English Language and Linguistics
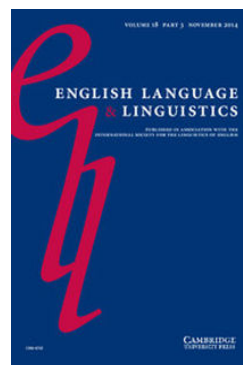
Additional services for *English Language and Linguistics:*

Email alerts: Click here
Subscriptions: Click here
Commercial reprints: Click here
Terms of use : Click here

Variation in English genitives across modality and genres

JASON GRAFMILLER

Link to this article: http://journals.cambridge.org/abstract_S1360674314000136

How to cite this article:

Request Permissions : Click here

Variation in English genitives across modality and genres[1]

J A S O N   G R A F M I L L E R
KU Leuven
(Received 4 January 2012; revised 10 March 2014)

The choice of genitive construction in English is conditioned by numerous semantic, syntactic and phonological factors. The present study explores the influence of these factors across different modalities (speech vs writing) and genres (e.g. press, fiction, etc.), and models the mediating effect of language-external variables on internal cognitive and linguistic factors within the context of a probabilistic grammar of genitive choice. The discussion revolves around debates concerning the driving force(s) behind recent changes in newspaper genitives, concluding that the trend reflects a push toward more economical modes of expression in reportage texts. Curiously, analysis finds few significant interactions with low-level processing-related factors, e.g. possessor frequency and lexical density – a surprising result in light of recent research. However, analysis further reveals significant inter-genre variability among several other crucial factors including possessor animacy and final sibilancy, which are significantly reduced in journalistic prose. These latter findings offer indirect evidence in favor of economization, and offer insight into the connections between external stylistic concerns, specific linguistic practices and internal probabilistic weights associated with specific grammatical constructions.

1   Introduction

The English genitive construction involves two well-known alternatives: the *s*-genitive (1), and the *of*-genitive (2).

(1) . . . and ran **the Grizzlies' winning streak** to four straight. <Brown Corpus, A13>
(2) He was **the sidekick of Gene Autry** I believe –
     <Switchboard Corpus, 2131>

These constructions encode a host of different semantic relations (e.g. Taylor 1996; Anschutz 1997; Rosenbach 2002; Kreyer 2003; Payne & Berlage 2011), and the range of contexts that allow alternation between the two types is quite diverse (Biber *et al.* 1999; Rosenbach 2002; Hinrichs & Szmrecsanyi 2007). Despite their semantic near equivalence, however, the two constructions do not occur with the same likelihood in such contexts.

No single factor has been shown to influence this choice more than possessor animacy (Rosenbach 2005, 2008). Animate, and especially human, possessors strongly

---

favor the *s*-genitive construction, in some cases nearly categorically (Tagliamonte & Jarmasz 2008). Other well-known factors include the length of the possessor and possessum phrases, the presence of a sibilant segment at the right edge of the possessor NP, possessor topicality or 'thematicity', and the semantic relation between the possessor and possessum (Osselton 1988; Anschutz 1997; Kreyer 2003; Rosenbach 2005; Hinrichs & Szmrecsanyi 2007; Tagliamonte & Jarmasz 2008). Information status (givenness) is also thought to be important (Biber *et al.* 1999; Shih *et al.* forthcoming), as is the lexical richness of the local context, at least for written English (Szmrecsanyi & Hinrichs 2008). Effects of phonological factors other than final sibilancy are less understood, though recent research hints at a small but significant influence of rhythmic structure (Ehret 2011; Shih *et al.* forthcoming). Finally, several studies have shown significant influence of social and stylistic variables, such as register (Rosenbach 2002), genre (Jucker 1993), and age and gender (Tagliamonte & Jarmasz 2008).

Analysis of data from a variety of sources has shown that these effects are reliable across time and region (e.g. Altenberg 1982; Rosenbach 2002; Hinrichs & Szmrecsanyi 2007; Tagliamonte & Jarmasz 2008; Szmrecsanyi 2010; Ehret 2011; Wolk *et al.* 2013). Studies have also observed variable patterns in the frequency of the two constructions across modality, where it has been noted that *s*-genitives are becoming increasingly more common in certain written styles, specifically journalistic prose (Jucker 1993; Leech & Smith 2006; Szmrecsanyi & Hinrichs 2008; Szmrecsanyi 2010, 2013). Parallels between spoken and press texts in the frequency of *s*-genitives suggest a kind of convergence of the two styles, but what is driving the trend toward an increased frequency of *s*-genitives is still not fully understood. In particular, the relation of genitive choice in press texts to that of other written styles, and of these various written styles to speech, is a topic in need of investigation, and is therefore the focus of this article.

The present study explores the nature of the relationship between the stylistic norms associated with different genres and modes of language production, and the internal linguistic factors known to influence the choice of genitive construction in American English. More narrowly, it focuses on a close examination of those interactions between factors which speak to recent debates over the forces driving changes in the frequency of newspaper *s*-genitives. Using multilevel logistic regression analysis of a combined corpus of six different styles of written and spoken genitives, the cross-genre stylistic variation in genitive construction choice is modeled in terms of systematic adjustments to the weights of specific features in a probabilistic grammar of genitive choice. In this spirit, the present study falls within the purview of cognitive sociolinguistics (e.g. Szmrecsanyi 2010) through its emphasis on corpus-driven analysis of the quantitative relation between internal cognitive and linguistic factors and 'speaker and situation related variation' (Geeraerts *et al.* 2010: 8) within a probabilistic model of linguistic knowledge, à la Bresnan *et al.* (2007) and Bresnan & Ford (2010). On this view, I take the variability found here in the weights of various factors across styles and genres to be a reflection of the usage-based linguistic knowledge internalized by experienced writers, e.g. journalists and academics, working within (relatively) narrow professional

communities. Such communities develop their own stylistic conventions over time, but I show here that stylistic conventions cannot easily be reduced to simple functional explanations.

2   Journalists' genitives: economization vs colloquialization

Studies of variation across English genres have shown that some written genres, e.g. fiction and letters, are susceptible to influence from spoken styles, while other genres, e.g. academic and newspaper prose, are more often subject to external stylistic constraints (Biber 1988, 1995; Biber & Finegan 1989). At the same time, newspaper prose is in fact very open to innovation, despite, or perhaps because of, such external pressures (Jucker 1993; Hundt & Mair 1999; Biber 2003). It is likely then, that in some genres, influence from spoken registers may be strong, but the evidence for this influence may be masked by other 'top-down' pressures – the need to save column inches, for example.

Turning to the genitive construction, the increasing frequency of the *s*-genitive in both speech and writing over the last half century is evident from real time studies of variation in both American and British English (Szmrecsanyi & Hinrichs 2008), and is further supported by apparent time trends in Canadian English (Tagliamonte & Jarmasz 2008). While it is not clear what is driving this change in speech (see Hinrichs & Szmrecsanyi 2007 for some discussion), explanations for the shift in writing have generally fallen into two camps. The 'colloquialization' account argues that changes in writing are due to an increasing tendency of written genres to more closely resemble spoken registers (Biber 1988; Biber & Finegan 1989; Jucker 1993; Hundt & Mair 1999; Leech & Smith 2006). It is hypothesized that in order to engage more readers, newspapers and news magazines have increasingly been adopting casual styles that parallel patterns in speech (Biber 2003). On the other hand, academic and technical publications have tended toward prose with more complex grammatical structures and larger specialized vocabularies, due to their targeting more specialized markets.

Alternatively, it may be that journalists' writing has come to parallel speech for independent stylistic reasons. An 'informational explosion' in certain written genres (e.g. newspaper reportage) has created 'pressure to communicate information as efficiently and economically as possible, resulting in compressed styles that depend heavily on tightly integrated [NP] constructions' (Biber 2003: 170). As the more compact construction, the *s*-genitive is naturally preferred to the *of*-genitive in these dense contexts. Szmrecsanyi & Hinrichs (2008) argue along these lines in their study of changes in genitive use in journalistic writing, basing their conclusions primarily on significant effects of high lexical density and possessor thematicity (measured as text frequency) in their written, but not spoken, data. They infer from their analysis that it is the need to achieve a maximal degree of 'compactness' in writing that is behind the significance of these factors. However, since the written data they examined came from newspaper texts only, it is still uncertain to what extent their findings extend to other written genres.

Table 1. *Sections of Brown Corpus*

Genre	Description	Section	No. of texts
PRESS	Newspaper reportage	A	16
NON-FICTION	Memoirs, Biography, Belles-lettres, etc.	G	28
LEARNED	Academic and technical writing	J	12
GENERAL FICTION	General fiction	K	17
WESTERN FICTION	Adventure and western fiction	N	16

The exact character of genitive choice across written styles is therefore still an open question. Genres vary considerably along numerous dimensions such as the level of concreteness, interactive focus, and aesthetic emphasis (Biber & Finegan 1989; Biber 1995), and it is likely that the strengths of different internal factors influencing grammatical choice will vary accordingly. For instance, given the slower, careful construction of professional writing, it may be that the influence of factors related to real-time processing, e.g. end weight, is diminished in certain styles. Other factors, e.g. rhythmic structure or semantic relation, might play a larger role, due perhaps to greater room for artistic playfulness in some genres (e.g. fiction). In what follows, I provide a framework for understanding cross-genre variation in the genitive choice in terms of the modulation of probabilistic grammatical weights associated with certain stylistic pressures and/or conventions. The results presented here also speak to debates over competing explanations for recent diachronic trends in genitive use.

