Dative and genitive variability in Late Modern English

Exploring cross-constructional variation and change*

Christoph Wolk[1], Joan Bresnan[2], Anette Rosenbach[3] and Benedikt Szmrecsanyi[4]
[1]Freiburg Institute for Advanced Studies / [2]Stanford University / [3]Tanagra Wines / [4]University of Leuven

We present a cross-constructional approach to the history of the genitive alternation and the dative alternation in Late Modern English (AD 1650 to AD 1999), drawing on richly annotated datasets and modern statistical modeling techniques. We identify cross-constructional similarities in the development of the genitive and the dative alternation over time (mainly with regard to the loosening of the animacy constraint), a development which parallels distributional changes in animacy categories in the corpus material. Theoretically, we transfer the notion of 'probabilistic grammar' to historical data and claim that the corpus models presented reflect past speakers' knowledge about the distribution of genitive and dative variants. The historical data also helps to determine what is constant (and timeless) in the effect of selected factors such as animacy or length, and what is variant.

Keywords: Late Modern English, corpus linguistics, dative alternation, genitive alternation, probabilistic grammar, cross-constructional analysis, animacy

* We express our thanks to the Freiburg Institute for Advanced Studies (FRIAS) for funding the project "Predicting syntax in space and time". We are also grateful to the audiences of the March 2010 workshop on "Probabilistic syntax: Phonetics, diachrony and synchrony" (hosted by the FRIAS) and the November 2010 workshop on "The development of syntactic alternations" (Stanford University) for constructive feedback. We acknowledge helpful comments and suggestions by Ludovic De Cuypere, Beth Levin and Joanna Nykiel, and owe gratitude to Stephanie Shih for advice on coding final sibilancy, and to Katharina Ehret, whose assistance has contributed significantly to the development of our research. Lastly, three anonymous referees provided us with extremely helpful and constructive comments and suggestions. The usual disclaimers apply. The third author is grateful to the University of Paderborn for providing her access to their on-line library resources while working on this article. This material is based upon work supported by the National Science Foundation under Grant No. BCS-1025602.

Diachronica 30:3 (2013), **382–419**. DOI 10.1075/dia.30.3.04wol
ISSN 0176–4225 / E-ISSN 1569–9714

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 techniques. We are concerned with the history of the dative alternation (Bresnan et al. 2007, Bresnan & Ford 2010), as in (1), and the genitive alternation (Rosenbach 2002, Hinrichs & Szmrecsanyi et al. 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)

Our study thus takes a fresh look at two comparatively well-studied alternations in the grammar of English. On the theoretical plane, we follow recent probabilistic approaches to language (Bod et al. 2003) and base our work on two crucial assumptions (both broadly in line with variationist theory in the spirit of Labov 1982, Tagliamonte 2001). First, we assume that grammatical variation — and change — is sensitive to probabilistic (rather than categorical) constraints, in that conditioning factors may influence linguistic choice-making in extremely subtle, stochastic ways (Bresnan & Hay 2008:246). Second, we premise that grammatical knowledge must have a probabilistic component, as the likelihood of finding a particular linguistic variant in a particular context in a corpus can be shown to correspond to the intuitions that speakers have about the acceptability of that particular variant, given the same context (Bresnan 2007, also Rosenbach 2003, Hinrichs & Szmrecsanyi 2007). It follows that we are not merely interested in modeling the distribution of linguistic variants in historical corpus data drawing on the mathematics of uncertainty. Instead, our work ultimately aims to illuminate

1. All examples are drawn from the ARCHER corpus (see §3) and are referenced by ARCHER text identifiers.

aspects of the linguistic knowledge that writers in the Late Modern English period must have had, and how this knowledge has evolved over time.

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 are common to both alternations: consider the length 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. By comparing the development of these variables across two constructions, we can assess whether an individual change is construction-specific or may point to more general changes in grammar. The way these factors affect syntactic choices is not specific to English but echoes cross-linguistic regularities (Bresnan et al. 2001, Aissen 2003, Bresnan & 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. These regression models (which correctly predict over 90% of the dative and genitive observations in ARCHER), along with a series of detailed univariate analyses, will guide our investigation into cross-constructional variation and change.

This paper is structured as follows. In §2, we give an overview of the development of the dative and genitive alternations in the history of English. §3 introduces the data source. §4 defines the variable contexts. In §5, we discuss overall dative and genitive frequencies in real time. §6 presents the predictor variables that we utilize to model dative and genitive variability in §7, in which we report on regression models. In §8, we discuss and interpret the results. §9 offers 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 (Altenberg 1982: 13). By the fourteenth century, the inflected genitive was increasingly confined to a functional niche of 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 (Rosenbach 2002, Hinrichs & Szmrecsanyi 2007, Szmrecsanyi & Hinrichs 2008), and appears now to be spreading (Potter 1969, Dahl 1971, Raab-Fischer 1995, Rosenbach 2003, Szmrecsanyi 2009). In the news genre specifically, Hinrichs & 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, not much is known about genitive variability in Late Modern English.

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 less well documented. 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 1990, McFadden 2002); De Cuypere (2010) shows that this 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 & van der Wurff 2006), which during the Middle English period developed into "a fully productive alternative" (Fischer & van der Wurff 2006: 166) to the ditransitive dative construction. Conventional wisdom (e.g., McFadden 2002, Fischer & van der Wurff 2006) 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 1999, 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 that is relatively open to innovation, according to Hundt & Mair 1999: 236) and letters section (a register that is considered fairly oral, at least for private letters; see Raumolin-Brunberg 2005: 57). The genitive subcorpus 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 describes the variable contexts and so defines interchangeable dative and genitive contexts. First, previous analyses have used different definitions, and 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 all types of alternations, an *a priori* useful criterion is interchangeability (Labov 1966a, 1966b), i.e. the condition that each observed token could, in principle, have appeared in the form of any alternative variant. This criterion, however, is problematic for many reasons. First, due to limited available data, intuitions will have to be used, and these do not necessarily match up with observable behavior (e.g., Bresnan & 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 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* [determinerless NP], as in (3);
- [full NP]*s* [determinerless NP], as in (4);
- [full NP]' [determinerless NP], as in (5);
- [full NP] *of* [full NP], as in (6).

At this stage we also handcoded the boundaries of the possessor and possessum NP phrases (indicated by square brackets), as well as the possessor NP head noun.

(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 demonstrative possessums (e.g. *her face is as ugly as that of a dog*) (see Kreyer 2003: 170), 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 these are the only possible alternatives to *s*-genitives, which render the whole possessive construction definite (e.g. Rosenbach 2002: 30). As the definiteness constraint was already established in seventeenth-century English (Altenberg 1982: 27–28) we may safely exclude *of*-genitive constructions not headed by a definite article also in our Late Modern English data.

An exception was made for constructions which, although they lack an initial definite article, very clearly had a definite meaning, as in (8), or constructions headed by a proper name, as in (9).

(8) *SUICIDE OF THE AUSTRIAN AMBASSADOR AT PARIS.*<1883tim1.n6>

(9) a. *Mr. Gronus of the Three Tun Tavern,*[…] <1723dai2.n3b>
b. *France's President Giscard d'Estaing* […]<1979obs1.n8b>

Constructions as in (8) were essentially part of headlines, where the initial article often is omitted. The type of *of*-genitive construction in (9a), with a proper noun as possessum and a possessor specifying location, is a construction so far not discussed in the literature (to our knowledge) but clearly also occurring with *s*-genitives, as in (9b). Further excluded from this study are *of*-constructions with modifying function, as in (10), as these never alternate with *s*-genitives (Rosenbach 2002: 31).

(10) […] *all persons of quality here present* […] <1654mer2.n2b>

Likewise, *of*-genitives expressing a clearly appositive relation, as in (11), were not included, either.

