Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change

Christoph Wolk
Joan Bresnan
Anette Rosenbach
Benedikt Szmrecsanyi

We present a cross-constructional approach to the history of the dative alternation and the genitive alternation in Late Modern English (AD 1650 to AD 1990), drawing on richly annotated datasets and modern statistical modeling techniques. We follow sociolinguistic theory and the recent literature on gradient grammatical constraints in assuming that syntactic variation and change is probabilistic rather than categorical in nature. In this spirit, we show that historical dative and genitive variability exhibits some theoretically interesting common traits, such as the fact that the effect of more or less animate recipients in dative constructions and more or less animate possessors in genitive constructions appears to vary in parallel. This we interpret against the cultural backdrop of, for example, overall distributional changes in animacy categories, and we offer that distributional fluctuations such as these can trigger changes in probabilistic grammars in the long term.

Acknowledgments

1 Introduction

This study presents a novel, cross-constructional approach to the analysis of language variation and change, drawing on richly annotated datasets and state-of-the-art multivariate analysis tech-

1

niques. We are concerned with the history of the dative alternation (Bresnan et al. 2007; Bresnan and Ford 2010), as in (1), and the genitive alternation (Rosenbach 2002; Hinrichs and Szmrecsanyi 2007), as in (2), in Late Modern English.

(1) a. SUN., JAN. 23 – M.'s birthday – **wrote [M.] [an earnest loving note.]** <1887gibs.j6a>[1]
         (the ditransitive dative construction)
     b. SUN., JAN. 30 – Much better today. **Wrote [a note] [to M.]** expressive of my good state of feeling. <1887gibs.j6a>
         (the prepositional dative construction)

(2) a. before [**the Seneschal**]s [**Brother**] could arrive, he was secured by the Governor of Newport <1682pro1.n2b>
         (the *s*-genitive)
     b. the Duke of Norfolk, having lately received another Challenge from [**the Brother**] of [**the Seneschal**], went to the place appointed <1682pro1.n2b>
         (the *of*-genitive)

On the theoretical plane we follow sociolinguistic theory (e.g. Labov 1982; Tagliamonte 2001) and recent probabilistic approaches to language (e.g. Bod et al. 2003; Bresnan and Ford 2010) in assuming that syntactic variation – and change – is often probabilistic rather than categorical in nature.

The corpus database we tap is ARCHER, *A Representative Corpus of Historical English Registers*. Through substantial hand-coding, we derive extensively annotated datasets which characterize each dative or genitive observation in the dataset by way of a multitude of explanatory variables. Crucially, some of these are common to both alternations: consider the weight of the recipient/theme or possessor/possessum (the principle of 'end weight'), animacy of the recipient or possessor, or definiteness of the recipient or possessor. The way these factors affect syntactic choices is not specific to English but echoes cross-linguistic regularities (Aissen, 2003; Bresnan et al., 2001; Bresnan and Nikitina, 2009). We subsequently fit two logistic regression models with mixed effects that predict writers' dative and genitive choices by jointly considering all of the explanatory variables while also allowing for idiolectal and lemma-specific random effects. The regression models we present correctly predict over 90% of the dative and genitive observations in ARCHER.

Overall, we find that the dative alternation is a bit more stable in real time – frequency-wise and also probabilistically – than the genitive alternation, which exhibits more variability. Nonetheless, we find some theoretically interesting communalities, such as the fact that the effect of more or less animate recipients in dative constructions and more or less animate possessors in genitive constructions appears to vary in parallel. We interpret cross-constructional similarities like this against the cultural backdrop of, e.g., overall distributional changes in animacy categories, and we conjecture that such distributional fluctuations can trigger changes in probabilistic grammars in the long term.

This paper is structured as follows. In Section 2, we canvas the development of the dative and genitive alternation in the history of English. Section 3 introduces the data source. Section 4 defines the variable contexts. In Section 5, we discuss overall dative and genitive frequencies in

---

1 All linguistic examples in this paper are drawn from the ARCHER corpus (see Section 3) and are referenced by ARCHER text identifiers.

real time. Section 6 presents the predictor variables that we utilize to model dative and genitive variability in Section 7, in which we report regression models. In Section 8, we discuss and interpret the empirical facts, supplementing regression findings with more fine-grained univariate analyses where appropriate. Section 9 offers some concluding remarks.

2 A very short history of genitive and dative variation in English

The history of the genitive alternation is rather well-documented. Historically the *of*-genitive is the incoming form, which appeared during the ninth century. According to Thomas (1931: 284) (cited in Mustanoja 1960: 75), the inflected genitive vastly outnumbered the periphrasis with *of* up until the twelfth 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), such that the inflected genitive appeared to be dying out (Jucker 1993: 121). By the fourteenth century, the inflected genitive was increasingly confined to a functional niche coding animate possessors, possessive/subjective genitive relations, and topical possessors (Rosenbach 2002: 180-181). The Early Modern English period, however, sees a revival of the *s*-genitive, "against all odds" (Rosenbach 2002: 184). In Present-Day English, the *s*-genitive is comparatively frequent (Hinrichs and Szmrecsanyi 2007; Rosenbach 2002; Szmrecsanyi and Hinrichs 2008), and appears to be spreading right now (Dahl 1971; Potter 1969; Raab-Fischer 1995; Rosenbach 2003; Szmrecsanyi 2009). In the news genre specifically, Hinrichs and Szmrecsanyi (2007) diagnose a spread of the *s*-genitive in late twentieth century press English which appears to be due to a process of "economization". Despite the sizable body of research on genitive variability, however, we note that not much is known about genitive variability in the Late Modern English period.

It is fair to say that the dative alternation is one of the most extensively studied alternations in the grammar of English. Yet in comparison to the genitive alternation its history is comparatively understudied. We know that for most of the Old English period, the prepositional dative construction was not widely available (Mitchell 1985; Traugott 1992), and word order, in what in Modern English we would call the 'ditransitive' construction, was variable (Kemenade 1987; Koopman 1931; McFadden 2002); De Cuypere (2010) shows that this word order variability was subject to some of the same factors (animacy, pronominality, and so on) that drive the dative alternation in Modern English. Late Old English texts see the emergence – albeit initially subject to lexical restrictions (Allen 2009) – of the prepositional dative construction (Fischer 1992; Fischer and van der Wurff 2006), which during the Middle English period developed into "a fully productive alternative" (Fischer and van der Wurff 2006: 166) to the ditransitive dative construction. Conventional wisdom (for example, Fischer and van der Wurff 2006; McFadden 2002) holds that the loss of case distinctions during the Middle English period triggered the emergence of the prepositional dative construction as a means to avoid ambiguity, although there are alternative explanations, such as language contact with French (see Visser 1963). In any event, word order of nominal (but not pronominal) objects in the ditransitive dative construction was fixed along the lines of the Modern English pattern by the late fourteenth century (Allen 2009). The development of the dative alternation in Early Modern English and Late Modern English is virtually unexplored.

3 Data

The present study's data source is ARCHER, *A Representative Corpus of Historical English Registers*, release 3.1 (Biber et al. 1994). ARCHER covers the period between 1650 and 1990, spans about 1.8 million words of running text, and samples eight different registers (drama, fiction, sermons, journals/diaries, medicine, news, science, letters) and the 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 for each text is typically also available . Coverage of American English is restricted to three of the seven periods.

To obtain sufficient token counts, our investigation of the dative alternation draws on the ARCHER corpus in its entirety (that is, all periods, registers, and both American and British texts). Genitives are substantially more frequent than datives, and so we restrict attention to alternating genitives in ARCHER's British English news (a fairly 'agile' genre according to Hundt and Mair 1999) and letters section, a sub-corpus that comprises 257 texts and totals roughly 242,000 words of running text spread out fairly evenly over the real time periods sampled in ARCHER.

4 The variable contexts

This section circumscribes the variable contexts and thus defines interchangeable dative and genitive contexts. We note, first, that previous analyses have used different definitions, and that the delineation of cases under investigation of course crucially depends on the research question. In any case, it is necessary to accurately define variable contexts; failing to do so would invalidate any quantitative results. For binary alternations, an a priori useful criterion is interchangeability (Labov 1966a,b), i.e. the condition that each observed token could, in principle, have appeared in the form of the alternative variant. This criterion, however, is problematic for many reasons. First, due to the limited amount of available data, intuitions will have to be used, and these do not necessarily match up with observable behavior (see, for example, Bresnan and Nikitina 2009). Furthermore, even strong intuitions against the possibility of alternation in certain cases may result purely from the combined influence of individual factors; removing such cases would weaken further reasoning about the relative effects of these factors. Second, even if intuitions were a more reliable measure, they would be difficult to apply to diachronic data, as we do not have access to the intuitions of writers who lived, say, three centuries years ago. Cases that may have alternated then need not do so now, and vice versa. Despite these caveats, we did our best to operationalize the interchangeability condition as described below.

4.1 The genitive alternation

In defining interchangeable genitive contexts, we proceeded as follows. Restricting attention to ARCHER's British English letters and news sections, we used *'s, of, and *s (the latter only in the first two periods, when spelling without an apostrophe was common) as search strings. We then manually extracted, in a strictly semasiological fashion, all occurrences matching the following patterns:

- [full NP]'s [full NP], as in (3);

- [full NP]s [full NP], as in (4);

- [full NP]' [full NP], as in (5);

- [full NP] *of* [full NP], as in (6).

