odds for the *s*-genitive increase

by a factor of "only" about 12.

- The importance of POSSESSUM LENGTH for predicting genitive choice is a late 20th century phenomenon. Between 1650 and 1950, the factor had virtually no effect. In ARCHER's 1950-1989 period, though, every additional graphemic character in the possessum increases the odds for the *s*-genitive by 21%. The growing importance of the factor can be interpreted as a diachronic strengthening of the principle of end-weight (Behaghel 1909/1910; Wasow 2002).
- POSSESSOR THEMATICITY is a factor that did not consistently favor the *s*-genitive prior to the second half of the 19th century (it did favor the *s*-genitive in the 1650-1699 period, however). So, in Present-Day English (1950-1989), a one-unit increase in a possessor head noun's local text frequency increases the odds for the *s*-genitive by 5%; the corresponding figure for the 1900-1949 period is 4%, and for the 1850-1899 period 9%. Conversely, in the 1700-1750 period for example, a one-unit increase in a possessor head noun's local text frequency actually *de*creased the odds for the *s*-genitive by 24%.

Other language-internal conditioning factors (for example, POSSESSOR LENGTH) do not interact significantly with real time.

8.3. Changing genitive grammars: interim summary

The logistic regression model reported in this section correctly predicts 96% of the genitive outcomes in the dataset. It accomplishes this high predictive accuracy by considering, among other things, four interaction terms between language-internal conditioning factors and real time. This is another way of saying that news writers' genitive choice grammars have changed in four principal ways. In regard to the overall most crucial predictor for genitive choice, GENITIVE RELATION, we observe that in 20th-century news texts, ownership relations hugely favor the *s*-genitive. This partly explains why *s*-genitives are comparatively frequent in late 20th-century news texts – more frequent, at any rate, than predicted by linear regression (cf. Figure 2). The status of POSSESSUM LENGTH and POSSESSOR THEMATICITY (the sixth and eighth most important factors in logistic regression, respectively) has a similar arc: Both factors have become potent during the past 150 years, and both favor the *s*-genitive. The aforementioned three factors, then, are likely to be implicated in the expansion of the *s*-genitive in 20th-century news texts. Why, then, was the slump in *s*-genitive frequencies around the first half of the 19th century so much more severe than one would have predicted on the basis of varying input frequencies alone (again, cf. Figure 2)? The answer is a change in the status of POSSESSOR ANIMACY, the overall third most important factor for predicting genitive choice. At the same time that input frequencies of human possessors collapsed, the animacy constraint was relaxed such that human possessors favored the *s*-genitive less. We are dealing here with a probabilistic grammar change that amplified the decline of *s*-genitive frequencies. Incidentally, this finding ties in with Rosenbach (2007: Figure 1 and Figure 2), who – in a study on determiner genitives – likewise reports a relaxation of the animacy constraint starting in the 19th century.

9. Discussion and conclusion

Our investigation into genitive variability in news texts during the Late Modern English period has the following principal findings. Somewhat contrary to expectations, we do not observe a gradual, consistent expansion of the *s*-genitive between 1650 and the Present-Day English period. What we see instead is a V-shaped frequency pattern: The *s*-genitive started out with a relatively high market share in the first two ARCHER periods. Its frequency then collapsed around the first half of the 19th century. However, it subsequently recovered and is more popular than ever in the last ARCHER period. What happened? On the one hand, linear regression analysis shows that slightly over half of this frequency fluctuation can be traced back to what this study has termed "environmental" factors: varying input frequencies of ownership relations and human possessors, as well as gradually increasing possessum lengths. Recall, for instance, that ownership relations – which favor the *s*-genitive – were particularly rare in the first half of the 19th century. On the other hand, though, the probabilistic machinery underpinning genitive choice grammars has been subject to change in the past 350 years or so. Logistic regression analysis demonstrated that in regard to four language-internal conditioning factors, news writers changed their ways of choosing genitives. As for the drop in *s*-genitive frequencies around 1800, we have seen that at roughly around the same time, the animacy constraint was relaxed considerably such that human possessors favored the *s*-genitive less and less. As for the *s*-genitive's subsequent recovery after 1850, the data suggest that the increase of *s*-genitives in (i) ownership relations, (ii) long possessums, and (iii) thematic possessors may all be implicated in the re-popularization of the *s*-genitive.

