Recent grammatical change in postcolonial Englishes: A real-time study of genitive variation in Caribbean and Indian newswriting

To be published in *Journal of English Linguistics*

Stephanie Hackert & Diana Wengler, LMU Munich

Abstract

This paper presents a diachronic analysis of genitive variation in five varieties of English. Based on a set of matching newspaper corpora from the 1960s and the early 2000s from the Bahamas, Jamaica, India, Great Britain, and the U.S., we look into variation and change in the underlying grammar of the genitive alternation, as defined by patterns of constraints affecting the variable. We employ random forests and Multifactorial Prediction and Deviation Analysis to analyze a richly annotated set of over 22,000 genitive tokens. Our analysis corroborates initial findings with regard to postcolonial Englishes, particularly in the Caribbean, that suggest that these varieties are partaking in American-led global trends in grammar toward, e.g., densification, without actually approximating American norms. We also notice that production-related constraints on genitive variation, such as syntactic weight or givenness, have increased their effects. While metropolitan and postcolonial Englishes share a core grammar of genitive variation, there is noticeable variation particularly with regard to semantic and socioculturally determined predictors such as text type. Overall, we see a widening gap between metropolitan and postcolonial Englishes. The case of Bahamian English is especially interesting, as it appears fairly American-oriented during colonial times but has aligned with other postcolonial Englishes since independence.

Keywords

genitive variation, diachrony, postcolonial varieties, Caribbean, India, BrE, AmE, newspaper, corpus, Multifactorial Prediction and Deviation Analysis, random forests, Americanization

1. Introduction

The past decade has seen an upsurge in work on recent grammatical change in English, much of it doubtlessly inspired by Leech et al.'s landmark volume on the Brown family of corpora (2009). So far, however, most of this work has focused on the two "metropolitan" standards (cf. Mesthrie & Bhatt 2008:3-4), i.e., British (BrE) and American English (AmE). The historical development of the grammars of postcolonial Englishes has received much less attention. Unfortunately, collections of historical texts from non-metropolitan varieties are rare; if they exist, they often cover so-called "native" varieties like Irish, New Zealand, or Australian English. For "non-native" postcolonial Englishes, the dearth of historical data is even more severe. Still, a small but growing number of such corpora are being compiled and studied, and Fuchs (2020:395) even speaks of a "third wave of world Englishes research" that has turned to analyzing processes of structural change based on truly diachronic research.

The present paper is part of this third wave of research into postcolonial Englishes. The variable we are interested in is genitive variation, i.e., the competition between the *s*-genitive and the *of*-construction, as in examples (1) and (2).

(1) The Ministry Paper showed that overall, *students' performance* has been steadily improving […]. (JAM 2012 JamRep214)

(2) But the real tragedy is *the consistently low performance of Jamaican students* in CSEC mathematics. (JAM 2012 JamEd002)

Genitive variation in English is extremely well studied, and the historical development of English genitive forms as well as the constraints governing their present-day distribution are well known. The

contemporary *'s* clitic descends from the Old English strong masculine genitive inflection *-es*, albeit not in direct continuation (cf. Rosenbach 2002:201-232). In Middle English, the *s*-genitive was almost ousted by the periphrastic *of*-construction. "[A]gainst all odds" (2002:184), it has risen again since Early Modern English, extending its domain of use from human possessors (e.g., *Mary's daughter*), to which it was virtually confined at the time (cf. Altenberg 1982:14), to collectives (e.g., *the government's policies*) and certain (often locative or temporal) inanimates (e.g., *yesterday's paper*) in Late Modern English. This process is continuing in present-day English. It is most noticeable in written, standardized language, and specifically in newspapers, where it is part of a more general trend toward "densification" (Leech et al. 2009:249-252), i.e., information compression. AmE has been found to be leading the spread of the *s*-genitive, which is why genitive variation has been linked with the purported "Americanization" (2009:252-259) of varieties of English worldwide.

For the reasons just sketched, diachronic investigations have figured prominently in genitive research, but like research on other grammatical changes in English, these investigations have so far focused on the metropolitan varieties and, to a lesser extent, "native" postcolonial Englishes (e.g., Hinrichs & Szmrecsanyi 2007; Hundt & Szmrecsanyi 2012; Szmrecsanyi et al. 2017). A few synchronic studies (Szmrecsanyi, Grafmiller, Heller & Röthlisberger 2016; Heller, Szmrecsanyi & Grafmiller 2017; Heller, Bernaisch & Gries 2017; Heller & Szmrecsanyi 2019) have involved other postcolonial Englishes, but to our knowledge there is only a single study that has investigated the development of genitive variation in such a variety based on real-time corpus data, i.e., Gries et al. (2018) on Singapore English. In what follows, we present a diachronic analysis of the genitive alternation in Caribbean and Indian English. We base our analysis on newspaper corpora from the 1960s and the early 2000s, which we compare to contemporaneous BrE and AmE as represented in the press sections of the Lancaster-Oslo-Bergen Corpus (LOB) and the Brown Corpus as well as BE06 and AE06 (cf. Baker 2009). In order to do so, we combine insights from corpus linguistics and variationist linguistics to present a model of the recent history of genitive variation in five varieties of

English, drawing on a large, richly annotated data set of over 22,000 genitive tokens and up-to-date statistical techniques. Specifically, we are interested in whether patterns of genitive alternation in Caribbean and Indian English evidence a shift in norm orientation from British-like during late colonial or early postcolonial times to American-like today.

This paper is structured as follows. We first present the research context in which our analysis is embedded (section 2) and then introduce our data and method (section 3). Sections 4 and 5 present the results of our statistical analyses, with section 4 focusing on the effects that the various constraints on the genitive alternation have in our sample and section 5 zooming in on varietal and historical differences. Section 6 discusses our results, and section 7 offers concluding remarks.

2. Research Context

As outlined above, the vast majority of work on grammatical change in English has dealt with BrE and AmE. This is surprising in view of the fact that the most influential recent model of World Englishes, i.e., Schneider's Dynamic Model (2003, 2007), is an explicitly historical model. In fact, Schneider (2003:238) himself suggests it should be "obvious" that "the sociolinguistic and linguistic scenarios in which New Englishes have evolved lend themselves to an investigation of such general questions as language variation and change." The Dynamic Model is well known and has been summarized, applied, and criticized numerous times. One of its main claims is that the coming-into-being of postcolonial Englishes always involves five phases: (1) foundation, (2) exonormative stabilization, (3) nativization, (4) endonormative stabilization, and (5) differentiation. We focus here on phase (4), i.e., endonormative stabilization, which is when local norms develop both structurally and with regard to the status and functions of varieties of English in the community.

The process of endonormative stabilization in the postcolonial Caribbean has been traced in a number of studies (for an overview, cf. Hackert 2016), which all agree that it fundamentally involves what Mair (2006:7) has described as "a three-way competition among a still powerful traditional

British model, the currently dominant American norm, and local usage." The phrase "local usage" here refers to creole features, which, even though they are overtly stigmatized, possess considerable covert prestige and may occur, albeit in limited form, even in newspapers (cf. Bruckmaier & Hackert 2011). What we are interested in here, though, is not the potential "creolization" of Caribbean standard English but the competition between BrE and AmE norms, as described by Schneider (2006:67):

> The vast majority of Outer-Circle world Englishes are products of British colonialism and traditionally in these countries British English […] used to be regarded as the linguistic norm and target of education. […] Today, however, an increasing impact of American English on practically all varieties of English around the globe can be observed.