3 The corpus

Genitive tokens were selected from two well-known corpora, primarily to take advantage of their extensive prior semantic and syntactic annotation. Spoken genitives were extracted from the manually parsed Penn Treebank portion of the Switchboard corpus of American English (Marcus *et al.* 1993; Godfrey *et al.* 1992), using selection methods and phonological annotations detailed in Shih *et al.* (2009, forthcoming), and animacy annotations derived from Zaenen *et al.* (2004). Written data were selected from the Boston University Noun Phrase Corpus of approximately 10,000 tokens of English genitive constructions (O'Connor *et al.* 2006), which comprised genitive tokens collected from five different sections of the Brown corpus of written American English, each representing a separate written genre (table 1).[2] Each NP token was further manually annotated for the animacy, definiteness, givenness, length, and syntactic category (proper noun, common noun, pronoun, etc.) of the possessor and possessum NPs (Garretson *et al.* 2004; O'Connor *et al.* 2006).

---

[2] Readers familiar with these corpora will no doubt note the temporal mismatch between the spoken and written data. This issue is addressed in section 5.3, which presents results from an apparent time simulation that comports with recent studies of real time change, e.g. Szmrecsanyi & Hinrichs (2008).

3.1   Selection criteria

For the final data set, a token was considered interchangeable if it could easily be converted from one variant to the other without the insertion or deletion of any lexical items from either the possessor or possessum. The one exception to this was the obligatory deletion of the determiner in the possessum, as in converting *the principles of the country* to *the country's principles*. Aside from this intuitive rule, several additional heuristics aided in classifying the genitive tokens. Genitives not meeting the interchangeability requirement fell into several classes, all of which have been discussed extensively in previous literature (e.g. Quirk *et al.* 1985; Biber *et al.* 1999; Rosenbach 2002, 2006; Kreyer 2003; Hinrichs & Szmrecsanyi 2007).

(a) *Elliptical genitives*. Constructions lacking an overt possessum phrase: *I'll meet you at Pat's*, *Her eyes are like a hawk's*
(b) *Descriptive genitives*. Constructions in which the possessor is not itself referential, but acts as a classifier that 'contributes to the denotation of the head noun' (Rosenbach 2006: 81): *women's magazines*, *smoker's cough*
(c) *Fixed expressions*. Phrases that have become conventionalized in one construction or the other: *arm's reach*, *Murphy's law*, *President of the United States*, *the law of the land*
(d) *Authored works*. Titles of books, films, musical pieces, etc., that are pre-modified by a possessor denoting their creator (Hinrichs & Szmrecsanyi 2007): *Allen Ginsburg's* Howl, *Racine's* Phèdre
(e) Any possessum headed by a determiner other than *the*, e.g. demonstratives, quantifiers and indefinite determiners: *nineteen years of age*, *this responsibility of the teacher*, *a characteristic of Trevelyan's prose*
(f) *Collective possessums*. *Of*-genitive constructions that refer to collections of individuals: *the group of students*, *the jury of eight women and four men*

Finally, all tokens involving pronominal possessors were excluded from the data set. Jucker (1993) observes in his corpus of newspaper prose that pronominal possessors are nearly categorical in their preference for the prenominal position, and this trend is characteristic of English more generally (Huddleston & Pullum *et al.* 2002; Rosenbach 2002). Furthermore, pronominal *of*-genitives are proscribed quite stringently in the prescriptive literature (Murphy 1997: 128; Quinion 2005). Such proscriptions may not play much of a role in casual speech, but in professionally edited writing they may be taken more seriously, resulting in a higher-level editorial purging of pronominal *of*-genitives.

3.2   Corpus results

Selection of the interchangeable genitives yielded a data set containing 3,612 genitive tokens (1,115 spoken; 2,497 written). In the present data, the *of*-genitive construction is more frequent than the *s*-genitive across both spoken and written varieties.[3] Within

---

[3] This pattern runs counter to the patterns reported by other recent studies of spoken English (Szmrecsanyi & Hinrichs 2008; Tagliamonte & Jarmasz 2008). However, after controlling for known factors, there is a

Table 2. *Distribution of genitive constructions across genres*

Genre	*of*-genitive		*s*-genitive		Total
SPOKEN	659	(59.1%)	456	(40.9%)	1,115
WRITTEN	1,612	(64.6%)	885	(35.4%)	2,497
Learned	314	(77.8%)	90	(22.2%)	404
Non-Fiction	623	(67.8%)	296	(32.2%)	919
General Fiction	227	(63.4%)	131	(36.6%)	358
Western Fiction	221	(55.3%)	178	(44.7%)	399
Press					
	227	(52.4%)	206	(47.6%)	433

the written data, the *of*-genitive is more frequent than the *s*-genitive for each of the five genres (table 2). Though the relative frequencies are similar, there are highly significant differences between the proportions of the two constructions in the GENERAL FICTION, $\chi^2\,(1,358) = 25.74$, $p < 0.0001$, NON-FICTION, $\chi^2\,(1,919) = 126.70$, $p < 0.0001$, and LEARNED genres, $\chi^2\,(1, 404) = 129.96$, $p < 0.0001$. Even in WESTERN FICTION there is a marginally significant preference for the *of*-genitive, $\chi^2\,(1, 399) = 4.63$, $p < 0.05$. No difference in the proportion of genitives emerges in the PRESS genre, $\chi^2(1, 433) = 1.02$, $p = 0.31$, suggesting there is something special about newspaper prose with respect to this particular linguistic variable – a finding that is certainly not new (Jucker 1993; Hundt & Mair 1999; Szmrecsanyi & Hinrichs 2008).

4   Conditioning factors

The factors included in the present analyses fall loosely into four classes: Semantic-pragmatic, Phonological, Processing and Economy-related factors (Hinrichs & Szmrecsanyi 2007). The annotation method for each factor is outlined in the following sections.

4.1   Semantic and pragmatic factors

ANIMACY. Rosenbach (2006: 105−6) notes that the genitive alternation is sensitive to at least a four-way animacy ranking (human > animal > collective > inanimate). This is the coding used by Hinrichs & Szmrecsanyi (2007) and Szmrecsanyi & Hinrichs (2008) in their analyses, which confirmed Rosenbach's rankings, while Tagliamonte & Jarmasz (2008) find similar results using a simple human/nonhuman distinction. Since factors other than animacy are the primary concern of the present study, animacy was coded as a binary living/non-living distinction: living beings (humans, animals

significant preference for the *s*-genitive in the spoken data here (see section 6.1.2). Furthermore, all of these studies examined different data sets collected from entirely different populations. The reader should therefore be cautious in interpreting cross-study discrepancies in these raw frequencies.

Table 3. *Categorization of semantic relations*

	Examples
PROTOTYPICAL	
Legal ownership	*Scotty's bed*, *the house of Francisco Chaves*
Kinship	*the children of these people*, *the Czarina's cousins*
Body part	*the fish's mouth*, *the back of the horse*, *the man's lashes*
Part–whole	*the car's starter*
NON-PROTOTYPICAL	*an employer's rights*, *the owner of the store*, *the bag's contents*

and organizations) were coded as animate, everything else was coded as inanimate. Representative examples are shown in (3) and (4).

(3) **Animate possessors**:
   Eileen's address, *the death of the officer*, *Papa's footsteps*, *the truest societies of Christians*, *the hoot of an owl*

(4) **Inanimate possessors**:
   the value of voting, *the edge of my chair*, *the winners of last year*, *today's society*, *Utah's weather*, *society's ills*

SEMANTIC RELATION. Classifying the various semantic relations encoded by the English genitive construction can be quite challenging (e.g. Taylor 1996: 339−48; Payne & Berlage 2011). Semantic relation was coded according to the schema used by Rosenbach (2002: 120−3) in her experiments on English genitives. Rosenbach collapsed several classes of relations into a single category of 'PROTOTYPICAL' genitives, which favor the *s*-genitive, and all others into a category marked simply as 'NON-PROTOTYPICAL'. Prototypical genitives were any examples that fell into one four subclasses, listed in table 3.

Every observed token of the genitive construction was manually annotated for the type of semantic relation, and the codings for each were cross-checked among three different annotators.

INFORMATION STATUS. The influence of possessor information status is still somewhat controversial. While some have claimed that given possessors favor the *s*-genitive (Biber *et al.* 1999), others have found no significant effect of possessor givenness on English genitives (Hinrichs & Szmrecsanyi 2007). To avoid the risk of (unintentional) bias from using automated coding (see Hinrichs & Szmrecsanyi 2007: 451) each possessor was manually annotated as either given or new. Givenness was defined as explicit reference to the possessor in any manner in the preceding ten lines.

THEMATICITY. Osselton (1988) examined the tendency of topical or 'thematic' possessors to favor the *s*-genitive construction, even when they are otherwise disfavored. Thematic possessors are those

of central concern to the writer's immediate theme: that is, in a book on phonetics, *sound* will get its [s-]genitive, in one on farming, *soil* will do so, and in a book on economics you can expect to find *a fund's success*, *the pound's strength, inflation's consequences*, and so on (Osselton 1988: 143).