One reviewer took issue with this study's liberal adoption of stock trading and market metaphors, submitting that historical change "is supposed to underlie some systematicity and be predictable in some way, rather than being subject to jitters and irrational usage by investors". As a matter of fact, we believe that there is probably a good deal of interaction between fluctuating environmental factors ('jitters') and changing genitive choice grammars ('systematicity'). Consider the factor possessor animacy: Between 1650 and 1750, about 38% of all possessors were human. After 1750, only about 28% of all possessors were human. It is indeed likely that this environmental change in the news genre habitat – an emerging preference for writing about things, collectives, and organizations, not persons – triggered the subsequent collapse of the animacy constraint. In a similar vein, from the middle of the 19th century onwards we are witnessing increasingly thematic possessors (cf. Table 9) and increasingly longer possessum phrases (cf. Table 6) across the board, a likely outgrowth of the "informational explosion" (Biber 2003: 170). It is thus not far-fetched to claim that the news genre's genitive choice system underwent some functional, and maybe even predictable, re-organization in the past 150 years or so: The *s*-genitive is increasingly attracted not by human possessors but by ownership relations, and it has come to be the preferred option to code thematic possessors and long possessums.

As for the morpho-syntactic status of the *s*-genitive in English, notice now that the *s*-genitive is regularly cited as a counterexample to the unidirectionality (less grammatical > more grammatical) of grammaticalization. This is because the *s*-genitive has developed from a fairly pristine inflection in Old English times to a more clitic-like marker in Present-Day English where it

can attach to, e.g., whole NPs (as in [*the Department of Transport*]$_{por}$'***s*** [*civil aviation policy directorate*]$_{pum}$ <ARCHER 1989tim2.n8b>). Against this backdrop, (parts of) the history of the English *s*-genitive is seen by many as one of those rare examples for "*de*grammaticalization" (cf., e.g., Janda 1980; Newmeyer 1998: section 5.3.4) or even "*anti*grammaticalization" (Haspelmath 2004). What picture would emerge if we applied customary (de)grammaticalization diagnostics to our longitudinal findings?

We begin by considering the workhorse diagnostic in the corpus-based grammaticalization literature: a construction's overall text frequency (see, for example, Krug 2000; Mair 2004; but cf. Hoffmann 2004). The idea is that "[l]ack of paradigmatic variability [...] accounts for the ubiquity of a feature in the texts of a language" (Lehmann 1995: 142), which is why "sheer textual frequency is prima facie evidence of degree of grammaticalization" (Hopper and Traugott 1993: 110). In this view, we would diagnose degrammaticalization between 1650 and approximately 1850 (which is the period that leads to the slump in *s*-genitive frequencies), and grammaticalization after 1850 – especially during the 20th century. The problem, of course, is that as we have seen overall text frequencies are a function of grammar-internal changes, which are expressive of a constructions's grammaticalization status, but also of environmental changes. The latter are a confounding factor that is *per se* unrelated to grammar changes and irrelevant to questions of (de)grammaticalization processes. In short, an interpretation that is based on grammar-internal changes only (as identified in logistic regression analysis; cf. Section 8) promises to be less *prima facie* and, indeed, more robust:

- Changes affecting the animacy constraint can be related to Lehmann's (1995) "paradigmatic variability" parameter, which is about "the freedom with which the language user chooses a sign" (Lehmann 1995: 137) and the dropping of selection restrictions (Lehmann 1995: 141) – such as the restriction that the *s*-genitive be used with human possessors only. We also know that in grammaticalization processes, it is common that "expressions for human concepts come to be used also for concepts that are inanimate" (Heine 1997: 87). By virtue of severely weakening a selection restriction, then, the collapse of the possessor animacy constraint in 19th century news prose may be seen as the symptom of a grammaticalization process.
- Consider next the increasing importance of possessor thematicity and possessum length in the 20th century. Also along the lines of Lehmann's paradigmatic variability parameter, the *s*-genitive's emerging 20th century (pragmatic) sensitivity to characteristics of the immediate as well as wider discourse context can be interpreted as a development towards a freer "choice of items according to communicative intents" (Lehmann 1995: 164). So, the emergence of the somewhat soft thematicity constraint and the strengthening of the end-weight-related possessum length factor appear to render genitive choice more pragmatic and thus less obligatory in a categorical sense. We thus diagnose degrammaticalization in the 20th century.
- We have also seen that in the 20th century, the *s*-genitive has come to increasingly attract ownership relations. This phenomenon falls squarely within the remit of Lehmann's "paradigmatic integrity" parameter: "[g]rammaticalization rips off the lexical features until only the grammatical features are left" (Lehmann 1995: 129), a phenomenon widely

known as 'semantic bleaching' or 'desemanticization'. Crucially, what we are observing here is the exact opposite process: lexical and/or semantic features are actually *added* to the *s*-genitive. This constitutes further evidence for *s*-genitive degrammaticalization in the 20th century.

