Stability and fluidity in syntactic variation world-wide

Heller, Benedikt; Szmrecsanyi, Benedikt; Grafmiller, Jason

DOI:
10.1177/0075424216685405

Document Version
Peer reviewed version

Citation for published version (Harvard):
Heller, B, Szmrecsanyi, B & Grafmiller, J 2017, 'Stability and fluidity in syntactic variation world-wide: The genitive alternation across varieties of English', *Journal of English Linguistics*, vol. 45, no. 1, pp. 3-27. https://doi.org/10.1177/0075424216685405

Journal of
English Linguistics

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

Journal:	*Journal of English Linguistics*
Manuscript ID	Draft
Manuscript Type:	Original Manuscript
Keywords:	World Englishes < Language Varieties, Genitive Alternation
Abstract:	How uniform is the grammar of English on a global scale? We investigate well-known syntactic variation between the s-genitive (Mr Barnsley's management) and the of-genitive (the management of Mr Barnsley) in international 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 that supports a Kachru-inspired distinction between Inner Circle and Outer Circle varieties: those constraints that tend to favor s-genitive usage tend to be weakened in Outer Circle varieties.

Abstract

How uniform is the grammar of English on a global scale? We investigate well-known syntactic variation between the *s*-genitive (*Mr Barnsley's management*) and the *of*-genitive (*the management of Mr Barnsley*) in international 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 that supports a Kachru-inspired distinction between Inner Circle and Outer Circle varieties: those constraints that tend to favor *s*-genitive usage tend to be weakened in Outer Circle varieties.

1 Introduction

Comparative perspectives on the grammar of worldwide varieties of English and the scope of grammatical differences between varieties have both been quite popular topics in the more recent literature (e.g. Gries & Deshors 2015; Kortmann & Wolk 2012; Szmrecsanyi & Kortmann 2009). We contribute to this discussion 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, which precedes the possessum in *of*-genitives. In the spirit of a number of recent studies of English genitive constructions (Ehret, Wolk & Szmrecsanyi 2014; Grafmiller 2014; Hundt & Szmrecsanyi 2012; Shih et al. 2015; Szmrecsanyi et al. 2016), we adopt a variationist perspective (Labov 1972) and restrict attention to genitive tokens where the alternative variant could also have been used. Thus, for example, genitive phrases as in *You know most of the lecturers* (ICE-SIN, s1a-001) are not considered, as the periphrasis with the *s*-genitive is not possible (**the lecturers' most*).

Genitive variation has been extremely well studied, of course, and the relevant literature 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 (more animate possessors favor the *s*-genitive), constituent length (thanks to the principle of end-weight, long possessors favor the *of*-genitive and long possessums the *s*-genitive), and final sibilancy (a final sibilant discourages the *s*-genitive). Crucially, however, none of these factors is deterministic: for example, animate possessors do favor the *s*-genitive, but if the possessum 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 research designs (Rosenbach 2005).

A number of studies have uncovered interesting interactions between language-internal constraints (e.g. animacy, weight, 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 we note that analysts have tended to restrict attention to the difference 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. Kachru's Three Circles model (Kachru 1985) categorizes varieties of English into an *Inner Circle*, an *Outer Circle*, and an *Expanding Circle*, the first two of which are important for this study. The Inner Circle contains those countries where English is the primary language (e.g. Great Britain and Canada). Outer Circle countries, on the other hand, have a colonial history; they were colonized by English-speaking settlers and in the process, English became important alongside other languages, as in India and Singapore, for example. In contrast to previous research, this paper offers an analysis that is not

Inner Circle-centric, but that also considers a number of Outer Circle varieties (specifically, Jamaican English, Singapore English, Indian English, Philippines English, and Hong Kong English). The Inner Circle varieties we investigate are British English, Irish English, Canadian English, and New Zealand 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 (Kachru 1985; Schneider 2007), and for that matter also not in characteristics of particular varieties of English. Rather, we will be 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:

1. To what extent do users of international varieties of English rely on the same or similar choice-making processes when it comes to choosing genitive variants?
2. Are cross-varietal differences random or can they be explained by variety type (e.g. Inner Circle versus Outer Circle) and/or sociohistorical factors?

