# Stability and Fluidity in Syntactic Variation World-Wide: The Genitive Alternation Across Varieties of English

Benedikt Heller, Benedikt Szmrecsanyi, Jason Grafmiller

KU Leuven

## Abstract

What are the scope and limits of syntactic variation within and across varieties of English? To address this question, we investigate well-known syntactic variation between the *s*-genitive (*Mr Barnsley's management*) and the *of*-genitive (*the management of Mr Barnsley*) in nine varieties of English. We specifically gauge the stability of constraints on this variation by analyzing a richly annotated dataset spanning $N = 10{,}558$ interchangeable genitives from nine components of the International Corpus of English. Regression modeling indicates that constraints such as possessor animacy, constituent length, final sibilancy of the possessor, as well as the effect of medium (spoken vs. written) as a language-external factor differ in strength across varieties. The language-internal constraints, however, never change effect direction. We conclude that the probabilistic grammar fueling genitive variation is surprisingly stable overall, but does exhibit some fluidity along the lines of a distinction between English as a native language (ENL) and English as a second language (ESL) varieties: those constraints that tend to favor *s*-genitive usage tend to be weakened in ESL varieties.

## Acknowledgements

**Introduction**

Establishing the scope of grammatical differences between varieties of English around the world in a comparative perspective is a quite popular research topic in the recent literature (e.g., Gries & Deshors 2015; Kortmann & Wolk 2012; Szmrecsanyi & Kortmann 2009). We contribute to this scholarship by exploring the constraints on syntactic variation in a comparatively large and typologically diverse sample of varieties of English. As a case study we specifically investigate the so-called "genitive alternation" between the *s*-genitive, as in (1), and the *of*-genitive, as in (2).

(1)     Parliament also removed additional powers granted to him last year to tackle [the country]$_{possessor}$'s [economic crisis]$_{possessum}$ under these powers (ICE-SIN, s2b-001)

(2)     Cement is one of the core raw materials for a developing country like India and it plays a vital role in the [economic growth]$_{possessum}$ of [the country]$_{possessor}$ (ICE-IND, w2a-031)

The two variants differ in their ordering of the possessor and possessum phrases, and also in an additional definite article preceding *of*-genitive possessums. In the spirit of recent work on the genitive alternation (Ehret, Wolk & Szmrecsanyi 2014; Grafmiller 2014; Shih et al. 2015, inter alia), we adopt a variationist perspective (Labov 1972; Jankowski & Tagliamonte 2014) and restrict attention to genitive tokens where the alternative variant could have been used. Thus, for example, *of*-constructions as in *You know most of the lecturers* (ICE-SIN, s1a-001) are not considered, as a periphrasis with the *s*-genitive (**the lecturers' most*) is not possible.

Genitive variation is extremely well studied. Historically, the *s*-genitive, after having been increasingly replaced by the *of*-genitive in Middle English, made a come-back in Early Modern English (e.g., Thomas 1931). In Late Modern English, the *s*-genitive has been on the rise primarily thanks to an expansion to inanimate noun classes (Wolk et al. 2013). The relevant

literature about the constraints that govern the choice between the *s*-genitive and the *of*-genitive is too voluminous to be reviewed here in much detail (see Rosenbach 2014 for an exhaustive overview). Suffice it to say that the determinants of genitive variation are numerous, multifactorial, and probabilistic. Constraints that are well known to influence genitive choice include, but are not limited to, possessor animacy (animate possessors favor the *s*-genitive), constituent length (long possessors favor the *of*-genitive and long possessums the *s*-genitive), and final sibilancy (a final sibilant in the possessor discourages the *s*-genitive). Crucially, however, none of these factors is deterministic: for example, animate possessors do favor the *s*-genitive, but if the possessor is long enough, the principle of end-weight may win out against possessor animacy. To come to terms with the multifactorial nature of genitive variation, analysts have long seen the need to use multivariate techniques for analyzing corpus data (Gries 2002), or to use experimental designs (Rosenbach 2005).

Research has uncovered interactions between language-internal constraints (e.g., animacy, length, etc.) and language-external factors such as genre, time, and variety (Hinrichs & Szmrecsanyi 2007; Grafmiller 2014; Wolk et al. 2013). However, as far as regional/geographic differences are concerned, analysts have tended to restrict attention to differences between British and American English, and we know next to nothing about genitive variation in the many other varieties of English spoken and written around the world—this is the gap in the literature that we seek to fill. In contrast to much previous research, this paper offers an analysis that is not solely focused on English as a native language (ENL) varieties, i.e., British, Irish, Canadian, and New Zealand English, but also considers a number of English as a second language (ESL) varieties such as Jamaican, Singapore, Indian, Philippine, and Hong Kong English.

That said, we would like to emphasize at the outset that our main interest does not lie in

assessing the aptness of particular models of variety categorization and genesis (e.g., Kachru 1985; Schneider 2007), or for that matter in characterizing particular varieties of English. Rather, we are concerned in this paper, in a more typologically inspired and thus abstractive spirit, with the scope and limits of syntactic variation within and across varieties of English around the world. Our investigation is guided by two research questions:

- To what extent do users of different varieties of English rely on the same or similar choice-making processes when it comes to choosing genitive variants?

- Are cross-varietal differences random or can they be explained by variety type (ENL vs. ESL)?

On the theoretical plane, we commit to the notion that grammar is the "cognitive organization of one's experience with language" (Bybee 2006:711) and apply the idea of a dynamic probabilistic grammar (e.g., Bybee & Hopper 2001; Gahl & Garnsey 2004) to the realm of variation across varieties. We specifically rely on the variation-centered, usage- and experience-based probabilistic grammar framework developed by Joan Bresnan and collaborators (e.g., Bresnan & Ford 2010). This framework makes three key assumptions: (i) Grammatical variation is sensitive to multiple and sometimes conflicting probabilistic constraints. Such constraints influence linguistic choice-making in subtle ways that may remain invisible unless analyzed quantitatively. (ii) Grammatical knowledge must have a probabilistic component, for the likelihood of finding a particular linguistic variant in a particular context in a corpus has been shown to correspond to the intuitions that speakers have about the acceptability of that particular variant, given the same context. (iii) This probabilistic knowledge is derived in large part from language experience, and so is subtly—but fluidly—(re)constructed throughout speakers' lifetimes.

What were our predictions and hypotheses prior to embarking on this study? It seems reasonable to assume that language users, whatever variety they speak or write, are subject to the same cognitive and processing constraints and are therefore prone to make overall similar syntactic choices. MacDonald (2013), for example, proposes a unified Production-Distribution-Comprehension (PDC) approach that explains how biases in language production lead to statistical patterns in production that language users implicitly learn and that subsequently guide comprehension. One of these biases in (incremental) language production is what MacDonald (2013:3) calls the "Easy First" principle: language users tend to place those constituents first that are comparatively easy to retrieve, so the execution of utterances can begin early.

Easy First is a more general account for well-known tendencies such as the principle of end-weight (Behaghel 1909; Wasow 1997), according to which longer constituents tend to follow shorter constituents – shorter constituents are "easier", hence the word order pattern. We expect to see end-weight effects in the genitive alternation, i.e., a preference for the variant that places long constituents after short constituents (as illustrated in 3 and 4), across the board, given their presumably strong links to the design of the human speech production system.

(3)    In [today]$_{possessor}$'s [retention-conscious climate of trying to keep numbers up and students in programmes]$_{possessum}$, I believe that this is a question that must be addressed (ICE-CAN, w1b-021)

(4)    but anyway it's uhm again the [story]$_{possessum}$ of [Bruny Surin and Donovan Bailey]$_{possessor}$ and it's the lead story in La Presse (ICE-CAN, s1b-021)

Thanks to the experience-based nature of grammar, and the different input(s) that users of different varieties receive over their lifetime, it is likely that there are subtle differences in the relative influence of preferences such as the principle of end-weight in particular speech

communities (see Bresnan & Hay 2008 for discussion). In this connection, Szmrecsanyi et al. (2016) have coined the term "probabilistic indigenization," which refers to the process through which probabilistic patterns of internal linguistic variation, such as end-weight effects, are reshaped by shifting usage frequencies in speakers of post-colonial varieties. We expected to see probabilistic indigenization effects not primarily with regard to constraints that are strongly processing-driven (such as, presumably, end-weight), but rather with constraints such as possessor animacy, or register/medium differences (register conventions are, after all, social conventions).