(11) […] *the number of 13000 men* […] <1697pos1.n2b>

The analysis further focused on determiner (specifying) *s*-genitive constructions, as only these alternate with *of*-genitive constructions (Rosenbach 2002: 31–32). Measure genitives as in (12), which share properties of both determiner and classifying *s*-genitives (Huddleston & Pullum 2002: 470) are included, and so are other *s*-genitive constructions which are ambiguous between a classifying and a determiner interpretation, as in (13), which allows both the interpretation of *a* [*minister's son*] (implying an unspecific possessor and thus a type of son) or [*a minister*]*'s son* (implying a specific possessor and thus the son of a specific minister). In contrast, (14) gives an example of an unambiguous classifying genitive as the adjective *busy* clearly refers to the possessum *New Year's Day* and not to the possessor *New Year* (*a busy* [*New Year's Day*]). The number of measure genitives and classifying genitives included was very small, however, and their inclusion or exclusion did not in any way affect the final results, as tested by running the analysis with and without these constructions.

(12) a. […] *for the purpose of having a day's shooting* […] <1822eva1.n5b>
b. *Wendesday the Gosport sailed out of Harbour to Spithead, and was paid a Month's Wages.* <1762pub1.n4b>

(13) *James Ker a minester's son who formerly went by the name of Harrisen* […] <1653merc.n2b>

(14) […] *and Post Office linesmen had a busy New Year's Day on repair work.* <1939man1.n7b>

Overall, we sought to establish criteria on the basis of previous research on genitive variation which would give us genitive constructions that are interchangeable in principle rather than relying on a coder's intuition, as we know that such intuitions can be rather subjective. The problem of interspeaker differences in judging the grammaticality of alternations is even more serious for historical data as we cannot possibly have intuitions about which specific genitive constructions could alternate in the past. That is, rather than asking "does this genitive alternate?" in every individual case, we relied on the application of clear and easily replicable inclusion/exclusion criteria that we assume may be reasonably applied to the period of investigation. Naturally, some of the genitives may be considered marginal thanks to their low likelihood of occurrence, which is due to some of the constraints kicking in forcefully. So, for example, a corresponding *s*-genitive to the *of*-genitive in (15) will probably be judged as ungrammatical by many speakers of English, but there is nothing in its formal or semantic properties that would rule it out in general. Rather, it is the length of the possessor which makes a corresponding *s*-genitive (15-b) highly improbable.

(15) a. […] *the pursuit of the bloodthirsty wretches who had perpetrated the foul deed.* <1872gla1.n6b>
b.???the bloodthirsty wretches who had perpetrated the foul deed's pursuit

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 that can appear with a dative object, drawing on the list of verbs used by Bresnan et al. (2007), which was derived by extracting all verbs that appear in either construction in the Penn Treebank versions of the Switchboard and *Wall Street Journal* corpora and filtered by requiring at least five attestations by native speakers in both constructions in Google (cf. Cueni 2004: 1–2). We then expanded the list as necessary given positive evidence in ARCHER. The dative verbs thus considered are the following:[2] *accord, afford, allocate, allot, allow, assess, assign, assure, award, bequeath, bet, bring, carry, cause, cede, charge, cost, deal, deliver, deny, extend, feed, fine, flip, float, funnel, get, give, grant, guarantee, hand, issue, lease, leave, lend, loan, mail, net, offer, owe, pay,*

2. The verbs *do*, *will* and *make* that were included in the original list are excluded here, as they are highly frequent overall, but very rare in dative contexts.

permit, prepay, present, promise, quote, read, refuse, reimburse, repay, resell, run, sell, send, serve, show, slip, submit, supply, swap, take, teach, tell, tender, trade, vote, wish, write. Subsequently, ARCHER was searched for all occurrences of these verbs using regular expressions that were kept as inclusive as possible in order to capture variant spellings. 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, we excluded the following dative contexts:

– **Benefactives**. Instances where usage of a dative form is likely to be benefactive, as in (16), were excluded, on the grounds that while the double-object realizations are interchangeable, the prepositional realization utilizes the preposition *for* instead of *to*.

(16) 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 (17):

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

On semantic grounds, (17) is clearly not a dative — his house can hardly be conceptualized as the recipient. We removed all such cases, again including rather than excluding ambiguous examples.[3]

– **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 (18)–(22):

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

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

3. 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.

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

(21) *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> (subject relative clause)

(22) *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 (18) and (19), we observe *per se* regular verb-recipient-theme or verb-theme-recipient orders. However, in (18) the recipient is marked with a preposition, in contrast to the prototypical ditransitive recipient-theme ordering, while in (19) — a variant also widely available in some British English dialects (cf. Siewierska & Hollmann 2007) — we do not find the preposition usually found in theme-recipient realizations. The factors determining this sort of alternative variability are different and beyond the scope of this paper. In (20) and (21), 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 elided recipients or themes, as in (22).

Coming back to our list of dative verbs, the vast majority of these appear in both constructions in the corpus. Of those appearing more than once in a suitable context, only the following are found in only one construction: *cost* (30 double-object datives), *submit* (18 prepositional datives), *extend* (10 prepositional datives), *issue* (8 prepositional datives), *assure* (3 double-object datives), *guarantee* (2 double-object datives) and *quote* (2 prepositional datives).[4] In all, then, it is clear that determining potentially alternating verbs in historical texts is difficult because we cannot rely on our intuitions. Yet we take the comparatively low numbers (see above) of potentially iffy attestations as an indication that erroneously

4. Some dative verbs, such as *charge*, are actually attested in both constructions, but only once: *he wouldn't charge you much* <1964berg.f8b> versus *I charge it to the firm* <1880jess.d6a>. The pronominal direct object *it* in the prepositional dative construction with *charge* could refer to the amount charged or the item purchased, but there is not necessarily real competition between the two interpretations. Instead we are probably dealing with gradient polysemies/metaphors and syntactic usage preferences.

Table 1. Interchangeable genitive frequencies (and their text percentages) by ARCHER period.

	of–genitive	*s*–genitive	Total	corpus size (words)
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
1950–1999	357 (62%)	221 (38%)	578 (100%)	34k
Total	2890 (76%)	934 (24%)	3824 (100%)	242k

included dative constructions should at worst constitute a very small part of the overall dataset. We may also be missing cases of alternation that have died out by now, but in designing the study we have made the simplifying assumption that constructional overlaps in current usage have had antecedents during the Late Modern English period.

Our dative dataset spans in all $N = 3093$ interchangeable datives.

5. A frequency overview

In this section, we survey genitive and dative frequencies over time. The genitive alternation exhibits robustly fluctuating variant proportions (Figure 1, Table 1). In Present-Day English (1950–1999), 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 (Szmrecsanyi 2013). The 1950–1999 period actually surpasses the first ARCHER period in terms of relative *s*-genitive frequencies. Note also that the V-shaped pattern manifests itself 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.

As for the dative alternation, Table 2 shows that dative proportions are fairly stable in real time. The share of ditransitive datives fluctuates modestly between

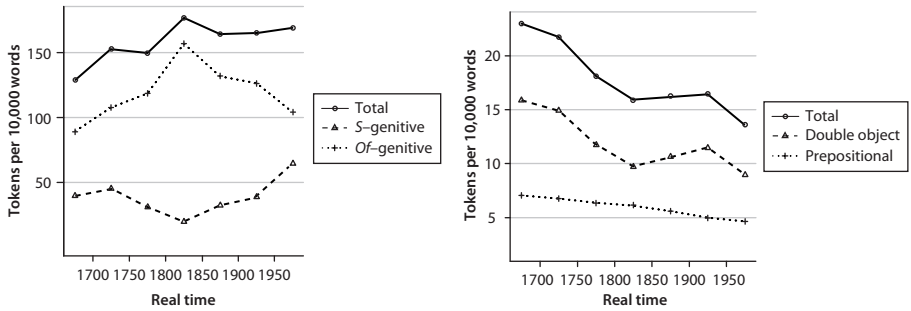

Figure 1. Mean interchangeable genitive (left) and dative (right) frequencies (normalized to frequency per 10,000 words) by ARCHER period.