Our jury consisting of four grammar-internal changes thus finds the *s*-genitive to have been subject to some grammaticalization in the 19th century (thanks primarily to a relaxation of the animacy constraint), and to somewhat massive *de*grammaticalization during the 20th century, due to a strengthening of the genitive relation constraint and the emergence of textlinguistic factors favoring its usage. Now, in this study we have admittedly not marshaled empirical measures that directly asses the grammatical status of the *s*-genitive, yet it is certainly legitimate to adopt the standard view that strong grammaticalization typically correlates with more inflection-hood and weak grammaticalization with more clitic-hood (Hopper and Traugott 1993: 7). Based on this premise, then, we venture that the *s*-genitive may have become more clitic-like and less inflection-like in Present-Day English than it has been for quite some time. It is interesting to observe that this verdict contradicts the account that would emerge from a mere consideration of text frequencies. What appears to have happened is that the *s*-genitive lost market share prior to the middle of the 19th century because it was bleeding human possessors to the *of*-genitive. In a similar vein, however, the *s*-genitive became increasingly popular again after the middle of the 19th century because it was (i) semantically enriched, attracting ownership genitives that might have previously been coded by the *of*-genitive, and because (ii) thematic possessors increasingly favored the *s*-genitive.

The present contribution has sought to track and interpret genitive variability in Late Modern English news texts. As always, much remains to be done – for instance, an analysis such as the one offered here that is focussed on genitive variability in news texts only is woefully selective and incomplete. We also need data on other alternations (such as the dative alternation) to gauge whether, say, the collapse of the animacy constraint is specific to the genitive alternation or, in fact, a more general phenomenon. But be that as it may, we hope to have demonstrated that we can learn a lot from combining the careful philological study of historical data with state-of-the-art quantitative analysis methods.

Notes

[1]Due to low token counts (44 hits in ARCHER news), the pattern [full NP]' [full NP], as in *the dyers' decision*, was ignored.

[2]Due to low token counts, possessor phrases ending in <dge> and <z> were not considered.

[3]The statistics environment R (RDevelopmentCoreTeam 2009) was used for linear regression modeling (function `lm`).

[4]The statistics environment R (RDevelopmentCoreTeam 2009) was used for mixed effect modeling: library `lme4`, function `lmer` (family = "binomial", method = "Laplace").

[5]In regard to the scalar predictors POSSESSOR LENGTH, POSSESSUM LENGTH, POSSESSOR THEMATICITY, and LEXICAL DENSITY, we tested whether these should be modeled linearly or logarithmically by considering model deviances. It turns out that linear modeling yields lower model deviances in all cases and is thus appropriate for all four predictors.

[6]We obtain this figure by multiplying the main effect of the predictor according to Table 12 (242.88) by the odds ratio of the interaction term according to Table 13 (.08). Hence, $242.88 * .08 = 19.43$.

References

Altenberg, B. (1982). *The Genitive v. the Of-Construction. A Study of Syntactic Variation in 17th Century English*. Malmö: CWK Gleerup.

Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge, New York: Cambridge University Press.

Behaghel, O. (1909/1910). Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen 25*(110-142).

Biber, D. (2003). Compressed noun-phrase structure in newspaper discourse: the competing demands of popularization vs. economy. In J. Aitchison and D. Lewis (Eds.), *New Media Language*, pp. 169–181. London/New York: Longman.

Biber, D., E. Finegan, and D. Atkinson (1994). ARCHER and its challenges: compiling and exploring A Representative Corpus of Historical English Registers. In U. Fries, G. Tottie, and P. Schneider (Eds.), *Creating and using English language corpora: papers from the Fourteenth International Conference on English Language Research and Computerized Corpora*, pp. 1–13. Amsterdam: Rodopi.

Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

Bod, R., J. Hay, and S. Jannedy (Eds.) (2003). *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Bresnan, J., A. Cueni, T. Nikitina, and H. R. Baayen (2007). Predicting the dative alternation. In G. Boume, I. Kraemer, and J. Zwarts (Eds.), *Cognitive Foundations of Interpretation*, pp. 69–94. Amsterdam: Royal Netherlands Academy of Science.

Crawley, M. J. (2005). *Statistics: an introduction using R*. Chichester, West Sussex, England: J. Wiley.

Dahl, L. (1971). The s-genitive with non-personal nouns in modern English journalistic style. *Neuphilologische Mitteilungen 72*, 140–172.

Gries, S. T. (2002). Evidence in Linguistics: three approaches to genitives in English. In R. M. Brend, W. J. Sullivan, and A. R. Lommel (Eds.), *LACUS Forum XXVIII: What Constitutes Evidence in Linguistics*, pp. 17–31. Fullerton, CA: LACUS.

Haspelmath, M. (2004). On directionality in language change with particular reference to grammaticalization. In O. Fischer, M. Norde, and H. Perridon (Eds.), *The nature of grammaticalization*, Typological Studies in Language, pp. 17–44. Amsterdam: Benjamins.

Hawkins, J. (1994). *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Heine, B. (1997). *Possession: cognitive sources, forces, and grammaticalization*. Cambridge: Cambridge University Press.

Hinrichs, L. and B. 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*(3), 437–474.

Hoffmann, S. (2004). Are Low-Frequency Complex Prepositions Grammaticalized? On the Limits of Corpus Data – and the Importance of Intuition. In C. Mair and H. Lindquist (Eds.), *Corpus Approaches to Grammaticalisation in English*, pp. 171–210. Amsterdam/Philadelphia: Benjamins.

Hopper, P. J. and E. C. Traugott (1993). *Grammaticalization*. Cambridge, New York: Cambridge University Press.

Hundt, M. and C. Mair (1999). 'Agile' and 'uptight' genres: the corpus-based approach to language change in progress. *International Journal of Corpus Linguistics 4*, 221–242.

Jaeger, T. F. (2006). *Redundancy and Syntactic Reduction in Spontaneous Speech*. Phd thesis, Stanford University.

Janda, R. D. (1980). On the Decline of Declensional Systems: the Overall Loss of OE Nominal Case Inflections and the ME Reanalysis of *-es* as *his*. In E. C. Traugott, R. Labrum, and S. Shepherd (Eds.), *Papers from the 4th International Conference on Historical Linguistics*, pp. 243–252. Amsterdam/Philadelphia: Benjamins.

Jucker, A. (1993). The genitive versus the of-construction in newspaper language. In A. Jucker (Ed.), *The Noun Phrase in English. Its Structure and Variability*, pp. 121–136. Heidelberg: Carl Winter.

Krug, M. G. (2000). *Emerging English modals: a corpus-based study of grammaticalization*. Berlin, New York: Mouton de Gruyter.

Labov, W. (1966a). The linguistic variable as a structural unit. *Washington Linguistics Review 3*, 4–22.

Labov, W. (1966b). *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language 45*, 715–762.

Labov, W. (1982). Building on empirical foundations. In W. Lehmann and Y. Malkiel (Eds.), *Perspectives on Historical Linguistics*, pp. 17–92. Amsterdam: Benjamins.

Lehmann, C. (1995). *Thoughts on Grammaticalization*. LINCOM Studies in Theoretical Linguistics. München, Newcastle: LINCOM EUROPA.

Mair, C. (2004). Corpus linguistics and grammaticalisation theory: statistics, frequencies, and beyond. In C. Mair and H. Lindquist (Eds.), *Corpus Approaches to Grammaticalisation in English*, pp. 121–150. Amsterdam/Philadelphia: Benjamins.

Mustanoja, T. F. (1960). *A Middle English syntax*, Volume I. Helsinki: Société Néophilologique.

Newmeyer, F. J. (1998). *Language form and language function*. Cambridge: MIT Press.

Osselton, N. (1988). Thematic Genitives. In G. Nixon and J. Honey (Eds.), *An Historic Tongue: studies in English linguistics in memory of Barbara Strang*. London: Routledge.

Potter, S. (1969). *Changing English*. London: AndrÃl' Deutsch.

Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik (1985). *A Comprehensive Grammar of the English Language*. London, New York: Longman.