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*]_{por}'s [late Distempers]**_{pum} having lasted but two days. <1697pos2.n2b>

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

(5)   Ministers are reluctant to use emergency powers and troops to move essential fuel supplies, particularly as there are signs that **[the tanker drivers]_{por}' [dispute]**_{pum} may be near a settlement. <1979obs1.n8b>

(6)   [. . .] and upon Saturday last soon after day break we heard great shooting, which assured us, that the two Fleets were met: whereupon Solemn Prayers were ordered in several Churches, for **[the good success]**_{pum} **of [our *Navy*]**_{por} <1665int2.n2b>

Crucially, we restricted attention to genitive constructions with non-pronominal possessors or possessums. As argued by Rosenbach (2002: 30), pronominal possessors are almost categorically realized by the *s*-genitive and therefore do not constitute genuine choice contexts. We further excluded determiner possessums (e.g. *her face is as ugly as that of a dog*) (see Kreyer 2003: 170 for a discussion), constructions that are clearly fixed expressions (e.g. *the Duke of Normandy*; . . . *by the name of* . . . , *King's College*) and partitive genitive relation contexts, as in (7).

(7)   To these I have added 3 chests *of* Wine, 1 Jarr *of* Rare Oyl, and another *of* as good Anchovies. <1667finc.x2b>

The analysis is further limited to *of*-genitive constructions headed by the definite article (as in *the use of the navy*), as they are the only possible alternatives to *s*-genitives, which render the whole possessive construction definite (see e.g. Rosenbach 2002: 30 for discussion). As this definiteness constraint was already established in seventeenth century English (Altenberg, 1982: 27-28) we may safely exclude them also in our Late Modern English data. The analysis further focussed on determiner (specifying) *s*-genitive constructions as only these alternate with *of*-genitive constructions (see Rosenbach 2002: 31–32 for discussion). Measure genitives (8), which share properties of both determiner and classifying *s*-genitives (Huddleston and Pullum 2002: 470) are included, and so are other *s*-genitive constructions which are ambiguous between a classifying and a determiner interpretation, as in *a farmer's servant* (1762publ.n4b).

(8)   [. . .] for the purpose of having a day's shooting [. . .] <1822eva1.n5b>

These coding guidelines yielded a dataset consisting of $N = 3824$ interchangeable genitives.

4.2 The dative alternation

For the dative alternation, we started with a list of verbs – drawing on the list of verbs used by Bresnan et al. (2007) – that can appear with a dative object, and expanded the list as necessary given positive evidence in ARCHER. The dative verbs thus considered are the following: *give, tell, sell, pay, offer, cost, send, take, show, bring, charge, owe, loan, write, feed, mail, hand, cause, leave, wish, allow, read, deny, serve, assign, allot, lend, promise, quote, afford, award, flip, float, swap, grant, issue, extend, lease, allocate, deliver, resell, teach, assure, cede, deal, fine, guarantee, permit, accord, assess, bequeath, bet, carry, funnel, get, net, prepay, present, refuse, reimburse, repay, run, slip, submit, supply, tender, trade, vote*. Subsequently, ARCHER was searched for all occurrences of these verbs, and each occurrence was then pre-classified and had its constituent boundaries identified utilizing a custom part-of-speech annotation and parsing process with subsequent manual post-screening. In this endeavor, we excluded the following dative contexts:

– *Benefactives*. Instances where usage of a dative form is likely to be benefactive, as in (9), were excluded, on the grounds that while the double-object realizations are interchangeable, the prepositional realization utilizes the preposition *for* instead of *to*.

(9)   a.   I'll ask Bella to [**make**] [**us**] [**some tea**]. <1938mccr.d7b>
        b.   to **make** [**room**] [**for the principal knight**] <1764walp.f4b>

Some verbs can take dative as well as benefactive complements, possibly even at the same time. In general, the roles of beneficiary and recipient can be difficult to distinguish. When in doubt we tended to include occurrences.

– *Locatives*. The preposition *to* is often used as a locative marker, resulting in an arrangement isomorphic to the prepositional dative. This generates ambiguities. Consider (10):

(10)   I told him that I would [**send**] [**it**] [**to his House**] <1780wood.j4b>

On semantic grounds, it should be clear that (10) is not a dative – *his house* can hardly be conceptualized as the recipient. We removed all such cases, again including rather than excluding ambiguous examples.[2]

– *Non-canonical constituent orderings*. In some cases, verbs are used in dative constructions that do not match with the prototypical constituent orders of either the ditransitive or prepositional dative. Such cases were not included in the analysis. Consider (11) – (15):

(11)   he would [**give**] [**to his dog**] [**whatever she gave him**] <1793hitc.f4a>
    (Heavy noun phrase shift)

(12)   he asking the aforesaid Cooks for some Broth, they [**gave**] [**it**] [**him**], upon which he fell sick <1682pro2.n2b>
    (reverse double object dative)

---

2   Exclusion experiments indicate that the results reported in the subsequent sections do not change qualitatively if the verbs driving this ambiguity – *send*, *bring*, and *take* – are completely removed.

	ditransitives		prepositional		Total		corpus (words)	size
1650-1699	286	(69%)	128	(31%)	414	(100%)		180k
1700-1749	265	(69%)	121	(31%)	386	(100%)		178k
1750-1799	421	(65%)	229	(35%)	650	(100%)		359k
1800-1849	176	(61%)	111	(39%)	287	(100%)		181k
1850-1899	380	(66%)	200	(34%)	580	(100%)		358k
1900-1949	203	(70%)	88	(30%)	291	(100%)		177k
1950-1989	319	(66%)	166	(34%)	485	(100%)		357k
Total	2050	(66%)	1043	(34%)	3093	(100%)		1,789k

Table 1: Interchangeable dative frequencies (not normalized) by ARCHER period.

(13)   by some mistake [**Melville**] [**was given**] [**an old Chevrolet**] <1951marq.f8a>
(passive)

(14)   That the imperfection of Voice, as well as the difficultie of swallowing were the effects
of the paralysis, may probably be allowed, & be a satisfactory reason, why the Person
Dr. Lister mention's, could not use the Quill [**which**] [**was given**] [**him**] to suck with
<1685howm.m2b>
(object relative clause)

(15)   The pleasure of riches is to be able, to [**give**] [ ] [**to those that deserve 'em**]
<1776fran.d4b>
(ellipsis)

As for heavy noun-phrase shift and the reverse double object dative in (11) and (12), we observe *per se* regular verb-recipient-theme or verb-theme-recipient orders. However, in (11) the recipient is marked with a preposition, in contrast to the prototypical ditransitive recipient-theme ordering, while in (12) – a variant also widely available in some British English dialects (cf. Haddican, 2010) – we do not find the preposition usual in theme-recipient realizations. The factors determining this sort of alternative variability are different and beyond the scope of this paper. In (13) and (14), we find one constituent in pre-verbal position due to passivization or relativization. Even if such a construction should alternate in a given case (e.g. by means of optional *to*), that alternation would not involve word order variability in a way comparable to prototypical dative variability, and thus the factors involved are not necessarily the same. A similar reasoning licenses the exclusion of datives with ellipsized recipients or themes, as in (15).

We thus obtained a dataset spanning $N = 3093$ interchangeable datives.

5 A frequency overview

In this section, we succinctly survey dative and genitive frequencies over time. As for the dative alternation, Table 1 shows that dative proportions are fairly stable in real time. The share of
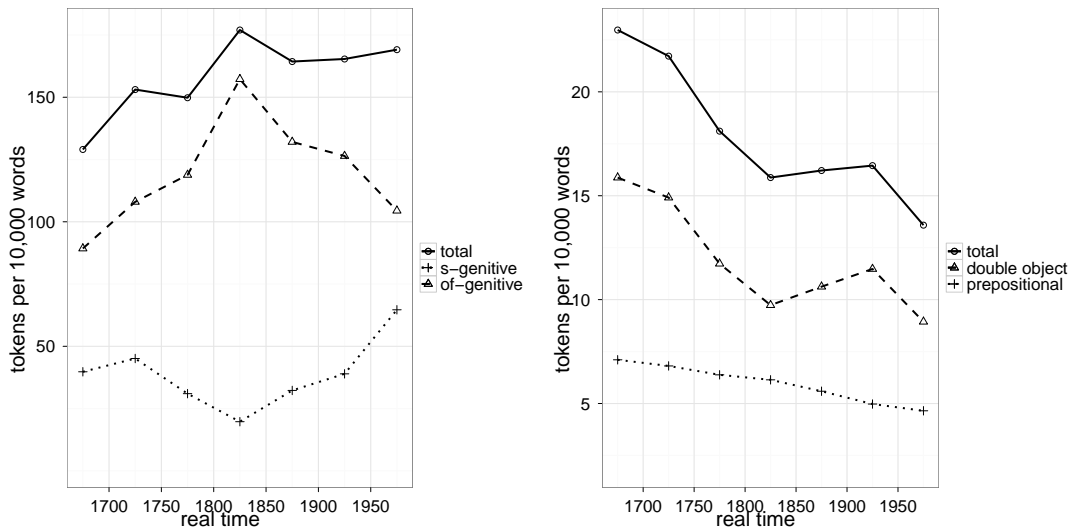
Figure 1: Mean interchangeable genitive (left) and dative (right) frequencies (normalized to frequency per 10,000 words) by ARCHER period.

ditransitive datives modestly fluctuates between 61% (1800-1849) and 70% (1900-1949). However, Figure 1 makes amply clear that the *absolute* frequency of dative constructions, ditransitive *or* prepositional, has steadily declined over time – from about 24 occurrences per ten thousand words (*pttw*) in the 1650-1649 period to about 13 occurrences *pttw* in the 1950-1989 period. In other words, datives with a theme *and* recipient argument have become a lot rarer, a development which is more marked in some ARCHER registers than in others. A fairly extreme text type is medical prose, where dative constructions have decreased from about 13 occurrences *pttw* (not a lot to start with) in the 1650-1649 period to about 1 occurrence *pttw* in the 1950-1989 period.