61% (1800–1849) and 70% (1900–1949). However, Figure 1 makes 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 1650–1699 period to about 13 occurrences *pttw* in the 1950–1999 period. This frequency decline of datives with a theme and recipient argument slot is primarily caused by a general frequency decline of the dative verbs under study here: the frequency of the five most frequent dative verbs (*give*, *tell*, *bring*, *send* and *pay*) has shrunk, for reasons whose discussion would go beyond the scope of the present study, from 60 occurrences *pttw* in the 1650–1699 period to 39 occurrences *pttw* in the 1950–1999 period.[5]

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

	ditransitives		prepositional		Total		corpus size (words)
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
1950–1999	319	(66%)	166	(34%)	485	(100%)	357k
Total	2050	(66%)	1043	(34%)	3093	(100%)	1,789k

5. These figures include intransitive or de-verbal usages (e.g. *The art of giving is the finest of the fine arts* <1799dunl.d4a>), particle verb usages (e.g. *her Company paid off* <1753bos1.n4a>), monotransitive usages (e.g. *the Jesuits sat round telling funny stories* <1943waug.x7b>) and passive usages (e.g. *You'll be sent to prison camps* <1962mann.f8b>).

6. Predictors

This section introduces the predictors (in variationist parlance: conditioning factors) on which we will draw to model the genitive and dative alternation. We first review predictors that are common to both alternations (§6.1), and subsequently discuss the predictors that are specific to the genitive alternation (§6.2).

6.1 Shared predictors

6.1.1 *Corpus metadata*

ARCHER provides the year of creation[6] 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 *Length*

A well-known factor influencing the ordering of constituents is the principle of 'end weight' (Behaghel 1909/1910, Wasow 2002), according to which in languages like English, 'heavier' (i.e. longer and/or more complex) constituents tend to follow 'lighter' ones.[7] To determine the constituent lengths of individual dative and genitive occurrences, the following procedure was used: First, the constituents of each token were manually identified, as described in §4. Genitive possessums, which may only start with a determiner in the *of*-genitive, subsequently had their first word removed from the count if that word was a determiner as there is no counterpart for such a determiner in the *s*-genitive. Then, word and character counts of each constituent were automatically obtained. Several operationalizations of length were explored, including both the number of words or characters for each individual constituent as well as aggregated measures, such as length differences

6. A small number of texts in ARCHER are not dated exactly; these were placed in the middle of a time segment, i.e. in year 5 of a given decade or year 25 of a 50-year period, as the case may be.

7. 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. See Hawkins (1994) for typological differences.

and ratios. All individual measures performed well and are actually very similar: Across all constituent types, word and character lengths correlate at r = 0.976.[8] We additionally extracted a sample of 200 constituent strings and coded them for the number of syllables they contained,[9] for the sake of comparing the other operationalizations to a more phonological measure. The correlation between syllable and word counts comes out as 0.978, and that between syllables and characters as 0.993. In the analyses to follow, we thus feel justified in using the measure that leads to the best model fits: orthographic character counts. As a bonus, character counts are sensitive to the weight of individual words; for example, names like *Apollonius* and *Sylvestre* are heavier than *Tom* and *Dick*, but word-based measures fail to take this variation into account. Technically, a natural logarithmic transformation was applied to the character counts to reduce skewness, and finally the values were centered around 50-year period means to reduce multicollinearity and to account for possible changes in average lengths.

Taking (23) 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 length scores are −0.92 and −0.35, respectively.

(23) *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***]$_{\text{recipient}}$ [***no peace***]$_{\text{theme}}$. <1958ocon.x8a>

6.1.3 *Animacy*

Previous studies have reported reliable and strong effects of animacy; especially for the genitive alternation, research has explored animacy as a locus of diachronic variability (cf. Rosenbach 2002, Hinrichs & Szmrecsanyi 2007, Jankowski 2009, Hundt & Szmrecsanyi 2012). 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 animacy of the possessor in the genitive data set and of both the recipient and theme in the dative data set. Both dative recipients and themes, however, had very few instances that were coded as collective, locative or temporal nouns, and these were collapsed into the inanimate category. The result is a binary categorization scheme that is identical to the one used in Bresnan et al. (2007).

8. Per-constituent correlations range from 0.961 for genitive possessum lengths to 0.983 for dative themes.

9. Syllables were determined using the Unisyn Lexicon, available at http://www.cstr.ed.ac.uk/projects/unisyn/. Syllable counts for words not available in the Unisyn Lexicon were coded manually.

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 (24a). Collective possessors, as in (24b), 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 (24c). Locatives are locations, including geographical states, e.g. *Russia*, *this kingdom*, *the seas*, as in (24d). All other concrete or non-concrete noun phrases were classified as inanimate (24e).

(24) Animacy categories
   a. animate: [***BISHOP ABEL MUZOREWA'S***]$_{\text{animate}}$ ***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]***$_{\textit{collective}}$ *have appointed Messieurs Cartigny, Saurin, Meyvaud, and another, to examine into the Structure of those Machines*. <1723dai2.n3b>
   c. temporal: *After* [***yesterday's***]$_{\text{temporal}}$ ***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***]$_{\text{locative}}$ *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***]$_{\text{inanimate}}$; *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 §4.1). The pronoun category consists of all definite pronouns, such as the recipient in (25).

(25) *What agreement was made, I know not; but at his return Bavia* ***gave*** [***him***]$_{\text{pronoun}}$ [***the promised jewel***]$_{\text{definite}}$ *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 (26) (*Wilhelm*), but also titles such as *the King of England* and names of institutions such as *the Medical Society*.

(26) *Rappaport did not offer* [***Wilhelm***]$_{\text{proper name}}$ [***a cigar***]$_{\text{indefinite}}$, *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 names 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 (25). All other constituents, crucially including those headed by an indefinite determiner such as the theme in sentence (26), 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 (Szmrecsanyi 2006, Szmrecsanyi & Hinrichs 2008, Grafmiller forthcoming). We used an automatic annotation process, relying on the Carnegie Mellon University Pronouncing Dictionary version 0.7a[10] for transcription and coding all possessor phrases ending in [s], [z], [ʃ], [ʒ], [tʃ] or [dʒ] as ending in a sibilant; this includes possessors ending with a plural *s*. Tokens not included in the dictionary were coded manually. (27) exemplifies a rare occurrence of an *s*-genitive with a possessor ending in a final sibilant.

(27) [***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. Most crucially, there is the problem of defining an exhaustive and mutually exclusive set of semantic relations (see Rosenbach 2002: 58–64 and Rosenbach & Vezzosi 2000: 292–294). Currently, no such generally accepted classification for the English genitives exists, which may be one reason why more often than not this factor is dismissed in corpus studies of English genitive variation. For present purposes we follow the binary distinction between prototypical and non-prototypical possessive relations adopted in Rosenbach (2002). Prototypical relations comprise legal ownership (28a), body part relations (28b), kinship relations

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

(28c) and part-whole relations (28d), while all remaining cases were coded as 'non-prototypical' (examples in (29)).[11] Prototypical relations have been shown to favor the *s*-genitive, while non-prototypical relations show a preference for the *of*-genitive. Valence relations, such as in subjective (29b) and objective (29c) genitives, strictly speaking fall outside this taxonomy but were included and subsumed under 'non-prototypical possessive relations' (e.g. Seiler 1982, Rosenbach & Vezzosi 2000).