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 (Bresnan 2007; 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 (Bresnan 2007). (iii) This probabilistic knowledge is derived in large part from language experience, and so is subtly—but fluidly—(re)constructed throughout speakers' lives.

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 data 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 across the board, given their presumably strong links to the design of the human speech production system. But at the same time it is likely – thanks to the experience-based nature of cognitive organization, and the different input(s) that users of different varieties are likely to receive over their lifetime – 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 expect to see probabilistic indigenization effects not primarily with regard to constraints that are strongly processing-driven (such as, presumably, end-weight), but more with constraints such as possessor animacy, or register/medium differences (register conventions are, after all, social conventions).

Given 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 logistic regression model with mixed effects fit to this dataset (dependent variable: genitive choice) correctly predicts 94 % of all genitive outcomes 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), for example; this is another way of saying that users of English wherever it is spoken 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 supports a Kachru-inspired distinction between Inner Circle and Outer Circle varieties: those constraints that tend to favor *s*-genitive usage tend to be downplayed in Outer Circle varieties, and those that favor *of*-genitive usage tend to be strengthened. We argue that this is an echo of the well-known SLA tendency to prefer analytic marking.

The remainder of this paper is structured as follows. Section 2 describes the data source,

the variable context, the constraints on genitive variation that we consider, and our methods. Section 3 presents the results. In Section 4 we discuss our findings, and Section 5 offers some concluding remarks.

2 Data and methods

2.1 Corpus data

The data for this study were extracted from ICE, the International Corpus of English (Greenbaum 1996). This corpus family is well suited for cross-varietal comparisons since all subcorpora follow the same design, which contains 600,000 words of spoken and 400,000 words of written language per corpus. These components are further subdivided into a wide variety of text types that were designed to offer a balanced reflection of the varieties (Nelson 1996). The present analysis explores nine subcorpora from the ICE family, which reflect the language use of people in (i) Great Britain, (ii) Ireland, (iii) Canada, (iv) New Zealand, (v) Jamaica, (vi) Singapore, (vii) India, (viii) the Philippines, and (ix) Hong Kong. Kachru's three circle model divides this selection into four Inner Circle varieties, (i)–(iv), and five Outer Circle varieties, (v)–(ix). From these nine corpora, a sample of 10,558 genitives was taken that proportionately reflects all text types in ICE. All of these genitives are interchangeable in terms of the variable context described below.

2.2 Defining the variable context

Following guidelines in Rosenbach (2002; 2014) and Wolk et al. (2013), interchangeable genitives were defined as the total number of genitives minus the categorical cases. In line with previous studies on the genitive alternation (Hinrichs & Szmrecsanyi 2007; Wolk et al. 2013), the following categorical contexts were excluded from our data: appositive genitives, classifying

genitives, double genitives, idiomatic/fixed genitives, and partitive genitives. Further, only cases with definite possessums were sampled.

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 over to Northern Ireland). Classifying genitives do not specify a possessum as a specific entity but express to which class it belongs. Example (6) shows the difference. 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). Idiomatic genitives are fixed patterns that only occur in one form, e.g. (3). Finally, partitive genitives express a measurement of some sort (e.g. time, distance, or value, see Biber et al. (1999) as in (8). 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 (9) were thus excluded. Bare plurals, like *Infections of the CNS* in (9) were also excluded for the same reasons.

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

(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)

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

(8) 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)

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

2.3 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 correction to ensure maximum precision and recall. 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, mode, and variety.

Animacy

Animacy of the possessor is arguably the most important constraint in the genitive alternation (Grafmiller 2014; Hinrichs & Szmrecsanyi 2007; Rosenbach 2005). It is so significant that it is used in many prescriptive grammars of English (e.g. Murphy 2012) as a rule on how to use English genitives. The rule dictates that if the possessor of a genitive construction is animate, the

s-genitive is to be used. 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 (Hundt & Szmrecsanyi 2012)

Table 1: Genitive choice an possessor animacy

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

The annotation of animacy was performed semi-automatically. Every possessor head's animacy status was automatically classified and manually corrected where necessary. After initially using a five-fold classification following guidelines in Zaenen et al. (2004), the classification was restricted to a binary distinction between animate and inanimate entities due to high collinearity in the resulting regression model ($\kappa = 28.1$), which came close to the "potentially harmful" threshold of 30 (Baayen 2008:182). 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.