The facts are exactly opposite in the case of the genitive alternation, in which we observe overall stable absolute frequencies (subject to a slight upward trend) but fluctuating variant pro-

	of-*genitive*		s-*genitive*		Total		corpus (words)	size
1650-1699	312	(69%)	139	(31%)	451	(100%)		35k
1700-1749	364	(71%)	152	(29%)	516	(100%)		34k
1750-1799	418	(79%)	109	(21%)	527	(100%)		35k
1800-1849	558	(89%)	70	(11%)	628	(100%)		35k
1850-1899	446	(80%)	109	(20%)	555	(100%)		34k
1900-1949	435	(76%)	134	(24%)	569	(100%)		34k
1950-1989	357	(62%)	221	(38%)	578	(100%)		34k
Total	2890	(76%)	934	(24%)	3824	(100%)		242k

Table 2: Interchangeable genitive frequencies (not normalized) by ARCHER period.

portions. Recall that for genitives, we restricted attention to ARCHER's news and letters section, and in these text types the overall frequency of genitive constructions has remained quite stable (Figure 1). We observe a low, 129 occurrences *pttw*, in the 1650-1699 period, and a high of 177 occurrences *pttw* in the 1800-1849 period. Having said that, notice that genitive *proportions* fluctuate substantially, as is evident from Table 2. In Present-Day English (1950-1989), the share of the *s*-genitive is 38%; in the 1800-1849 period, its share amounted to no more than 11%. More specifically, the *s*-genitive started out with a share of 31% in the 1650-1699 period. *S*-genitive frequencies then started to decline in the 1750-1799 period, reaching their low point in the 1800-1849 period but recovering subsequently (see Szmrecsanyi to appear for a detailed discussion). The 1950-1989 period actually surpasses the first ARCHER period in terms of relative *s*-genitive frequencies. Note also 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 ARCHER's middle periods is not any lower than, e.g. in the starting period. We conclude that despite the phenomenal comeback of the *s*-genitive "against all odds" (Rosenbach 2002: 184) in the Early Modern English period, we do not see a further gradual linear increase in *s*-genitive frequencies in the late Modern English period.

We stress again that overall dative frequencies (ditransitive or prepositional) are fairly volatile in real time while overall genitive frequencies (*s* or *of*) are rather stable. Why is this the case? We note that English provides fewer syntactic means to manipulate NP constructions such as the genitive than VP constructions such as the dative. Specifically, genitives have two argument slots – possessor and possessum – none of which can be easily omitted except in some fairly rare contexts. By contrast, dative constructions come with one verb argument slot in addition to two NP or PP argument slots, recipient and theme. Crucially, the recipient in particular may often be omitted in the interest of economy, resulting in monotransitive usages (which we do not track in the present study because there is no positional variability). Consider thus (16-a), where the recipient of a drug (in popular diction, the patient) is made explicit in a ditransitive dative construction (*him*), and (16-b), an example exhibiting a monotransitive usage of the verb *administer* where the recipient is not made explicit.

(16) a. But his friends, anxious for relief, wished something still to be tried, although there was no prospect of success. I *gave* [**him**] therefore [**a grain of Tartar Emetic**], in one of the spoonfuls of soaked bread, with a view to ease the oppression at his breast <1775bath.m4b>

   b. Since Dr Gibb introduced this preparation of ammonium, various practitioners have **administered** [**it**], and have spoken favourably of it as a remedial agent in pertussis. <1864ritc.m6b>

We speculate, then, that the frequency decline of recipient-plus-theme dative constructions in ARCHER may be seen against the backdrop of the 'informational explosion' in modernity (Biber 2003: 180), and the concomitant drift towards more economy in written styles. This economy presumably advantages monotransitive usages of dative verbs, at the expense of ditransitive and prepositional dative constructions with an explicit recipient.

6 Predictors

6.1 Shared predictors

6.1.1 Corpus metadata

ARCHER provides the year of creation[3] of each corpus file. To ease the assessment of diachronic changes and make statistical analyses more reliable, the individual dates were centered around 1800 and converted to centuries, so that a text from 1651 would count as $^{1651-1800}/_{100} = -1.49$, and a text from 1931 as $^{1931-1800}/_{100} = 1.31$. ARCHER also yields a register classification for each text; previous studies (Bresnan et al., 2007) have obtained slight but reliable register differences, which is why we take register into account when studying the dative alternation (recall that the genitive dataset only draws on two registers, which do not make a significant difference).

6.1.2 Syntactic weight

One of the most well-known factors that influence the ordering of constituents is the principle of 'end weight' (Behaghel, 1910; Wasow, 2002), according to which 'heavier' (i.e. longer and/or more complex) constituents tend to follow 'lighter' ones. To determine the constituent weights of individual dative and genitive occurrences, the following procedure was used: First, the constituents of each token were manually identified, as described in Section 4. Then, word and character counts of each constituent were automatically obtained. Several operationalizations of weight were explored, including both the number of words or characters for each individual constituent as well as aggregated measures, such as weight differences and ratios. All of these performed well, but using the individual number of characters provided the best results. Then, a logarithmic transformation was applied to the resulting values to reduce skewness, and finally the values were centered around 50-year period means to reduce multicollinearity and account for possible changes in average lengths.

Taking (17) as an example, the number of characters in the recipient is 2 and in the theme, including the space between the two words, it is 8. After logarithmic transformation and centering, the weight scores are -0.93 and -0.36, respectively.

(17)   No credit is owing to me for taking the bath at Lourdes. Sally went along with us and she was determined that I take it and **gave** [**me**]$_{recipient}$ [**no peace**]$_{theme}$. <1958ocon.x8a>

6.1.3 Animacy

Previous studies have reported reliable and strong effects of animacy, and especially for the genitive alternation, research has explored animacy as a locus of diachronic variability (cf. Rosenbach, 2002; Hinrichs and Szmrecsanyi, 2007). For the present study, our operationalization of animacy was based on a simplified version of the guidelines in Zaenen et al. (2004). Five animacy categories were distinguished: animate, collective, inanimate, locative, and temporal nouns (see Rosenbach 2008 for a similar categorization). We coded the possessor in the genitive data set and

---

3   A small number of texts in ARCHER is not dated exactly; these were placed in the middle of a time segment, i.e. in year five of a given decade or year 25 of a 50-year period, as the case may be.

both recipient and theme in the dative data set. Because of data sparsity, the dative themes and recipients were subsequently reduced, first by collapsing the less frequent locative and temporal nouns with the inanimate category for a tripartite scheme, used for the analysis of recipients. For dative themes, we contrast animate with all other themes, a binary scheme which is identical to the one used in Bresnan et al. (2007).

Let us now discuss the animacy categories in more detail. Animate possessors comprise humans, higher animals and sentient human-like beings such as gods, e.g. *king*, *horse*, *god*, or *John*, as in (18-a). Collective possessors, as in (18-b) are organizations such as *administration* or *church*, as well as temporally stable groups of humans with potentially variable concord, such as *delegation*, *family* or *enemy*. Temporal nouns consist of both points in time and durations, for example *February* or *moment*, as in (18-c). Locatives are locations, including geographical states, e.g. *Russia*, *this kingdom*, *the seas*, as in (18-d). All other concrete or non-concrete noun phrases were classified as inanimate (18-e).

(18)   Animacy categories

　　　a.　*animate*: [**BISHOP ABEL MUZOREWA'S**]$_{animate}$ **personal security squad** has been enlarged to 30 men following the discovery by police that the new Black Prime Minister of Zimbabwe – Rhodesia was among the names on an assassination list. <1979stm2.n8b>

　　　b.　*collective:* . . . and **the Gentlemen of** [**the Academy of Sciences**]$_{collective}$ have appointed Messieurs Cartigny, Saurin, Meyvaud, and another, to examine into the Structure of those Machines. <1723dai2.n3b>

　　　c.　*temporal*: After [**yesterday's**]$_{temporal}$ **outbreaks** police toured farms within ten miles of the stricken area warning farmers not to move their cattle. <1967stm1.n8b>

　　　d.　*locative*: **The inhabitants of** [**this island**]$_{locative}$ were reported to be very ferocious, and no wonder. <1872gla1.n6b>

　　　e.　*inanimate*: People of a low, obscure education cannot stand **the rays of** [**greatness**]$_{inanimate}$; they are frightened out of their wits when kings and great men speak to them; . . . <1748ches.x3b>

6.1.4 Definiteness and nominal expression

Our definiteness annotation comprises four levels: indefinite, definite, proper name and (definite) pronoun. For datives, both recipient and theme were annotated according to the full scheme. For genitives only the possessor was considered and only a reduced scheme was applied, as tokens involving pronominal phrases or indefinite possessums were removed at the outset (see Section 4.1). The pronoun category consists of all definite pronouns, such as the recipient in (19).