As a metric of thematicity, the (log) frequency of the head noun of the possessor NP in each text or conversation was used (Hinrichs & Szmrecsanyi 2007).

4.2 Phonological factors

FINAL SIBILANT. It is well established that the presence of a final sibilant in the possessor NP disfavors the use of the *s*-genitive (Quirk *et al.* 1985; Biber *et al.* 1999; Kreyer 2003; Hinrichs & Szmrecsanyi 2007, inter alia).[4] All possessor NPs were automatically annotated for the presence or absence of a final sibilant, using the phonetic transcriptions in the Carnegie Mellon University (CMU) Pronunciation Dictionary (v. 0.7a). Words that were not found in the CMU dictionary were subsequently hand-coded. Six phones were classified as sibilant: [s] as in *house*, [z] as in *news*, [ʃ] as in *Bush*, [tʃ] as in *avalanche*, [dʒ] as in *bridge* and [ʒ] as in *garage*.

RHYTHMICITY. The comparative rhythmicity between the two genitive constructions was coded according to the methods developed by Shih *et al.* (forthcoming), who demonstrated that preferred genitive construction will be the one resulting in a stress pattern that most closely accords with an ideal weak–strong stress alternation. Each genitive token was coded for the lexical stresses of each word, obtained from the CMU dictionary. For each token, two stress distance measures were calculated: one for the *s*-genitive, and one for the *of*-genitive. For the *s*-genitive construction, the distance was the number of unstressed syllables between the final stressed syllable of the possessor NP and the first stressed syllable of the possessum noun (phrase). Conversely, the distance in the *of*-genitive alternate was the number of syllables between the final stress of the possessum NP and the first stress of the possessor NP. Within our data, *s*-genitive distance ranged from 0 to 4, while *of*-genitive distance ranged from 1 to 6.

From these two measures, a measure of the comparative rhythmicity between the two constructions was computed using the formula in (5) (see Shih *et al.* forthcoming).

(5) **Rhythm measure**:

RHYTHM $= |$ *of*-genitive DISTANCE $- 1 | - |$ *s*-genitive DISTANCE $- 1|$

When the *of*-genitive distance is greater than the *s*-genitive, RHYTHM is positive, and the greater the value, the more rhythmic the *s*-genitive is relative to the *of*-genitive. When RHYTHM is negative, the *s*-genitive distance is greater than the *of*-genitive, and the smaller the negative value of RHYTHM, the more rhythmic the *of*-genitive is compared to the alternative *s*-genitive. In short, the higher RHYTHM is above 0, the more the *s*-genitive should be preferred, and the lower RHYTHM is below 0, the more the *of*-genitive should be favored. When RHYTHM is 0, neither construction is preferred to the other.

[4] Note that the *s*-genitive can also pose an orthographic problem for possessors ending in <s>, as writers may struggle with how to appropriately represent the possessive marker, i.e. as <'s> or <'>. Opinions regarding the correct methods for writing are still divided (see, e.g., Kaye 2004), and uncertainty may lead writers to avoid the *s*-genitive in such instances.

4.3 Processing-related factors

END-WEIGHT. There is a rich literature examining the effects of 'end-weight' (Wasow 2002) on syntactic in a variety of syntactic phenomena (e.g. Altenberg 1982; Hawkins 1994; Rosenbach 2005; Bresnan *et al.* 2007). In the genitive construction, it is known that heavier, i.e. longer, possessors favor the *of*-genitive construction, since the possessor is placed second in such instances. Following Bresnan & Ford (2010) and Grafmiller & Shih (2011), a comparative measure of end weight was calculated for each token using the ratio of the number of orthographic words in the possessor NP to the number of words in the possessum.

PERSISTENCE. The phenomenon known as 'structural parallelism' (Weiner & Labov 1983) or 'persistence' (Szmrecsanyi 2006) refers to the influence of speakers' prior choice of a given construction on their uses of the same construction in subsequent contexts. Examples of *s*-genitive persistence in spoken and written language can be seen in (6).

(6) (a) Some of the, some of **the women's roles**, I think, are almost for the worse, because we're losing out on some things, going back to work, but I think if we can, if we can expand **the men's roles** at the same time, . . .<Switchboard A.2370>

  (b) . . . which was his nickname for a messenger who had worked in the White House since **Teddy Roosevelt's administration**, and discuss the welfare of some one of the animals. It was part of **Little Jack's work** to look after the dogs. <Brown G41>

Following Hinrichs & Szmrecsanyi (2007), each genitive token was coded for the use of an *s*-genitive in the first genitive choice context – if any – immediately preceding it.

4.4 Economy-related factors

TYPE–TOKEN RATIO. Being the shorter, more 'compact' alternative, the *s*-genitive construction tends to be preferred in environments of high lexical density (Biber *et al.* 1999; Biber 2003; Hinrichs & Szmrecsanyi 2007; Szmrecsanyi & Hinrichs 2008). To measure lexical richness, the type–token ratio (TTR) over the five sentences preceding and following each token was calculated. With the exception of western fiction, all genres exhibit a significant tendency for *s*-genitives to occur more in lexically dense environments (One-sided Mann-Whitney test: SPOKEN $U = 129721, p < 0.0001$; LEARNED $U = 11964, p < 0.05$; NON-FICTION $U = 73728, p < 0.001$; PRESS $U = 20289.5, p < 0.01$; GEN. FICTION $U = 13229.5, p < 0.05$; WEST. FICTION $U = 19553.5, p = 0.459$).

5 Analysis and results

The contributions of each of these factors were investigated using multilevel, mixed-effects logistic regression modeling which estimates the combined contribution of a set of conditioning factors in predicting an outcome (Pinheiro & Bates 2000; Harrell 2001; Gelman & Hill 2007). In the case of binary logistic regression, the outcome

Table 4. *Model 1: summary statistics*

N	3564	Log likelihood	−1096
df	60	*AIC*	2312
κ	12.42	Adjusted D_{xy}	0.927
% correct	89.3(62.9)	*c* index	0.963

is the probability of observing one of two discrete alternatives – in this study the likelihood of the *s*-genitive construction. Multilevel regression analysis estimates the effect size and direction of each individual predictor, and provides a measure of the variability in the data explained by the predictors. Importantly, it not only allows us to control for systematic variation along known parameters in ways that significance tests over univariate data cannot, but it also enables the partial pooling of data across specific groups of interest to adjust for idiosyncratic variation within those groups (the so-called 'random' effects). In essence, '[mixed-effects regression analysis] is the closest a corpus linguist can come to conducting a controlled experiment' (Hinrichs & Szmrecsanyi 2007: 459).

To examine the interrelation between external stylistic factors and internal linguistic ones in genitive choice, two analyses were conducted. The first involved fitting a model to the full corpus of spoken and written data, henceforth Model 1, and the second involved fitting a similar model to only the written subset, henceforth Model 2. Both models initially included interactions of the six levels of GENRE (five in the case of Model 2) with all other predictors discussed previously. In light of recent observations regarding the moderating effect of animacy on other factors (e.g. Tagliamonte & Jarmasz 2008; Shih *et al.* forthcoming), interactions of animacy with all other factors were also included. As step-wise variable selection methods have several known drawbacks (see, e.g., Harrell 2001: 56–60), all predictors and interactions were left in the models unless the absolute value of the coefficient was less than the standard error. For the random effects, the models included the maximal structure justified by model comparison via likelihood ratio tests (Baayen 2008: 253–6). This process justified a by-speaker/author varying intercept as well as a by-speaker varying slope for POSSESSOR ANIMACY in Model 1, and an additional by-author varying slope for TYPE−TOKEN RATIO in Model 2. Both models were evaluated for data multi-collinearity and extreme leverage (48 tokens, 1.3%, removed from Model 1; 33 tokens, 1.3%, from Model 2), and validated against over-fitting using bootstrap re-sampling with random replacement (10,000 runs).

5.1   *Model 1: the combined data*

5.1.1   *Model summary*

Summary statistics of Model 1 are presented in table 4. Predictive performance of the model is excellent, as indicated by its concordance probability (*c* index), computed by
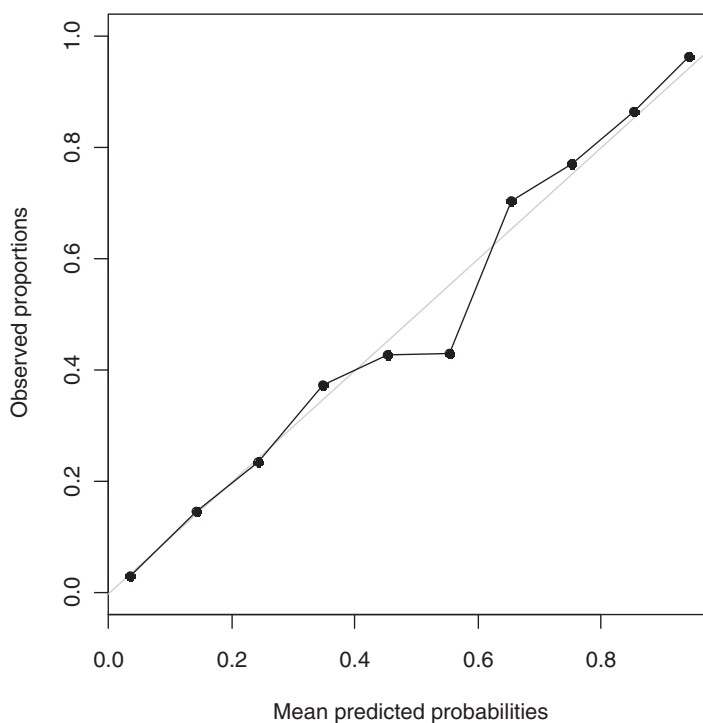
Figure 1. Model 1 predictions by observed corpus tokens

taking the proportion of all possible -*s*-genitive and -*of*-genitive pairs in the data for which the model correctly assigns a higher probability to the *s*-genitive.[5]

The predictive accuracy, presented visually in figure 1, provides a more intuitive measure of model performance. The *x* axis represents the probabilities assigned by Model 1 (binned into deciles), and the *y* axis the proportions of *s*-genitives in the data for each of the bins predicted by the model. The closer the points are to a straight line, the more accurate the model.