Final sibilancy of the possessor

If a sibilant (i.e. one of the sounds in 11) is present at the end of the possessor phrase, as in (12), *s*-genitive usage is less likely for articulatory reasons (Zwicky 1987).

(11) [s], [z], [ʃ], [tʃ], [ʒ], [dʒ]

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

Table 2 shows that in the dataset under investigation, *s*-genitives occur only in about 10 % of the

cases where the possessor phrase ends in a sibilant as opposed to about 30 % if there is no final sibilant. Final sibilancy was annotated automatically by referring to the CMU Pronunciation Dictionary[1]. If a word could not be found, the annotation relied on orthography.

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 %)

Givenness of the possessor

Givenness captures whether or not information has been mentioned before. For this study, we coded a possessor as *given* if the lemma of the possessor head can be found in the previous context of the genitive construction, otherwise it was marked as *new*. Many previous studies have shown that *given* possessor heads favor *s*-genitive use, but not always significantly so (Hinrichs & Szmrecsanyi 2007; Grafmiller 2014). Other studies have referred to the same concept as, e.g., information status (Jankowski 2013) or discourse accessibility (Gries & Bernaisch 2015). The annotation of givenness was performed automatically by a script that searched the previous context of each genitive case and used a lemma list[2] to determine if the

[1]Available at: http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

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

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 %) than discourse-new possessors (22.4 %).

Thematicity of the possessor

Thematicity 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 (Osselton 1988). The total number of mentions was counted automatically and then normalized by the usual corpus text size of 2,000 words. In our data, possessor heads in *s*-genitives are more thematic than possessor heads in *of*-genitives (Table 4).

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

Overall frequency of the possessor

To our knowledge, the overall frequency of the possessor head has not been included in multivariate analyses of the genitive alternation so far, but has been shown to significantly

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, and similarly so for the other varieties (Table 5).

Length of the constituents

Syntactic weight is predicted to influence genitive choice along the lines of the principle of end-weight, which was first described as "das Gesetz der wachsenden Glieder" by Behaghel (1909: 139) and is defined as "the tendency for long and complex elements to be placed towards the end of a clause" (Biber et al. 1999). Therefore, the longer the possessor, the higher the probability of an *of*-genitive since possessors are placed last in *of*-genitives. Different ways of measuring syntactic weight (e.g. number of characters, number of words, number of syntactic nodes) correlate very highly (Rosenbach 2014:227); we choose to establish the number of orthographic characters of the two genitive constituents, following Wolk et al. (2013).

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

(14) the power of the Chinese Government (ICE-HK, s1a-021)

(15) laser's potential medical uses (ICE-IND, w2b-031)

Examples (14) and (15), in which the longer constituents are placed last, illustrate the principle of end-weight. That their respective counterparts—*the Chinese Government's power* and *the potential medical uses of laser*—are usually less likely, is well documented (Rosenbach 2002; Rosenbach 2005; Wolk et al. 2013). To determine the length of the constituents, a simple regular expression search counted the number of characters in the possessor and possessum phrases. Whitespaces and special characters (e.g. hyphens) were not counted. Sometimes, end-weight effects are measured by means of weight differences or ratios; the present study, however, follows studies such as Hinrichs & Szmrecsanyi (2007) that use two separate length measures for the possessor and the possessum. Table 6 and Table 7 show that possessors tend to be shorter in *s*-genitives and longer in *of*-genitives. Possessums show a complementary pattern, being on average longer in *s*-genitives and shorter in *of*-genitives.