(19)   What agreement was made, I know not; but at his return Bavia **gave** [**him**]$_{pronoun}$ [**the promised jewel**]$_{definite}$ and was put on board the other ship, which brought her to Jamaica. <1720pitt.f3b>

The category 'proper name' includes prototypical proper nouns, as in (20) (*Wilhelm*), but also titles such as *the king of England* and names of institutions such as *the Medical Society*.

(20)   Rappaport did not **offer** [***Wilhelm***]$_{proper\ name}$ [**a cigar**]$_{indefinite}$, but, holding one up, he asked, "What do you say at the size of these, huh? They're Churchill-type cigars." <1951bell.f8a>

We classified as proper name only noun phrases that could be considered proper names in Present-Day English and that were capitalized in the text, excluding clear common noun uses. As definite we coded all noun phrases headed by a definite determiner, *that* or an *s*-genitive, such as the theme in sentence (19). All other constituents, and especially those headed by an indefinite determiner such as the theme in sentence (20), were classified as indefinite.

6.2 Genitives only

6.2.1 Final sibilancy

The literature suggests a clear, presumably phonologically motivated preference for using the *of*-genitive with possessors ending in a sibilant, an effect that is reliable across a multitude of corpora covering both spoken and written language (for example, Grafmiller submitted; Szmrec-sanyi 2006; Szmrecsanyi and Hinrichs 2008). We used an automatic annotation process, relying on the Carnegie Mellon University Pronouncing Dictionary version 0.7a[4] for transcription and coding all possessor phrases ending in [s], [z], [ʃ], [ʒ], [tʃ] or [dʒ] as ending in a sibilant. Tokens not included in the dictionary were coded manually. (21) exemplifies one of the rare occurrences of an *s*-genitive with a possessor ending in a final sibilant.

(21)   [**Alice**]$_{+\text{final sibilant}}$ **'s** [**child**] is to be called Victoria Alberta Elisabeth Matilde Marie, and will be called Victoria – the first of our grand-children that will be called after either of us. <1863qvic.x6b>

6.2.2 Semantic relation

Genitives may encode a wide range of different relations, which are notoriously difficult to classify. For the purpose of the present paper we follow the binary distinction between prototyp-ical and non-prototypical possessive relations as adopted in Rosenbach's (2002) study, which in turn is based on Koptjevskaja-Tamm's (2001; 2002) typological classification of possessives in European languages. Prototypical relations comprise legal ownership (22-a), body parts (22-b), kinship (22-c), and part-whole (22-d) while all remaining cases were coded as 'non-prototypical' (see the examples in (23)). Prototypical relations have been shown to favor the *s*-genitive, non-prototypical relations show a preference for the *of*-genitive. Valence relations, such as in sub-jective (23-b) and objective (23-c) genitives, strictly speaking fall outside this taxonomy but were included in this study and subsumed under 'non-prototypical possessive relations' (see e.g. Rosenbach and Vezzosi 2000; Seiler 1982 for discussion).

(22)   Semantic relations considered prototypical
    a.   *ownership*: RHODESIAN forces have increased security measures in and around [**Mr Ian Smith's cattle ranch and farm at Selukwe**]$_{+\text{prototypical}}$ after a sharp upsurge of guerrilla activity in the Midlands region of the country. <1979stm1.n8b>
    b.   *body parts*: The Irish came in to the house pul'd the man out of bed from his wife and murdered him; then took all the rest of the houshold, led them to the seaside, and threw them off the rocks; one of the Children hung about one of [**the murderers legs**]$_{+\text{prototypical}}$, yet was pull'd off and thrown after the rest. <1653merc.n2b>

---

4   Available online at `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`.

  c. *kinship*: It's said [**the Duke of Berwick's Son**]$_{+prototypical}$ is in one of the Ships, and Perth's two Sons in the other. <1715eve1.n3b>

  d. *part-whole*: [**The Hull of a Ship**]$_{+prototypical}$ was seen floating between Blackness and Point and Calais and Ambeleteuse; <1735rea1.n3b>

(23) Semantic relations considered non-prototypical

  a. Christian sources in Egypt say that President Sadat has gone back on a pledge he gave some years ago not to allow Islamic law to become [**the law of the country**]$_{-prototypical}$. <1979stm1.n8b>

  b. THE new drama, 'John Garth,' produced at Wallack's Theatre, New York, is spoken of by the press as the best work ever written for the American stage. [**Mr. Wallack's acting as the hero**]$_{-prototypical}$ is greatly admired. <1872gla1n6b>

  c. However, this rule is sometimes dispensed with; and particularly since the signing of the Preliminaries of Peace, our Government has permitted [**the granting of such passports**]$_{-prototypical}$, provisionally, for the space of a year, to ships built out of the Republic, provided that they entirely belong to natives of this country, and also fitted out here. <1802joh2.n5b>

  d. The Supreme Educational Council had given instructions to the school-masters which had established religious neutrality, and a request by the Council-General of the Seine that [**the name of God**]$_{-prototypical}$ should never be uttered in school had been rejected. <1883tim2.n6b>

6.3 Lexical effects

It is well-known that dative verbs differ in their likelihood to be used in either construction (for example, Gries and Stefanowitsch 2004). All dative tokens were thus coded for their verb lemma. Genitives have no carrier for lexical effects that is as clear; we decided to use the lemma of the possessor head noun to test for by-item effects. This, however, leads to the difficulty of having too many types with only one observation. To simplify the analysis, we collapsed all nouns that did not reach a threshold of at least four observations. The same procedure was then also applied to the last word of the theme in datives, to control for idiomatic preferences of frequently occurring themes.

7 Regression analysis

Logistic regression is a statistical analysis technique related to VARBRUL analysis customary in variationist sociolinguistics (Sankoff and Labov, 1979). The technique permits quantification of the effect of individual explanatory factors on a binary dependent variable, such as dative or genitive outcomes. We are utilizing a modern refinement of logistic regression analysis known as *mixed-effects logistic regression* (Pinheiro and Bates, 2000).[5] In addition to so-called *fixed effects* – which are classically estimated predictors suited for assessing the reliability of the effect of repeatable characteristics – mixed-effects modeling allows for *random effects* that are well suited to capture variation dependent on open-ended, potentially hierarchical and unbalanced groups. For example, consider idiolectal variation and author idiosyncrasies, operationalized here by means of corpus file ID (we adopt the reasonable assumption that each corpus file has

---

5 We utilized R version 2.12 (R Development Core Team 2010) and lme4 version 0.999375-33.

a different author). Now, it is certainly possible that individual authors differ in their genitive or dative preferences. However, traditional estimation of these idiosyncrasies via fixed effects is not viable.[6] Yet as the individual observations are not statistically independent (as assumed by the bare logistic regression procedure), it would not be advisable to leave this information out of the model. Furthermore, the issue whether the behavior of, say, a given author is statistically significantly different from another author – the question ultimately answered by fixed effect modeling – is not relevant for present purposes. Random effects, then, provide a sophisticated yet elegant method for taking such variation into account, making sure that the estimation of the interesting variables can proceed unaffected by this noise and that the results are easy to generalize.

The following procedure was used for model fitting: First, we constructed models containing all predictors and all putatively relevant interactions. These models were then reduced by removing predictors and interactions that did not have reliable effects, and the new models were compared to the fuller ones by means of the Akaike Information Criterion. Random effects were evaluated by means of likelihood ratio tests. Finally, the models underwent bootstrap validation to assess the possibility of overfitting. More precisely, the individual observations were repeatedly randomly resampled with replacement and the model was fit to this new data set. To ensure that each fifty year period has a sufficient number of observations in each run, the total number of observations per period was kept constant. All results reported as significant below are also stable under bootstrap validation.

7.1 The genitive alternation

Table 3 reports fixed effects in the genitive model; the predicted odds are for the *s*-genitive. The classification accuracy is excellent – the model achieves a Somers' D_{xy} value of 0.93 and correctly predicts 92.1% of all genitive tokens, a considerable increase over baseline (75.6%) consistently predicting the overall most frequent token (in the present case, the *of*-genitive). Multicollinearity is not an issue, as the model's condition number ($\kappa = 6.7$) is well below the customary threshold of 15 that indicates medium collinearity.

To make the interpretation of this table more accessible, let us walk through some of the entries in the table. Consider definiteness of the possessor: The default level of this factor is 'definite'; given two contexts identical but for their definiteness classification, one being definite and the other a proper name, the model estimates a so-called 'odds ratio' of $exp(1.57) = 4.81$.[7] In other words, vis-à-vis a definite noun, a proper name is 4.81 times more likely to appear in the *s*-genitive. The standard error ('SE') column indicates how confident we can be in these values: with 95% certainty the true coefficient will lie within the range of the reported coefficient plus/minus twice the SE. If that range does not include zero, the coefficient is statistically significant. The column labeled '*p*' indicates the customary significance thresholds reached by the individual predictors. In addition to such main effects, logistic models can specify interac-

6   First, the number of texts is quite large, and the distribution of observations across texts and lemmas is skewed. Together with low token numbers for many of these this leads to severe technical problems, such as non-identifiability or overfitting, for classical estimation.