This binary classification is grossly simplified but we use it for the following reasons: First, it is theoretically well motivated by typological work on possession (cf. Heine 1997, Koptjevskaja-Tamm 2001, 2002) rather than being a set of ad hoc categories. Second, it can be rather straightforwardly applied in a corpus analysis. The prototypical possessive relations, i.e. legal ownership, kinship, body parts and part-whole are easy to identify, and one can then simply assign the remaining cases to the 'non-prototypical' category as a kind of elsewhere category. In sum we adopt a theoretically motivated, well-defined and practically feasible categorization.[12]

(28) 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 tooke 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>
   c. kinship: *It's said* [***the Duke of Berwick's Son***]$_{+\text{prototypical}}$ *is in one of the Ships, and Perth's two Sons in the other.* <1715eve1.n3b>
   d. part-whole: [***The Hull of a Ship***]$_{+\text{prototypical}}$ *was seen floating between Blackness and Point and Calais and Ambeleteuse*; <1735rea1.n3b>

(29) 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***]$_{-\text{prototypical}}$. <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***]$_{-\text{prototypical}}$ *is greatly admired.* <1872gla1n6b>

11. See Rosenbach (2002: 120–123) for arguments motivating that distinction.

12. See also Grafmiller (forthcoming, §4.1.2) for a similar approach and argument.

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***]$_{\text{-prototypical}}$, *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 schoolmasters which had established religious neutrality, and a request by the Council-General of the Seine that* [***the name of God***]$_{\text{-prototypical}}$ *should never be uttered in school had been rejected*. <1883tim2.n6b>

6.3 Lexical effects

It is well known that dative verbs differ in the likelihood that they will be used in either construction (e.g., Gries & Stefanowitsch 2004). All dative tokens were thus coded for verb lemma. Genitives have no clear carrier for lexical effects; 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 instance. 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 & Labov 1979). The technique permits quantification of the simultaneous effect of multiple individual explanatory factors[13] on a binary dependent variable, such as dative or genitive outcomes. We utilize a modern refinement of logistic regression analysis, 'mixed-effects logistic regression' (Pinheiro & Bates 2000).[14] In addition to so-called 'fixed effects' — which are classically estimated predictors suited for assessing the reliability of

13. For all categorial predictors, default levels were chosen so that the most frequent category is the default, which removes unnecessary collinearity from the model (cf. Wissmann et al. 2007). As a consequence, similar predictors may have different default levels both within and across models; in our data, this affects 'Animacy of Theme' in the dative model, which in contrast to all other animacy levels has 'inanimate' as the default, and the definiteness predictors.

14. We utilized the implementation of generalized linear mixed effects models in the lm4 library (Bates et al. 2011).

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 make the reasonable assumption that each corpus file has 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.[15] 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 of 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. Contrast this with the effect of e.g. animacy or the variety of English used; we consider these variables to be of primary interest, and in order to assess their reliability they need to be included as fixed effects. Where this is not the case, random effects provide a sophisticated yet elegant method for taking grouped variation into account, making sure that the estimation of the relevant variables can proceed unaffected by this noise and that the results are easy to generalize beyond the particular sample of authors and texts.

The following procedure was used for model selection and validation: 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[16] 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.

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

16. All random effects in our models are crossed, i.e. the information in each random effect does not by design determine the value of any other random effect, with one exception: ID is nested under REGISTER, as each file is assigned to exactly one REGISTER.

7.1 The genitive alternation

Table 3 reports fixed effects in the genitive model; the predicted odds are for the *of*-genitive. The classification accuracy is excellent — the model achieves a Somers' D_{xy} value of 0.93 and correctly predicts 91.9% of all genitive tokens, a considerable increase over baseline (75.6%) consistently predicting the overall most frequent realization (here, the *of*-genitive). Multicollinearity is not an issue, as the model's condition number ($\kappa = 8.4$) is well below the customary threshold of 15, which indicates medium collinearity.

To make this table more accessible, let us walk through some of the entries. 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'[17] of $e^{-1.54} = 0.21$. In other words, vis-à-vis a definite noun, a proper name is only a fifth (0.21 times) as likely to appear in the *of*-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 or 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 interaction 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 more likely to occur with the *of*-genitive (as shown by the positive sign of the coefficient). There is no reliable real-time change independent of animacy, as evidenced by the small, non-significant coefficient for 'centuries since 1800'. However, there are significant interaction effects between real time and collective, locative and temporal possessors. All interaction coefficients are negative, indicating that in real time, these three types become increasingly less likely to occur with the *of*-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 $e^{2.34+(-0.18*0)+(-0.47*0)} = 10.38$ times more likely to appear with the *of*-genitive than an animate possessor, while in 1950 it would be $e^{2.34+(-0.18*1.5)+(-0.47*1.5)} = 6.05$ times as likely. In addition, the effect of another predictor turns out to vary significantly as a function of real

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

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 *of*-genitive.

	Coefficient	SE	p
(Intercept)	-0.49	0.24	*
Animacy of possessor (default: animate)			
collective	2.34	0.32	***
inanimate	3.88	0.35	***
locative	3.26	0.39	***
temporal	1.92	0.35	***
Definiteness of possessor (default: definite)			
proper name	-1.54	0.17	***
indefinite	0.28	0.22	
Constituent length			
possessum length	-1.02	0.15	***
possessum length, squared	-0.67	0.13	***
possessor length	1.37	0.16	***
possessor length, squared	0.93	0.16	***
Semantic relation (default: non-prototypical)			
prototypical	-0.68	0.15	***
Final sibilant in possessor (default: no final sibilant)			
possessor has final sibilant	0.72	0.18	***
Real time			
centuries since 1800	-0.18	0.13	
Interactions			
animacy of possessor: collective (I: centuries since 1800)	-0.47	0.22	*
animacy of possessor: inanimate (I: centuries since 1800)	0.13	0.31	
animacy of possessor: locative (I: centuries since 1800)	-0.77	0.29	**
animacy of possessor: temporal (I: centuries since 1800)	-0.75	0.25	**
possessum length (I: centuries since 1800)	-0.32	0.11	**

* significant at p < .05, ** p < .01, *** p < .001

Table 4. Random effects in the genitive model.

Group	N_{groups}	Variance	StdDev
Text	242	0.95	0.97
Possessor head lemma	193	1.44	1.2

time: The linear component of the quadratic effect of possessum length is gradually increased as time progresses. We return to such real-time changes in §8.

Table 4 shows the variance of each random effect.[18] Both corpus text ID and possessor head noun show 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.

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.

	Coefficient	SE	p
Intercept	-1.32	0.48	**
Animacy of theme (default: inanimate)			
animate	1.73	0.49	***
Animacy of recipient (default: animate)			
inanimate	2.24	0.29	***
Definiteness of recipient (default: pronoun)			
definite	2.19	0.35	***
proper name	2.16	0.33	***
indefinite	3.66	0.4	***
Definiteness of theme (default: definite)			
proper name	1.13	0.65	.
indefinite	-1.17	0.21	***
pronoun	1.82	0.59	**
Constituent length			
theme length	-1.24	0.16	***
recipient length	1.33	0.18	***
Variety and real time			
American English (default: British English)	-0.37	0.23	
centuries since 1800	-0.38	0.16	*
text is from the twentieth century	0.48	0.34	
Interactions			
animacy of recipient is inanimate (I: twentieth century)	-1.12	0.46	*
length of theme (I: American English)	-0.72	0.28	*

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

18. Including weight as both a linear and quadratic predictor does not introduce collinearity problems, because the predictor is centered prior to modeling.

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 94.0% of all dative outcomes (baseline: 66.1%). Again, multicollinearity is not a problem ($\kappa = 7.2$).