5.1.2 *Main effects*
Statistics for the main effect sizes and explanatory power of the individual predictors in Model 1 are presented in table 8. Model predictions are for the *s*-genitive, so factors with positive coefficients favor the *s*-genitive, while negative coefficients reflect a preference for the *of*-genitive.

---

[5] *C* values above 0.8 are considered indicative of good model fit. The adjusted D_{xy} (Somers' D_{xy}) is another measure of model fitness derived from *c*. Both statistics are considered to be better suited than R^2 for evaluating logistic models (Harrell 2001: 247) . The statistic κ is a measure of data (multi-)collinearity (Baayen 2008: 181–2).

Table 5. *Main effects of individual predictors in Model 1. Model predictions are for the* s-*genitive*

Predictor	Coefficient	Std error	Wald Z	p-value
(Intercept)	− 0.591	0.164	− 3.61	0.000
GENRE (base = SPOKEN):				
GENERAL FICTION	− 1.115	0.394	− 2.83	0.005
LEARNED	− 2.053	0.450	− 4.56	0.000
NON-FICTION	− 1.106	0.272	− 4.07	0.000
PRESS	0.580	0.252	2.30	0.021
WESTERN FICTION	− 1.127	0.410	− 2.75	0.006
ANIMATE POSSESSOR	4.365	0.247	17.66	0.000
RHYTHM	0.061	0.229	0.27	0.791
FINAL SIBILANT (present)	− 1.512	0.360	− 4.20	0.000
GIVEN POSSESSOR	− 0.431	0.297	− 1.45	0.147
SEMANTIC RELATION (prototypical)	0.913	0.435	2.10	0.036
POSS'R / POSS'M LENGTH	− 1.861	0.348	− 5.34	0.000
TYPE-TOKEN RATIO	0.480	0.167	2.87	0.004
POSSESSOR TEXT FREQ	0.053	0.270	0.20	0.843
PRECEDING S-GENITIVE	0.121	0.123	0.99	0.323
Random effects	*Variance*	*Std deviation*		
SPEAKER (Intercept, N = 770)	1.183	1.088		
SPEAKER × POSSESSOR ANIMACY	1.430	1.196	*p* < 0.000	

For levels of GENRE, the model coefficients are to be interpreted relative to the baseline level SPOKEN, which is to say, the model's main effects represent those for the spoken data. Numerical predictors were centered and standardized by dividing by twice the standard deviation. This protects against harmful effects of data multi-collinearity, and normalizing by two standard deviations enables direct comparison between effect sizes (coefficients) of numerical and binary predictors (Gelman 2008). For ease of comparison, model estimates (in linear log odds scale) for the significant main predictors are shown in figure 2.

Of all the factors, ANIMACY has the largest effect size by far. Averaging over the entire dataset, animate possessors are almost sixty times more likely to be used in the *s*-genitive.[6] Prototypical genitive relations, e.g. kinship or part-whole relations, also favor the *s*-genitive, though by a much smaller factor of about 2.5. The presence of a FINAL SIBILANT on the possessor has a significant effect, though in the opposite direction. Possessor NPs ending in a sibilant are roughly 4 times *less* likely to occur in the *s*-genitive. Similarly, increasing the possessor/possessum LENGTH RATIO also disfavors the *s*-genitive. For example, moving from a ratio of 1/2 to a ratio of 3 results in a nearly

---

[6] The odds ratio is computed by exponentiating the coefficient β by base e: e^{β}. Probabilities correspond to the inverse-logit of the coefficient, calculated as $P = e^{\beta}/(1 + e^{\beta})$.
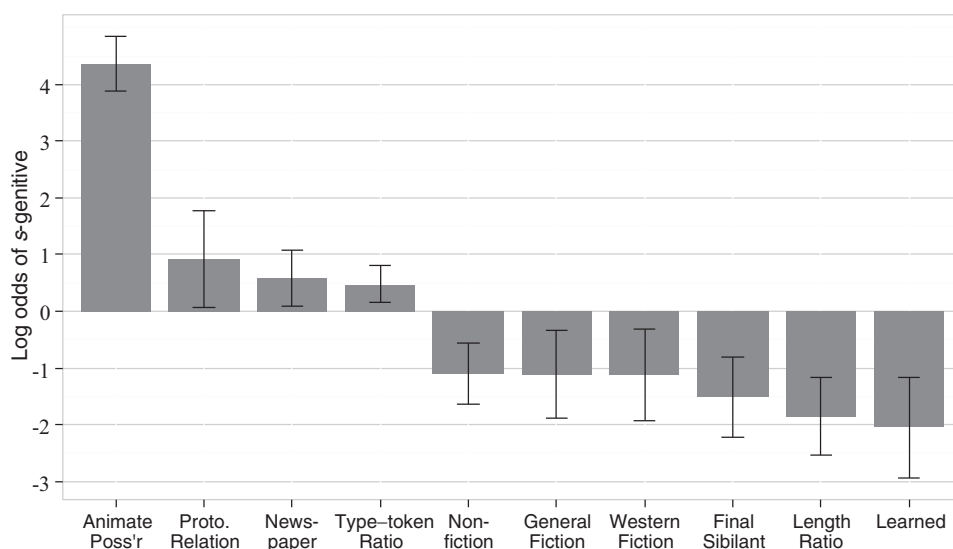
Figure 2. Estimates of significant main effects of Model 1 (with 95 percent confidence
intervals). Genre levels are relative to baseline level SPOKEN

two-and-a-half-fold decrease in *s*-genitive likelihood.[7] Lexically richer environments,
as measured by TTR, also significantly favor the use of the *s*-genitive, though the small
effect is difficult to interpret meaningfully.

When it comes to the stylistic effects of GENRE, the most important finding is that,
in general, individuals are significantly less likely to use *s*-genitives in writing than in
spoken conversation. The lone exception to this is PRESS. A potential genitive is about
1.8 times *more* likely to be an *s*-genitive in newspaper writing than it is in speech.
Newspaper prose therefore exhibits a stark contrast to other genres in its preference for
the *s*-genitive construction, in accordance with previous findings (Jucker 1993; Leech
& Smith 2006; Szmrecsanyi & Hinrichs 2008).

Turning to the non-significant predictors, there is no main effect of RHYTHM,
PERSISTENCE or POSSESSOR TEXT FREQUENCY in the model. Recent studies have found
significant, though small, effects for each of these factors (e.g. Hinrichs & Szmrecsanyi
2007; Shih *et al*. forthcoming), however, the present study differs from these (and
others) in its use of mixed-effects models, which treated each individual speaker/author
as a source of random variability in the data. Individual speaking and writing styles
can vary quite considerably, and so it is possible that the idiosyncratic patterns with
respect to individual users would account for enough variability to mask any aggregate
effects of these particular predictors here. Nevertheless, despite the lack of significant

---

[7] Due to the normalizing process used here, the units of numerical variables are no longer on their natural scales.
Rather they are on a 2 standard deviation scale such that a 1 unit increase in the model input represents a shift
from one standard deviation below, to one standard deviation above the mean of the raw variable. In the case of
the length ratios, the mean is approximately 1.75 with a standard deviation of 1.25.

Table 6. *Significant interaction effects in Model 1. All genre interactions are
interpreted relative to the baseline genre level* SPOKEN

Predictor	Coefficient	Std error	Wald Z	p-value
Genre × POSS'R ANIMACY:				
GENERAL FICTION	− 1.252	0.546	− 2.29	0.022
NON‑FICTION	− 1.015	0.404	− 2.51	0.012
PRESS	− 2.462	0.422	− 5.84	0.000
Genre × POSS'R GIVENNESS:				
GENERAL FICTION	1.545	0.578	2.67	0.008
LEARNED	2.810	0.653	4.30	0.000
NON‑FICTION	1.277	0.382	3.35	0.001
PRESS	1.549	0.413	3.76	0.000
WESTERN FICTION	1.887	0.606	3.12	0.002
Genre × LENGTH RATIO				
GENERAL FICTION	2.477	0.658	3.76	0.000
LEARNED	1.773	0.573	3.10	0.002
WESTERN FICTION	2.277	0.617	3.69	0.000
Genre × TYPE‑TOKEN RATIO:				
LEARNED	1.158	0.585	1.98	0.048
Genre × POSS'R TEXT FREQ:				
GENERAL FICTION	1.325	0.489	2.71	0.007
LEARNED	0.977	0.469	2.08	0.037
ANIMACY × RHYTHM	− 0.571	0.240	− 2.38	0.017
ANIMACY × SEMANTIC RELATION	1.388	0.420	3.31	0.001

main effects for these predictors, there are significant interactions involving these and
other factors, to which I turn now.