7   The present study typically reports fixed effect sizes as logarithmically transformed odds ratios (column 'Coefficient' in Tables 3 and 5).

	Coefficient	SE	p
Intercept	−.89	.24	***
Possessum length (in 1800)	.30	.13	*
Possessum length, squared	.68	.14	***
Possessor length	−2.66	.21	***
Possessor length, squared	−1.05	.22	***
Possessor is proper noun	1.57	.17	***
Possessor is indefinite	−.32	.22	
Animacy of possessor: collective	−2.51	.32	***
Animacy of possessor: inanimate	−3.90	.35	***
Animacy of possessor: locative	−3.39	.41	***
Animacy of possessor: temporal	−2.22	.37	***
Centuries since 1800 (Time)	−.01	.12	
Semantic relation is prototypical	.80	.16	***
Possessor has final sibilant	−.77	.18	***
Animacy of possessor: collective (I: Time)	.54	.23	*
Animacy of possessor: inanimate (I: Time)	−.12	.31	
Animacy of possessor: locative (I: Time)	.79	.30	**
Animacy of possessor: temporal (I: Time)	.83	.26	**
Possessum length (I: Time)	.28	.11	*

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

Table 3: Fixed effects in the minimal adequate mixed-effects logistic regression model for genitive variation in ARCHER. "I" indicates interactions. Predicted odds are for the *s*-genitive.

15

	N$_{\text{groups}}$	Variance	StdDev
Text	242	1.00	1.00
Possessor head lemma	188	1.79	1.34

Table 4: Random effects in the genitive model.

tion terms which allow for changes in the effects of predictors depending on the values of other predictors. Such interactions can be used, for instance, to test for diachronic changes. An example can be found in Table 3 in the segment 'Animacy of Possessor'. The non-interaction (i.e. main effect) coefficients compare animate possessors to the other types for the year 1800, with all other types reliably less likely to occur with the *s*-genitive (as shown by the negative sign of the coefficient). There is no reliable real-time change independent of animacy, as evidenced by the small, non-significant coefficient for 'Year'. However, there are significant interaction effects between real time and collective, locative, and temporal possessors. Notice that all interaction coefficients are positive, indicating that in real time, these three types become increasingly likely to occur with the *s*-genitive. To quantify the size of this change, the relevant coefficients are simply multiplied with their numeric values and summed, so that in 1800 and all other things being equal, a collective possessor is $exp(-2.51 + (-0.01 * 0) + (0.54 * 0)) = 0.08$ times more likely to appear with the *s*-genitive than an animate possessor, while in 1950 it would be $exp(-2.51 + (-0.01 * 1.5) + (0.54 * 1.5)) = 0.18$ times as likely. In addition to this interaction, the effect of another predictor turns out to vary significantly as a function of real time: The linear component of the quadratic effect of possessum length is gradually reduced as time progresses. We will return to the issue of these real-time changes in Section 8.

Let us now consider the random effects. Table 4 shows the variance of each random effect. Both corpus text ID and possessor head noun show very comparable amounts of variation. It is also possible to identify individual groups particularly attracted to one of the realizations. For example, *people*, *parliament* and *lord* appear more often than expected with the *of*-genitive, while *company*, *enemy*, and *China* tend to prefer the *s*-genitive. Concerning individual corpus files (and thus, by inference, author idiosyncrasies), we find that texts 1819mor1.n5b and 1819mor2.n5b, both from the 1810 *Morning Chronicle*, favor most strongly the *of*-genitive. Texts 1979obs1.n8b and 1979obs2.n8b, both from the 1979 *Observer*, attract the *s*-genitive most robustly.

7.2 The dative alternation

Table 5 details fixed effects in the dative model; the predicted odds are for the prepositional dative. The classification accuracy surpasses that of the genitive model, achieving a Somers' D$_{xy}$ value of 0.97 and correctly predicting 94.1% of all dative outcomes (baseline: 66.1%). Again, multicollinearity is not a problem ($\kappa = 8.6$).

Due to the comparatively low number of tokens in the non-animate categories, we were unable to confirm a linear interaction effect between real time and the effect of animacy on dative choice. Adding an additional indicator for dative tokens from the twentieth century, however,

16

	Coefficient	SE	*p*
Intercept	-1.10	.49	*
Animacy of theme: animate	1.75	.49	***
Animacy of recipient: collective	1.38	.52	**
Animacy of recipient: inanimate	2.50	.32	***
Text is from the twentieth century (indicator)	.32	.34	
Theme length	-1.28	.17	***
American English	$-.27$	.22	
Centuries since 1800 (Time)	$-.55$	.22	*
Recipient length	1.58	.21	***
Definiteness of recipient: definite	1.77	.38	***
Definiteness of recipient: proper name	1.78	.37	***
Definiteness of recipient: indefinite	3.78	.47	***
Definiteness of theme: definite	1.07	.66	
Definiteness of theme: indefinite	-1.21	.21	***
Definiteness of theme: pronoun	1.81	.59	**
Recipient: collective (I: twentieth century)	$-.44$	.96	
Recipient: inanimate (I: twentieth century)	-1.35	.53	*
Theme length (I: American English)	$-.68$	.28	*
Recipient length (I: Time)	$-.36$	.17	*
Recipient: definite (I: Time)	.78	.34	*
Recipient: proper noun (I: Time)	.62	.32	.
Recipient: indefinite (I:Time)	$-.14$	.41	

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

Table 5: Fixed effects in the minimal adequate mixed-effects logistic regression model for dative variation in ARCHER. "I" indicates interactions. Predicted odds are for the prepositional dative.

17

	N$_{\text{groups}}$	Variance	Std.Dev.
Text	741	0.35	0.59
Theme	171	1.27	1.13
Verb lemma	49	4.86	2.20
Register	8	0.16	0.40

Table 6: Random effects in the dative model.

	adjustment
Fiction	−0.33
Drama	−0.31
Journal	−0.25
Letters	−0.04
Medicine	0.12
Science	0.22
Sermons	0.22
News	0.45

Table 7: Intercept adjustments for random effect 'register' in the dative model.

yields a robust real-time change: In comparison to the previous periods, the disfavoring effect of inanimate recipients towards the double-object dative is less pronounced after 1900. Moving on to the linear changes in real time, we find that longer recipients are less strongly attracted to the prepositional dative as time progresses, while the differences between pronominal and definite or proper name recipients are becoming more pronounced. Lastly, we find a significant difference between American and British English in that theme length has a stronger effect in American English. This is another way of saying that while the probability of realization as double object dative for longer themes is greater than for shorter themes in both varieties, this difference is more pronounced in American English.

Table 6 shows the variances of the random effects in the dative model. The individual groups show much more variability than in the genitive model, with verb lemma accounting for a large amount of variation and register only for a rather small amount; the effects of theme and text lie between those two extremes. As for verb lemmas, we find that *cost*, *tell*, and *allow* are strongly attracted to the double object dative, while *present*, *extend*, and *take* show the opposite pattern. Table 7 lists intercept adjustments for all registers. More oral registers tend to favor the double object dative, while more literate genres tend to use more prepositional datives, a result that matches with the difference between spoken and written materials reported in Bresnan et al. (2007).

8 Discussion

We now turn to the discussion and interpretation of the regression models reported in the previous section. Right at the outset, we emphasize that as main effects, the language-internal predictors considered in the present study generally behave as advertised in the literature (subject to the error margins inevitable in statistical analysis and differences due to slight operational differences). In other words, there are no surprises concerning how factors such as syntactic weight, animacy, definiteness, and nominal expression bear on genitive and dative outcomes. By this token, our study diagnoses a good deal of *probabilistic stability*. We thus focus in the remainder of this section on the interactions between language-internal variables (animacy and

syntactic weight) and language-external variables (real time and variety). We specifically rely on statistical significance of interaction terms as identified in regression modeling as a criterion to diagnose genuine interrelationships, and we subsequently explore the exact nature of these interrelationships, drawing on univariate visualization techniques. While such techniques cannot account for the influence of other explanatory variables included in regression analysis, univariate plots (unlike e.g. partial effects plots) straightforwardly visualize the distribution of actual corpus attestations while yielding a high resolution.