To assume American linguistic influences appears entirely plausible for the Caribbean in light of its geographical proximity and current socioeconomic (and, in the case of the Bahamas, strong sociohistorical) ties with the North American mainland. A number of studies have looked at the impact of AmE on Caribbean Englishes (e.g., Mair 2002; Oenbring 2010; Bruckmaier & Hackert 2011; Hänsel & Deuber 2013; Hackert 2015; Hackert & Deuber 2015; Deuber et al. fc.). The picture that emerges from these studies is one in which the amount of features found varies widely and depends not only on linguistic level, with vocabulary generally much more open to American items than spelling, but also on a community's sociohistorical, cultural, and economic ties with the U.S. Thus, countries like Jamaica or Trinidad and Tobago possess a strong British normative base; they appear more conservative than other Caribbean countries in linguistic terms. Even though their British colonial background is identical, the Bahamas' population structure and linguistic makeup were crucially shaped by large-scale migration from North America (cf. Hackert & Huber 2007), with a more recent but probably no less influential impact through institutional contacts, temporary migration, education, tourism, media, and lifestyles (cf. Craton & Saunders 1998:275-434).

The evidence at the grammatical level is mixed. With regard to some features, such as verb and negative contractions, Caribbean journalists have been following AmE-led trends currently observable in news writing around the world, but they have been doing so much more reluctantly than British journalists, and with regard to other features, such as the *be*-passive, we actually see an increasing gap between metropolitan and Caribbean usage (cf. Hackert & Deuber 2015:396-397). All in all, news writing from the region retains a particular, often decidedly conservative flavor, and it appears to be amiss to speak of large-scale linguistic Americanization.

Economic and ideological Americanization, in the form of consumerism, "can-do ambition and […] entrepreneurial spirit" (Kapur 2012), seems to exist in India, too, resulting in an unprecedented strife for AmE accents among Indian professionals and service workers in globally active businesses such as call centers (Nadeem 2011). At other linguistic levels, IndE still appears thoroughly British-oriented. Thus, Gonçalves et al. (2018) find that, with regard to vocabulary and spelling, IndE clusters together with southern-hemisphere settler Englishes such as South African, Australian, and New Zealand English and closest to BrE (and Irish English) of all varieties tested. Evidence from grammatical studies is scarce but once more indicates participation in global trends such as shifts in modal and semi-modal usage or the expansion of the progressive (e.g., Laliberté 2016; Fuchs 2020) but no approximation to American norms.

Genitive variation has been linked with the question of linguistic Americanization, as it has been found that AmE is leading BrE not only in the overall return to the *s*-genitive but also in the form's expansion to inanimate possessor head nouns (Hinrichs & Szmrecsany 2007:450), in the tendency to encode thematic, i.e., high-frequency, NPs by means of the *s*-genitive, and in the trend toward its occurrence with longer possessum NPs (2007:468). The expansion of the *s*-genitive is not an isolated phenomenon but part of a more general, discourse-based trend toward densification, which, in turn, is related to the modern "informational explosion" (Biber 2003:180). Densification, i.e., the tendency to compact ever more information into a given amount of words, involves not just

the spread of the *s*-genitive but also an increase in lexical density in general (Hinrichs & Szmrecsanyi 2007:468), a rise in the frequency of nouns as well as of noun + common noun sequences (Leech et al. 2009:257), and the success of so-called "pseudotitles," i.e., determinerless structures of the type *Yankees legend Joe DiMaggio*, which first occurred in *Time* magazine in the 1920s but have spread worldwide since then (Meyer 2002) and also occur in Caribbean journalistic writing (Hackert 2015; Hackert & Deuber 2015). Densification can be observed in all informational registers but appears to be strongest in journalistic writing (Leech et al. 2009:234), possibly owing to economic pressure. Not unexpectedly, as the "characteristically more compact" (Biber et al. 1999:300) of the two competing genitive variants, the *s*-genitive has seen a steeper incline in news writing than in other text types (Hinrichs & Szmrecsanyi 2007:457-458).

All of the phenomena just described, i.e., endonormative stabilization, Americanization, and densification, are diachronic processes. Still, a lot of studies on the development of postcolonial Englishes have taken a synchronic approach, often based on the data contained in the International Corpus of English or, more recently, the Corpus of Global Web-Based English (Davies & Fuchs 2015), and indeed, the equation of BrE with the "parent" variety and the "new" English under investigation with the "younger" dialect resembles the sociolinguistic apparent-time construct. There is at least one massive complication, however, which is language change. Thus, contemporary, present-day, BrE is not necessarily equivalent to contemporaneous, colonial, BrE, and the same holds for other varieties, so the inference of diachronic developments from present-day data rests on shaky ground theoretically.

In this paper, we describe real-time developments in grammar with a view to detecting possible changes in norm orientation in two Caribbean varieties, i.e., Jamaican English (JamE) and Bahamian English (BahE). We compare these developments to what is observable in Indian English (IndE) and in BrE and AmE as "norm-providing" varieties (Kachru 1985:16). IndE is often considered a typical "New English," i.e., one of a number of increasingly autonomous second-

language varieties of English (ESL) functioning not only as interethnic lingua francas in highly multilingual postcolonial situations but also enjoying special prestige in public, formal domains such as education, administration, and the media. BahE and JamE constitute the local forms of standard English co-existing with Caribbean English-lexifier creoles in a continuum situation. They have been described as varieties of English as a second dialect (ESD), to take account of the fact that the Caribbean creoles "are sociolinguistically in a similar position to dialects in Europe, serving as the spoken everyday language as against the formal/written standard" (Görlach 2002:54).

3. Data and Method

The present study is based on ten newspaper collections matched for size and composition. We chose news writing not just for its comparatively easy accessibility, but also because "newspaper prose seems to be the most promising genre to analyze in any study of language change in progress, given its openness to innovation" (Hinrichs & Szmrecsanyi 2007:441). This is a factor well worth considering in the sociolinguistic situation obtaining in postcolonial anglophone countries, where, as described above, there is competition between norms with regard to the status, functions, and structure of the varieties of English used in the community. For each postcolonial country, we compare a newspaper corpus from 1968 with one from the early 2000s. While the Bahamas remained a British colony until 1973, Jamaica had been independent since 1962, so the late 1960s mark the transition period between late colonial and early postcolonial times in the Caribbean. For reasons of comparability, we also decided on 1968 as the year of publication for our Indian historical newspaper sample, even though India had long begun decolonization, having become independent in 1947. British and American newspaper data from the 1960s and the 2000s are contained in the Brown family of corpora. LOB and Brown have become standard reference corpora in the comparative study of varieties of English worldwide; BE06 and AE06 are extensions of the Brown family to contemporary BrE and AmE and, just like the former, available via the CQPweb server at Lancaster

University (https://cqpweb.lancs.ac.uk/). In addition to the observation of diachronic change for every individual variety spanning roughly a generation of speakers, this setup permits the synchronic comparison of metropolitan and postcolonial language use for the decades covered.

With regard to contents, for all our self-compiled corpora, we only sampled news that had appeared in the national or local sections of the relevant papers, in order to not accidentally include pieces distributed by international news agencies. We also used exclusively editorials with local relevance. The contemporary data were downloaded from newspapers' websites. All historical data were purchased as scanned PDF files from the Library of Congress, the British Library, or Newspaper Archive (www.newspaperarchive.com). Because the print quality of these PDFs was often too poor to be suitable for efficient OCR, most of them had to be transformed into .txt files manually.

[TABLE 1 HERE]

Table 1 presents an overview of our data. With the exception of the contemporary Jamaican corpus, which was compiled at the University of Münster, and the Indian component of the South Asian Varieties of English Corpus (IND SAVE), which is housed at the University of Gießen, all of the postcolonial corpora were compiled at LMU Munich. All samples consist of news reports and editorials roughly in a two-to-one proportion; thus, Brown Press, LOB Press, etc. contain only sections A and B of each corpus but not section C, which is the press reviews.