5.1.3 Interactions
Table 6 lists the significant interactions in Model 1, and figure 3 shows a visual profile
of the variation in main effects grouped by genre.

While the model provides little evidence for any interaction effects of phonological
or economy-related predictors across spoken and written genitives in general, specific
genre interactions offer some insight into the importance of stylistic pressures on
two key factors, namely end weight and possessor animacy. For LENGTH RATIO, the
model reports a negative main effect ($\beta = -1.86$), interpreted as the effect size of the
baseline level SPOKEN. The interaction term for each other genre is positive, indicating
a reduced effect of end weight, though this effect is significant only for the GENERAL
FICTION, WESTERN FICTION, and LEARNED genres. Moreover, for the two fiction genres,
the direction of the effect is reversed: a larger possessor/possessum length ratio actually
favors the *s*-genitive. A plausible explanation for this is that these genres are simply less
susceptible to the processing demands underlying weight effects. This seems especially
true of academic prose (LEARNED), as the genre is well-known (and sometimes parodied)
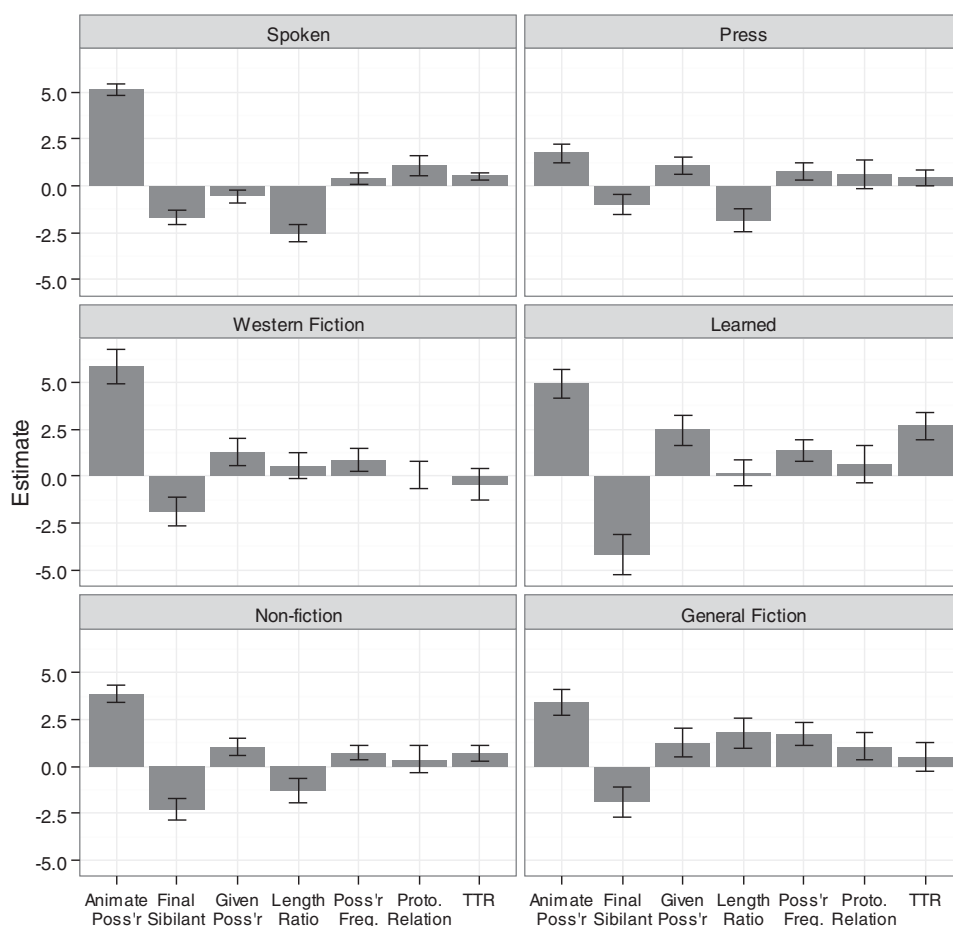for its complex syntax and larger vocabulary. The other two genres comprise examples

Figure 3. Profile of significant effects by genre

of narrative fiction, a diverse category that not only covers a wide range of individual styles – compare, say, Hemingway vs Faulkner – but also often emphasizes aesthetic expression over communicative economy. For these genres, efficiency therefore might take a backseat to artistic style. Furthermore, Hinrichs & Szmrecsanyi (2007: 465) suggest that the influence of parsing efficiency is operative primarily in more colloquial styles, which would explain the fact that newspaper prose and narrative non-fiction (the latter being perhaps slightly less prone to aesthetic flourishes than fiction) do not differ significantly from speech in this regard.

In contrast to end weight, the effect of animacy is qualitatively the same across all genres, i.e. it has the same direction, but it is quantitatively weaker within some genres, especially newspaper prose, when compared to spoken English. But a closer look reveals that the reasons for this weaker effect are not the same across written genres. GENERAL FICTION and NON-FICTION show a marked decrease in log odds for
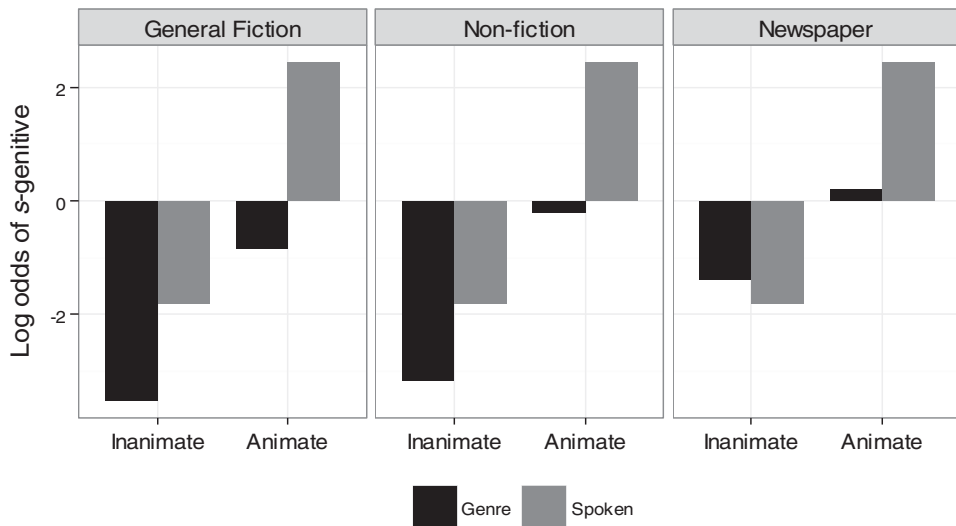
Figure 4. Significant interactions of GENERAL FICTION, NON-FICTION and PRESS genres and POSSESSOR ANIMACY in Model 1

both animate and inanimate possessors, reflecting a general bias for *of*-genitives in these genres. But this is not the pattern found in newspaper texts. There is indeed a decrease in *s*-genitive likelihood for animate possessors, but also a marked *increase* in *s*-genitive likelihood for inanimate possessors when compared to speech (figure 4). Together with the main effect preference for *s*-genitives in PRESS styles, the overall reduction in the influence of animacy on genitive choice in newspaper writing suggests a strong stylistic bias in favor of the *s*-genitive in these texts. While in other written genres, writers strongly prefer inanimate possessors in *of*-genitives, and animates in *s*-genitives somewhat less so, journalists are (implicitly) less discriminating. This result is compatible with the notion that reporters' need for economy of expression is reflected in unconscious adjustments to their probabilistic knowledge (i.e. factor weights).

Also of interest are those measures thought to be indicative of either the shifting of the written norm toward more a conversational style, i.e. RHYTHM, or of increased stylistic pressure to use more compact expressions, i.e. TEXT FREQUENCY (thematicity) and TTR (Szmrecsanyi & Hinrichs 2008). No significant interaction of GENRE × RHYTHM was found, though when it comes to the thematicity and TTR, results from Model 1 point to academic prose (LEARNED) and not newspaper prose as the truly odd duck. GENERAL FICTION also exhibits a significant interaction with text frequency, though why this should be so is uncertain. It may be due to sampling differences – fiction texts are excerpts of longer material, while newspaper texts constitute an entire article – though this is true of all the non-press genres. It could also simply be random chance that the particular set of texts exhibit this pattern. Further investigation is certainly needed.

Finally, there are the significant interactions of ANIMACY × RHYTHM and ANIMACY × SEMANTIC RELATION. The negative interaction coefficient for ANIMACY × RHYTHM

($\beta = -0.57$) shows that as we move from inanimate to animate possessors, the influence of rhythm significantly decreases. In fact, there is no significant influence of RHYTHM among animate possessors. The model does no better than chance when predicting the effects of RHYTHM when the possessor is an animate entity. The opposite pattern obtains for the ANIMACY × SEMANTIC RELATION interaction.