8.1 Interactions involving syntactic weight

Our regression models have uncovered a set of interactions between syntactic weight and real time, and between syntactic weight and variety (British vs. American). For one thing, in the genitive model, we find fairly complex, nonlinear relationships of constituent lengths and genitive choice. In short, end weight does not work as it should for very short constituents – for example, the shortest possessums in our dataset are actually less likely to appear in the *of*-genitive than slightly longer possessums. Only after a certain minimum threshold of about eight to twelve characters do we observe the expected pattern of *s*-genitive probability increasing with possessum length. As an additional twist, this non-linearity is subject to diachronic change, in that the linear component becomes steeper and thus gains influence over time. In plain English, then, weight is more well-behaved in later ARCHER periods. Figure 2 is an attempt to come to terms with this complexity. The Figure plots the distribution of actually observed genitive realizations (*y*-axis) against possessum lengths (*x*-axis), dividing the dataset into two halves: an early one containing genitive observations before 1820, and a later one containing all observations after 1820. Due to the continuous nature of log lengths, we next segment the total range of lengths into fifty bins. The non-linearity discussed above is clearly visible in both non-parametric regression curves, but we observe that the non-linearity is more pronounced for the early genitive tokens. We note that this non-linearity is not documented in the literature, and we presume that it may be rooted in the fact that we modeled possessum and possessor weight separately, a modeling decision which – although justified in terms of model goodness-of-fit measures – may not do full justice to the possibly very complex interplay between relative and absolute weights. Notice also that the effect of possessum weight in particular has proven more delicate to capture than other weight phenomena in previous research (see, for example, Szmrecsanyi 2010), a fact that additionally suggests that there are aspects to the data that current regression modeling approaches have trouble with. We finally wish to emphasize that while the existence of weight effects is well-known, the jury is still out on best operationalization (cf. Grafmiller and Shih 2011). We therefore must defer this issue to future work.

In the dative model, we also find an interaction between syntactic weight and real time, here involving the dative recipient. The direction of the interaction is opposite to that of the main effect, indicating a weakening of the relative effect of syntactic weight over time.

Finally, we observe an interaction between theme length and variety type in the dative model, such that in American English increasing theme length decreases the probability of a prepositional dative more robustly than it does in British English. Figure 3 plots observed dative realizations against theme length in fifty bins per variety. The curves are indistinguishable for short themes, but beginning at theme lengths of about 15 characters they increasingly diverge.
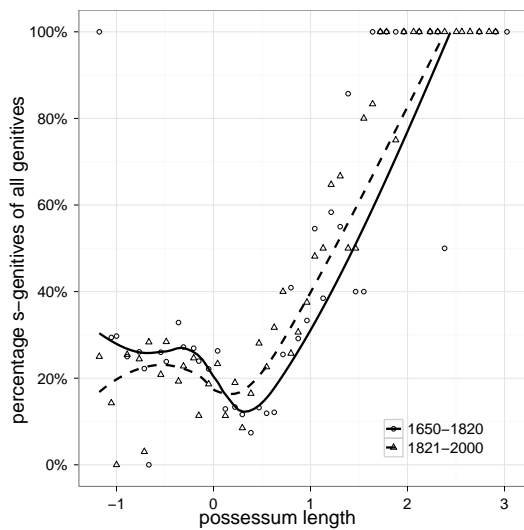
Figure 2: *S*-genitive rates (*y*-axis) as a function of possessum length (*x*-axis; binned period-centered *log*-transformed possessum length) and real time slice (heavy smoother: 1650–1820; dotted smoother: 1821–2000).
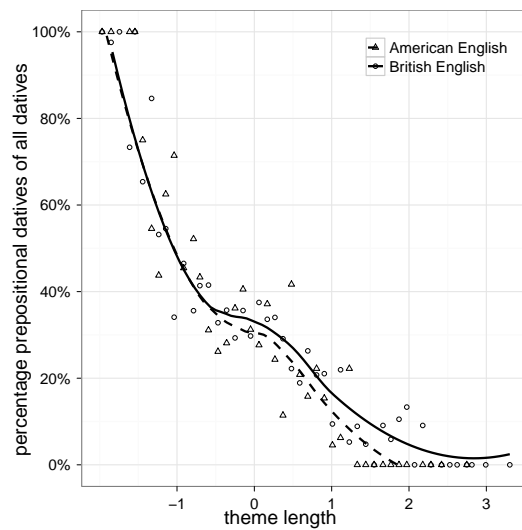
Figure 3: Prepositional dative rates (*y*-axis) as a function of theme length (*x*-axis; binned period-centered *log*-transformed theme length) and variety (heavy smoother: British English; dotted smoother: American English).

This finding matches up well with the psycholinguistic experiments reported in Bresnan and Ford (2010), who – starting at theme lengths of four words – found longer reaction times for long themes in double object constructions for American subjects as compared to Australian participants in a continuous lexical decision task.

In all, the cumulative weight of cross-constructional evidence suggests that syntactic weight, despite its putative roots in the human speech processing system (Hawkins 1994), is not a stable factor, synchronically or diachronically. In other words, while the findings presented here do show that end weight is a generally good predictor that works in the expected direction (in line with what processing considerations would lead one to expect), the factor nonetheless appears to be remarkably plastic and subject to modulation by individual speech communities.

8.2 Interactions involving animacy

In both the genitive and the dative model, the effect that (some) animacy categories have on syntactic choices interacts significantly with real time. In the genitive model, the *s*-genitive becomes less strongly disfavored with collective, locative and temporal possessors over time. The dative model suggests that inanimate recipients are coded significantly more often with the double object dative during the twentieth century than in earlier periods.

Figure 4 displays observed proportions of dative and genitive realizations per 50-year period and animacy category. From these plots, it is clear that the changes that happened are not as linear as our regression might seem to suggest. Going through the categories in turn, we find that
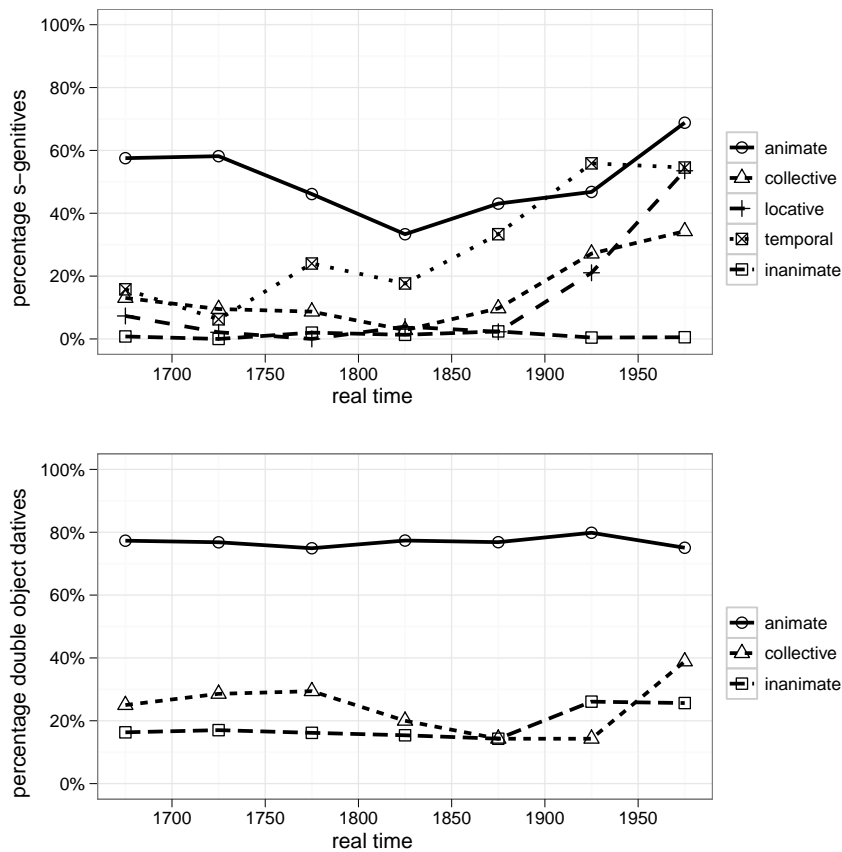
Figure 4: Animacy and real time. *S*-genitive rates (upper plot) and ditransitive dative rates (lower plot), on *y*-axis, as a function of ARCHER time slice (*x*-axis) and several animacy categories.

while inanimate possessors are stable in their dispreference for the *s*-genitive (as indicated by regression analysis), animate possessors actually show a more V-shaped pattern: these halve their proportion of *s*-genitives between 1750 and 1850, and then regain the lost *s*-genitive proportion continuously in the following 150 years.[8] Temporal possessors show a rather consistent upward trend. Locative and collective possessors exhibit stability for 250 years; from 1900 onwards, though, they exhibit a marked increase and subsequent growth in *s*-genitive rates. In the case of the dative alternation, we find long-term stability in proportions for animate and inanimate recipients, with collectives slowly but steadily appearing relatively more often in prepositional datives. Then all of a sudden, again at around 1900, inanimate recipients start to appear more often as double-object datives, and subsequently collectives appear to reverse their long-term trend.

---

8   A large part of this change, however, is due to the frequency decline of the *s*-genitive in the period between1800 and 1850, a period that behaves oddly as well according to other measures, such as raw frequencies.

8.3 On animacy

Why are we seeing these real-time, cross-constructionally parallel fluctuations in the effect of animacy? In what follows we consider several distinct explanations that could each account for the patterns we have seen. Let us first explain the assumptions on which these explanations are based. We will state these as sets of preferences that we consider to be probabilistic, i.e. quantified by a (variable) parameter determining how likely the satisfaction of the preference is. First, we shall assume that both the ditransitive dative construction and the *s*-genitive are preferred over the prepositional dative and the *of*-genitive with probability p_1 for encoding a special meaning of *(potential) possession*. Second, we assume that *(potential) possession* strongly prefers at least one *animate* entity in the role of possessor, quantified by probability p_2, and that the *animate* property is preferably applied to humans and higher animals, quantified by probability $p_{2'}$. Third, we assume that entities are distributed throughout discourse in such a way that the entity-specific value of $p_{2'}$[9] is preferred to be greater than a certain fixed value (which will be rather large), quantified by parameter p_3; in other words, most things are reliably classifiable as animate or not. Let us now utilize these assumptions to present the different explanations.