Based on these considerations, we created a variationist dataset covering N = 10,558 interchangeable *s*- and *of*-genitives drawn from nine components of the International Corpus of English. A mixed effects logistic regression model (dependent variable: genitive choice) correctly predicts 94 percent of all genitive outcomes in the dataset and indicates that constraints such as possessor animacy, constituent length, final sibilancy of the possessor, as well as the effect of medium as a language-external factor differ in strength across varieties. Crucially, however, the language-internal constraints do not change effect direction, as predicted by Easy First (MacDonald 2013). This is another way of saying that users of English consistently place long constituents after short constituents, avoid the *s*-genitive when the possessor ends in a sibilant, and prefer the *s*-genitive when the possessor is animate. This overall stability notwithstanding, the subtle fluidity that we find does seem to distinguish ENL from ESL varieties: those constraints that tend to favor *s*-genitive usage tend to be downplayed in ESL varieties, and those that favor *of*-genitive usage tend to be strengthened. We will argue that in a synchronic perspective, this is an echo of the well-known SLA preference for analytic marking (an exploration of the alternative explanation that ESL varieties lag behind ENL varieties with

regard to the diachronic drift towards more *s*-genitive usage is reserved for another occasion – see the conclusion section for discussion).

The remainder of this paper is structured as follows. Section two describes the data source, the variable context, the constraints on genitive variation that we consider, and our methods. Section three presents the results. In section four we discuss our findings, and section five offers some concluding remarks.

## Data and Methods

### Corpus Data and Varieties Covered

The data for this study were extracted from the International Corpus of English (ICE) (Nelson 1996), a corpus family that is well suited for cross-varietal comparisons since all subcorpora follow the same design. They each contain 600,000 words of spoken and 400,000 words of written language and are further subdivided into a wide variety of text types to offer a balanced reflection of the varieties (Nelson 1996). The spoken component contains dialogs and monologs (face-to-face-conversations, phonecalls, classroom lessons, broadcast discussions, etc.); the written component distinguishes non-printed (e.g., student writing) and printed (e.g., academic writing) texts (for an overview of all text categories in ICE see The ICE Project 2009).

The present analysis explores nine ICE subcorpora, which represent (i) British, (ii) Irish, (iii) Canadian, (iv) New Zealand, (v) Jamaican, (vi) Singapore, (vii) Indian, (viii) Philippine, and (ix) Hong Kong English. Although this selection was motivated by a desire to cover a wide range of varieties, some existing corpora could not be included (i.e., African varieties, or ICE-US) due to limited availability or format issues at the time of study design.

The varieties can be classified as ENL, or L1 (i–iv), and ESL, or L2 (v–ix) varieties. This

particular classification is the customary one adopted in well-known reference works (Crystal 2004; Kortmann & Szmrecsanyi 2004; Kortmann & Lunkenheimer 2012). We are aware that this simple distinction is rather crude, and that more fine-grained approaches could have been taken, such as distinguishing between low-contact varieties and high-contact varieties (Trudgill 2009), or taking more account of the distinction between language shift varieties such as Irish English and other varieties (Hickey 2010). Further, some varieties are difficult to place in either category. Singapore English, for example, is said to be on its way to becoming an ENL variety (Pakir 2001; Schneider 2007). It is also subject to debate whether Jamaican English is an ESL variety (we classify JamE as ESL since it is Jamaican Creole that is the native language of most of the population, not Jamaican English, which is taught in schools; see Hinrichs 2006; Patrick 2008). In spite of the imperfections of the classification, an ENL-ESL distinction seems to offer the appropriate level of granularity regarding the goals of the present study. We note in passing that the ENL-ESL distinction also translates into the terminology of customary models such as Kachru's (1985) Three Circle Model (where ENL corresponds to the Inner Circle, and ESL to the Outer Circle), or Schneider's (2007) Dynamic Model (where ENL varieties occupy stage five, and ESL varieties occupy earlier stages), but we again hasten to add that the present study is not primarily concerned with models of variety classification or evolution.

Unlike other syntactic phenomena, genitive constructions are textually relatively frequent. A pilot study indicated that 10 percent of texts in ICE yield enough observations for reliable statistical analysis, so a sample was created that contains text one, text eleven, text twenty-one, etc. The selection proportionately reflects all text categories. After automatic extraction of all genitive markers (i.e., *of*, *'s*, and word-final *s'*), false positives were manually discarded, which resulted in a sample of 10,558 interchangeable genitives.

*Defining the Variable Context*

Following previous studies (e.g., Rosenbach 2002, 2014; Wolk et al. 2013), we considered as interchangeable any genitive token that did not fall into one of the following categories: appositive genitives, classifying genitives, double genitives, idiomatic/fixed genitives, partitive genitives, and any genitives with indefinite possessums.

In appositive genitives, the *of*-phrase is a post-modification whose head is co-referential with the head of the preceding noun phrase and usually describes it further (Biber et al. 1999). The expression in (5), for example, is not about a group supporting a US envoy's idea (this reading would be interchangeable), but the *of*-phrase describes the idea that this group supports (i.e., sending a US envoy to Northern Ireland).

(5)     The group supports the idea of a special US envoy being sent to Northern Ireland (ICE-IRE, w2c-001)

Classifying genitives do not specify a possessum as a specific entity but express to which class it belongs. Example (6) shows the difference.

(6)     England's chairman of selectors, Ted Dexter, has bowed to the wishes of Botham who, rightly enough, is playing the king in Jack and the Beanstalk, a part, it appears, that might have been written into the old children's story for him (ICE-NZ, w2e-001)

While the first genitive, *England's chairman of selectors*, denotes a specific person, the second one, *the old children's story*, does not refer to a specific story of a specific group of old children, but specifies the class of story (i.e., a story for children). Double genitives contain two genitive markers at once, as in (7).

(7)     A painting of Pete's also forms part of this type (ICE-JA, s2b-041)

(8)     Florists could lose out on sales opportunities this Valentine's Day (ICE-JA, w2c-

001)

Idiomatic genitives are fixed patterns that only occur in one form, e.g., (8). Finally, partitive genitives express a measurement of some sort (e.g., time, distance, or value; see Biber et al. 1999) as in (9).

(9)     He's calculated that a dollar's worth of trees planted in nineteen forty-seven would produce sixty dollars worth of logs and lumber in nineteen ninety-five (ICE-CAN, s2b-031)

Further, all *of*-genitives that did not contain a definite possessum were considered categorical and were therefore excluded. The reason for this is that in *s*-genitives, the clitic *'s* has a determiner function, which results in definite possessums in all *s*-genitives. For *of*-genitives to be comparable to that, an additional definite article is needed (recall examples 1 and 2). Cases like *a major focus of our neurosurgical practice* in (10) were thus excluded. Bare plurals, like *Infections of the CNS* in (10), were also excluded for the same reasons.

(10)     Infections of the CNS will certainly remain a major focus of our neurosurgical practice (ICE-HK, w2a-021)

By defining interchangeability in this way, we assume that the norm of what constitutes the choice context is comparable across varieties. We believe this to be a valid assumption for various reasons. First, the assumption of a constant interchangeability norm is a crucial prerequisite for variationist/probabilistic analysis, and enables researchers to see the big picture. If we assumed different norms, varieties would not be comparable. Second, careful manual screening of all cases in both the choice context and the categorical context did not reveal any patterns that suggest different norms. Further, our regression model (see section Data Analysis below) obtained very good diagnostics, which lets us confidently attribute the variation to the

various language-internal and language-external constraints that we study.

*Constraints and Annotation*

The extraction of interchangeable genitive occurrences from the corpus material was accomplished in three steps: (i) Automatic extraction of all text units that contain one of the genitive markers *of*, *'s*, or word-final *s'*; (ii) automatic filtering according to lexical, part-of-speech, and grammatical constraints; (iii) manual checking to correct errors in the automatic filtering process. After that, possessor and possessum phrases of each genitive case were manually annotated and their nominal heads were extracted. Finally, all cases were annotated for the language-internal and language-external predictors presented in the following sections: animacy, final sibilancy, givenness, thematicity, overall frequency, length, lexical density, medium, and variety.

Animacy of the possessor is arguably the most important constraint in the genitive alternation (e.g., Rosenbach 2005). It is so important that it is used in many prescriptive grammars of English (e.g., Murphy 2012) as a rule on how to use English genitives (*s*-genitive if possessor is animate, otherwise *of*-genitive). Previous research, however, has demonstrated that the animacy constraint can be overpowered by syntactic weight (Rosenbach 2005), that it is subject to diachronic change (Wolk et al. 2013), and that its strength can differ by variety (Hinrichs & Szmrecsanyi 2007; Hundt & Szmrecsanyi 2012).

The annotation of animacy was performed semi-automatically. Every possessor head's animacy status was automatically classified[1] and manually corrected where necessary. After initially using a five-fold classification following Wolk et al.'s (2013) approach (i.e., distinguishing animate, collective, temporal, locative, and inanimate entities), the classification

was simplified to a binary distinction between animate and inanimate entities due to a high degree of correlation between the more fine-grained animacy categories and several other predictors.

Table 1 shows that in our data inanimate possessors are predominantly realized as *of*-genitives. Animate possessors, on the other hand, are almost equally likely to occur in the *s*- as in *of*-genitive.