5.1.4 Interim summary

Model 1 examined the ways in which the linguistic factors driving the genitive alternation vary between speech and several different styles of writing. By comparing spoken and written data together in the same model, we are able to see exactly how different factors vary between styles. The results, especially for the comparison of spoken and press English, largely accord with recent work on journalists' genitives which foregrounds the role of the economizing pressures of this style (Szmrecsanyi & Hinrichs 2008; Szmrecsanyi 2013). It would seem from Model 1, however, that the differences between newspaper genitives and those of other written genres lie in the (unconscious) moderation of the primary influences on genitive choice, e.g. animacy and end weight, and not in the differential effects of other factors more directly associated with processing efficiency or informational density in newspaper texts.

What Model 1 does not do, though, and what has not been done by others, is compare the effects of linguistic factors in press writings to other written genres directly. The second part of this study therefore examines the forces driving the spread of the s-genitive in newspaper writing, by directly comparing the probabilistic patterns of genitive use in this genre to other styles of written English. We want to know: how different is press writing from other writing with respect to this linguistic variable?

5.2 The written data

To answer this question, a second model (Model 2) was run on only the written data, treating PRESS as the baseline level of a five-level GENRE factor. All other predictors in the model were the same as for Model 1.

Before moving on to the results of Model 2, I return to the distribution of genitives among the different written genres, presented graphically in figure 5. With the exception of press texts, all genres contain significantly higher proportions of of-genitives than s-genitives, with the largest disparity in the LEARNED genre.

Writing in these texts tends toward more deliberate word and construction choice, and unlike with other genres of written English, this trend toward greater complexity – and by implication, greater formality – has been growing stronger in academic writing over time (Biber 1988, 1995). Between the two extremes lie the three other genres: GENERAL FICTION, WESTERN FICTION and NON-FICTION. Biber (1988) situates all these genres (including PRESS and LEARNED) along an 'informational vs involved focus' dimension, which marks 'high informational density and exact informational content versus affective, interactional, and generalized content' (107). Yet, despite

Table 7. *Model 2: summary statistics*

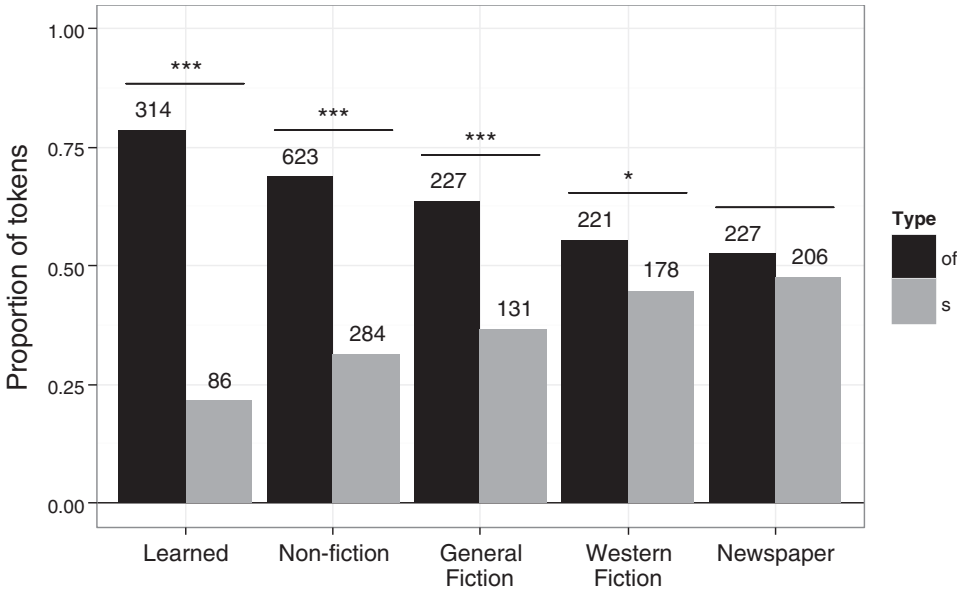| N	2497	Log likelihood	−906.1
df	49	*AIC*	1910
κ	14.19	Adjusted Nagelkerke R^2	0.543
% correct	86.22	Adjusted D_{xy}	0.848
(baseline)	64.56	*c* index	0.924

Figure 5. Distribution of written genitives by genre (* = p < .05, ** = p <.01, *** = p < .001)

their similar informational focus, the largest divergence in the distribution of genitives and other constructions (see Biber 2003) is between academic and journalistic prose styles. As I show in section 6.2.2, this divergence is reflected in adjustments to the probabilistic weights of several key factors affecting genitive choice. These results fit with the intuition that the *of*-genitive is the more formal variant (Rosenbach 2002: 39; Tagliamonte & Jarmasz, 2008).

5.2.1 Model 2 summary and main effects
Statistics for Model 2 are presented in table 7. As with Model 1, Model 2 performs quite well, as indicated by the high *c* statistic. Model 2 accurately predicts 86 percent of the possible outcomes (see figure 6), and bootstrap cross-validation confirms that it does not over-fit the data. Overall, the explanatory power and fit of the model is excellent.

Results from Model 2 are mostly consistent with those of Model 1. ANIMACY is again the most important predictor of genitive choice by far, followed by significant

Table 8. *Statistics of individual predictors in Model 2. Model predictions are for the*
s-*genitive*

Predictor	Estimate	Std error	Wald Z	p-value
(INTERCEPT)	0.354	0.216	1.64	0.102
GENRE (base = PRESS)				
GENERAL FICTION	− 2.420	0.461	− 5.25	0.000
LEARNED	− 3.297	0.550	− 6.00	0.000
NON-FICTION	− 2.155	0.344	− 6.26	0.000
WESTERN FICTION	− 2.873	0.556	− 5.17	0.000
POSSESSOR ANIMACY	2.160	0.389	5.55	0.000
RHYTHM	− 0.200	0.288	− 0.70	0.486
FINAL SIBILANT	− 1.146	0.334	− 3.43	0.001
POSSESSOR GIVENNESS	1.150	0.310	3.71	0.000
SEMANTIC RELATION	0.543	0.537	1.01	0.312
POSS'R / POSS'M LENGTH	− 1.630	0.354	− 4.60	0.000
TYPE-TOKEN RATIO	0.280	0.300	0.94	0.350
POSSESSOR TEXT FREQ	0.940	0.381	2.47	0.014
PERSISTENCE	0.104	0.155	0.67	0.502
Random effects	*Variance*	*Std deviation*		
AUTHOR (Intercept, $N = 153$)	1.207	1.099		
AUTHOR × POSSESSOR ANIMACY	1.926	1.388	$p < 0.000$	
AUTHOR × TYPE-TOKEN RATIO	0.419	0.647	$p = 0.041$	

Table 9. *Interaction effects in Model 2. Non-significant interactions not shown*

| Predictor	Coefficient	Std error	Wald Z	p-value
Genre × POSS'R ANIMACY:				
GENERAL FICTION	1.693	0.715	2.37	0.018
LEARNED	2.708	0.819	3.31	0.001
NON-FICTION	2.055	0.554	3.71	0.000
WESTERN FICTION	4.098	0.896	4.57	0.000
Genre × FINAL SIBILANT:				
LEARNED	− 2.835	1.019	− 2.78	0.005
NON-FICTION	− 1.132	0.510	− 2.22	0.027
Genre × POSS'R GIVENNESS:				
LEARNED	1.708	0.746	2.29	0.022
Genre × LENGTH RATIO:				
GENERAL FICTION	2.085	0.606	3.44	0.001
LEARNED	1.550	0.548	2.83	0.005
WESTERN FICTION	1.988	0.598	3.33	0.001
Genre × TYPE-TOKEN RATIO:				
LEARNED	1.514	0.627	2.42	0.016
ANIMACY × SEMANTIC RELATION	2.101	0.621	3.39	0.001
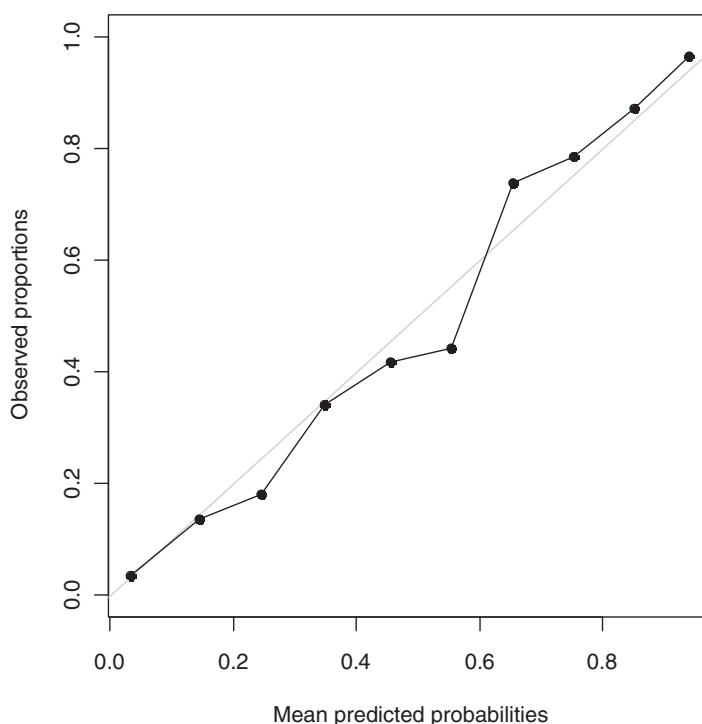
Figure 6. Predictions for Model 2 fixed effects by observed corpus tokens

contributions of GENRE, the presence of a FINAL SIBILANT, LENGTH RATIO, possessor GIVENNESS, and finally the TEXT FREQUENCY of the possessor (figure 7). The details of the main effects are shown in table 11.