1. *Grammar is changing*

   This explanation makes an argument that genitive and dative grammars have come to feature fewer selection restrictions. Couching this in grammaticalization terms, we would say that the semantics of the constructions involved are subject to bleaching and that their host class expands. Going back to our assumptions, this could clearly be formalized as a decrease of p_1 – in other words, there is an increase in the relative number of cases where a double object dative or an *s*-genitive does not encode a meaning that comes within the remit of *(potential) possession* in a strict sense.

2. *Semantics are changing*

   There are two distinct, but related explanations involving semantic changes.

   a) *Relational semantics*:

   This explanation holds that grammar (i. e. p_1) is stable, but that the relational semantic input feeding into the grammar is subject to diachronic change. More specifically, the relation of possession is hypothesized to go against the preference for *animacy* more easily, i.e. p_2 decreases. The intuitive interpretation for this would be that possession, while traditionally restricted to human beings, is something that can now also be associated with other entities, for example corporations. Nobel Prize laureate and professor of Economics and International Affairs Paul Krugman recently pointed out[10] how 'transformational technologies', particularly railroad transportation, enabled the rise of abstract forms of legal ownership at the expense of concrete ownership during the nineteenth century. Once such forms of possession become common enough, it would not be surprising to see an extension of general possession to classes of possessors that are not animate.

---

9 For example, a ship or computer may more often be associated with animate properties than, say, a rock; hence $p_{2'}(ship)$ and $p_{2'}(computer)$ will be smaller than $p_{2'}(rock)$.

10 For the blog post, see `http://krugman.blogs.nytimes.com/2011/01/30/transformational-technologies/`.

b) *Ontological semantics*:

Alternatively, instead of the semantics of the relation changing, the ontological semantics of the world may change, leading to the same local result. To stay within the example, a corporation possessing an asset could be licensed by an extension of *(potential) possession* to entities not considered *animate*, or by extending *animacy* to corporations – in terms of our assumptions, a decrease of $p_{2'}$. This does not necessarily mean that speakers become confused about the 'true' animacy of the entities they encounter. To give an example, vehicles such as ships are clearly inanimate, but can appear in several relations normally restricted to animate contexts (consider reference by the personal pronoun *she*, which is attested in Early Modern English already). Observe here that several of the *s*-genitive usages with pure inanimates in our data involve ships as possessors.

3. *Environmental context is changing*

The ontological semantics explanation above leads us directly to the final explanation we will consider. Let us assume that $p_{2'}$ stays constant while different (classes of) entities – such as ships – have separate values for $p_{2'}$. If the distribution of entities in the environment changes such that those that have lower $p_{2'}$ values become more frequent (i.e. p_3 decreases), it will appear that any categories between which these entities oscillate become less distinct, even without any change in the linguistic system itself. As an example, consider a city council. This entity would clearly be classified as a collective, yet a speaker may have interacted with a good number of its members, and may well construe it as more animate – and thus as a better (potential) possessor – than she would construe a large, somewhat faceless collective such as, say, an army. If discourse contained a large proportion of *city council* type recipients or possessors, collectives would show a pattern closer to clear animates, and if *army* type recipients or possessors were frequent in discourse collectives would rather tend towards clear inanimates. While one could combat this issue by making the classification scheme more detailed, doing so increases the chance of choosing the wrong class, as we cannot directly access the representations used by speakers. We would thus just move the problem.

How do these explanations fare on the empirical results presented here? We first note that a pure version of explanation 1 does not satisfactorily explain the cross-constructional parallelisms we observed, in contrast to explanations 2a/b and 3. However, explanation 1 does seem most consistent with the observed increase in temporal *s*-genitives, a change that is difficult to reconcile with strong *possessorship* constraints. Explanation 3 predicts that clearly inanimate cases should not change, which is supported by our data set; in addition, many of the *s*-genitive with inanimate possessors in the data (such as *ships* and animacy-related notions like *life* and *soul*) are good examples for oscillation between categories. This may be a limitation of the size and composition of the current data set though – Rosenbach (2003) found a clear age-grading effect in an experimental study, such that younger speakers rated *s*-genitives with clearly inanimate possessors more acceptable than older speakers did, which would argue against explanation 3. Moving on, the semantic explanations, and especially explanation 2b, would lead one to predict similar changes happening to other constructions. One study finding a qualitatively similar result

is Hundt (2004), who presents evidence from ARCHER for a real-time spread of inanimate subjects in the progressive construction (as in (24)), which had previously been limited to animate subjects.

(24)   a.   *I was just leaving* these Lodgings <1737anon.f3b> (Hundt 2004: 51) (human subject)

         b.   I had never given up my opinion that *an abscess was gathering* <1868bowd.m6a> (Hundt 2004: 62) (non-human subject)

The animacy change in progressives that Hundt (2004) diagnoses appears to have started about a century earlier than in genitives and datives, however.

All together, there is circumstantial evidence for and against each of the three explanations we have offered. While we have presented them as analytically distinct beasts, they are not by necessity mutually exclusive: the observed pattern could well have resulted from any combination of underlying changes. From the viewpoint of probabilistic, experience-based grammar it is not implausible that the underlying causes go 'hand-in-hand', with individual changes enabling and facilitating others. For example, let us assume, as in explanation three, an increase in the discourse frequency of entities oscillating between categories (i.e. a decrease in p_3). As a result, the individual categories will, for existing speakers, seem less distinct even without any actual change in the grammar or semantics. Now consider new speakers entering the speech community. The input they are exposed to contains more conflicts than that of past speakers, and thus their hypotheses about the association between grammar, possession and animacy are likely to be less strong, which would formalize to decreases in p_1, p_2 and/or $p_{2'}$. This would then lead to even more actual usage violating the original constraint, feeding back into the process.

Several of the explanations considered in the previous discussion presume an underlying change in the discourse environment. Being able to observe such changes in ARCHER would strengthen our case considerably. So, for the sake of describing the population of nominal animacy categories as a function of real time, we created two *general* noun samples (which are not limited to genitive or dative NPs), one each for ARCHER's British letters and news sections and each sampling approximately 5,000 random nouns spread out evenly over ARCHER's time periods. We next coded the nouns in these samples for animacy according to the guidelines in Zaenen et al. (2004), subsequently collapsing categories as necessary to match those described in Section 6.1.3. The area plots in Figure 5 depict the distribution of animacy categories in real time. ARCHER's letter section (right diagram) is fairly stable over time, and we will thus concentrate on the news section (left diagram) in what follows. Observe first that there is no straightforward relationship between the distribution depicted and the frequency of genitive and dative outcomes in the data: place nouns become less frequent, time nouns stay rather constant, and collective nouns become more frequent – yet all three categories have become more likely, as we have seen, to appear e.g. as possessors of *s*-genitives. That said, the increase in the frequency of collective nouns, which started during the 1850-1899 period, is consistent with our conjecture about environmental and cultural changes due to industrialization and the transformational technologies that have accompanied it. We finally note that in the 1800-1849 period, the frequency of animate nouns in both news and letters decreases, coinciding with – and partially accounting for – the substantial drop in *s*-genitive frequencies at that time (cf. Figures 1 and 4).
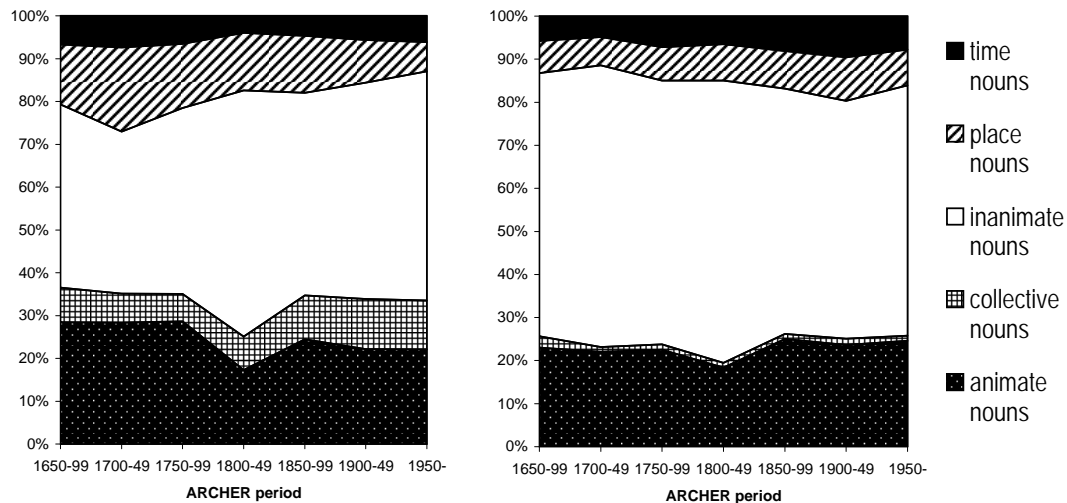
Figure 5: General distribution of animacy categories in ARCHER's British news (left) and letter sections (right) per 50-year period, based on random noun samples.