[TABLE 1 HERE]

If a sibilant (i.e., [s], [z], [ʃ], [tʃ], [ʒ], and [dʒ]) is present at the end of the possessor phrase, as in (11), *s*-genitive usage is less likely for articulatory reasons (Zwicky 1987).

(11)    The paradox's conclusion (ICE-NZ, w2b-021)

Table 2 shows that in the dataset under investigation, *s*-genitives occur only in about 10 percent of the cases where the possessor phrase ends in a sibilant as opposed to about 30 percent if there is no final sibilant. Final sibilancy was annotated automatically by referring to the CMU Pronunciation Dictionary.[2] If a word could not be found in the dictionary, the annotation relied on orthography.

[TABLE 2 HERE]

Givenness, also sometimes referred to as information status (Jankowski 2013), captures whether or not information has been mentioned in the preceding context. Many previous studies have shown that *given* possessor heads favor *s*-genitive use, but not always significantly so (e.g., Grafmiller 2014). For this study, we coded a possessor as *given* if the lemma of the possessor head occurred anywhere in the respective corpus text prior to the observed genitive token. The annotation of givenness was performed automatically by a script that searched the previous context of each genitive case and used a lemma list[3] to determine if the lemma in question had

been mentioned before. In the present data set (see Table 3), given possessor heads are more likely to be realized as *s*-genitives (28.9 percent) than discourse-new possessors (22.4 percent).

[TABLE 3 HERE]

Thematicity (Osselton 1988) reflects the degree to which a possessor head constitutes a central topic of a corpus text. It is defined here as the frequency with which a possessor head occurs in the entire corpus text in question: if a possessor head is highly thematic, it is more likely to take the *s*-genitive (Hinrichs & Szmrecsanyi 2007). The total number of mentions was counted automatically and then normalized by the usual corpus text size of 2000 words. In our data, possessor heads in *s*-genitives are more thematic than possessor heads in *of*-genitives (Table 4).

[TABLE 4 HERE]

To our knowledge, the overall frequency of the possessor head has not been included in genitive alternation research so far, but frequency has been shown to influence other grammatical choices (Hilpert 2008). To determine how often possessor heads are used overall, we automatically counted their occurrence in the Corpus of Global Web-based English (GloWbE corpus), a large-scale 1.9 billion word corpus of online English, which samples a multitude of different varieties and mirrors the ICE family in its 60/40 division of less formal and more formal texts (Davies & Fuchs 2015). Frequencies were retrieved from the respective components of GloWbE that represent the variety in question—possessor heads from Indian English, for example, were counted in the Indian component of GloWbE.

The summary in Table 5 shows that the possessor heads in *of*-genitives are more frequent on average than the possessor heads in *s*-genitives. It also shows a high standard deviation, which indicates that the frequencies values deviate from the mean value quite a bit.

[TABLE 5 HERE]

The length of the possessor and possessum is predicted to influence genitive choice along the lines of the principle of end-weight, which was first described by Behaghel (1909) and which is defined as "the tendency for long and complex elements to be placed towards the end of a clause" (Biber et al. 1999). In the genitive alternation, this translates to relatively long possessors, such as *the Chinese Government* in (12), favoring the *of*-genitive, and relatively long possessums, such as *potential medical uses* in (13), favoring the *s*-genitive (Rosenbach 2002; Rosenbach 2005; Wolk et al. 2013).

  (12) the [power ]possessum of [the Chinese Government]possessor (ICE-HK, s1a-021)

  (13) [laser]possessor's [potential medical uses]possessum (ICE-IND, w2b-031)

There are different ways of measuring syntactic weight (e.g., number of characters, number of words, number of syntactic nodes), which all correlate very highly (Rosenbach 2014:227); following Wolk et al. (2013), we chose to use the number of orthographic characters. Whitespaces and special characters (e.g., hyphens) were not counted.  In this study, we include separate measures of the length of possessor and possessum (e.g., Hinrichs & Szmrecsanyi 2007), rather than a ratio of the two (e.g., Bresnan & Ford 2010).

Table 6 shows that possessors tend to be shorter in *s*-genitives and longer in *of*-genitives. Possessums show the reverse pattern (Table 7).

[TABLE 6 HERE]

[TABLE 7 HERE]

Lexical density, here measured as type-token ratio, represents how many unique words occur in a given text relative the total number of words in that text. Grafmiller (2014) shows that lexically dense environments favor the use of *s*-genitives. Since type-token ratio is very sensitive

to overall text length, it was measured in the immediate one-hundred-word environment of the genitive instance in question. Table 8 shows that the average type-token ratio is slightly higher in *s*-genitives than in *of*-genitives.

[TABLE 8 HERE]

The literature also reports effects that language-external factors (such as medium and variety) have on genitive choice. Grafmiller (2014), for example, found that spoken texts favor *s*-genitives more than written texts (with the exception of press texts), and Hundt & Szmrecsanyi (2012) found that in different varieties (i.e., early British English and early New Zealand English), possessor animacy varies in strength. The present study focuses on variety, but also controls for medium (i.e., spoken vs. written language). It seeks to understand how genitive distributions differ across varieties and how variety interacts with the language-internal constraints discussed above.

As can be seen in Table 9, there are no obvious differences in genitive frequency between spoken and written texts when aggregated over varieties. Regarding variety differences, Table 10 shows that *s*-genitive rates range from 16.3 percent in Jamaican English to 34.7 percent in Canadian English

[TABLE 9 HERE]

[TABLE 10 HERE]

*Data Analysis*

In order to statistically model genitive choice in varieties of English, we utilized mixed-effects logistic regression modeling, which models multifactorial phenomena that are simultaneously influenced by multiple constraints. While some constraints (see section *Constraints and*

*Annotation*) favor the *s*-genitive (e.g., animate possessors), others favor the *of*-genitive (e.g., the presence of a final sibilant). Without controlling for multiple factors at once and estimating the contributions of each individual predictor while holding other predictors constant, it is hard to pinpoint which role individual constraints play. Additionally, there might be interactions between the constraints, such that the influence of one constraint might change depending on another constraint. The present investigation is particularly interested in how well-studied constraints (such as animacy and length) interact with VARIETY (e.g., is the animacy constraint stronger in certain varieties than in others?). By using logistic regression modeling, we can both estimate the contributions of the individual constraints and account for possible interactions. The resulting model provides coefficients for every constraint and every interaction, and these reflect the degree to which the constraints influence genitive choice.

In order to improve the estimation of these coefficients and the generalizability of the model, we included random effects, which have been shown to make the estimation of the constraints that are of primary interest (fixed effects) more accurate (Gries 2015). Random effects capture the influence of constraints that are particular to the sample, i.e., not repeatable, or not of primary interest to the analysis (Baayen 2008:241; Gelman & Hill 2007). For example, idiolectal preferences are particular to the sample since a different sample from another corpus would most likely sample production of different language users. Other constraints that were modeled as random effects are the ICE-specific text categories *monologue*, *dialogue*, *printed*, and *non-printed* (see Nelson 1996), as well as the head lemmas of the possessor and possessum phrases. Idiolectal differences were modeled as nested into text category (referred to as "Text category/Language user" in Table 11), accounting for the structure of ICE, which does not sample individual language users in more than one text category. All of these constraints were

modeled as random intercepts.

The model selection process followed the guidelines in Zuur et al. (2009) and Gries (2015), i.e., proceeding from a full model and gradually deleting unnecessary components, first in the random effects structure and then in the fixed effects structure. Model diagnostics were monitored in the process. The remaining coefficients were validated using a bootstrapping approach as outlined in Baayen (2008:283). In this approach, 100 random subsets of the data are taken, and the same regression model is fit to the subset each time. The coefficients will be slightly different every time, which enables us to estimate how variable the coefficients are and thus deduce a range in which we are 95 percent sure the true value of the respective coefficients lies. This bootstrap analysis showed that these ranges never include 0, which means that we can be confident that our results are not due to chance. All calculations were performed with R (R Core Team 2015) and the lme4 package (Bates et al. 2014).

The resulting minimal adequate regression model classifies 94.14 percent of all cases correctly (baseline: 75.32 percent *of*-genitives). The concordance statistic C, which ranges between 0.5 (random classification) and 1 (perfect classification) and which is independent of the baseline, is 0.981. C-values above 0.8 indicate that the model is good enough to explain the variation patterns in the data (Tagliamonte & Baayen 2012:156). The model shows medium multicollinearity ($\kappa = 18.06$), which is well below the "potentially harmful" threshold of thirty (Baayen 2008:182). Thus we may assume that the estimation of a given predictor's coefficient is not distorted by the influence of any other predictor.