Variationist studies crucially rely on accurate definition of the envelope of variation, i.e., the set of contexts in which a particular linguistic variable occurs. For the present study, we restricted this context to the binary alternation between the *'s* clitic and the periphrastic *of*-construction. We are aware of the fact that these two forms may also alternate with noun-noun structures, as in *the university's budget* vs. *the budget of the university* vs. *the university budget* (Feist 2012:266). As not all noun-noun structures can be freely interchanged with both overtly marked genitive constructions

(e.g., *corn fields* vs. *fields of corn* vs. ?*corn's field*s) and we are interested here not in expressions of genitive meaning in general but view the genitive alternation as a window on the recent history of Caribbean and Indian English, we focus on its two "major variants" (Rosenbach 2014:222), i.e., the *s*-genitive and the *of*-construction. For these variants, the envelope of variation had to be restricted to interchangeable cases, i.e., tokens that would have also permitted the alternative construction. This excludes, among others, descriptive genitives (e.g., *mother's milk*; AE06_B26), measurement NPs (e.g., *the percentage of knife-related violent crime*; BE06_B17), or conventionalized phrases (e.g., *the President of the United States*). Non-interchangeable genitives are described in detail and illustrated by Hinrichs & Szmrecsanyi (2007:446-447), and we followed their list of exclusions exactly. For reasons of space, we refrain from repeating it here.

The extraction of interchangeable genitives involved three steps. First, we used WordSmith Tools 6.0 (Scott 2012) to automatically retrieve all constructions containing the genitive markers *'s*, *s'*, and *of*. We then employed lexical, syntactic, and part-of-speech characteristics of the two genitive variants, such as the presence of an indefinite article in the possessum NP or of a numeral heading the possessor NP, to semi-automatically filter these concordances and exclude constructions that did not represent genitives. Finally, we manually weeded out all remaining non-interchangeable genitives according to the criteria specified in Hinrichs & Szmrecsanyi (2007:446-447). The selection of interchangeable genitives involved two coders. In order to assess inter-rater agreement, a small subset of 100 potential *s*-genitives and 200 *of*-constructions were coded by both of them. Next, Cohen's κ, a coefficient measuring interrater agreement including chance correction, was calculated using the "irr" package (Gamer et al. 2019) in R (R Core Team 2020). The *of*-genitive yielded an agreement rate of 92.5%, and thus a Cohen's κ of 0.82. For the *s*-genitive, an agreement rate of 97% and a Cohen's κ of 0.92 were obtained.

We coded, first, for animacy.[1] Possessor noun phrase (NP) animacy is probably the best-known factor influencing genitive variation in English: the more animate this NP, the more likely it is

to take the *s*-genitive. While it is uncontroversial that the *s*-genitive has been spreading from animate to non-animate possessors in Late Modern English, some authors (e.g., Jespersen 1909-49/VII:327-328) have claimed that this is owed to inanimates proper, while others (e.g., Mair 2006:146-147) have maintained that it is predominantly collectives which have become more favorable to the *s*-genitive. Locative and temporal expressions have also been found to favor the form (cf. Rosenbach 2014:226). Clearly, thus, "the choice between the two genitives in English is sensitive to graduated animacy values" (Rosenbach 2008:153), and we heeded this piece of advice, distinguishing between the five different animacy categories displayed in Table 2.

[TABLE 2 HERE]

We also looked at the various semantic relations expressed by genitive constructions and the effects that these relations have on the realization of the genitive. They are often subsumed under the heading of "possession," and the two constituents of a genitive construction are referred to as "possessor" and "possessum," even though they often do not actually express possession. The exact delimitation of an exhaustive, mutually exclusive set of genitive meanings in English has proven difficult (cf. Rosenbach 2014:229). If they actually consider genitive semantics, many studies (e.g., Wolk et al. 2013:397-402; Grafmiller 2014:477-482; Szmrecsanyi et al. 2017:5-11) assume a binary division into prototypical vs. non-prototypical relations, with the former comprising legal or permanent ownership (e.g., *Morgan's premises*; JAM 2012 JamRep313), kinship terms (e.g., *Strachan's mother*; BAH 2012 Pr_85), body parts (e.g., *the hearts of ordinary men, women and children*; IND SAVE SM_2004-10-05), and part-whole relations (e.g., *the cathedral's doors*; BAH 2012 Pr_86). Because initial inspection of our data had shown differences with regard to genitive preference among the four prototypical relations, we coded for each of them separately but did not further subdivide non-prototypical genitives.

The length of the possessor and the possessum in any English genitive also influences its formal realization. This happens in accordance with the principle of end weight, which predicts that heavier elements will be placed toward the end of any syntactic unit (Behaghel 1909:139). With regard to genitive choice, this would mean that if the possessor is heavy, the *of*-genitive should be preferred because it places the possessor last; if the possessum is heavy, there should be a preference for the *s*-genitive because it places the possessum last. Examples (3) and (4) illustrate.

(3) On Thursday at *the official openings of the Free National Movement (FNM) Southern Shores and Tall Pines constituency offices*, Prime Minister Hubert Ingraham revealed […]. (BAH 2012 Ed_153)

(4) Harris made this announcement during *last Wednesday's launch of the sixth annual Disaster Awareness Quiz and Poster Competition* at the St. James Parish Council. (JAM 2012 JamRep098)

In order to measure syntactic weight, we employed orthographic characters (excluding spaces and special characters like hyphens). In addition to separate possessor and possessum counts, we also set up a length difference measurement, subtracting the length of the possessum phrase from possessor length for each token. This procedure results in a positive syntactic weight if the possessor is longer than the possessum and a negative one if the reverse is the case.

We now turn to definiteness. A definite NP generally encodes information that is known or can be identified by hearers or readers from textual or contextual cues. Because it is more easily accessed by speakers or writers, it tends to precede new discourse elements. Definite possessor NPs, as in example (5), have been found to favor the *s*-genitive (cf. Rosenbach 2014:232), since it places the possessor first and thus realizes old-before-new word order.

(5) […] public protest drove Prime Minister Kamla Persad-Bissessar into a hurried repeal of a controversial Section 34 of *the country's new, but in-limbo, Administration of Justice (Indictable Offences) Act*. (JAM 2012 Ed 024)

In order to operationalize possessor definiteness, we employed Garretson's (2004:19-23) two-level coding scheme, which makes use of the "existential *there*" test to distinguish definite from indefinite NPs, with each category typically co-occurring with particular articles, pronouns, and quantifiers.[2]

If a possessor referent has been mentioned explicitly in the previous discourse, it is considered given. Givenness is an indicator of information status (Hinrichs & Szmrecsanyi 2007:451), which, just like syntactic weight and definiteness, stipulates a certain ordering of discourse constituents. Given information tends to precede new information; hence, given possessors tend to favor the *s*-genitive (cf. Biber et al. 1999:305-306). In operationalizing givenness, we followed Szmrecsanyi, Grafmiller, Heller & Röthlisberger (2016:113) and counted explicit mentions of the possessor head noun in the 100 words preceding any token. This was done automatically by means of a Python script.

We also coded for persistence, i.e., the presence of an identical genitive construction in the immediately preceding discourse. With regard to genitive variation, persistence was identified as an influential factor by Altenberg (1982:290) and first investigated in multivariate analysis by Szmrecsanyi (2006:87-107), but the general principle is well-known: language users tend to recycle material they have heard or used before, which leads to a variety of phenomena, from priming in discourse to patterns of historical present tense use in narrative. In order to operationalize persistence in genitive marking, we used an R script to identify, for each token, whether the immediately preceding token employed the same genitive form or not.