As the findings of Model 1 suggested, there is a great deal of variation between newspaper prose and other writing styles when it comes to genitive choice. Simply knowing the genre in which a possessor–possessum pair occurs tells us more about the author's likely choice of genitive construction than any other single factor besides possessor animacy. In the most extreme example, a genitive is over twenty times more likely to be an *s*-genitive in a newspaper article than in an academic journal (LEARNED). As expected, FINAL SIBILANT and LENGTH RATIO have significant negative effects on *s*-genitive likelihood, while the TEXT FREQUENCY of the possessor has a significant positive influence, unlike in Model 1. Moving from tokens featuring possessors in the lowest frequency range (1 occurrence) to those with possessors in the highest range (> 9 occurrences), the likelihood of the *s*-genitive increases by a factor of roughly 2.5. Possessor GIVENNESS also plays a role in predicting written genitives, with given possessors being 3.26 times more likely to occur in the *s*-genitive. This is not surprising given the significant positive interactions of GIVENNESS and GENRE observed in Model 1. Lastly, SEMANTIC RELATION, PERSISTENCE, RHYTHM and TTR each failed to achieve significance as main effects.
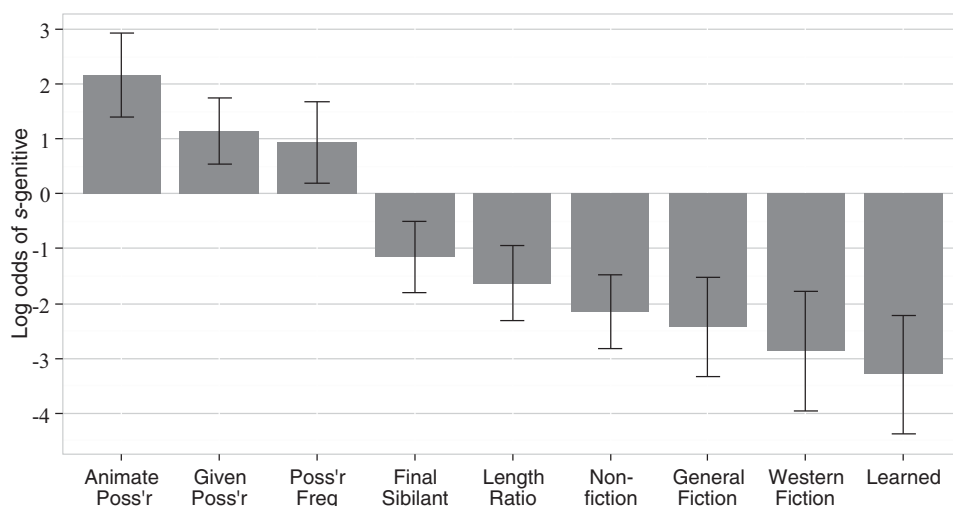
Figure 7. Estimates of significant main effects of Model 2 (with 95 percent confidence
intervals). Genre levels are relative to baseline level PRESS

5.2.2 Interactions

With a couple exceptions, the significant interactions in Model 2 largely accorded
with the findings of Model 1. In comparing written genres to speech, Model 1 reported
significant positive interactions of the genres GENERAL FICTION, LEARNED, and WESTERN
FICTION with LENGTH RATIO, but no interaction with PRESS, suggesting that newspaper
texts do not differ significantly from speech in this regard. Not surprisingly, Model 2
also reports a significant positive interaction with each of these genres when compared
to the baseline PRESS. Similarly, Model 2 finds significant positive interactions of
LEARNED × TYPE-TOKEN RATIO and ANIMACY × SEMANTIC RELATION, just as one would
expect given the previous model's results.

Two other sets of interactions deserve additional attention. First, unlike Model 1,
every written genre significantly interacts with possessor ANIMACY. The direction of
the effect does not change, but the influence of animacy greatly increases for every
genre relative to PRESS. Not only is the effect of animacy significantly weaker in most
written genres compared to speech, but it is significantly weaker still for newspaper
writing. This pattern can be represented as a ranking along a scale according to how
strong the effect of possessor animacy is for each genre.

(5) Importance of Possessor Animacy on genitive choice:
       Speech > Fiction, Narrative non-fiction, Academic > Press

This pattern is hinted at in Model 1 by the large negative interaction term of PRESS ×
ANIMACY, but only confirmed through direct comparison in Model 2.

Second, two genre interactions with FINAL SIBILANT also have significant effects ($p <$
0.05). Among non-press genres, the presence of a final sibilant on the possessor has the

expected effect: possessors with final sibilants are much more likely to be used in the *of*-genitive. In newspaper prose, however, the presence of a final sibilant has no effect on the choice of construction; either construction is equally likely. Taken together, these results show that the influences of factors which tend to favor the *of*-genitive in other genres are significantly attenuated in journalistic English.

5.3   A note on the corpora

Before moving on, it should be mentioned that the two corpora used in this study are not only split in modality, but in time as well. The Brown and Switchboard corpora comprise data from the early 1960s and early 1990s respectively, and it has been shown that changes in genitive frequencies have occurred in just that brief period (Leech & Smith 2006; Hinrichs & Szmrecsanyi 2007). It is therefore uncertain how much comparing the two directly can tell us about general trends in spoken versus written English, though it is possible to get an inkling of what a genuine synchronic comparison might reveal. Using birthdate information from Switchboard, a simulated apparent time study compared Brown data to Switchboard data from speakers who would have already been adults at the time Brown was collected (1960–1). Results from the time-adjusted corpus ($N = 2759$; 295 spoken) are very similar to those of the full corpus. Compared to speech, *s*-genitives in the time-adjusted corpus are significantly more likely in newspaper texts ($\beta = 0.99, p < 0.01$), but significantly less likely only in the LEARNED genre ($\beta = -1.38, p < 0.05$). This contrasts with results from the full model showing the *s*-genitive to be significantly *less* likely in all written genres (other than PRESS) when compared to conversational speech. Together, these results support previous findings that spoken use of the *s*-genitive has been increasing relative to most written genres over the last few decades (Szmrecsanyi & Hinrichs 2008), though in this case it has not quite reached the level in journalistic prose.

The interaction of genre with possessor ANIMACY demonstrates this even more clearly. Only press texts show a significant reduction in the influence of animacy in the time-adjusted corpus ($\beta = -2.44, p < 0.0001$), similar to what was found in the full model. With the exception of these texts, older speakers in the Switchboard data do indeed appear to pattern more like writers in the Brown corpus in their choice of genitive construction. In other respects, the time-adjusted model reports the very same significant interactions of GENRE × GIVENNESS and GENRE × LENGTH RATIO, adding confidence to the interpretations of Model 1.

6   Discussion and conclusion

Through an examination of forces influencing the choice of genitive construction in American English, the present study demonstrates how certain language-external factors, e.g. stylistic norms, unconsciously shape linguistic usage. The results presented here are entirely compatible with recent work (e.g. Szmrecsanyi & Hinrichs 2008; Szmrecsanyi 2013) pointing to certain genre-specific pressures on journalists' writing,

namely the need for maximizing compactness and informational density. Szmrecsanyi & Hinrichs (2008) argue for an economization explanation based primarily upon evidence of the significant influence of low-level processing-related factors such as lexical density and text frequency (thematicity). The present study finds no effects of these factors in newspaper prose, and they are found to be only marginally influential in other genres. The present results thus offer little evidence that lexical density and thematicity are uniquely influential on journalists' genitives. The high frequency of *s*-genitives in press texts may indeed be driven by pressures of economy, but these particular factors may not be reliable diagnostics of economic pressures in this regard.

But could 'economization' be manifested in other ways? The interactions of genre with other factors in this study are of particular relevance here. Specifically, in press texts, inanimate possessors and possessors with final sibilants, both of which typically favor the *of*-genitive, are much more likely to be used in the *s*-genitive. The influences of these factors, which are quite strong themselves, are significantly moderated by the stylistic demands of this genre. I interpret this is as indirect evidence for economization, but economization at higher, perhaps conscious, level. I suggest this is how external stylistic concerns, e.g. the need to be concise, translate to specific linguistic practices via speakers' (writers') unconscious adjustments to the internal probabilistic weights associated with specific grammatical constructions. There is ample evidence showing that newspaper texts are highly compact (e.g. Biber 1988, 2003), but there is also evidence that journalistic prose is uniquely flexible as well (Hundt & Mair 1999). Journalists are undoubtedly under pressure to maximize concision in the service of space limitations, which would be reflected in, among other things, the use of as many *s*-genitives as possible. The results presented here show that this stylistic shift can be understood in terms of the moderation of certain internal grammatical constraints, in particular, the weakening of those features that negatively influence *s*-genitive use: inanimate possessors or possessors with final sibilants. Journalists' genitives and genitives of conversational speech are influenced by these factors to significantly different degrees. This provides little evidence that newspaper prose is aligning with speech in the way that the colloquialization account would predict, and is compatible with the hypothesis that the driving force behind the rise in press *s*-genitives is something like economization.