[TABLE 11 HERE]

Table 11 shows the significant random effects in the resulting regression model. The higher the variance value, the more do members of the respective group exhibit a bias toward

one of the genitive variants. Possessor head lemma has a higher variance than possessum head lemma, which indicates that there are more possessor head lemmas with a bias toward either the *s*-genitive or the *of*-genitive than possessum head lemmas. Examples include temporal and locative possessor head lemmas like *week* and *Jamaica*, which tend to be realized as *s*-genitives, and possessor head lemmas like *God* and *department*, which are part of phrases like *The kingdom of God* or *the head of the department*, which tend to be realized as *of*-genitives. In the case of the possessum heads it is job titles (e.g., *director* in *director of administration*, and *president* in *president of committee*) that favor the *of*-genitive, and it is often words that are linked to a prototypical relation of possession that tend toward *s*-genitives (e.g., *hat* in *woman's hat*, and *house* in *friend's house*). Language user idiosyncrasies nested into text category show less variance, which indicates that idiolectal variation is comparatively subtle.

## Results

Table 12 shows the effects that the individual constraints have on genitive choice. The coefficients (column b) and the odds ratios are effect sizes, which reflect how the presence/absence of the constraints (e.g., the presence of a final sibilant) or, in the case of numerical predictors, a one-unit increase (e.g., +1 character in possessor length) changes the odds for *s*-genitive occurrence. If $b$ is positive (e.g., 4.31 for the change of POSSESSOR ANIMACY from *inanimate* to *animate*), the level change in question makes *s*-genitives more likely (e.g., animate possessors facilitate *s*-genitives); if it is negative, *s*-genitive use is less likely. The odds ratios show similar information, but have a lower bound at zero, and a value of one indicates an equal chance of either variant. If the odds ratio is greater than one, *s*-genitives are more likely than *of*-genitives; if it is between zero and one, they are less likely. Odds ratios can be interpreted as factors with which the odds of an *s*-genitive realization are increased or decreased given

certain predictor changes. For example, the presence of a final sibilant in the possessor phrase changes *s*-genitive odds by a factor of 0.35, i.e., makes *s*-genitive usage only 0.35 times as likely in the presence of a final sibilant as compared to its absence.

[TABLE 12 HERE]

We can see that possessor animacy makes a huge difference in genitive choice, with animate possessors increasing *s*-genitive odds by a factor of seventy-four vis-à-vis inanimate possessors.[4] Possessor length is also highly significant, and every one-character increase in length makes *s*-genitives less likely. Possessum length affects genitive choice in the opposite direction: for every additional character an *s*-genitive outcome becomes more likely. With a negative coefficient for possessor length and a positive coefficient for possessum length, the predictions are in line with the principle of end-weight. The effects of final sibilancy and givenness are also as expected, with final sibilancy disfavoring and givenness favoring *s*-genitive use. MEDIUM, as a main effect, does not make a significant difference in the model (p > 0.05), but was not excluded since it is part of a higher-order interaction (see Table 13 below). Thematicity and type-token ratio also increase *s*-genitive use and thus influence the alternation as expected. Possessor head frequency, however, slightly but significantly discourages an *s*-genitive outcome, which is unexpected. Finally, there are five significant level changes in VARIETY, all of which favor *s*-genitives in comparison to the British English baseline. They are, in order of effect size: Hong Kong English, New Zealand English, Canadian English, Philippines English, and Singapore English. Indian English, Jamaican English, and Irish English do not significantly differ from British English as far as the overall probability of the *s*-genitive is concerned.

[TABLE 13 HERE]

Table 13 reports significant interaction terms involving VARIETY. The interaction of

POSSESSOR ANIMACY and VARIETY shows negative coefficients across all varieties, significantly and most strongly so for Philippines English, Hong Kong English, and New Zealand English. This indicates that the tendency to use the *s*-genitive with animate possessors is strongest in British English. On the other hand, while the main effect of POSSESSUM LENGTH, reflecting the constraint's strength in British English, is barely significant, it is considerably stronger in Irish, Philippine, Hong Kong, Singapore, and Canadian English, indicating that this factor is rather weak in British English. FINAL SIBILANCY also shows stronger effects in other varieties, significantly so in Hong Kong and Indian English. MEDIUM does not seem to make a difference in British English (main effect not significant), but a change from written to spoken language does make a significant difference in Philippine and Hong Kong English. Remaining interaction terms between VARIETY and other predictors not listed in the table were not significant.

To further highlight the nature of the interaction terms in the model, we now move on to present effect plots, which show fitted values on a probability scale. Effect plots visually summarize information from the regression model by systematically altering the values of individual predictors (e.g., animacy and variety) while holding the values of all other predictors in the model constant (see Fox 2003 for a detailed explanation). The effects plots in figures 1–4 provide a visual representation of the significant interactions in the regression model, reflecting how users of different varieties are expected to choose variants for typical genitive contexts that only differ in the levels of the predictors involved in the respective interaction.

Figure 1 shows the interaction of variety and possessor animacy. The probability of *s*-genitive use in the animate condition is higher for the five varieties on the left side, i.e., the ENL varieties and Singapore English. Estimates for the other ESL varieties are considerably lower. Figure 2 shows the interaction of variety and possessum length. While users of Irish English

make sure to use the *s*-genitive with very long possessums, users of the other varieties show a weaker tendency to do so. Users of British English and Indian English, on the other hand, seem unaffected even by very long possessums. Figure 3 shows the interaction of variety and final sibilancy. Users of Hong Kong and Indian English are very unlikely to use *s*-genitives in the presence of a final sibilant in the possessor phrase. Final sibilancy seems to make the least difference in Philippine English, where the estimates for *present* and *absent* are closest. Finally, Figure 4 shows the interaction of variety and medium. *S*-genitive predictions are slightly elevated in spoken Canadian, New Zealand, and British English, while medium does not seem to make a difference in Irish, Singapore, Jamaican, and Indian English. In Philippine and Hong Kong English, predictions for *s*-genitive use are clearly higher in written texts.

[FIGURE 1 HERE]

[FIGURE 2 HERE]

[FIGURE 3 HERE]

[FIGURE 4 HERE]

## Discussion

Our analysis of the scope and limits of syntactic variation between the *s*-genitive and the *of*-genitive was guided by two questions: first, what is the extent to which users of different varieties of English rely on the same or similar choice-making processes when it comes to selecting genitive variants? Second, are cross-varietal differences random or can they be explained by variety type (ENL vs. EFL), based on a classification that was derived from standard reference works (Crystal 2004; Kortmann & Szmrecsanyi 2004; Kortmann & Lunkenheimer 2012)? With regard to the first question, we found that language-internal constraints in genitive variation are fairly stable across varieties, with degrees of fluidity that

remain within certain limitations. As to the second question, we argued that our results point to a probabilistic distinction between ENL and EFL varieties, especially with regard to the effect of possessor animacy.

Recall that *s*-genitive proportions (see Table 10) are lower in some ESL varieties (Jamaican, Indian, and Philippine English) than in the ENL varieties we studied (British, New Zealand, Irish, and Canadian English). It turns out that similar to the ENL varieties, Singapore and Hong Kong English (which are ESL) also show high *s*-genitive rates, a finding that is robust under multivariate control in a regression analysis (not reported here) without interactions. How to account for these frequency differentials? Given that language learners tend to avoid inflectional marking and prefer analyticity (Klein & Perdue 1997:311) and that languages with a high proportion of language learners tend to lose case marking (Bentz & Winter 2013), we would like to argue that users of ESL varieties tend to favor the more analytic, periphrastic, and thus more transparent *of*-genitive in contexts where ENL speakers might be more likely to use the *s*-genitive. Hence we find higher frequencies of *of*-genitives in ESL varieties – in other words, the of-genitive is a comparatively stronger default option in ESL varieties than in ENL varieties.

We move on to discussing the conditioning of genitive variants in multivariate analysis. The nature of the main effects conforms across the board with the Easy First principle (MacDonald 2013), according to which language users tend to place constituents that are more easily retrievable earlier for the sake of optimizing utterance planning. Indeed, we consistently find that shorter genitive constituents are placed first, and that animate, given and highly thematic possessors, which are also easier to retrieve (MacDonald 2013:3–6), favor the *s*-genitive. There is one constraint in our analysis that *prima facie* violates Easy First: frequent possessors, which should be easier to recall from memory (MacDonald 2013:3), were predicted

to favor the *s*-genitive, but in fact were found to discourage *s*-genitive use. We suspect that this may be due to the fact that proper noun possessors, as in (14), which are low in frequency, are mostly realized as *s*-genitives.