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 a further indicator for dative tokens from the twentieth century, however, 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. Furthermore, 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 such that as theme length increases, the probability of realization as a double-object dative rises more quickly.

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 genres closer to the written norm tend to use more prepositional datives, a result that matches the difference between spoken and written materials in Bresnan et al. (2007).

Table 6. Random effects in the dative model

Group	N_{groups}	Variance	StdDev
Text	741	0.38	0.62
Theme	171	1.27	1.13
Verb lemma	49	4.86	2.21
Register	8	0.14	0.37

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

	adjustment
Drama	–0.27
Fiction	–0.27
Journal	–0.26
Letters	–0.03
Medicine	0.08
News	0.38
Science	0.19
Sermons	0.24

8. Discussion

We now turn to the interpretation of the regression models reported in the previous section. 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 length, animacy, definiteness and nominal expression bear on genitive and dative outcomes. By this token, our study diagnoses a good deal of 'probabilistic stability'. We focus now on the interactions between language-internal variables (animacy and constituent length) 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 show the distribution of actual corpus attestations while yielding a high resolution.

8.1 Interactions involving length

Our regression models have uncovered a set of interactions between length and real time, and between length 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 expected 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 8–12 orthographic 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, length is better behaved in later ARCHER periods. Figure 2 attempts 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, and plot smoother curves to highlight the trend in the data. The non-linearity discussed above is visible in both non-parametric regression curves, but we observe that it 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 length separately, a decision which — although justified in terms of model goodness-of-fit measures, which ultimately indicate how well the model captures linguistic structure — may not do full justice to the possibly complex interplay between relative and absolute lengths. Also, the effect of possessum length in particular has proven more difficult to capture than other length 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.

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. This finding confirms the relative importance of theme length as a locus of probabilistic differences between varieties of English, also found in a psycholinguistic experiment contrasting American and Australian participants (Bresnan & Ford 2010).

As seen in §6.1.2, while the existence of length effects is well known, the jury is still out on the best operationalization (see also Grafmiller & Shih 2011). Do the observed effects depend on our particular metric — orthographic character counts? To address this question, we also applied the final model structure using, as length measure, the number of words instead of the number of characters. The quadratic effect of genitive possessum and possessor length remained significant in this model, and the same is true for the difference between British and American English regarding dative theme lengths. The diachronic strengthening of the influence of genitive possessum length failed to achieve statistical significance. However, an interaction of real time and possessor length emerged as significant. More specifically, the effect of possessor length when measured in

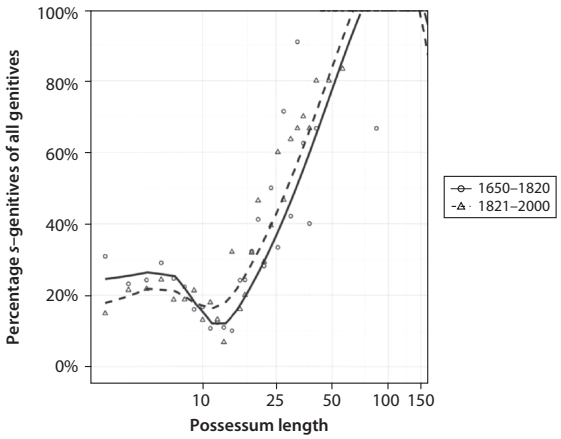

Figure 2. *S*-genitive rates (*y*-axis) as a function of possessum length (*x*-axis; binned uncentered possessum length on a log scale) and real time slice (heavy smoother: 1650–1820; dotted smoother: 1821–2000).

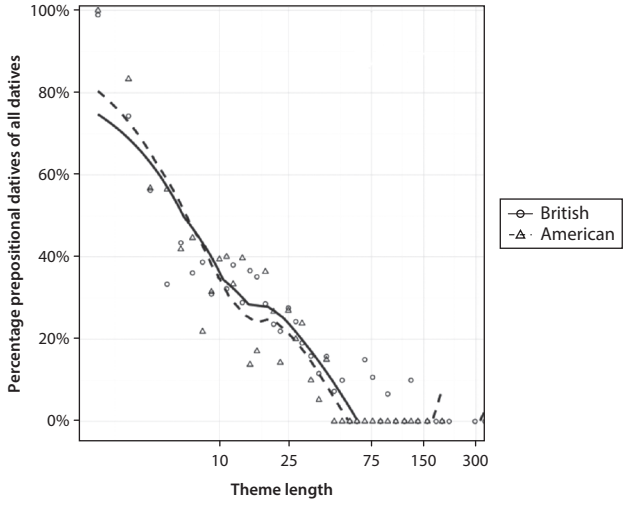

Figure 3. Prepositional dative rates (*y*-axis) as a function of theme length (*x*-axis; binned uncentered theme length on a log scale) and variety (heavy smoother: British English; dotted smoother: American English).

words increases in real time. These results suggest a genuine diachronic change in the effect of length in the genitive alternation. That said, operationalizations which treat each constituent in isolation are able to capture aspects of the complex nature of length effects only. Thus, future work on optimal length measurements is indispensable.

In all, the cumulative weight of cross-constructional evidence suggests that length, 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 show that length is generally a good predictor that works in the expected direction (in line with what processing considerations would lead one to expect), the factor nonetheless appears remarkably variable 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 in the twentieth century than in earlier periods.

Figure 4 displays observed proportions of dative and genitive realizations per 50-year period and animacy category. To make both plots more comparable we show the collective dative recipients separately from the other inanimates. Figure 4 makes 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 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 over the next 150 years.[19] 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. Because there is such a huge literature on the genitive alternation, we contextualize our findings in a bit more detail. Closing an empirical gap in the history of the genitive alternation in the Late Modern English period, our study shows that the *s*-genitive, which was basically restricted to animate possessors previously, extends to non-animate noun classes, a development starting in the seventeenth century (cf. Rosenbach 2007: 154–160). Previous research gives ample evidence for a contemporary spread of *s*-genitives to inanimate noun classes in the latter part of the twentieth century (e.g. Raab-Fischer 1995, Hundt 1997, Rosenbach 2002, 2003) but so far the precise

19. A large part of this change, however, is due to the frequency decline of the *s*-genitive in the period between 1800 and 1850, a period that also behaves oddly according to other measures, such as raw frequencies.

time-course of this extension has not been known. Rosenbach's (2007) study of the distribution of *s*-genitives (according to animacy) in Late Modern English points already to an early spread of the *s*-genitive to temporal nouns from the eighteenth century and a boost in the extension to collective and locative nouns from 1900, but her study looked at the distribution of *s*-genitives only. The present study corroborates and supplements these findings empirically by offering a comprehensive model of genitive variation for that period based on the interaction of various factors. 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, again around 1900, inanimate recipients start to appear more often as double-object datives, and subsequently collectives reverse their long-term trend.

While these developments are not identical across both constructions, they are remarkably similar: In both cases, the percentage of *s*-genitive or double object dative realizations of possessors and recipients that do not match the animate category increase during the twentieth century. Inanimate *s*-genitives do not exhibit such a pattern; however, this may be a limitation of the data set, as 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. Similarly, while animate recipients do not fluctuate as much as possessors do, 78.2% of animate recipients are pronouns, but only 6.2% of collectives and 10.7% of inanimates are.

As pronominal recipients strongly favor the double-object dative, animate recipients are biased toward stability in a way that the other recipients and possessors are not. To make both plots more comparable, we plot an additional line depicting only non-pronominal animate recipients. This line exhibits a V-shaped pattern similar to the one for animate possessors, with the exception of the 1850–1900 period (see Figure 4). When these facts are taken into account, the similarities in the development of the two constructions become even more striking. Comparable changes can be observed in 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 (30)), which had previously been limited to animate subjects.

(30) 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.