If possessor NPs end in a final sibilant, as in *the niece's boyfriend* (JAM 2012 JamRep 308) or *Dr. Gomez's achievements* (BAH 2012 Pr_187), the *s*-genitive is usually inhibited. This effect has been described by Hinrichs & Szmrecsanyi (2007:452) as a phonological instantiation of the principle of *horror aequi*, i.e., "the widespread (and presumably universal) tendency to avoid the repetition of identical and adjacent […] elements or structures" (Rohdenburg 2003:205). A syntactic *horror aequi* effect emerges when "language users avoid two identical genitive constructions in the same NP" (Hinrichs & Szmrecsanyi 2007:456). While Hinrichs & Szmrecsanyi observe a significant increase of

the *s*-form when genitives are embedded within *of*-genitives, they also note that, "[o]n the whole, […] nested *s*-genitives are quite rare" (2007:456). This was also the impression we gained whilst coding our data, which is why we decided instead to look at all genitives occurring as the prepositional complement in an *of*-phrase, whether this *of*-phrase realizes another (interchangeable) genitive, as in example (6), or not, as in (7).

(6) Mr A K Hafizka […] stated that the BPCC was not opposed to *the acceptance of the legitimate demands of the BEST employees* […]. (IND 1968 Ln_114)

(7) Singhvi and Lall directed the petitioner to place *a copy of the order of the chief justice* on record before the matter could be further examined. (IND SAVE TI_37459)

We retained such embedded genitives as a separate category.

Finally, we coded for text type, i.e., the distinction between news reports and editorials. We take this distinction to be economy-related, because the need to encode a maximum of information in a given amount of words is higher in reportage than in editorials, which express opinions and seek to persuade, which per se requires "a less dense and more transparent means of expression" (Biber et al. 1999:302).[2] For Brown and LOB and their 1990s updates Frown and FLOB, in fact, Hinrichs & Szmrecsanyi (2007:461) find that the occurrence of "a genitive slot […] in editorials […] instead of in the reportage section significantly decreases the odds for an *s*-genitive by 33 percent," and we also marked off editorials from news reports proper.

4. Constraints and Their Effects

Our final data set comprises 22,411 tokens, of which 8,266 are *s*-genitives and 14,145 *of*-constructions. Figure 1 shows absolute numbers as well as proportions of genitive types by variety and time. In all varieties, *s*-genitive usage increased in the late twentieth century, albeit to massively different extents. In the historical data, we see a clear split between BrE, AmE, and BahE on the one hand, which all feature very similar proportions of the form (38%, 37%, and 39%, respectively), and

JamE and IndE on the other, which show much lower rates (20% and 26%). In the modern corpora, AmE has by far the highest rate of *s*-genitives (61%), tailed only by BrE (56%). Our data thus confirm the oft-described "follow-my-leader" situation (Leech et al. 2009:253), in which AmE is more or less closely ahead of BrE in a process of parallel change. The contemporary postcolonial data, interestingly, show rates of *s*-genitive usage that resemble those of the 1960s metropolitan data (42%, 36%, and 32%, respectively). While they are following the AmE-led drift toward the form, thus, journalists from the Caribbean and from India appear to be doing so much more hesitantly than British journalists, which results in a noticeable gap between metropolitan and postcolonial varieties in present-day English. Bahamian journalists, finally, show hardly any change in patterns of genitive usage, and while their *s*-genitive frequencies were AmE-like in the 1960s, BahE appears like a "proper" postcolonial English in the 2000s.

[FIGURE 1 HERE]

As outlined in the previous section, genitive variation in English is subject to "numerous, multifactorial, and probabilistic" constraints (Heller, Szmrecsanyi & Grafmiller 2017:4). Moreover, some of these constraints correlate strongly, such that, for example, animate possessors tend to be shorter than inanimate ones. Still, "so far none of the factors suggested in the literature has been shown to be completely reducible to another one" (Rosenbach 2014:230), and even a strong factor such as animacy may be overridden by other factors under certain circumstances. To come to terms with such "messy" patterns, variationist linguists have long employed probabilistic modeling in the form of multivariate analysis. We made use of two fairly recent developments in this field of statistics, i.e., random forests and Multifactorial Prediction and Deviation Analysis. Random forests offer several advantages over more established techniques such as fixed- or mixed-effects modeling, the most important of which in the present context is their ability to handle collinear predictors. We

employed R's "party" package (e.g., Hothorn et al. 2006) to run separate random forest analyses for each data set. These analyses achieved classification accuracies between 80% and 87%, which was always significantly better than chance prediction or choice based on the more frequent of the two alternating forms, i.e., the *of*-genitive. The resulting *C* values ranged from 0.89 to 0.94.

In a first step, we extracted probabilities from these analyses (which we then subjected to Wilcoxon rank sum tests for pairwise comparisons and to Kruskal-Wallis rank sum tests in the case of more than two predictor levels) in order to shine a light on the behavior of each of the constraints on genitive variation outlined in section 3. We did this for two reasons. On the one hand, while these constraints' impact is amply documented for the metropolitan varieties and for "native" postcolonial Englishes, genitive variation has never been investigated for BahE and is clearly understudied in JamE and IndE. On the other hand, this enabled us to detect idiosyncracies in our data, such as they always affect corpora of naturally occurring language and hence the statistics based on them.

Turning first to animacy, Figure 2 shows *s*-genitive probabilities by animacy and data set, with significance levels set at $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***). We see that animates, locatives, and temporals had fairly high *s*-genitive probabilities in all varieties even in the 1960s. In AmE, BrE, and JamE, these categories have seen statistically highly significant growth in the use of the form. In BahE, we see a decline of the *s*-genitive among locatives and temporals. This is likely owed to data composition: while in BAH 1968 about 22% (43/197) of locative possessor NPs are tokens such as *the Bahamas*, *the Commonwealth of the Bahamas*, or *the Bahama Islands*, whose final sibilant inhibits the *s*-genitive, in BAH 2012, such tokens account for a full 42% of locatives (47/112). Among temporals, the decline in *s*-genitives is possibly linked to the number of possessor NPs referring to years (e.g., *by the end of 1968*; BAH 1968 0307_07), which account for 13% (19/142) in the historical vs. 21% (15/70) in the modern BahE data and co-occur almost exclusively with the *of*-construction. The same effect may at least in part explain the disproportionate rise of the *s*-genitive among JamE temporals. In sum, in the Caribbean data, locatives and temporals appear to be fairly

strongly affected by their lexical composition, which, in turn, is owed to topic preferences in the texts sampled.

Returning to general trends, the two largest animacy classes, i.e., animates and inanimates, quite clearly separate the metropolitan varieties from the postcolonial ones. Whereas we see significant growth in the use of the *s*-genitive in both categories in the former, in BahE and IndE the observable changes are much smaller, partly inconsistent, and not always statistically significant. JamE follows the metropolitan varieties in showing growth in the *s*-genitive with both animates and inanimates, albeit at much lower levels. Only collectives show statistically highly significant growth in *s*-genitive probabilities in all varieties. Thus, whereas among metropolitan varieties the *s*-genitive is rising across animacy categories, in the postcolonial ones, it appears to be collective possessor NPs that are driving the form's expansion.