9 Concluding remarks

The present paper was concerned with the development of two sites of syntactic variation, the dative and genitive alternation, which share a number of explanatory factors constraining the choice between variant constructions. In covering the period between 1650 and 1990, the study fills a gap in the literature on the history of the genitive and the dative alternation, i.e. the late Modern English period, whose grammar has been notoriously understudied until recently (see e.g. Hundt and Lenker 2006 for a discussion). We have shown that while the explanatory factors included in our analysis (such as end weight or animacy) all have the expected effect directions, the specific strengths of these effects are historically variable. Most notably, we saw that the animacy effect has weakened over time, a development that started in the course of the nineteenth century and affected the genitive and dative alternation at roughly the same time. In particular, we saw an increasing frequency of *s*-genitives and ditransitive dative construction with collective nouns (which waver between an animate and an inanimate interpretation), locative nouns and temporal nouns, rather than a sweeping extension to just any inanimate possessors or recipients.

The very fact that the observed weakening of the animacy constraint is not construction-specific suggests that this development is to be seen against the backdrop of general changes going on at the time, be it environmental (read: cultural) changes, variation in the conceptualization of animacy categories, or drifts in 'the grammar' (by which we mean simultaneous shifts in the semantics and/or syntax of the constructions). The timing of this change, which started during the heyday of the Industrial Revolution, strongly suggests that environmental changes are at least partly responsible for the extension of the *s*-genitive and ditransitive dative construction to non-animate nouns. The rise of organizations and companies naturally also triggered more frequent reference to such entities by collective nouns, which in turn boosted the frequency of *s*-genitives with collective referents. We suggested that this development went hand in hand with

a change in the conceptualization of animacy and possibly also of possession, which bestowed 'animate' attributes on organizations and (some) inanimate entities, licensing typical possessors to be less animate and thus increasing popularity of the *s*-genitive with such referents. It is important to keep in mind, then, that semantic categories such as animacy and possession are not invariant, timeless categories but essentially subject to conceptualization, as for example evidenced by considerable cross-linguistic variability and neurolinguistic experiments (see e.g. the discussion of animacy in Rosenbach 2008: Section 3 or the typological literature on possession, e.g. Seiler 1982). Accordingly, semantic categories such as animacy or possession must be assumed to have the potential to change over time. Changes in grammar, such as the shift of possessive '*s* from a purely inflectional marker to an element in the newly evolving definiteness system (which happened presumably in the late Middle English period, cf. Rosenbach 2004: 83–85) also contributed to the extension of the *s*-genitive to locative and temporal nouns, which are known to be good referential anchors, a property important for prenominal definite determiners (see Rosenbach 2008: Section 3 for discussion). Our analysis of the dative alternation also uncovered variability in the strength of the 'weight' factor in British and American English, similar to the differences between American and Australian subjects reported in Bresnan and Ford (2010). The present study thus lends further empirical support to Bresnan and Ford's argument that syntactic probabilities may be variable in varieties of English.

In short, the present study has highlighted how *effect strengths* are inherently variable across space and time while *effect directions* are stable. In other words, whatever underlies the workings of factors such as animacy or weight is potentially subject to variability and change, and thus is plastic. But not anything goes – there appear to be time- and placeless, basic constraints in the sense that synchronically and diachronically, animate entities always tend to be positioned before inanimate ones such that a change toward 'inanimate-before animate' is highly unlikely. Similarly, short elements usually tend to precede longer constituents in English [11] and not the other way round. This stability sits well with the probabilistic grammar approach adopted in, for instance, Bresnan and Ford (2010). In the same spirit we consider the regression models reported in the present study to be models of language users' internalized knowledge about the strength of the factors that constrain (or 'predict') syntactic choices. As such regression models are excellent tools to track down subtle changes in factor strengths (and their interactions) in a cognitively realistic way, although we hasten to add that the details of the precise cognitive mechanisms underpinning these models and the dynamics of the changes they describe are still not exceedingly well understood. As our study shows, however, adopting a cross-constructional approach to syntactic variation and change may point to general changes in grammar which could remain elusive when looking at some specific alternation in isolation. Cross-constructional analysis thus opens up yet another window into the processes driving the dynamics of syntactic alternations.

---

11 Generally, more complex phrases are ordered peripherally to less complex ones. The ordering is language-sensitive, with some languages – particularly VO languages such as English – preferring 'short-before-long' order and other languages, especially OV languages such as Japanese or Korean, favoring 'long-before-short' order.

References

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory 21*, 435–483.

Allen, C. L. (2009). Case syncretism and word order change. In A. v. Kemenade and B. Los (Eds.), *The Handbook of the History of English*, pp. 201–223. Chichester: Wiley-Blackwell.

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

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.

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.

Bresnan, J., S. Dingare, and C. D. Manning (2001). Soft constraints mirror hard constraints: Voice and person in English and Lummi. In *Proceedings of the LFG 01 Conference*, pp. 13–32. CSLI Publications.

Bresnan, J. and M. Ford (2010). Predicting Syntax: Processing Dative Constructions in American and Australian Varieties of English. *Language 86*(1), 186–213.

Bresnan, J. and T. Nikitina (2009). The gradience of the dative alternation. In L. Uyechi and L. Hee Wee (Eds.), *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, pp. 161–184. Stanford: CSLI Publications.

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

De Cuypere, L. (2010). A discourse based account of the Old English double object alternation. *Sprachwissenschaft*.

Fischer, O. (1992). Syntax. In N. Blake (Ed.), *The Cambridge History of the English Language Volume 2, 1066–1476*, pp. 207–408. Cambridge: Cambridge University Press.

Fischer, O. and W. van der Wurff (2006). Syntax. In R. Hogg and D. Denison (Eds.), *A History of the English Language*, pp. 109–198. Cambridge: Cambridge University Press.

Grafmiller, J. (sub.). Variation in English genitives across modality and genre.

Grafmiller, J. and S. Shih (2011). Weighing in on end weight. *Talk given at the LSA 2011 Annual Meeting, 6-9 January 2011, Pittsburgh, Pennsylvania*.

Gries, S. and A. Stefanowitsch (2004). Extending collostructional analysis: A corpus-based perspective on alternations. *International Journal of Corpus Linguistics 9*(1), 97–129.

Haddican, W. (2010). Theme–goal ditransitives and theme passivisation in British English dialects. *Lingua 120*, 2424–2443.

Hawkins, J. (1994). *A Performance Theory of Order and Constituency*. 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.

Huddleston, R. and G. K. Pullum (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Hundt, M. (2004). Animacy, agentivity, and the spread of the progressive in Modern English. *English Language and Linguistics 8*, 47–69.

Hundt, M. and U. Lenker (2006). Modern English in the making: Introduction. In C. Houswitschka, G. Knappe, and A. Müller (Eds.), *Anglistentag 2005 Bamberg: Proceedings*, pp. 439–445. Trier: Wissenschaftlicher Verlag.

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.

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.

Kemenade, A. v. (1987). *Syntactic case and morphological case in the history of English*. Dordrecht, Providence: Foris.

Koopman, W. (1931). *Word Order in Old English*. PhD thesis, University of Amsterdam.

Koptjevskaja-Tamm, M. (2001). Adnominal possession. In M. Haspelmath, E. König, W. Oesterreicher, and W. Raible (Eds.), *Language Typology and Language Universals*, Volume 2 of *Handbooks of Linguistics and Communciation Science*, pp. 960–70. Berlin, New York: Mouton de Gruyter.

Koptjevskaja-Tamm, M. (2002). Adnominal possession in European languages: form and function. *Sprachtypologie und Universalienforschung 55*, 141–172.

Kreyer, R. (2003). Genitive and of-construction in modern written English. Processability and human involvement. *International Journal of Corpus Linguistics 8*(2), 169–207.

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

McFadden, T. (2002). The rise of the *to*-dative in Middle English. In D. Lightfoot (Ed.), *Syntactic Effects of Morphological Change*, pp. 107–123. Oxford: Oxford University Press.

Mitchell, B. (1985). *Old English syntax*. Oxford, New York: Clarendon Press.

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

Pinheiro, J. C. and D. M. Bates (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

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

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

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.

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*, pp. 379–412. Berlin, New York: Mouton de Gruyter.

Rosenbach, A. (2004). The English s-genitive: A case of degrammaticalization? In O. Fischer, H. Perri-

don, and M. Norde (Eds.), *Up and down the Cline — The Nature of Grammaticalization*, pp. 73–96. Amsterdam, Philadelphia: Benjamins.

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.

Sankoff, D. and W. Labov (1979). On the use of variable rules. *Language in Society 8*, 189–222.

Seiler, H.-J. (1982). *Possession as an Operational Dimension of Language*. Tübingen: Narr.

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. (to appear). The great regression: genitive variability in Late Modern English news texts. In K. Börjars, D. Denison, and A. Scott (Eds.), *Morphosyntactic categories and the expression of possession*. Amsterdam, Philadelphia: Benjamins.

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, Philadelphia: 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.

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

Traugott, E. C. (1992). Syntax. In R. Hogg (Ed.), *The Cambridge History of the English Language Volume 1, The Beginnings to 1066*, pp. 168–289. Cambridge University Press: Cambridge.

Visser, F. T. (1963). *An historical syntax of the English language*. Leiden: Brill.

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

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.