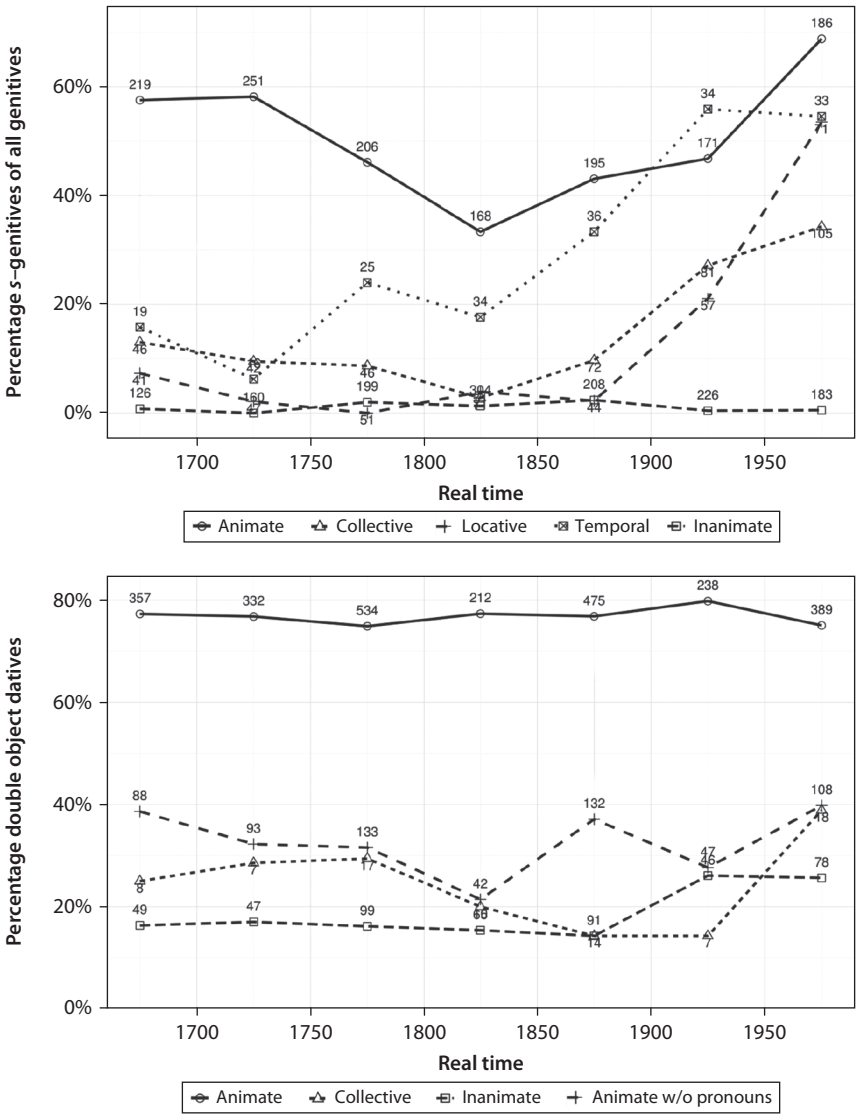

Figure 4. Animacy and real time. *S*-genitive rates (upper plot) and ditransitive dative rates (lower plot), on *y*-axis, as a function of ARCHER time slice (*x*-axis) and several animacy categories. Numbers in plot indicate absolute dative and genitive frequencies per time slice and animacy category.

Can input frequency fluctuations explain the observed patterns? 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 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 §6.1.3. The area plot in Figure 5 depicts the distribution of animacy categories in ARCHER's news section (we add that ARCHER's letter section is diachronically more stable, but exhibits roughly the same trends). First, 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 in *s*-genitives. That said, collective nouns, for example, have become relatively more frequent in the sample, as Table 8 highlights numerically. Thus, in the 1650–1699 period, collective nouns constituted only 8% of all nouns; in the 1950–1999 period, they constituted 11% of all nouns. The differential between the first ARCHER period and the last may seem subtle, but it approaches statistical significance, according to a chi-square test of independence ($p = 0.052$); the difference between the second ARCHER period and the last is, in any event, significant at $p = 0.004$. The bottom line is that the increase in the frequency of collective nouns, which started during the 1850–1899 period, is consistent with the time course of changes affecting the ability of collective nouns to serve as *s*-genitive possessors, suggesting that environmental and cultural changes may have played a role. Finally, in the 1800–1849 period (see Figure 5), the frequency of animate nouns in

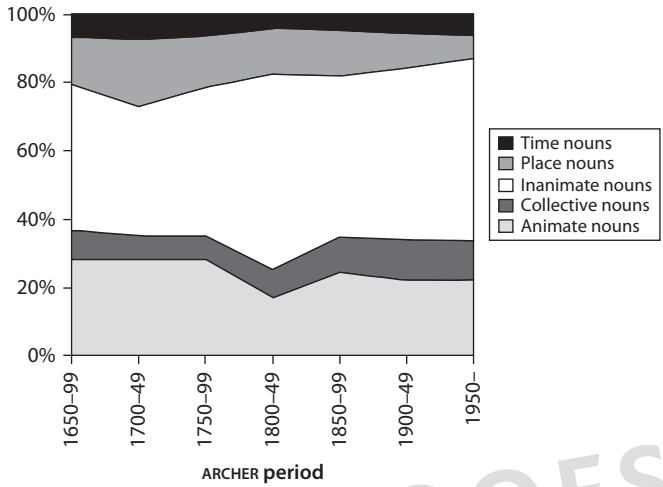

Figure 5. Area plot depicting the general distribution of animacy categories (*y*-axis, in %) against real time (*x*-axis) in ARCHER's British news section, based on a random noun sample ($N_{total} = 5{,}174$).

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).

In all, we would like to emphasize that the genitive and dative alternations share (1) distributional similarities (the generally similar probabilistic determinants of realization choice), (2) formal similarities (constituent order) and (3) a common core of meaning ('[potential] possession' in recipient/possessor-first order) (see Bresnan & Nikitina 2009). A change in any of these shared properties would account for the empirically observable parallelisms, yet still allow for construction-specific developments such as the observed increase in temporal *s*-genitives, which is difficult to reconcile with strong possessorship constraints.

Table 8. Occurrences of collective nouns in the ARCHER-news-based random noun sample depicted in Figure 5: raw number of collective noun occurrences, sample size per period and percentage of collective nouns.

	N collective nouns	sample size per period	% collective nouns
1650–1699	61	750	8%
1700–1749	49	729	7%
1750–1799	48	747	6%
1800–1849	59	757	8%
1850–1899	77	752	10%
1900–1949	85	723	12%
1950–1999	82	716	11%

9. Concluding remarks

The present paper investigates two cases of syntactic variation over time, i.e. the genitive and dative alternation, during the Late Modern English period — a period whose grammar in general has been notoriously understudied until recently. Apart from filling a descriptive gap in the history of the genitive and dative alternations, this study makes two major contributions.

9.1 Cross-constructional comparison

While variation studies usually focus on individual syntactic alternations, the present paper provides a large-scale quantitative study of two cases of syntactic variation and compares how they develop over time (from 1650 to 1999). The genitive and dative alternation, though different types of syntactic variation operating within different syntactic domains (nominal vs. verbal), still have quite a few things in common: Both are cases of word order variation and as such share a

number of explanatory constraints (most notably the factors of animacy, definiteness and length or weight), which have been shown to be effective in word order choice on both the nominal and verbal plane. So, what are the advantages of taking a bird's eye view and looking at the history of two alternations instead of focusing on one? One of our main empirical findings is the observation that the effect of animacy weakens over time in both the genitive and dative alternation, which parallels an increasing frequency, in some genres, of expressions referring to inanimate and collective entities. Adopting a cross-constructional approach to syntactic variation and change may thus point to general changes in grammar which could remain elusive when looking at a given specific alternation in isolation.