[FIGURE 2 HERE]

The most important result of the analysis displayed in Figure 3 is that the choice of genitive construction in our data not actually run along the prototypical vs. non-prototypical boundary but follows a gradient, with "proper" possessive constructions, i.e., those referring to legal or permanent ownership, most likely to come in the form of an *s*-genitive. Even though we see a rise of the *s*-genitive in almost all semantic relations between the 1960s and today (which reflects the general rise of the *s*-genitive displayed in Figure 1), in BahE and IndE the smaller classes of body parts and part-whole relationships show either non-significant growth or even decline. (As in Figure 2, significance levels were set at $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***)). Again, this finding may at least in part be owed to corpus idiosyncracies related to local peculiarities. While BAH 1968, for example, contains 16 tokens of *Nassau* in possessor position among part-wholes, of which a full 75% are instantiations of the *s*-genitive, *Nassau* occurs only once in this context in BAH 2012. In IND 1968,

we find a disproportionately high frequency of *s*-marked tokens of *Bombay* (10/15) and *India* (9/13), while in the modern data *India* is both considerably less frequent and less frequently *s*-marked (1/4), and *Bombay*'s contemporary realization, *Mumbai*, occurs only once.

[FIGURE 3 HERE]

Based on what we have outlined in section 3, we expected the *s*-genitive to be correlated with lighter or negative syntactic weights (e.g., *Eleuthera's* [9] *eastern shore* [12] -> [-3]; BAH 1968 0312_03) and the *of*-construction with heavier or positive ones (e.g., *the coast* [8] *of Eleuthera* [9] -> [1]; BAH 1968 0312_01). Figure 4, which displays *s*-genitive probabilities by data set and z-scored syntactic weight, shows that this is actually the case: there is an inverse correlation between the likelihood with which *s*-genitives occur and the difference between possessor and possessum length throughout our sample.[3] We also see a shift toward the *s*-genitive between historical and modern data sets. This shift is not surprising; it simply mirrors what we already saw in Figure 1. However, the smoothers have not simply moved upward in parallel fashion but have also changed shape. Whereas for the historical data we see a more or less linear connection between decreasing syntactic weight and increasing *s*-genitive probability, the smoothers representing the modern sets form *s*-curves. After a certain cutoff point in terms of syntactic weight, each of these curves shows a rapid upswing, which compresses the transition zone in which *s*-genitives and *of*-constructions are more or less equally likely. This appears to indicate a focusing process, with the two genitive constructions moving in opposite directions with regard to syntactic weight. Cutoff points and upswing curvature angles are variety-specific, but AmE is once more ahead in terms of the former, while both BrE and AmE appear almost equally advanced in terms of the focusing process, and *s*-genitives become next to impossible at much lower syntactic weights than in the postcolonial varieties.

[FIGURE 4 HERE]

Syntactic weight has been identified as the most prominent of a number of factors affecting genitive variation in English that are related to the production of language, i.e., to memory retrieval, serial action planning, and motor control. Recently, the view has gained currency that both the typology of languages and language-internal variation in syntactic structures is crucially shaped by language users' attempts at minimizing the computational efforts arising from utterance planning. An important principle governing linguistic behavior in this respect is "Easy First" (MacDonald 2013:3), which stipulates that elements that are "more frequent, shorter (both number of words in a phrase, and number of syllables in a word), less syntactically complex, more important or conceptually salient to the speaker, and previously mentioned ('given') in the discourse" are easier to activate during language production than their respective counterparts and therefore tend to be placed earlier in utterances. Such a production-related approach not only encompasses the principle of end weight within a larger, cognitive framework but also helps to explain other constraints on genitive variation, such as definiteness, givenness, or persistence. We treat these and additional factors in turn now, summarizing our results in Figure 5.

As outlined in section 3, if possessor NPs are definite, we would expect them to favor the *s*-genitive, and Figure 5.1 shows that this is the case in all our data sets. Even though the difference between contexts is statistically highly significant throughout, it is consistently smaller in the historical than in the contemporary data. As was the case with syntactic weight, it appears as if a difference related to the production of language might have grown between the 1960s and today. The same holds for givenness, which, as noted above, is also part of the principle of Easy First. Figure 5.2 shows that given possessors favor the *s*-genitive in all our data sets, but the gap between contexts is consistently larger in the 2000s than in the 1960s.

Moving on to persistence, Figure 5.3 shows that *s*-genitives are attracted by preceding *s*-genitives, to a statistically highly significant degree, in all our data sets.[4] On a theoretical plane, persistence can be related to language production constraints if viewed as an instantiation of the general principle of "Plan Reuse" (MacDonald 2013:4). Importantly, while Easy First promotes syntactic flexibility (2013:3), which, in the case of the genitive alternation, results in variation in the order of possessor and possessum, Plan Reuse fosters syntactic rigidity (2013:4), in the form of repetitive choice behavior.

The disfavoring effect that sibilant-final possessors have on the *s*-genitive emerges in Figure 5.4. With regard to nested genitive constructions (Figure 5.5), our hypothesis was that, regardless of their function, such constructions would favor the *s*-genitive, because *of*-phrases within *of*-phrases would create *horror aequi*. And indeed, with the exception of 1960s AmE and contemporary IndE, this hypothesis is confirmed, and statistically highly significantly so. Both the sibilancy constraint and the avoidance of embedded *of*-constructions may again be viewed as related to principles governing the production of language, in this case "Reduce Interference." According to MacDonald (2013:4), "elements interfere with one another in memory during the short interval between their presentation and recall, with increasing interference when the elements share similarity in sound, meaning, spatial location, or other dimensions." Like the production-related constraints investigated previously, both dimensions of Reduce Interference have become stronger in our data between the 1960s and the 2000s, albeit to varying extents in different varieties and with minor exceptions.

With regard to the effect of text type, which is displayed in Figure 5.6., we see the expected pattern of news reports over editorials in both metropolitan varieties and in BahE. Maybe somewhat surprisingly, JamE and IndE show higher *s*-genitive probabilities in editorials than in news reports, but then we are dealing here with discourse-pragmatic conventions, which are not just diachronically variable but also, and perhaps primarily, culture-specific. Apparently, metropolitan conventions dictate that news reports must be close-packed and concise, while editorials give more leeway in this

respect. That said, the difference between text types has become small in contemporary AE06, which may indicate that AmE is reaching a saturation point with regard to the compacting of information in reportage. In JamE and IndE, the two text categories have become more similar in the 2000s than they were in the 1960s, which may hint at a gradual approximation to metropolitan norms with regard to this predictor variable.

[FIGURE 5 HERE]

5. Cross-Variety Differences and the Question of Norm Orientation

Having ascertained the effect that the various constraints have on the genitive alternation in our data, we computed variable importance measurements from the set of random forests we had grown previously. These measures, visualized in Figure 6, consistently show the same two predictors at the top of the list, generally with a substantial margin against all remaining ones: possessor head animacy and syntactic weight. Syntactic weight has caught up with animacy in all modern data sets; in AE06 Press (as well as in BAH 1968), it ranks even higher than the latter. In other words, animacy was and still is the most important predictor of genitive choice in varieties of English, but syntactic weight has substantially gained in importance since the 1960s. With the exception of text type, which more or less consistently gets demoted, the importance of other factors remains consistently low and rather variable intra- and intervarietally between the 1960s and today.

While variable importance measurements enable us to establish the ranking of constraints in individual data sets and to compare these rankings, they do not yet help us in determining the overall extent of similarities and differences between sets, which, however, is necessary in order to answer the question of whether there has been a shift in norm orientation in Caribbean and Indian varieties of English. Such a shift would imply a greater overall similarity between BrE and the non-metropolitan varieties in the 1960s but a closer fit between the latter and AmE in the 2000s. In other words, the

question is which data sets "share an underlying grammar, and to what extent" (Tagliamonte 2013:161), which implies not just similarities but also dissimilarities with regard to the two constructional choices.