(14)    In Chong Fung-yuen's case, it might be said, the CFA has politely put its foot down (ICE-HK, w2b-011)

So the probabilistic constraints we considered are overall well behaved, but how does our analysis shed light on the issue of stability versus fluidity in probabilistic grammars? The fact of the matter is that the majority of language-internal constraints under study do not significantly interact with VARIETY – and this is another way of saying that we are observing a good deal of cross-varietal stability. Users of English, whatever their regional and/or cultural background, choose genitives in similar ways likely because they are subject to the same cognitive constraints in language production. This, of course, is not surprising. But the model also does show degrees of fluidity, which remain within the limitations dictated by an overall stable pattern of constituent ordering choice driven by utterance planning. The language-internal constraints that do interact with VARIETY (i.e., possessor animacy, possessum length, and final sibilancy) do not affect genitive choice in different directions across varieties (e.g., animate possessors favoring the *s*-genitive in variety A, but favoring the *of*-genitive in variety B)—we merely find different degrees of strength in the same direction.

Psycholinguistically inspired proposals in the spirit of MacDonald's (2013) PDC account do not easily explain these differences in effect strength. Relying on this account, one would have predicted that language users in ESL varieties show *stronger* utterance planning biases since in many cases they constitute language learners who use English alongside one or more other languages (Kachru 1985:12). Findings from previous research suggest that the memory-

related mechanism of Easy First should have a stronger effect in L2 production than in L1 production since in L2 processing there is constant interference from the L1 language (MacWhinney 1997; Szmalec, Brysbaert & Duyck 2012:89). Therefore, one would have expected to find that language users of ESL varieties more often choose *s*-genitives, e.g., with animate possessors. What we do in fact find, however, is that ESL users, similar to their overall preference as represented in the main effects, more often choose the *of*-genitive with animate possessors. This trend is amply clear from Figure 1. Supplementary regression analysis (not reported here), in which the nine levels of the factor VARIETY were conflated into two levels *ENL* vs. *ESL*, shows that the effect of possessor animacy is significantly weaker in ESL varieties than in ENL varieties. An Easy First-inspired explanation at any cost would have to suggest that in ESL varieties, animate possessors are not "easier" than inanimate possessors to the same degree as in ENL varieties—hence the weaker showing of possessor animacy in the regression models—but there is no evidence that we know of that would support this claim. The more likely explanation is that ESL users simply tend to weaken constraints that favor the *s*-genitive, and instead strengthen constraints that favor the *of*-genitive. In this connection, we note that the subtle probabilistic differences we have uncovered provide evidence for what Szmrecsanyi et al. (2016) term "probabilistic indigenization", i.e., a gradual shift in usage patterns in post-colonial varieties. Since we observe differences in the strength of constraints that influence the choice between the two genitive variants, we conclude that probabilistic genitive grammars are subtly but measurably distinct from each other as well as, in the case of postcolonial varieties, from their input varieties.

We reiterate that the direction of the effect of language-internal constraints is stable across varieties; so long constituents, for example, always follow—not precede—short

constituents, etc. There was one predictor in the model, however, whose effect changed direction across varieties: MEDIUM. Its positive coefficient 0.31 in Table 12 and the slightly elevated estimates in Figure 4 for the spoken condition of some varieties indicate a slight, however not statistically significant, tendency toward greater *s*-genitive usage in spoken texts. In Philippine and Hong Kong English, on the other hand, the spoken medium in fact *dis*favors *s*-genitive usage when compared to the written medium. We suspect that in Philippine and Hong Kong English, which are less advanced on the evolutionary path of Schneider's (2007) Dynamic Model, there is an even stronger need to go for the safer, more analytic *of*-genitive in spoken language, considering the online processing constraints that this medium entails. But be that as it may, it does of course stand to reason that the predictor that is most unstable in the analysis, MEDIUM, is one that is language-external in nature and thus defined culturally. In this light, the hypothesis that we should see instability primarily with predictors that are not strongly processing-driven is borne out by the data.

Lastly, it seems worth mentioning that those constraints we find to be subject to probabilistic indigenization across varieties—possessor animacy, final sibilancy, and possessum length—are well-known to be unstable in diachrony. As for possessor animacy, the literature reports that the *s*-genitive has come to be increasingly used with collective, locative, and temporal possessors in some written genres (Wolk et al. 2013; Ehret, Wolk & Szmrecsanyi 2014). Final sibilancy has been found to increasingly disfavor *s*-genitive use over time (Hinrichs & Szmrecsanyi 2007), and longer possessums are now, according to the literature, more likely to take the *s*-genitive than they used to (Wolk et al. 2013; Ehret, Wolk & Szmrecsanyi 2014). The bottom line is that as so often in work on language variation and change, synchrony and diachrony mirror each other.

**Conclusion**

The big question that this paper has sought to address concerns the extent to which the grammar of English is stable or fluid on a global scale. By tackling this issue, our work comes under the remit of research on English as a world language, but we add variationist/probabilistic rigor and new theoretical twists to this line of scholarship. By way of a case study, we investigated well-known syntactic variation between the *s*-genitive and the *of*-genitive in nine varieties of English. To establish the (in-)stability of constraints on this variation, we investigated a richly annotated dataset consisting of $N = 10{,}558$ interchangeable genitives from nine components of the International Corpus of English (ICE). Analysis showed that constraints such as possessor animacy, constituent length, final sibilancy of the possessor, as well as the effect of medium (spoken vs. written) are somewhat unstable in a cross-variety perspective. The language-internal constraints, however, never change effect direction. We are thus led to conclude that the probabilistic grammar regulating genitive variation in varieties of English is surprisingly stable overall. To the extent that we find fluidity and probabilistic indigenization, we see evidence for a distinction between ENL and ESL varieties: those constraints that tend to favor *s*-genitive usage tend to be weaker in ESL varieties than in ENL varieties.

As always, there are many ways in which the analysis reported in the present paper could be extended. For one thing, work is underway to annotate the dataset for additional language-internal constraints such as genitive relation, the definiteness of the possessor or noun phrase expression type, or persistence/priming. We are currently also working on extracting genitives from the Corpus of Global Web-based English (GloWbE) (Davies & Fuchs 2015), with the goal of adding web-based language to the array of text types sampled in ICE. As regards the type of evidence we consider, corpus analysis has taken center stage in the present paper, but we are

working on spot-checking the cognitive robustness of the corpus-derived probabilities via rating experiments along the lines of Bresnan and Ford (2010), who showed that language users' acceptability ("naturalness") intuitions about syntactic choices match probabilities as calculated in a corpus-based regression model.

Further research is encouraged to investigate predictions generated by our analysis. First, since we found possessor animacy to regulate genitive choice differently in ENL and ESL varieties, we predict that genitive use in EFL varieties (e.g., Edwards 2014) deviates from ENL genitive use even more than ESL varieties do. Second, we tentatively suggest that the strength of the possessor animacy constraint is proportional to how advanced a variety is according to Schneider's (2007) Dynamic Model. Third, we predict that since animacy is a well-known constraint in a number of other syntactic alternations (e.g., the English dative alternation, relativizer choice, the active-passive alternation, etc.), it influences these alternations more strongly in ENL varieties than in ESL varieties: in ESL varieties, recipient animacy supposedly favors the prepositional dative less strongly, animate referents are less rigidly referred to by "who," and animate objects are used less often in passive constructions. Last but not least, note that we took a strictly synchronic perspective in the present study. Research with a more explicitly diachronic twist based on longitudinal data sources is needed to investigate if the probabilistic preference for the *of*-genitive that we find in ESL varieties may (also) be due to historical patterns such that ESL varieties lag behind ENL varieties in the drift towards more *s*-genitive usage, for example thanks to a less enthusiastic expansion of the *s*-genitive to inanimate possessor categories.

**Notes**

1. The automatic classification was performed with a Perl script relying on annotations in the Manchester possessive -*s* database (Denison, Scott & Börjars 2010), and in the WordNet database (Fellbaum 1998).

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

3. Word lemmas were derived from Yasumasa Someya's lemma list, available at http://lexically.net/downloads/BNC_wordlists/e_lemma.txt.

4. N.B.: The coefficients in Table 12 have to be interpreted with the interactions in Table 13 in mind. Since there is also an interaction of POSSESSOR ANIMACY and VARIETY, the coefficient of POSSESSOR ANIMACY in Table 12 is conditional on VARIETY being at its baseline value (Jaccard 2001). What we see here is, therefore, the effect of POSSESSOR ANIMACY in British English only.