Lexical density

Lexical density, here measured in 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 100-word environment of the genitive instance in question. If instance token was located in the beginning of a text, and therefore there

were less than 50 words of previous context available, the number of words taken from the following context was increased until a total of 100 words was reached. If instance token was located towards the end, words were added from the preceding context. Table (8) shows that the average type-token ratio is slightly higher in *s*-genitives than in *of*-genitives.

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 modality

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

Language-external factors

The literature reports effects that language-external factors (such as mode, time, 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 modality (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 we aggregate over varieties. As far as variety differences are concerned, Table 10 shows that while the *s*-genitive is used in 1/4 of the cases on average (24.7 %), it ranges from slightly below 1/6 in Jamaican English (16.3 %) to slightly above 1/3 in Canadian English (34.7 %).

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 %)

2.4 Data analysis

In order to statistically model genitive choice in varieties of English, we utilized mixed-effects logistic regression modeling. Random effects can be used to capture the influence of predictors that are particular to the sample, i.e. not repeatable (Baayen 2008: 241), and that are not of primary interest to the analysis (Gelman & Hill 2007), which makes the estimation of the fixed effects more accurate (Gries 2015). In this analysis, random effects were implemented to account for idiosyncrasies of individual language users, genres (i.e. monologue, dialogue, printed, and

non-printed), and the possessor and possessum head noun lemmas. 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 predictors and their coefficients were validated using a bootstrapping approach as outlined in Baayen (2008: 283), which showed that none of the confidence intervals that were obtained for the coefficients included zero, which underscored our model's validity. All calculations were performed with R (R Core Team 2015) and the mixed-effects model was fit using the lme4 package (Bates et al. 2014).

The minimal adequate regression model classifies 94.14 % of all cases correctly (baseline: 75.32 % *of* -genitives). Measures that are independent of baseline accuracy also show high values (C = 0.981, Dxy = 0.961). The model has medium collinearity (κ = 18.06), which is still below the "potentially harmful" threshold of 30 (Baayen 2008: 182).

Table 11: Variance accounted for by random effects

Group	Variance
Possessor head lemma	3.86
Possessum head lemma	2.12
Genre/Language user	1.53

Table 11 shows that the head noun lemmas of the constituents received the heaviest adjustments. For the possessor heads, high negative adjustments occur with entities that are animate but often used in the *of*-genitive (e.g. *God* as in *the kingdom of God*, and *staff* as in *member of staff*). High positive adjustments were made for temporal and locative heads (e.g. *today* as in *today's market*, and *Jamaica* as in *Jamaica's situation*). In the case of the possessum

heads it is job titles (e.g. *director* as in *director of administration*, and *president* as in *president of committee*) that favor the *of*-genitive, and it is often words that are linked to a prototypical relation of possession that received positive adjustments (e.g. *hat* as in *woman's hat*, and *house* as in *friend's house*). Adjustments for language user idiosyncrasies show less variance, but were still significant.

3 Results

Table 12 shows the regression coefficients of the individual predictors in presence of higher-order interactions, which are displayed in Table 13. The coefficients reflect changes in *s*-genitive likelihood caused by changes in the predictors. Column *b* shows these coefficients on a log odds scale. If *b* is positive (e.g. when POSSESSOR ANIMACY changes from *inanimate* to *animate*), the level change in question makes *s*-genitives more likely; if it is negative, *s*-genitive use is less likely. The "Odds ratio" column shows similar information, but the threshold that distinguishes between increased and decreased *s*-genitive probability is 1. If the odds ratio is greater than 1, *s*-genitives are more likely; if it smaller than 1, it is 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 0.36, i.e. makes *s*-genitive usage less likely.

We can see that possessor animacy makes a huge difference for genitive choice, with animate possessors increasing *s*-genitive odds by a factor of 74, vis-à-vis inanimate possessors.[3]

[3]These coefficients have to be interpreted with the interactions in mind. Since there is also an interaction of POSSESSOR ANIMACY and VARIETY, the coefficient of POSSESSOR ANIMACY 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.