9.2 Probabilistic grammar and historical data

Theoretically, we adopt the idea of a dynamic probabilistic grammar (e.g. Bybee & Hopper 2001, Bod et al. 2003, Gahl & Garnsey 2006, Gahl & Yu 2006), applied to the domain of syntactic variation (e.g. Bresnan 2007, Bresnan & Ford 2010) and transfer it to the domain of historical data. Modern statistical techniques such as logistic regression allow us to model speakers' predictive abilities in very precise ways, and we claim that the corpus models for genitive and dative variation from 1650 to 1999 presented in this study represent implicit probabilistic knowledge of past speakers. We are not the first to use statistical methods to track grammatical variation and change (e.g., Kroch 1994, Nevalainen 2003, Gries & Hilpert 2010), but our approach is novel in adopting a probabilistic grammar framework. The main evidence for probabilistic grammar comes from studies which show that present-day speakers' intuitions about probabilities of grammatical variants in experimental studies match those found in corpora, which suggests that speakers have some sort of predictive knowledge of the distribution of grammatical choices, i.e. a probabilistic grammar (Arnold et al. 2000, Gries 2002, 2003, Bresnan 2007 and Bresnan & Ford 2010). If the main evidence for probabilistic grammar comes from experimental (and psycholinguistic) studies, how can we transfer it to past speakers of English, which are no longer available for experimental testing? For past stages of English the only direct evidence available is the corpus data, after all. The factors of animacy and length (weight) constitute some of the major constraints on choice of construction in cases of word order variation, and their effects seem to derive from cognitive or processing constraints (e.g. Bock et al. 1992, McDonald et al. 1993, Hawkins 1994, Wasow 2002). Having no access to past speakers' intuitions, we can still apply the uniformitarian principle and may reasonably assume that the cognitive mechanisms underlying present-day probabilistic patterns also underlie past variation (see Jäger & Rosenbach 2008 for

applying the uniformitarian assumption to psycholinguistics).[20] And indeed we have seen that the factors of animacy and length (weight) show diachronic stability in that their general effects remain constant. Our data demonstrates a stable preference over time for placing animate referents first and for a short-before-long preference. What changes is the strength of the effects, most notably and clearly for animacy, and somewhat less clearly so for length (weight). The effect direction, however, does not change. In this respect, the historical material also provides evidence for the idea of probabilistic grammar. Historical data is just another piece of evidence for how the mind works, another window to the mind.[21] In the same way that cross-varietal and typological evidence gives us an idea about the constraints on the scope of variation, so does historical data — it helps us understand what is stable and invariable as opposed to what is variable. The stability of the general effect of animacy over time, as noted, may thus be taken as another piece of evidence for their cognitive grounding. The variability in the strength of effects, on the other side, indicates that probabilistic knowledge is not fixed but essentially plastic in nature, as evidenced by the fluctuation of frequencies of variants over time. While the regression models used in this study allow us to track down subtle changes in factor strengths (and their interactions) in a cognitively realistic way, the details of the precise cognitive underpinning of these models and the dynamics of the changes they describe are still not well understood and will be, we hope, the subject of future studies. In this endeavor evidence from historical data may provide another window into the dynamics of syntactic variation and change and thus probabilistic grammar.

References

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

Allen, Cynthia L. 2009. Case syncretism and word order change. In Ans van Kemenade & Bettelou Los (eds.), *The handbook of the history of English*, 201–223. Chichester, UK: Wiley-Blackwell.

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

20. For other approaches using psycholinguistic factors to account for language change see e.g. some of the work in grammaticalization theory (Diessel 2007 offers an overview).

21. Using historical data as a window to the mind is an idea first expressed by Kiparsky (1968) for the domain of mental representations within generative linguistics and later adopted also by functional approaches, e.g. Mayerthaler (1981), Stein (1988) or Rosenbach (2002: 269–277).

Arnold, Jennifer E., Thomas Wasow, Anthony Losongco & Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76(1). 28–55.

Bates, Douglas, Martin Maechler & Ben Bolker. 2011. lme4: Linear mixed-effects models using S4 classes. R package version 0.999 375–42.

Behaghel, Otto. 1909/1910. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.

Biber, Douglas, Edward Finegan & Dwight Atkinson. 1994. ARCHER and its challenges: Compiling and exploring a representative corpus of historical English registers. In Udo Fries, Gunnel Tottie & Peter Schneider (eds.), *Creating and using English language corpora: Papers from the Fourteenth International Conference on English Language Research and Computerized Corpora*, 1–13. Amsterdam: Rodopi.

Bock, J. Kathryn, Helga Loebell & Randal Morey. 1992. From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review* 99. 150–171.

Bod, Rens, Jennifer Hay & Stefanie Jannedy (eds.). 2003. *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base*, 77–96, Berlin: Mouton de Gruyter.

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

Bresnan, Joan, Shipra Dingare & Christopher D. Manning. 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. *Proceedings of the LFG 01 Conference*, 13–32.

Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1). 186–213.

Bresnan, Joan & Jennifer Hay. 2008. Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua* 118(2). 245–259.

Bresnan, Joan & Tatiana Nikitina. 2009. The gradience of the dative alternation. In Linda Uyechi & Lian Hee Wee (eds.), *Reality exploration and discovery: Pattern interaction in language and life*, 161–184. Stanford, CA: CSLI Publications.

Bybee, Joan & Paul Hopper. 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Cueni, Anna. 2004. *Predicting the outcome of the choice between the dative constructions of English*. Ms., Stanford University.

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

De Cuypere, Ludovic. 2010. A discourse based account of the Old English double object alternation. *Sprachwissenschaft* 35. 337–368.

Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25(2). 108–127.

Fischer, Olga. 1992. Syntax. In Norman Blake (ed.), *The Cambridge history of the English language* vol. 2, *1066–1476*, 207–408. Cambridge: Cambridge University Press.

Fischer, Olga & Wim van der Wurff. 2006. Syntax. In Richard Hogg & David Denison (eds.), *A history of the English language*, 109–198. Cambridge: Cambridge University Press.

Gahl, Susanne & Susan Marie Garnsey. 2006. Knowledge of grammar includes knowledge of syntactic probabilities. *Language* 82(2). 405–410.

Gahl, Susanne & Alan C.L. Yu (eds.). 2006. Special theme issue: Exemplar-based models in linguistics. *The Linguistic Review* 23(3).

Grafmiller, Jason. Forthcoming. Variation in English genitives across modality and genre. *English Language and Linguistics*.

Grafmiller, Jason & Stephanie Shih. 2011. Weighing in on end weight. Talk given at the LSA 2011 Annual Meeting, 6–9 January 2011, Pittsburgh, PA.

Gries, Stefan Th. 2002. Evidence in linguistics: Three approaches to genitives in English. In Ruth M. Brend, William J. Sullivan & Arle R. Lommel (eds.), *LACUS Forum XXVIII: What constitutes evidence in linguistics?*, 17–31. Fullerton, CA: LACUS.

Gries, Stefan Th. 2003. Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1. 1–27.

Gries, Stefan Th. & Martin Hilpert. 2010. Modeling diachronic change in the third person singular: A multifactorial, verb- and author-specific exploratory approach. *English Language and Linguistics* 14(3). 293–320.

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

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

Heine, Bernd. 1997. *Possession: Cognitive sources, forces, and grammaticalization*. 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). 437–474.

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

Hundt, Marianne. 1997. Has BrE been catching up with AmE over the past thirty years? In Magnus Ljung (ed.), *Corpus-based studies in English: Papers from the seventeenth International Conference on English Language Research on Computerized Corpora*, 135–151. Amsterdam: Rodopi.