**TABLE 1**

Genitive Choice and Possessor Animacy

| Possessor animacy | *s*-genitives | *of*-genitives |
|---|---|---|
| inanimate | 743 (11.4 %) | 5780 (88.6 %) |
| animate | 1869 (46.0 %) | 2195 (54.0 %) |

**TABLE 2**

Genitive Choice and Final Sibilancy of Possessor

| Final sibilancy | *s*-genitives | *of*-genitives |
|---|---|---|
| absent | 2,331 (29.6 %) | 5,535 (70.4 %) |
| present | 279 (10.4 %) | 2,413 (89.6 %) |

**TABLE 3**

Genitive Choice and Givenness

| Possessor givenness | *s*-genitives | *of*-genitives |
|---|---|---|
| new | 1,531 (22.4 %) | 5,299 (77.6 %) |
| given | 1,079 (28.9 %) | 2,649 (71.1 %) |

**TABLE 4**

Genitive Choice and Thematicity of Possessor (Mean Frequency Per 2,000 Words of Running Text)

| Thematicity | *s*-genitives | *of*-genitives |
|-------------|---------------|----------------|
| mean | 10.8 | 9.1 |
| std. dev. | 12.8 | 12.5 |

**TABLE 5**

Genitive Choice and Overall Frequency of the Possessor Head (Frequency Per Million Words in the GloWbE Corpus)

| Head frequency | *s*-genitive | *of*-genitive |
|---|---|---|
| mean | 215.6 | 232.9 |
| std. dev. | 460.4 | 614.6 |

**TABLE 6**

Genitive Choice and Possessor Length (Number of Orthographic Characters)

| Possessor length | *s*-genitives | *of*-genitives |
|---|---|---|
| mean | 8.9 | 18.2 |
| std. dev. | 4.5 | 17.4 |

**TABLE 7**

Genitive Choice and Possessum Length (Number of Orthographic Characters)

| Possessum length | *s*-genitives | *of*-genitives |
|---|---|---|
| mean | 15.0 | 9.6 |
| std. dev. | 15.6 | 5.7 |

**TABLE 8**

Genitive Choice and Type-Token Ratio (Type-Token Ratio in the Immediate Environment of +/- 50 Words)

| Type-token ratio | *s*-genitives | *of*-genitives |
|---|---|---|
| mean | 0.82 | 0.81 |
| std. dev. | 0.07 | 0.07 |

**TABLE 9**

Genitive Choice and Medium

| Medium | *s*-genitives | *of*-genitives |
|---|---|---|
| written | 1,569 (25.3 %) | 4,632 (74.7 %) |
| spoken | 1,041 (23.9 %) | 3,316 (76.1 %) |

**TABLE 10**

Genitive Choice and Variety

| Variety | *s*-genitives | *of*-genitives |
| --- | --- | --- |
| Great Britain | 284 (23.3 %) | 937 (76.7 %) |
| Canada | 380 (34.7 %) | 714 (65.3 %) |
| Ireland | 264 (26.5 %) | 731 (73.5 %) |
| New Zealand | 327 (25.5 %) | 957 (74.5 %) |
| Singapore | 310 (30.2 %) | 717 (69.8 %) |
| Jamaica | 173 (16.3 %) | 889 (83.7 %) |
| Hong Kong | 318 (27 %) | 859 (73 %) |
| India | 247 (18.5 %) | 1,088 (81.5 %) |
| Philippines | 307 (22.5 %) | 1,056 (77.5 %) |
| Total | 2,610 (24.7 %) | 7,948 (75.3 %) |

**TABLE 11**

Variance Accounted for by Random Effects

| Group | Variance |
|---|---|
| Possessor head lemma | 3.86 |
| Possessum head lemma | 2.12 |
| Text category/Language user | 1.53 |

**TABLE 12**

Regression Coefficients and Odds Ratios of Individual Predictors in the Minimal Adequate

Model. Predictions are for the *S*-Genitive. The Baseline for the Factor VARIETY is British

English.

| Predictor | b | Odds ratio | p |
|---|---|---|---|
| (Intercept) | -4.84 | 0.01 | < 0.001 |
| POSSESSOR ANIMACY | | | |
| inanimate → animate | 4.31 | 74.19 | < 0.001 |
| POSSESSOR LENGTH (logged) | | | |
| one unit increase | -2.40 | 0.09 | < 0.001 |
| POSSESSUM LENGTH (logged) | | | |
| one unit increase | 0.55 | 1.74 | 0.026 |
| FINAL SIBILANCY | | | |
| absent → present | -1.04 | 0.35 | 0.002 |
| POSSESSOR GIVENNESS | | | |
| new → given | 0.25 | 1.29 | 0.017 |
| MEDIUM | | | |
| written → spoken | 0.31 | 1.36 | 0.446 |
| POSSESSOR HEAD FREQUENCY (logged) | | | |
| one unit increase | -0.11 | 0.90 | < 0.001 |
| POSSESSOR THEMATICITY (logged) | | | |
| one unit increase | 0.16 | 1.17 | 0.003 |
| TYPE-TOKEN RATIO | | | |
| one unit increase | 2.83 | 17.01 | < 0.001 |
| VARIETY | | | |
| Great Britain → Canada | 1.24 | 3.44 | 0.010 |
| Great Britain → Hong Kong | 1.52 | 4.55 | 0.002 |
| Great Britain → New Zealand | 1.30 | 3.67 | 0.007 |
| Great Britain → Philippines | 1.16 | 3.20 | 0.013 |
| Great Britain → Singapore | 1.02 | 2.76 | 0.034 |

**TABLE 13**

Regression Coefficients and Odds Ratios of Predictor Interactions in the Minimal Adequate Model. Predictions are for the *S*-Genitive. The Baseline for the Factor VARIETY is British English. (Note: Non-Significant Factor Levels are not Shown.)

| Interaction | b | Odds ratio | p |
|---|---|---|---|
| POSSESSOR ANIMACY : VARIETY | | | |
| animate + Hong Kong | -0.88 | 0.41 | 0.040 |
| animate + New Zealand | -0.88 | 0.41 | 0.034 |
| animate + Philippines | -1.65 | 0.19 | < 0.001 |
| POSSESSUM LENGTH (logged) : VARIETY | | | |
| one unit increase + Canada | 0.65 | 1.92 | 0.046 |
| one unit increase + Hong Kong | 0.71 | 2.03 | 0.034 |
| one unit increase + Ireland | 1.10 | 3.02 | 0.001 |
| one unit increase + Philippines | 0.73 | 2.08 | 0.017 |
| one unit increase + Singapore | 0.70 | 2.01 | 0.037 |
| FINAL SIBILANCY : VARIETY | | | |
| present + Hong Kong | -2.73 | 0.06 | < 0.001 |
| present + India | -1.31 | 0.27 | 0.022 |
| MEDIUM : VARIETY | | | |
| spoken + Hong Kong | -1.38 | 0.25 | 0.014 |
| spoken + Philippines | -1.83 | 0.16 | 0.001 |

**Figure 1:** Effects Plot of the Interaction Between VARIETY and POSSESSOR ANIMACY.

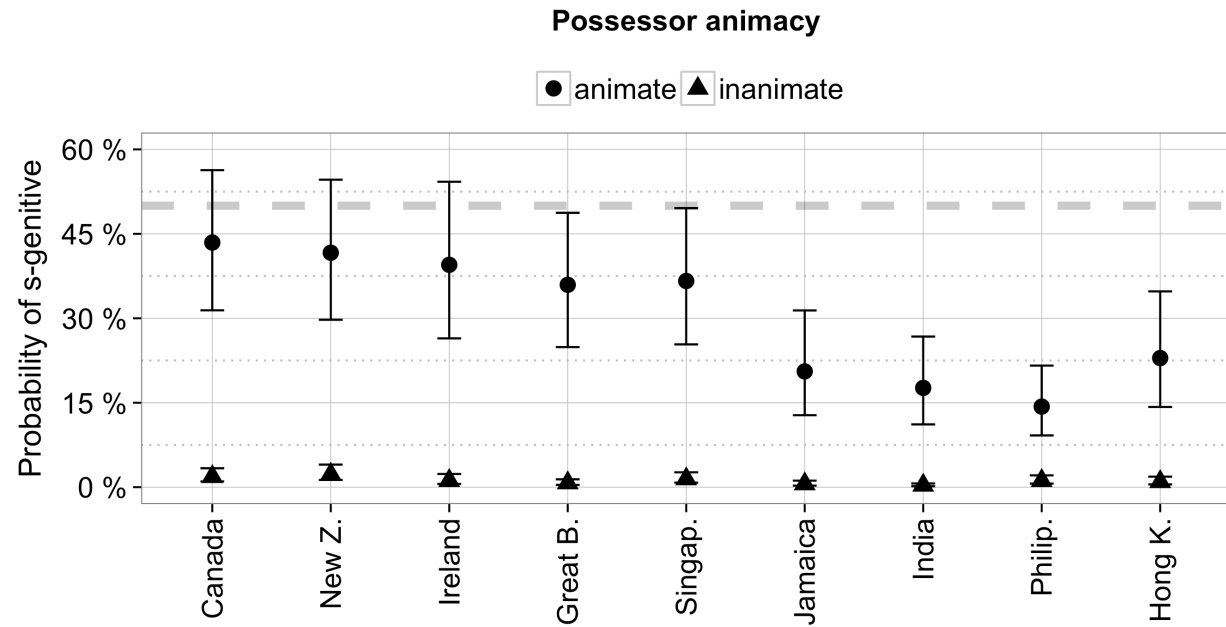Predictions are for the *S*-Genitive. The Dashed Horizontal Line Represents the 50 % Mark.