Possessor length is also highly significant, and every increase in number of characters makes *s*-genitives less likely. Possessum length affects genitive choice in the opposite direction: for every one-unit increase 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*-genitives. Mode, as a main effect, does not make a significant difference in the model but was not excluded since it is part of a higher-order interaction (see Table 13 below). The remaining numerical predictors of 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 12: Regression coefficients of individual predictors in the minimal adequate model. Predictions are for the *s*-genitive. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

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				
false → true	-1.04	0.35	0.002	**
POSSESSOR GIVENNESS				
new → given	0.25	1.29	0.017	*
MODE				
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 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. FINAL SIBILANCY also shows stronger effects in other varieties, significantly so in Hong Kong and Indian English. Mode 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 differences in Philippine and Hong Kong English. For the interactions between MODE and VARIETY we can also observe a difference between Inner Circle and Outer Circle varieties: coefficients for Inner Circle varieties are rather small, whereas Outer Circle varieties receive higher adjustments.

Interaction terms betweem VARIETY and the following other predictors were not significant at all: POSSESSUM ANIMACY, POSSESSOR LENGTH, GIVENNESS, POSSESSOR HEAD FREQUENCY, THEMATICITY, and TYPE-TOKEN RATIO.

Table 13: Regression coefficients of predictor interactions in the minimal adequate model. Predictions are for the s-genitive. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. 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				
true + Hong Kong	-2.73	0.06	< 0.001	***
true + India	-1.31	0.27	0.022	*
MODE : VARIETY				
spoken + Hong Kong	-1.38	0.25	0.014	*
spoken + Philippines	-1.83	0.16	0.001	**

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. For interactions, they visually summarize information from the intercept, the coefficients of the two predictors involved, and their interaction term while holding all other coefficients in the model at their respective average values (see Fox 2003 for a detailed explanation). Thus the effects plots in Figure 1 can be seen to reflect how language users of different varieties are expected to deal with typical genitive constructions that only differ in the levels of the predictors involved in the respective interaction.

The effects plots provide a visual representation of the significant interactions in the regression model, which not only indicate deviation from the British English baseline, but also give a sense of the relation between other varieties (e.g. Singapore English and Hong Kong English). Figure 1a 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 Inner Circle varieties and Singapore English. Estimates for the other Outer Circle varieties are considerably lower. Figure 1b 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 1c 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 *true* and *false* are closest. Finally, Figure 1d shows the interaction of variety and mode. *S*-genitive predictions are slightly elevated in spoken Canadian, New Zealand, and British English, while mode does not seem to make a difference in Irish and Singapore English. In Jamaican and Indian English, we see very slight differences in the opposite direction. In Philippine and Hong Kong English, predictions for *s*-genitive use are clearly higher in written texts.