It was also thought that prosodic influences on variation across genres would provide another dimension along which to evaluate the colloquialization account. Presumably, in genres where space is not an issue, one would not expect economy-related factors to show much of an influence. At the same time, the freedom from such constraints might provide space for other effects to show through. Looking at the effects of linguistic features (thought to be) more intimately related to speech might provide more direct evidence for or against the colloquialization account. Shih *et al.* (forthcoming) showed that rhythmic structure did indeed have a significant influence in speech, but the present study finds no main effect of rhythm on the choice of genitive construction in the full corpus, nor an interaction of rhythm with any genres. I conclude that this variable is simply too weak to use as a test of this hypothesis, though it is possible that other

measures of rhythmic optimization may provide a clearer picture (see Shih *et al.* forthcoming).

　Throughout this article, I have characterized variation in genitive use across spoken and written styles in terms of the modulation of factor weights in the probabilistic grammar underlying genitive choice in English. Such grammars give rise to gradient patterns in language use, and it is hoped that future work will continue to explore how the subtleties of culturally and socially conditioned variation can be captured via quantitative models of the kind presented here.

Author's address:

Department of Linguistics
KU Leuven
Blijde-Inkomsstraat 21
B-3000, Leuven
Belgium
Jason.Grafmiller@arts.kuleuven.be

References

Altenberg, B. 1982. *The genitive v. the* of-*construction: A study of syntactic variation in 17th Century English.* Malmö: CWK Gleerup.

Anschutz, Arlea. 1997. How to choose a possessive noun phrase construction in four easy steps. *Studies in Language* 21, 1–35.

Baayen, Harald. 2008. *Analyzing linguistic data*. Cambridge: Cambridge University Press.

Baayen, R. H., D. J. Davidson & D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 390–412.

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

Biber, Douglas. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.

Biber, Douglas. 2003. Compressed noun-phrase structures in newspaper discourse. In J. Aitchison & D. M. Lewis (eds.), *New media language*, 169–81. London: Routledge.

Biber, Douglas & Edward Finegan. 1989. Drift and the evolution of English style: A history of three genres. *Language* 65(3), 487–517.

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

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Baayen Harald. 2007. Predicting the dative alternation. In G. Boume, I. Kraemer & J. Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1), 168–213.

Ehret, Katharina. 2011. The role of rhythm as a factor in historical genitive variability. Masters thesis. University of Freiburg.

Garretson, Gregory, M. Catherine O'Connor, Barbora Skarabela & Marjorie Hogan. 2004. Coding practices used in the project Optimality Typology of Determiner Phrases. http://npcorpus.edu/documentation/index.html.

Geeraerts, Dirk, Gitte Kristiansen & Yves Peirsman. 2010. Introduction. Advances in cognitive sociolinguistics. In Dirk Geeraerts, Gitte Kristiansen & Yves Peirsman (eds.), *Advances in cognitive sociolinguistics*, 1–19. Berlin and New York: Mouton de Gruyter.

Gelman, Andrew. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* 27(15), 2865–73.

Gelman, Andrew & Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Godfrey, John J., Edward C. Holliman & Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of ICASSP-92*, 517–20.

Grafmiller, Jason & Stephanie Shih. 2011. New approaches to end weight. *Variation and Typology: New Trends in Syntactic Research*. Helsinki, Finland, 26.Aug.

Harrell, Frank. 2001. *Regression modeling strategies*. New York: Springer.

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hinrichs, Lars & Benedikt Szmrecsanyi. 2007. Recent changes in the function and frequency of standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language and Linguistics* 11(3), 437– 74.

Huddleston, Rodney & Geoffrey K. Pullum *et al.* 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Hundt, Marianne & Christian Mair. 1999. 'Agile' and 'uptight' genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4, 221–42.

Jucker, Andreas. 1993. The genitive versus the of-construction in newspaper language. In Andreas Jucker (ed.), *The noun phrase in English: Its structure and variability*, 121–36. Heidelberg: Carl Winter.

Kaye, Alan S. 2004. On the bare genitive. *English Today* 20(3), 57–8.

Kreyer, Rolf. 2003. Genitive and of-construction in modern English. Processability and human involvement. *International Journal of Corpus Linguistics* 8, 169–207.

Leech, Geoffrey & Nicholas Smith. 2006. Recent grammatical change in English 1961–1992: Some preliminary findings of a comparison of American with British English. In Antoinette Renouf & Andrew Kehoe (eds.), *The changing face of corpus linguistics*, 185–204. Amsterdam: Rodopi.

Marcus, Mitchell P., Mary Ann Marcinkiewicz & Beatrice Santorini. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19, 313–30.

Murphy, Raymond. 1997. *English grammar in use with answers: A self-study reference and practice book for intermediate students of English*. Cambridge: Cambridge University Press.

O'Connor, M. Catherine, Joan Maling, Arto Anttila, Vivienne Fong, Gregory Garretson, Barbora Skarabela, Marjorie Hogan & Fred Karlsson. 2006. *Boston University Noun Phrase Corpus*. http://npcorpus.bu.edu.

Osselton, Noel E. 1988. Thematic genitives. In Graham Nixon & John Honey (eds.), *An historic tongue: Studies in English linguistics in memory of Barbara Strang*. London: Routledge.

Payne, John & Eva Berlage. 2011. The effect of semantic relations on genitive variation. *International Society for the Linguistics of English 2*. Boston 17–21 June.

Pinheiro, José C. & Douglas M. Bates. 2000. *Mixed effects models in S and S-PLUS*. New York: Springer.

Quinion, Michael. November 12, 2005. Double possessive. *World Wide Words*. Retrieved 9 February 2010. From www.worldwidewords.org/qa/qa -dou3.htm.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London and New York: Longman.

Rosenbach, Anette. 2002. *Genitive variation in English: Conceptual factors in synchronic and diachronic studies*. Berlin and New York: Mouton de Gruyter.

Rosenbach, Anette. 2003. Aspects of iconicity and economy in the choice between the *s*-genitive and the *of*-genitive in English. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*, 379–412. Berlin and New York: Mouton de Gruyter.

Rosenbach, Anette. 2005. Animacy versus weight as determinants of grammatical variation in English. *Language* 81(3). 613–44.

Rosenbach, Anette. 2006. Descriptive genitives in English: A case study on constructional gradience. *English Language and Linguistics* 10(1), 77– 118.

Rosenbach, Anette. 2008. Animacy and grammatical variation: Findings from English genitive variation. *Lingua* 118, 151–71.

Schlüter, Julia. 2005. *Rhythmic grammar: The influence of rhythm on grammatical variation and change in English*. Berlin and New York: Mouton de Gruyter.

Shih, Stephanie, Jason Grafmiller, Richard Futrell & Joan Bresnan. Forthcoming. Rhythm's role in genitive and dative construction choice in spoken English. In Ralf Vogel & Rubem van der Vijver (eds.), *Rhythm in phonetics, grammar, and cognition*. Berlin: De Gruyter

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

Szmrecsanyi, Benedikt. 2010. The English genitive alternation in a cognitive sociolinguistics perspective. In Dirk Geeraerts, Greta Kristiansen & Yves Peirsman (eds.), *Advances in cognitive sociolinguistics*, 141–66. Berlin and New York: Mouton de Gruyter.

Szmrecsanyi, Benedikt. 2013. The great regression: Genitive variability in Late Modern English news texts. In Kersti Börjars, David Denison & Alan Scott (eds.), *Morphosyntactic categories and the expression of possession*, 59–89. Amsterdam and Philadelphia: John Benjamins.

Szmrecsanyi, Benedikt & Lars Hinrichs. 2008. Probabilistic determinants of genitive variation in spoken and written English: A multivariate comparison across time, space, and genres. In Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta & Minna Korhonen, (eds.), *The dynamics of linguistic variation: Corpus evidence on English past and present*, 291– 309. Amsterdam and Philadelphia: John Benjamins.

Tagliamonte, Sali & Lidia Jarmasz. 2008. Variation and change in the English genitive: A sociolinguistic perspective. Paper presented at the 82nd annual meeting of the Linguistic Society of America. Chicago, 4 January.

Taylor, John R. 1989. Possessive genitives in English. *Linguistics* 27, 663–86.

Taylor, John R. 1996. *Possessives in English*. Oxford: Clarendon Press.

Wasow, Thomas. 2002. *Postverbal behavior*. Stanford: CSLI Publications.

Weiner, E. Judith & William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19(1), 29–58.

Wolk, Christoph, Joan Bresnan, Anette Rosenbach & Benedikt Szmrecsanyi. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica* 30(3), 382–419.

Zaenen, Annie, Jean Carlette, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, M. Catherine O'Connor & Tom Wasow. 2004. Animacy encoding in English: Why and how. In Bonnie Webber & Donna Byron (eds.), *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, 118–25. East Stroudsburg, PA: Association for Computational Linguistics (ACL).