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

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

Hundt, Marianne & Benedikt Szmrecsanyi. 2012. Animacy in early New Zealand English. *English World-Wide* 33(3). 241–263.

Jankowski, Bridget. 2009. Grammatical and register variation and change: A multi-corpora perspective on the English genitive. Paper presented at the American Association for Corpus Linguistics (AACL) 2009.

Jäger, Gerhard & Anette Rosenbach. 2008. Priming and unidirectional language change. *Theoretical Linguistics* 34(2). 85–113.

van Kemenade, Ans. 1987. *Syntactic case and morphological case in the history of English*. Dordrecht, Netherlands: Foris.

Kiparsky, Paul. 1968. Linguistic universals and linguistic change. In Emmon Bach & Robert T. Harms (eds.), *Universals in linguistic theory*, 171–210. London: Holt, Rinehart & Winston.

Koopman, Willem. 1990. *Word order in Old English*. Amsterdam: University of Amsterdam dissertation.

Koptjevskaja-Tamm, Maria. 2001. Adnominal possession. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals*, vol. 2, 960–970. Berlin: Mouton de Gruyter.

Koptjevskaja-Tamm, Maria. 2002. Adnominal possession in European languages: Form and function. *Sprachtypologie und Universalienforschung* 55. 141–172.

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

Kroch, Anthony. 1994. Morphosyntactic variation. In Katherine Beals, Jeanette Denton, Robert Knippen, Lynette Melnar, Hisami Suzuki & Erica Zeinfeld (eds.), *Papers from the 30th Regional Meeting of the Chicago Linguistics Society*, vol. 2, *The parasession on variation in linguistic theory (CLS 30)*, 180–201. Chicago: Chicago Linguistic Society.

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

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

Labov, William. 1982. Building on empirical foundations. In Winfred Lehmann & Yakov Malkiel (eds.), *Perspectives on historical linguistics*, 17–92. Amsterdam: John Benjamins.

Mayerthaler, Willi. 1981. *Morphologische Natürlichkeit*. Wiesbaden, Germany: Athenaion.

McDonald, Janet L., J. Kathryn Bock & Michael H. Kelly. 1993. Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology* 25. 188–230.

McFadden, Thomas. 2002. The rise of the *to*-dative in Middle English. In David Lightfoot (ed.), *Syntactic effects of morphological change*, 107–123. Oxford: Oxford University Press.

Mitchell, Bruce. 1985. *Old English syntax*. Oxford & New York: Clarendon Press.

Mustanoja, Tauno F. 1960. *A Middle English syntax*, vol. I. Helsinki: Société Néophilologique.

Nevalainen, Terttu. 2003. *Socio-historical linguistics: Language change in Tudor and Stuart England*. London: Longman.

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

Potter, Simeon. 1969. *Changing English*. London: André Deutsch.

R Development Core Team. 2010. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/.

Raab-Fischer, Roswitha. 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.

Raumolin-Brunberg, Helena. 2005. The diffusion of subject *you*: A case study in historical sociolinguistics. *Language Variation and Change* 17. 55–73.

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

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

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

Rosenbach, Anette & Letizia Vezzosi. 2000. Genitive constructions in Early Modern English: New evidence from a corpus analysis. In Rosanna Sornicola, Erich Poppe & Ariel Shisha-Halevy (eds.), *Stability, variation and change of word-order patterns over time*, 285–307. Amsterdam: John Benjamins.

Sankoff, David & William Labov. 1979. On the use of variable rules. *Language in Society* 8. 189–222.

Seiler, Hans-Jakob. 1982. *Possession as an operational dimension of language*. Tübingen: Narr.

Siewierska, Anna & Willem Hollmann. 2007. Ditransitive clauses in English with special reference to Lancashire dialect. In Mike Hannay & Gerard J. Steen (eds.), *Structural-functional studies in English grammar: In honour of Lachlan Mackenzie*, 83–102. Amsterdam: John Benjamins.

Stein, D. 1988. Psycholinguistic issues and determinants of syntactic change. *Papiere zur Linguistik* 39(2). 31–48.

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

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

Szmrecsanyi, Benedikt. 2010. The English genitive alternation in a cognitive sociolinguistics perspective. In Dirk Geeraerts, Gitte Kristiansen & Yves Peirsman (eds.), *Advances in cognitive sociolinguistics*, 141–166. Berlin: 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*, 59–88. Amsterdam: 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: John Benjamins.

Tagliamonte, Sali. 2001. Comparative sociolinguistics. In Jack Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *Handbook of language variation and change*, 729–763. Malden, MA: Blackwell.

Thomas, Russel. 1931. *Syntactical processes involved in the development of the adnominal periphrastic genitive in the English language*. Ann Arbor, MI: University of Michigan dissertation.

Traugott, Elizabeth Closs. 1992. Syntax. In Richard Hogg (ed.), *The Cambridge history of the English language*, vol. 1, *The beginnings to 1066*, 168–289. Cambridge: Cambridge University Press.

Visser, Frederikus Theodorus. 1963. *An historical syntax of the English language*. Leiden: Brill.

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

Wissmann, Malte, Helge Toutenburg & Shalabh. 2007. Role of categorical variables in multicollinearity in the Linear Regression Model. Technical report 008, Department of Statistics, University of Munich.

Zaenen, Annie, Jean Carlette, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden et al. 2004. Animacy encoding in English: Why and how. In Donna Byron & Bonnie Webber (eds.), *Proceedings of the 2004 ACL Workshop on Discourse Annotation, Barcelona, July 2004*, 118–125. East Strouddsburg, PA: Association for Computational Linguistics.

Résumé

Nous présentons une approche interconstructionnelle de l'histoire de l'alternance du génitif et du datif en anglais moderne tardif (1650 à 1999) en nous appuyant sur des bases de données annotées et des techniques de modélisation statistique actuelles. Nous identifions les similitudes interconstructionnelles que l'on retrouve dans le développement de l'alternance du génitif et du datif au fil du temps (concernant principalement l'affaiblissement de la contrainte animé/inanimé), un développement qui correspond aux changements de distribution des catégories animé/inanimé dans le corpus. D'un point de vue théorique, nous transférons la notion de "grammaire probabiliste" à des données historiques et posons que les modèles extraits du corpus indiquent les connaissances de ces locuteurs du passé sur la distribution des variantes du génitif et du datif. Les données historiques aident également à établir les éléments constants (et invariables au cours du temps) des effets de facteurs ciblés, tels que le caractère animé/inanimé ou la longueur, par opposition à leurs éléments variables.

Zusammenfassung

Das Papier stellt eine konstruktionsvergleichende Analyse der diachronen Entwicklung der Genitiv- und Dativalternation im Spätneuenglischen (AD 1650 bis AD 1999) vor, die auf detaillierte annotierte Datensätze und moderne Techniken der statistischen Modellierung zurückgreift. Es wird gezeigt, dass sich diese beiden Alternationen ähnlich entwickeln, vor allem in Bezug auf den Faktor Belebtheit, der an Einfluss verliert. Dieser Wandel verläuft parallel zu Änderungen in der relativen Frequenz der Belebtheitskategorien im Korpus. Auf der theoretischen Ebene überträgt die Studie die Idee der probabilistischen Grammatik auf historische Daten und schlägt vor, dass die vorgestellten Modelle das sprachliche Wissen historischer Sprecher zu diesen Alternationen abbilden. So helfen historische Daten zu identifizieren, welche Effekte der ausgewählten Faktoren konstant sind, und was variieren kann.

Corresponding author's address

Christoph Wolk
Freiburg Institute for Advanced Studies (FRIAS)
Albert-Ludwigs-Universität Freiburg
Albertstraße 19
79104 Freiburg i.Br., Germany

christoph.wolk@frias.uni-freiburg.de