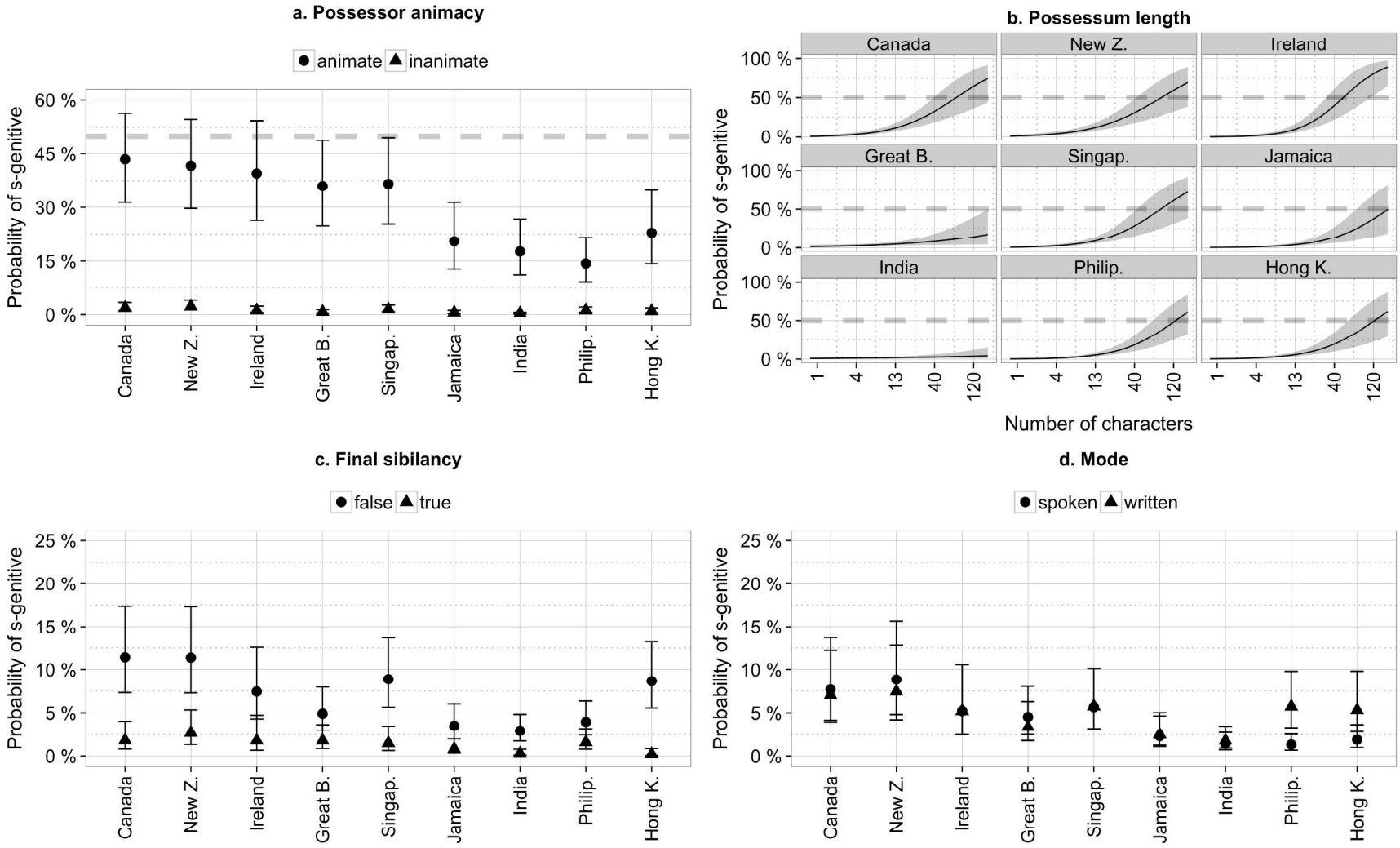

Figure 1: Effects plots of significant interactions with VARIETY. Predictions are for the *s*-genitive. Dashed horizontal lines indicate the 50 % mark.

4 Discussion

Our analysis of patterns of variation between the *s*-genitive and the *of*-genitive was guided by two questions: First, what is the extent to which users of international 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 (e.g. Inner Circle versus Outer Circle) and/or sociohistorical factors? With regard to the first question, we saw 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 support a Kachru-inspired distinction between Inner Circle and Outer Circle varieties.

Let us consider first the fact that genitive frequencies (see Table 10) are lower in some Outer Circle varieties (Jamaican, Indian, and Philippine English) than in the Inner Circle varieties we study (British, New Zealand, Irish, and Canadian English). It turns out that similar to the Inner Circle varieties, Singapore and Hong Kong English, which are Outer Circle, also show high *s*-genitive rates, a finding that is robust under multivariate control. 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 have been shown to lose case marking (Bentz & Winter 2013), we assume that for language users of most Outer Circle varieties, the *of*-genitive is, thanks to its analyticity, the more entrenched and thus safer option. It remains unclear why the *s*-genitive is comparatively frequent in Singapore and Hong Kong English. Inspired by Brunner (2014), who found evidence that noun phrase modification patterns (including genitives) in Singapore English are influenced

by indigenous languages, we speculate that the higher *s*-genitive ratio in these varieties might be due to an influence of Mandarin and Cantonese Chinese, which both use an *s*-genitive-like possessor-possessum order to express possession (Li & Thompson 1981:24; Matthews & Yip 1994:107) .