[FIGURE 6 HERE]

In order to look into the latter, we opted for Multifactorial Prediction and Deviation Analysis with Regressions (MuPDAR). MuPDAR was developed primarily for learner corpus research (e.g., Gries & Deshors 2014; Wulff & Gries 2015) but has also been applied to the comparison of postcolonial Englishes, particularly to the question of degrees of nativization as well as to the idea of linguistic epicenters (e.g., Gries & Bernaisch 2016; Heller, Bernaisch & Gries 2017; Gries et al. 2018). We employed the non-parametric variant of this approach, which is again based not on fixed- or mixed-effects modeling but on random forest analysis, i.e., MuPDAR-F.

In contrast to other versions of regression analysis, MuPDAR-F provides a principled way of comparing data sets by establishing a yardstick against which all others are measured. This happens "by answering the following question: 'In the situation that the target variety speaker is in now, what would the reference variety speaker have done, and if the two choices differ, how so and why?'" (Heller, Bernaisch & Gries 2017:122). In detail, the following steps are necessary:

- fitting a regression/random forest $R(F)_1$ that predicts the choices that speakers of the source/reference level […] make with regard to the phenomenon in question;
- applying the results of $R(F)_1$ to the other/target speakers in the data […] to predict for each of their data points what the […] speaker of the source/reference variety would have done in their place;
- fitting a regression/random forest $R(F)_2$ that explores how the other speakers' choices differ from those of the speakers of the source/reference variety (2017:122).

Transferred to the research question at hand, our reference varieties are represented by LOB, Brown, BE06 and AE06. If there had been a shift in norm orientation among postcolonial Englishes from BrE to AmE, LOB should make better predictions for BAH 1968, JAM 1968, and IND 1968 than Brown, whereas the regression model generated for AE06 should have a higher predictive accuracy for BAH 2012, JAM 2012, and IND SAVE than that obtained for BE06.

In detail, we proceeded in the following manner. We employed the initial random forest analyses of the four metropolitan data sets displayed in Figure 6 as RF_1s and for each of them computed predicted *s*-genitive probabilities for the contemporaneous postcolonial data sets. We then used these predictions to calculate deviations for each data point. These deviations indicate the difference in probability for each token where reference variety users would have made a different choice in terms of genitive construction than users of the postcolonial variety under investigation. Variants correctly predicted were coded as zero; for all other variants, if RF_1 predicted an *s*-genitive but the token in question showed an *of*-construction, its deviation value was negative; if RF_1 predicted an *of*-construction but the actual data point showed an *s*-genitive, this resulted in a positive deviation value. Thus, deviation values range between -0.5 to +0.5; higher absolute numbers indicate greater deviation. On the basis of these deviations, we ran a second set of random forest analyses, i.e., RF_2s, on our four reference data sets. These analyses included data set as an additional factor but were restricted to the wrongly predicted tokens, i.e., tokens whose deviation value was not zero.[5] Figure 7 visualizes the steps previously described.

[FIGURE 7 HERE]

Since the goodness-of-fit values for all RF_2s were encouraging, we proceeded further to (a) compute variable importance measures again and (b) explore deviations by variety and predictor.[6] In step (a), animacy and syntactic weight once more turned out as the most powerful predictors. These two

factors thus not only possess the most explanatory power in each individual data set, metropolitan or postcolonial, as seen in Figure 6, but they are also primarily responsible for the deviation values that indicate where users of the postcolonial varieties differ in their choice of genitive construction from those of metropolitan Englishes. We then (b) averaged predicted deviations for animacy and syntactic weight by reference corpus and centered those values around the mean deviations of each reference corpus.[7] Figures 8 and 9 show the results of these calculations.

[FIGURE 8 HERE]

Figure 8 displays the effect of animacy on genitive choices made in BahE, JamE, and IndE in comparison to those made in contemporaneous BrE and AmE. In comparison to present-day usage, as displayed on the right, the choices made by users of postcolonial varieties in the 1960s, as shown in the left half of Figure 8, were much closer to those made by reference variety users, either BrE or AmE, as evidenced in the denser clusters of data points around the zero line. Among the historical data, JAM 1968 emerges as the most conservative set, in that it shows a greater preference for *of*-constructions in almost all animacy contexts than both reference corpora, i.e., LOB Press and Brown Press. IND 1968 shows a strong preference for the *s*-genitive among temporals but otherwise clusters largely with the metropolitan varieties; BAH 1968 even outperforms the more *s*-genitive-friendly AmE reference corpus among inanimates, locatives, and temporals. In the modern data, by contrast, postcolonial writers generally emerge as more reluctant to use *s*-genitives than writers of metropolitan Englishes. With exceptions among locatives (which, however, have been shown in section 4 to be most strongly affected by topic-related corpus idiosyncracies), their choices are a lot more similar to those made in BrE than in AmE, which flatly contradicts the Americanization hypothesis.

[FIGURE 9 HERE]

With regard to syntactic weight (Figure 9), we see, once again, that in the 1960s, Bahamian, Jamaican, and Indian journalists were much closer in terms of genitive choice to their British and American colleagues than they are today, as evidenced by the flatter curves in the left graph. What is even more interesting in the present context is the fact that the similarity to LOB Press is greater than that to Brown Press, with the smoothers for BAH 1968, JAM 1968, and IND 1968 all running along the zero line up to a centered syntactic weight mean of roughly zero. In the contemporary data, the gap between the metropolitan and the postcolonial varieties has increased noticeably. This means, apparently, that the principle of end weight works less strongly in contemporary BahE, JamE, and IndE than in present-day BrE and AmE; the historical data sets are more alike in this respect. That said, in general, the smoothers for all contemporary postcolonial data sets are equally dissimilar to both reference corpora, except with very long possessors, where we see a trend toward AmE usage.

In a final step, using R's "ape" package (Paradis & Schliep 2019), we carried out a phylogenetic cluster analysis based on varieties' deviations in terms of *s*-genitive probabilities. The analysis proceeded in a round-robin fashion, with each data set's RF_1 used to predict all other sets' genitive tokens in turn. All predictions entered into the construction of a two-dimensional distance object, which was then subjected to the actual cluster analysis (cf. Baayen 2008:148-160), whose results are visualized in Figure 10. We also computed predictive accuracies for relevant pairs of metropolitan and postcolonial varieties; these are displayed in Table 3.

[FIGURE 10 HERE]

The picture that emerges in Figure 10 clearly shows two clusters of varieties, one centering around LOB Press and Brown Press and comprising not just all other historical data sets but also the present-day Jamaican and Indian ones. JAM 1968 is closest to IND 1968; together, they share a split with

IND SAVE. LOB Press and Brown Press are in between, closest to BAH 1968, which, in turn, shares a branch with JAM 2012. At the other end of the cluster, we find AE06 and BE06 and, slightly further down the same branch, BAH 2012. In both subclusters, BrE is right next to AmE, with the distance between the historical and modern BrE-AmE pairs being fairly large. The two JamE data sets are also quite far apart from each other, whereas the distance between historical and contemporary data is comparatively small for both BahE and IndE. Table 3, finally, shows that BrE fares better than AmE with regard to predicting genitive variation for the postcolonial varieties in both time periods investigated, which once more runs counter to the Americanization hypothesis. Overall, predictive accuracy has gone down; this, in turn, supports the assumption of growing distance between postcolonial and metropolitan varieties.