**Figure 2:** Effects Plot of the Interaction Between VARIETY and POSSESSUM LENGTH. Predictions are for the *S*-Genitive. The Dashed Horizontal Line Represents the 50 % Mark.
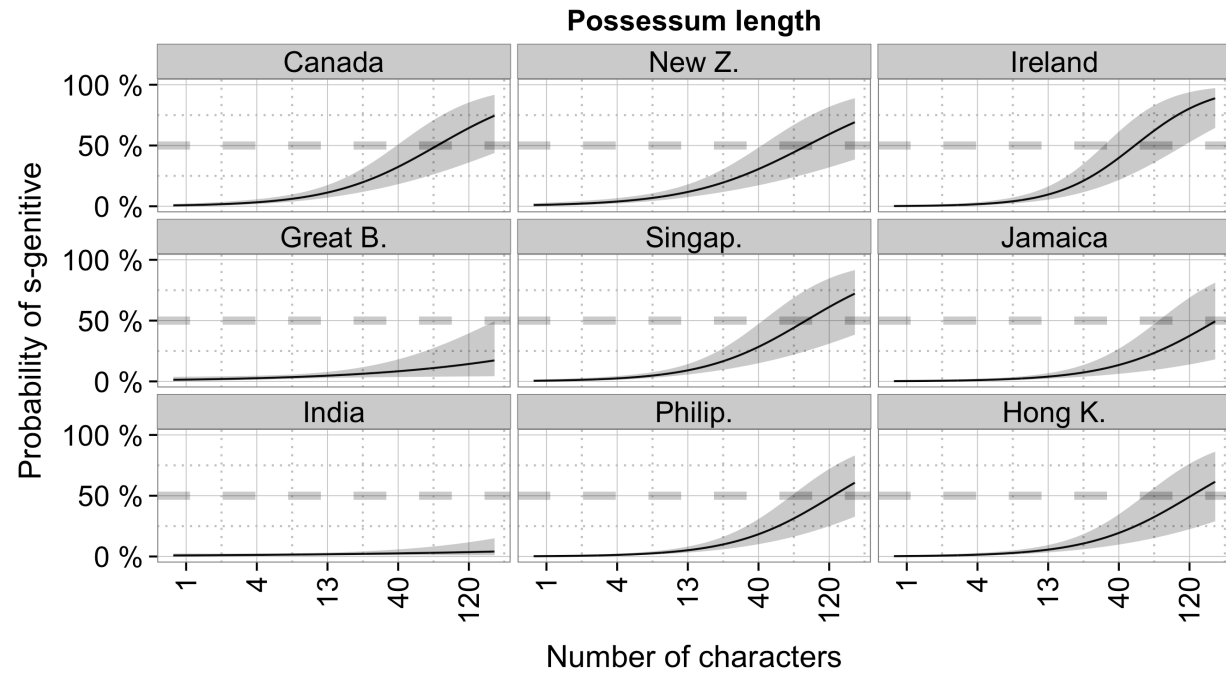
**Figure 3:** Effects Plot of the Interaction Between VARIETY and FINAL SIBILANCY. Predictions are for the *S*-Genitive. The Dashed Horizontal Line Represents the 50 % Mark.
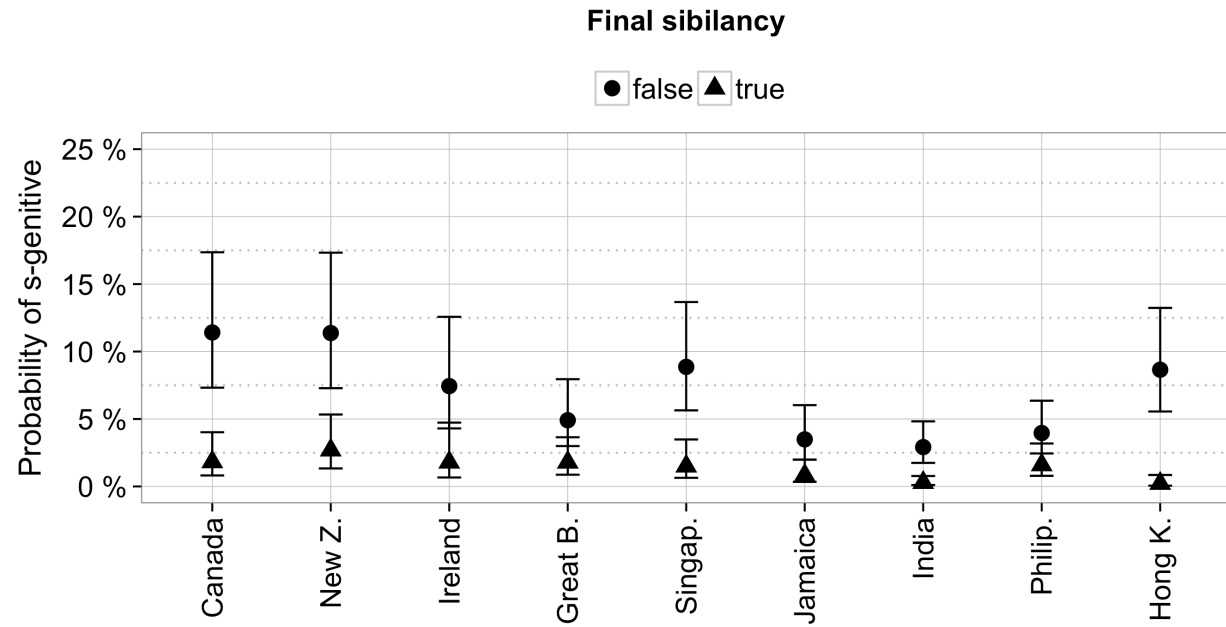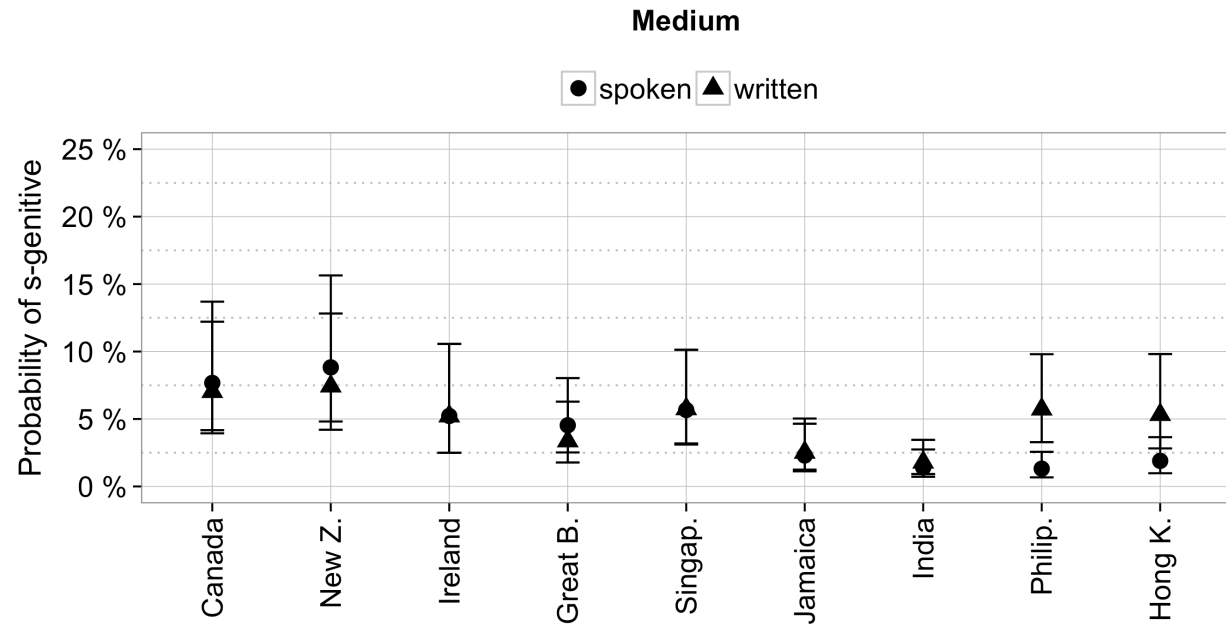
**Figure 4:** Effects Plot of the Interaction Between VARIETY and MEDIUM. Predictions are for the *S*-Genitive. The Dashed Horizontal Line Represents the 50 % Mark.

**Author Biographies**

- Benedikt Heller is a PhD student at KU Leuven (research group "Quantitative Lexicology and Variational Linguistics") and investigates the genitive alternation in varieties of English. He obtained his master's degree in 2013 from the University of Giessen (Germany).