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 place constituents that are more easily retrievable earlier for the sake of optimizing utterance planning. We indeed 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: the overall frequency of the possessor head, which—according to Easy First—should favor highly frequent possessors in *s*-genitives since highly frequent possessors are easier to recall from memory (MacDonald 2013:3), was found to discourage *s*-genitive use with high values. We are likely to see here interference effects with noun phrase expression type: proper names, as in (16), which are low in frequency, are mostly realized as *s*-genitives.

(16) 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 predictors 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. Hence language users, whatever their regional and/or cultural

background, chose genitives in similar ways, and thus 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 (possessor animacy, possessum length, and final sibilancy) do not affect genitive choice in different directions across varieties (e.g. animate possessors favor in 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 Outer Circle 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 native L1 language (MacWhinney 1997; Szmalec, Brysbaert & Duyck:89). If this were true, we would have found that language users of Outer Circle varieties more often choose *s*-genitives, e.g., with animate possessors. What we do in fact find, however, is that Outer Circle 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 1a. Additional regression analysis, for which the nine levels of the factor VARIETY were conflated into two levels of the factor KACHRU CIRLCE (*inner* vs. *outer*), shows that the effect of possessor animacy is significantly weaker in Outer Circle varieties than in Inner Circle varieties (see Table 14 in the appendix). An Easy

First-inspired explanation at any cost would have to suggest that in Outer Circle varieties, animate possessors are not "easier" than inanimate possessors to the same degree as in Inner Circle 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 Outer Circle users simply tend to weaken constraints that favor the *s*-genitive, and strengthen constraints the *of*-genitive.

We reiterate that the direction of the effect of language-internal constraints is stable across varieties; so, for example, long constituents always follow—not precede—short constituents, and so on. There was one predictor in the model, however, whose effect changed sign across varieties: MODE. It turns out that while spoken texts favor *s*-genitive usage in three varieties (Canadian, New Zealand, and British English), the spoken mode in fact slightly *dis*favors *s*-genitive usage in Indian English and, particularly, in Philippine and Hong Kong English (Figure 1d) (MODE is not significant in Irish, Singapore, and Jamaican English). We suspect that in Indian, 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 mode entails. But be that as it may, it does of course stand to reason that the predictor that is most unstable in the analysis, MODE, 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.

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. How big are these differences in an aggregate perspective? In an effort to offer an admittedly coarse-grained but nonetheless, we believe, informative quantification of the "probabilistic distance", as it were, between British English and the other varieties, we added up the absolute values of all regression coefficients associated with the respective varieties in our model, and exercise which yielded the following distances between British English and the varieties in question:[4]

Jamaican English (2.0)

Indian English (2.5)

Irish English (2.8)

New Zealand English (3.3)

Singapore English (3.5)

Canadian English (3.6)

Philippines English (5.5)

Hong Kong English (7.2)

Ignoring the outlier Hong Kong English, this ranking supports a distinction between varieties that are, for reasons of their colonial histories, oriented toward British English, and those that are rather oriented toward American English. The Englishes of Jamaica, India, Ireland, and New Zealand exhibit the smallest distances to British English. Singapore, which also used to be a British colony but which is increasingly oriented towards the US (Schneider 1999:196–197),has

[4]This includes the respective main effects of the factor VARIETY (see Table 12), and the interaction terms (see Table 14).

an intermediate distance to British English. Canadian English, which is closely related to American English, and Philippine English, a variety sparked by American colonization, are most distant from British English.

Lastly, it seems worth mentioning that those constraints – possessor animacy, final sibilancy, and possessum length – that we find to be subject to probabilistic indigenization across varieties 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; Szmrecsanyi et al. 2014; 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 mirrors diachrony and vice versa.

5 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 international 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). The regression analysis indicated

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 Kachru-inspired distinction between Inner Circle and Outer Circle varieties: those constraints that tend to favor *s*-genitive usage tend to be weaker in Outer Circle varieties than in Inner Circle 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 going to spot-check the cognitive robustness of the corpus-derived probabilities via rating experiments along the lines of Bresnan (2007) and 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 Inner Circle and Outer Circle varieties, we predict that genitive use in the expanding circle (e.g. Edwards 2014) deviates from Inner Circle genitive use even more than Outer Circle 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, active-passive alternation, etc.), it influences these alternations more strongly in Inner Circle varieties than in Outer Circle varieties: In Outer Circle 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.