[TABLE 3 HERE]

6. Summary and Discussion

All research on genitive variation in present-day English has shown a continued rise of the *s*-genitive, and so it does not come as a surprise that we see, in Figure 1, precisely such a rise in BrE, AmE, BahE, JamE, and IndE journalistic writing between the 1960s and the 2000s, with AmE leading this process but followed very closely by BrE. Two of the postcolonial varieties covered here, i.e., JamE and IndE, show the same development, albeit at much lower frequency levels. BahE, finally, patterned with the metropolitan varieties in the 1960s but has changed sides, so to speak, and now shows an *s*-genitive frequency that closely resembles those of the other two postcolonial Englishes, which, in turn, evidence frequencies that look very much like those found in the 1960s metropolitan data sets.

With regard to patterns of constraints, our findings corroborate those of Szmrecsanyi, Grafmiller, Heller & Röthlisberger (2016:132), who identify a "core probabilistic grammar" of

genitive variation in contemporary English, with factors influencing the variable and their effects largely the same across varieties and time. In all our data sets (as well as throughout the English-speaking world), animacy and syntactic weight are by far the two most powerful determinants of genitive choice, with animates favoring the *s*-genitive and inanimates disfavoring it. Shorter possessors attract it; longer ones repel it.

As for animacy, the data displayed in Figure 2 demonstrate that the loosening of this constraint, which began centuries ago (cf., e.g., Wolk et al. 2013:408-410), is continuing in present-day English, but only collective possessors are expanding massively in terms of *s*-genitive occurrence in all of the varieties covered here. Animates and inanimates are either not growing at all or doing so at substantially lower levels in the postcolonial varieties than in the metropolitan ones. Figure 2 thus provides further evidence not only with regard to competing theories concerning animacy effects (cf. section 2) but also with regard to the increasing separation between BrE and AmE on the one hand and BahE, JamE, and IndE on the other.

With regard to syntactic weight, we observe two independent if interrelated developments. First, Figure 3 has shown a growing distance between, or focusing of, genitive constructions in terms of relative possessor and possessum length. The MuPDAR-F analysis presented in Figure 9 additionally shows that journalists from the Bahamas, Jamaica, and India consistently prefer *of*-constructions where newswriters from Britain and the U.S. exhibit a tendency toward *s*-genitive usage, i.e., when possessums are longer than possessors, and vice versa. In other words, postcolonial writers are less influenced by syntactic weight in their choice of genitive construction than writers from Britain or the U.S. While this trend was evident even in the 1960s, it has become considerably more pronounced since then, which offers support for the idea that the above-described focusing process has not affected postcolonial varieties to the same extent as the metropolitan ones.

This runs counter to the expectations of a production-related approach to syntactic variation. As noted by Heller, Szmrecsanyi & Grafmiller (2017:21), users of ESL should show "stronger

utterance planning biases," since "in L2 processing there is constant interference from the L1." However, this expectation is not borne out by their data, either, as the constraints that show significant interaction with variety (i.e., animacy, syntactic weight, and sibilancy) do so in the "wrong" direction. Instead, "ESL users simply tend to weaken constraints that favor the *s*-genitive, and instead strengthen constraints that favor the *of*-genitive" (2017:22), which is exactly what we have just described for our data. While Heller, Szmrecsanyi & Grafmiller (2017:7) interpret this tendency as a synchronic "echo of the well-known second-language acquisition preference for analytic marking," we favor the historical explanation they also allude to (2017:23-24), which claims that ESL varieties lag behind ENL varieties not just in the overall expansion of the *s*-genitive but also in terms of the trends characterizing individual constraints. This explanation aligns well with what we have been describing so far, but we return to the question of ENL vs. ESL (and ESD) below.

Second, as shown in Figure 6, syntactic weight has gained substantially in explanatory power vis-à-vis all other factors, including animacy. This indicates that the principle of end weight, though already operative in earlier periods of Germanic (Behaghel 1909), is still gaining ground in present-day English. Like definiteness and givenness, end weight may be seen as an instantiation of the Easy First principle. Together with the principles of Plan Reuse and Reduce Interference, which, as claimed in section 4, influence genitive variation by way of persistence and possessor sibilancy, Easy First has been described as a major force shaping the distribution of linguistic forms both typologically and with respect to language-internal syntactic variation (MacDonald 2013). All three principles derive from requirements of language production, but they compete: while Easy First results in variability in syntactic structure, Plan Reuse promotes word order rigidity. Still, they may converge on the same form, which may be either the *s*-genitive (when possessors are relatively short, do not end in a final sibilant, and are preceded by another *s*-genitive) or the *of*-construction if the opposite is the case.

Figure 5 has shown that all of the constraints on genitive variation that are production-related, i.e., definiteness, givenness, persistence, sibilancy, and, to a lesser extent, embedding, have increased their effect, in the sense that the differences between contexts with regard to genitive choice have increased across varieties between the 1960s and today. Apparently, thus, there has been a focusing of the two genitive variants not just by syntactic weight but also according to all other production-related factors tested here. Even though the ranking of factors below animacy and syntactic weight displayed in Figure 6 may have to be taken with a grain of salt, as they all have comparatively little explanatory power, we frequently see a demotion of the two remaining non-production related predictors, i.e., semantic relation and text type, vis-à-vis production-related factors between historical and contemporary data sets.

The existence of a core grammar does not preclude variation according to extralinguistic dimensions such as variety or time. Even production-related factors, which are ultimately cognitively grounded, can be molded by language users at least to a certain extent. The limits of this kind of variation appear to be set by predictor strength and explanatory power, with the direction of effects generally remaining stable. Of course, we would like to know what prompts variety-specific or diachronic shifts in variable linguistic phenomena, which takes us to functional explanations and, ultimately, sociocultural interpretation. As indicated in section 1, the continued spread of the *s*-genitive has been described as a case of densification, which, in turn, has been interpreted as a function of linguistic economization. Such phenomena are at least in part subject to community-specific conventions, which vary across space and time. As Figure 5.6 has shown, the expected pattern of higher *s*-genitive probabilites in news reports than in editorials holds only in the metropolitan varieties and in BahE. In JamE and IndE, we find an editorials-over-reports pattern. Interestingly, this pattern emerges as statistically significant in the historical Jamaican and Indian data but no longer in the contemporary sets, which might point to these varieties having embarked on a trajectory toward metropolitan conventions.

We have mentioned the ENL-ESL-ESD distinction before but discuss its significance for the present study in more detail now. As stated in section 2, the varieties we chose for investigation represent all three types. The status of BrE and AmE as "native" varieties is generally unquestioned. IndE is also uncontroversially an ESL variety, even though some Indians self-identify as "mother-tongue" speakers (Census of India 2011:8). Where our classification differs from others (e.g., Heller, Szmrecsanyi & Grafmiller 2017) is in the assumption of ESD (rather than ESL) status for anglophone Caribbean countries, as the history and sociolinguistics of English there are very different from those found in India today (cf. Schneider 2007:161-173, 219-238). Still, we also find important similarities between the two regions in terms of English learning and use, and we suggest that it is these similarities, together with the differential impact that AmE has had, which help explain the patterns of genitive choice that we see in our data.