Raab-Fischer, R. (1995). Löst der Genitiv die of-Phrase ab? Eine korpusgestützte Studie zum Sprachwandel im heutigen Englisch. *Zeitschrift für Anglistik und Amerikanistik 43*(2), 123–132.

RDevelopmentCoreTeam (2009). R: a language and environment for statistical computing.

Rosenbach, A. (2002). *Genitive variation in English: conceptual factors in synchronic and diachronic studies*. Berlin, New York: Mouton de Gruyter.

Rosenbach, A. (2003). Aspects of iconicity and economy in the choice between the s-genitive and the of-genitive in English. In G. Rohdenburg and B. Mondorf (Eds.), *Determinants Of Grammatical Variation in English*, Topics in English Linguistics, pp. 379–412. Berlin, New York: Mouton de Gruyter.

Rosenbach, A. (2005). Animacy versus weight as determinants of grammatical variation in English. *Language 81*(3), 613–644.

Rosenbach, A. (2007). Emerging variation: determiner genitives and noun modifiers in English. *English Language and Linguistics 11*(1), 143–189.

Rosenbach, A. (2008). Animacy and grammatical variation – findings from English genitive variation . *Lingua 118*(2), 151–171.

Rosenbach, A. and L. Vezzosi (2000). Genitive constructions in Early Modern English: New evidence from a corpus analysis. In R. Sornicola, E. Poppe, and A. Shisha-Halevy (Eds.), *Stability, Variation and Change of Word-Order Patterns over Time*, pp. 285–307. Amsterdam, Philadelphia: Benjamins.

Sakamoto, Y. and H. Akaike (1978). Analysis of cross classified data by AIC. *Annals of the Institute of Statistical Mathematics 30*(1), 185–197.

Szmrecsanyi, B. (2004). On Operationalizing Syntactic Complexity. In G. Purnelle, C. Fairon, and A. Dister (Eds.), *Le poids des mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis. Louvain-la-Neuve, March 10-12, 2004*, Volume II, pp. 1032–1039. Louvain-la-Neuve: Presses universitaires de Louvain.

Szmrecsanyi, B. (2006). *Morphosyntactic persistence in spoken English: A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin/New York: Mouton de Gruyter.

Szmrecsanyi, B. (2009). Typological parameters of intralingual variability: grammatical analyticity versus syntheticity in varieties of English. *Language Variation and Change 21*(3), 1–35.

Szmrecsanyi, B. (2010). The English genitive alternation in a cognitive sociolinguistics perspective. In D. Geeraerts, G. Kristiansen, and Y. Peirsman (Eds.), *Advances in Cognitive Sociolinguistics*, pp. 141–166. Berlin, New York: Mouton de Gruyter.

Szmrecsanyi, B. and L. Hinrichs (2008). Probabilistic determinants of genitive variation in spoken and written English: a multivariate comparison across time, space, and genres. In T. Nevalainen, I. Taavitsainen, P. Pahta, and M. Korhonen (Eds.), *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*, pp. 291–309. Amsterdam: Benjamins.

Tagliamonte, S. (2001). Comparative sociolinguistics. In J. Chambers, P. Trudgill, and N. Schilling-Estes (Eds.), *Handbook of Language Variation and Change*, pp. 729–763. Malden and Oxford: Blackwell.

Taylor, J. (1989). Possessive Genitives in English. *Linguistics 27*, 663–686.

Thomas, R. (1931). *Syntactical processes involved in the development of the adnominal periphrastic genitive in the English language*. Phd thesis, University of Michigan.

Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change 9*, 81–105.

Wasow, T. (2002). *Postverbal behavior*. Stanford, CA: CSLI Publications.

Weinreich, U., W. Labov, and M. I. Herzog (1968). Empirical foundations for a theory of language change. In W. Lehmann and Y. Malkiel (Eds.), *Directions for historical linguistics*, pp. 96–195. Austin: University of Texas Press.

Zaenen, A., J. Carlette, G. Garretson, J. Bresnan, A. Koontz-Garboden, T. Nikitina, M. C. O'Connor, and T. Wasow (2004). Animacy encoding in English: why and how. In D. Byron and B. Webber (Eds.), *Proceedings of the 2004 ACL workshop on discourse annotation, Barcelona, July 2004*, pp. 118–125.

Zwicky, A. M. (1987). Suppressing the Zs. *Journal of Linguistics 23*, 133–148.