References

Baayen, Harald. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412. doi:10.1016/j.jml.2007.12.005.

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. 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*, 75–96. Berlin: Mouton de Gruyter.

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.

Brunner, Thomas. 2014. Structural nativization, typology and complexity: noun phrase structures in British, Kenyan and Singaporean English. *English Language and Linguistics* 18(1). 23–48.

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.

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.

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.

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.

Greenbaum, Sidney. 1996. Introducing ICE. In Sidney Greenbaum (ed.), *Comparing English Worldwide: The International Corpus of English*, 3–12. Oxford: Clarendon.

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 & Tobias Bernaisch. 2015. Exploring epicenters empirically: Focus on South Asian Englishes. *English World-Wide* 37(1).

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.

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

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 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 & 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.

Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: a Functional Reference Grammar*. University of California 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 A.M.B. De Groot & Kroll (eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, 113–142. Mahawa, NJ: Lawrence Erlbaum Associates.

Matthews, Stephen & Virginia Yip. 1994. *Cantonese: A Comprehensive Grammar*. Routledge.

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.

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. 1999. Notes on Singaporean English. In Uwe Carls & Peter Lucko (eds.), *Form, Function and Variation in English: Studies in Honour of Klaus Hansen*, 193–205. Frankfurt am Main: Peter Lang.

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

Shih, Stephanie, Jason Grafmiller, Richard Futrell & Joan Bresnan. 2015. 8 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/9783110378092/9783110378092.207/9783110378092.207.xml (17 September, 2015).

Szmalec, Arnaud, Marc Brysbaert & Wouter Duyck. 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.

Szmrecsanyi, Benedikt, Anette Rosenbach, Joan Bresnan & Christoph Wolk. 2014. Culturally conditioned language change? A multivariate analysis of genitive constructions in ARCHER. In Marianne Hundt (ed.), *Late Modern English syntax*, 133–152. (Studies in English Language). Cambridge, New York: Cambridge University Press.

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.

Zaenen, Annie, Jean Carletta, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden,

Tatiana Nikitina, Mary Catherine O'Connor & Thomas Wasow. 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.

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.

A Appendix

A.1 Regression model with inner circle/outer circle distinction

Table 14: Regression model in which VARIETY was conflated to NATIVITY (levels: *inner circle* and *outer circle*). Predictions are for the *s*-genitive. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

Predictor	b	Odds ratio	p	
(Intercept)	-3.99	0.02	< 0.001	***
POSSESSOR ANIMACY				
inanimate → animate	3.85	47.17	< 0.001	***
POSSESSOR LENGTH (logged)				
one unit increase	-2.39	0.09	< 0.001	***
POSSESSUM LENGTH (logged)				
one unit increase	1.14	3.14	< 0.001	***
FINAL SIBILANCY				
false → true	-1.47	0.23	< 0.001	***
POSSESSOR GIVENNESS				
new → given	0.27	1.31	0.010	*
MODE				
written → spoken	0.11	1.11	0.622	
POSSESSOR HEAD FREQUENCY (logged)				
one unit increase	-0.12	0.89	< 0.001	***
POSSESSOR THEMATICITY (logged)				
one unit increase	0.15	1.16	0.004	**
TYPE-TOKEN RATIO				

one unit increase	2.80	16.50	< 0.001	***
NATIVITY				
inner circle → outer circle	-0.05	0.95	0.827	
POSSESSOR ANIMACY : NATIVITY				
animate + outer circle	-0.42	0.65	0.035	*
POSSESSUM LENGTH (logged) : NATIVITY				
one unit increase + outer circle	-0.07	0.93	0.631	
FINAL SIBILANCY : NATIVITY				
true + outer circle	-0.29	0.75	0.252	
MODE : NATIVITY				
spoken + outer circle	-0.74	0.48	0.010	*