- Benedikt Szmrecsanyi holds a research professorship at the Department of Linguistics of KU Leuven (research group "Quantitative Lexicology and Variational Linguistics"). His research interests focus on variation studies (synchronic & diachronic), probabilistic grammar, language complexity, geolinguistics, and dialect typology. Recent books include *Grammatical Variation in British English Dialects* (2013, CUP), and *Aggregating dialectology, typology, and register analysis* (2014, ed. with Bernhard Wälchli, de Gruyter).

- Jason Grafmiller is a post-doctoral researcher at KU Leuven (research group "Quantitative Lexicology and Variational Linguistics"). He obtained his PhD in 2013 from Stanford University.

# References

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, New York: Cambridge University Press.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2014. *lme4: Linear mixed-effects models using Eigen and S4*. http://CRAN.R-project.org/package=lme4.

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

Bentz, Christian & Bodo Winter. 2013. Languages with More Second Language Learners Tend to Lose Nominal Case. *Language Dynamics and Change*. 1–27.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1). 168–213. doi:10.1353/lan.0.0189.

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. doi:10.1016/j.lingua.2007.02.007.

Bybee, Joan. 2006. From Usage to Grammar: The Mind's Response to Repetition. *Language* 82(4). 711–733.

Bybee, Joan & Paul Hopper. 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins.

Crystal, David. 2004. *The Cambridge Encyclopedia of the English Language*. Second Edition. Cambridge: Cambridge University Press.

Davies, Mark & Robert Fuchs. 2015. Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide* 36(1). 1–28. doi:10.1075/eww.36.1.01dav.

Denison, David, Alan K. Scott & Kersti Börjars. 2010. The real distribution of the English "group genitive." *Studies in Language* 34(3). 532–564. doi:10.1075/sl.34.3.02den.

Edwards, Alison. 2014. English in the Netherlands. Functions, forms and attitudes. University of Cambridge Ph.D. Thesis.

Ehret, Katharina, Christoph Wolk & Benedikt Szmrecsanyi. 2014. Quirky quadratures: on rhythm and weight as constraints on genitive variation in an unconventional data set. *English Language and Linguistics* 18(02). 263–303. doi:10.1017/S1360674314000033.

Fellbaum, Christiane. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Fox, John. 2003. Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software* 8(15).

Gahl, Susanne & Susan Garnsey. 2004. Knowledge of Grammar, Knowledge of Usage: Syntactic Probabilities Affect Pronunciation Variation. *Language* 80. 748–775.

Gelman, Andrew & Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. (Analytical Methods for Social Research). Cambridge ; New York: Cambridge University Press.

Grafmiller, Jason. 2014. Variation in English genitives across modality and genres. *English Language and Linguistics* 18(03). 471–496. doi:10.1017/S1360674314000136.

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. 2015. The most under-used method in corpus linguistics: multi-level (and mixed-effects) models. *Corpora* 10(1). 95–125. doi:10.3366/cor.2015.0068.

Gries, Stefan Th. & Sandra C. Deshors. 2015. EFL and/vs. ESL?: A multi-level regression modeling perspective on bridging the paradigm gap. *International Journal of Learner Corpus Research* 1(1). 130–159. doi:10.1075/ijlcr.1.1.05gri.

Hickey, Raymond. 2010. The Englishes of Ireland: Emergence and transportation. In Andy Kirkpatrick (ed.), *The Routledge Handbook of World Englishes*, 76–95. London and New York: Routledge.

Hilpert, Martin. 2008. The English comparative – language structure and language use. *English Language and Linguistics* 12(03). 395. doi:10.1017/S1360674308002694.

Hinrichs, Lars. 2006. *Codeswitching on the Web. English and Jamaican Creole in e-mail communication*. Amsterdam, Philadelphia: John Benjamins.

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. 437–474. doi:10.1017/S1360674307002341.

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

Jaccard, James. 2001. *Interaction Effects in Logistic Regression*. Sage.

Jankowski, Bridget L. & Sali A. Tagliamonte. 2014. On the genitive's trail: data and method from a sociolinguistic perspective. *English Language and Linguistics* 18(02). 305–329. doi:10.1017/S1360674314000045.

Jankowski, Bridget Lynn. 2013. A variationist approach to cross-register language variation and

change. University of Toronto. https://tspace.library.utoronto.ca/handle/1807/43604 (20 November, 2015).

Kachru, Braj B. 1985. Standards, codification and sociolinguistic realism: the English language in the outer circle. In Randolph Quirk & Henry G. Widdowson (eds.), *English in the World: Teaching and Learning the Language and Literatures*, 11–30. Cambridge: Cambridge University Press.

Klein, Wolfgang & Clive Perdue. 1997. The basic variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13. 301–347.

Kortmann, Bernd & Kerstin Lunkenheimer (eds.). 2012. *The Mouton World Atlas of Variation in English*. Berlin: Mouton de Gruyter.

Kortmann, Bernd & Benedikt Szmrecsanyi. 2004. Global synopsis: morphological and syntactic variation in English. In Bernd Kortmann, Edgar Schneider, K. Burridge, R. Mesthrie & C. Upton (eds.), *A Handbook of Varieties of English*, vol. 2, 1142–1202. Berlin/New York: Mouton de Gruyter.

Kortmann, Bernd & Christoph Wolk. 2012. Morphosyntactic variation in the anglophone world: A global perspective. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The Mouton World Atlas of Variation in English*. Berlin, Boston: DE GRUYTER. http://www.degruyter.com/view/books/9783110280128/9783110280128.906/9783110280128.906.xml (13 February, 2016).

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.

MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4. 1–16. doi:10.3389/fpsyg.2013.00226.

MacWhinney, Brian. 1997. Second Language Acquisition and the Competition Model. In Anette

M.B. De Groot & Judit F. Kroll (eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, 113–142. Mahawa, NJ: Lawrence Erlbaum Associates.

Murphy, Raymond. 2012. *English Grammar in Use. A self-study reference and practice book for intermetdiate learners of English.* 4th ed. Cambridge University Press.

Nelson, Gerald. 1996. The Design of the Corpus. In Sidney Greenbaum (ed.), *Comparing English Worldwide: The International Corpus of English*, 3–12. Oxford: Clarendon.

Osselton, Noel. 1988. Thematic Genitives. In Graham Nixon & John Honey (eds.), *An Historic Tongue: studies in English linguistics in memory of Barbara Strang*, 138–144. London: Routledge.

Pakir, Anne. 2001. The voices of English-knowing bilinguals and the emergence of new epicentres. In Vincent B. Y. Ooi (ed.), *Evolving Identities: The English Language in Singapore and Malaysia*, 1–11. Times Academic Press.

Patrick, Peter L. 2008. Jamaican Creole: morphology and syntax. *Varieties of English 2. The Americas and the Caribbean*, 609–644. Berlin, New York: Mouton de Gruyter.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

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

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

Rosenbach, Anette. 2014. English genitive variation – the state of the art. *English Language and Linguistics* 18(02). 215–262. doi:10.1017/S1360674314000021.

Schneider, Edgar. 2007. *Postcolonial English: Varieties Around the World*. Cambridge:

Cambridge University Press.

Shih, Stephanie, Jason Grafmiller, Richard Futrell & Joan Bresnan. 2015. Rhythm's role in genitive construction choice in spoken English. In Ralf Vogel & Ruben Vijver (eds.), *Rhythm in Cognition and Grammar*. Berlin, München, Boston: DE GRUYTER. http://www.degruyter.com/view/books/9783110378092/9783110378092.207/978311037 8092.207.xml (17 September, 2015).

Szmalec, Arnaud, Marc Brysbaert & Wouter Duyck. 2012. Working memory and (second) language processing. In Altarriba J & Isurin L (eds.), *Memory, language, and bilingualism: Theoretical and applied approaches*, 74–94. Cambridge: CUP.

Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations: modeling syntactic variation in varieties of English. *English World-Wide* 37(2).

Szmrecsanyi, Benedikt & Bernd Kortmann. 2009. The morphosyntax of varieties of English worldwide: A quantitative perspective. *Lingua* 119(11). 1643–1663. doi:10.1016/j.lingua.2007.09.016.

Tagliamonte, Sali & Harald Baayen. 2012. Models, forests and trees of York English: "Was/were" variation as a case study for statistical practice. *Language Variation and Change* 24. 135–178.

The ICE Project. 2009. The Design of ICE Corpora @ ICE-corpora.net. *International Corpus of English*. http://ice-corpora.net/ice/design.htm (20 June, 2016).

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

Trudgill, Peter. 2009. Vernacular universals and the sociolinguistic typology of English dialects.

In Marrku Filppula, Juhani Klemola & Heli Paulasto (eds.), *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*, 302–329. London: Routledge.

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

Wolk, Christoph, Joan Bresnan, Anette Rosenbach & Benedikt Szmrecsanyi. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica* 30(3). 382–419. doi:10.1075/dia.30.3.04wol.

Zuur, Alain F., Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer.

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