First, with regard to language education and the acquisition of standard English, even though most Caribbean nationals acquire a creole as their vernacular, Jamaicans and Bahamians traditionally regard themselves as "English-speaking," with additional command of a particular local "dialect," and some speakers deny anything but the "Queen's English" in either their own or their community's repertoire. This perception has an objective base in the existence of the continuum situation described in section 2; it is captured linguistically in the classification of the region as ESD-using. As in India, the school context is traditionally the environment where Caribbean children first come into contact with the standard variety, even though particularly middle-class parents these days place great emphasis on early exposure to "proper" English. At the same time, "the pedagogical emphasis remains […] in large part on students' mechanical correctness in Standard English" (Oenbring & Fielding 2014:29-30), and skills in English grammar and composition are still often measured in terms of "school-taught rules" of the type "when it is more than one, you use the plural" (Devonish & Thomas 2012:193) or, for that matter, the textbook rule governing the choice of genitive constructions that "[n]ames of non-living things cannot be used as possessive nouns" (Nongrum 2017:35). In India,

traditional teaching methods appear even more mechanical, entailing primarily "rote-oriented, text book based routine exercises, loud reading, dictation, and question-answer sessions" (Majumdar & Mooij 2012:226). It is precisely such acquisition patterns that might ultimately underlie the variety-type specific constraint patterns that we have found in our data. They also tie in with the general conservatism that these varieties have been found to exhibit with regard not just to genitive variation but also to other grammatical features, as outlined in section 2.

As for language use in public, formal domains, the "mediascapes" (Appadurai 1990) of all three postcolonial countries surveyed here also show interesting similarities that have their origins in long-standing patterns of external, i.e., metropolitan, dependency and control. These patterns translated into particular local setups with regard to ownership, composition of staff, and training that crucially influenced linguistic norms. In both the Caribbean and India, until way after the end of formal colonialism, the view dominated that "the newspaper office alone offers suitable training" (Rau 1958:144; cf. Storr 2016:46-47). Aspiring local journalists generally did not hold degrees or diplomas, as systematic and institutional education in journalism did not exist. Even today, many senior journalists and staff feel that training on the job is more important than formal qualifications, as extant programs equip journalism students only inadequately and insufficiently (cf. Salwen & Garrison 1991:45; Storr 2016:180).

This is where the idea of "norm-providing" varieties (Kachru 1985:16) comes into play. Even though almost all postcolonial countries are traditionally BrE-oriented, AmE gained ground with the ascent of the U.S. to global superpower status. Our data both confirm and modify the picture drawn in section 2 as well as in studies of British-American linguistic relationships. The metropolitan varieties clearly exhibit the by now famous "follow-my-leader" pattern (Leech et al. 2009:253), with AmE ahead of BrE in a pattern of parallel change. The postcolonial Englishes investigated here are following this pattern, too, albeit reluctantly and with a gap of about a generation. We might, in analogy with another famous cross-variety pattern, term this observation "postcolonial lag." That

said, we need to explain the strange case of BahE, which, in the 1960s appeared thoroughly Americanized but in the 2000s largely aligns with JamE and IndE.

That BahE exhibited American-like grammatical patterns even during colonial times has been shown before (Hackert 2015; Hackert & Deuber 2015). Its behavior with regard to genitive variation therefore does not come as a surprise, and we can explain it easily by recourse to longstanding institutional patterns and personal links. Even though its basis was British colonial, the Bahamian media system has always had close ties with the U.S. (cf. Storr 2000:335-337). In fact, the *Nassau Guardian and Bahamas Observer*, from which our 1968 Bahamian newspaper sample is drawn, came under American ownership in 1967. *Guardian* journalists themselves were largely local and trained on the job, and it appears likely that the stylistic patterns they adopted during that training were modeled on other material extant in the paper. A cursory examination of this other material revealed that all of the news reports dealing with international affairs came from United Press International, an American news agency. U.S. news often appeared in single-authored articles that we were able to trace back to American journalists writing for U.S. papers; these articles must have also been bought overseas. Finally, at the time, the *Guardian* also featured the syndicated advice column "Ann Landers," a pen name used by *Chicago Sun-Times* columnist Ruth Crowley.

Despite all similarities in terms of geographical location and political, economic, and sociocultural history, the linguistic situation in Jamaica at the time must have differed from the Bahamian one. Jamaica never experienced the massive influx of settlers from North America that has shaped Bahamian demographics (cf. section 2) and that may well have prepared the ground for later personal and institutional contacts and thus renewed linguistic influence. Additionally, decolonization began earlier in Jamaica (independence in 1962) than in the Bahamas (independence in 1973). It has been surmised that this earlier decolonization process may have led to the newly emerging local standard being more firmly based in the traditional BrE model (cf. Deuber et al. fc.), and, in fact, the competing AmE model had barely begun its global ascent, which moreover happened not via actual

speaker contact but by means of modern media phenomena. India, finally, became independent even earlier (in 1947), but the two newspapers investigated here were still thoroughly British-oriented in the 1960s, with the *Statesman* owned by a British family until 1963 (Merrill & Fisher 1980:297) and the *Times of India* "aimed primarily at the English educated intelligentsia" of the country (1980:331). Both also had British or British-trained editors and managers (1980:297, 332-333).

In the 2000s, as seen particularly clearly in Figure 10, journalistic writing in the Bahamas is no longer as different from that practiced in the other two postcolonial countries investigated here. JamE, especially, is much closer to BahE than it was in the 1960s. The development leading up to this result could be described as a case of region-specific "convergent change" (Leech et al. 2009:253), but it clearly does not imply that JamE is following BahE, even though JAM 2012 is located right between BAH 1968 and BAH 2012 in Figure 10. This would be far-fetched, given the pan-Caribbean influence that Jamaican varieties (but clearly more so the creole than standard English) have had via cultural exports such as reggae and Rastafarianism (cf. Mair 2013:263-265). What has happened is simply that BahE has not changed as much as JamE during decolonization, at least with regard to genitive variation. It began as comparatively AmE-like and still is, whereas JamE was more BrE-like in the 1960s but resembles 1960s AmE today. Like BahE, IndE has experienced comparatively little change, but other than the latter, it was far more conservative to begin with, and has remained comparatively so.

7. Conclusion

In this paper, we have investigated recent changes in genitive variation in five varieties of English based on a set of matching historical and contemporary newspaper corpora. Employing up-to-date statistical modeling, we have presented exactly the kind of "[r]esearch with a more explicitly diachronic twist based on longitudinal data sources" that Heller, Szmrecsanyi & Grafmiller (2017:23) call for in their synchronic comparison of cross-varietal differences in genitive variation. Our findings

corroborate the continued expansion of the *s*-genitive in present-day English that has been described in so many studies as well as initial findings with regard to postcolonial Englishes, particularly in the Caribbean, that suggest that these varieties are partaking in AmE-led global trends in grammar toward, e.g., densification, without actually approximating American norms. While JamE and IndE present examples of postcolonial lag, BahE looks more like a case of linguistic decolonization, appearing fairly American-oriented during colonial times but having aligned with other postcolonial Englishes after independence.

Other than a lot of previous work on the AmE impact on postcolonial varieties, the present study went beyond "surface" phenomena and frequency differences. We attempted to uncover variation and change in the underlying grammar of the genitive alternation, as defined by patterns of variability in the constraints affecting the variable. The varieties investigated here share a core grammar of genitive variation, to the extent that the direction of language-internal, and particularly production-related, constraints on the phenomenon is similar across varieties and time. Where our data sets differ is in the strength of these constraints and the effects that external factors such as text type exert. Whether the gap that appears to have opened up between metropolitan and postcolonial varieties in this respect will be found for other features remains to be seen. Clearly, more diachronic work is needed in order to answer this question.

Declaration of Conflicting Interests

The author(s) declared no potential conflict of interests with respect to the research, authorship, and/or publication of this article.

Funding

Data collection and preparation were partly funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) as part of the project "Standardization and language change in postcolonial speech communities: The Bahamas" (Hackert, April 2016 – March 2019, HA 3514/3-1).

Acknowledgements

We are very grateful to Dagmar Deuber and Tobias Bernaisch for allowing us to use JAM 1968 and IND SAVE, to Simone Bauer, Oliver Deck, Laura Herbig, Annette Hillers-Chen, Alexander Laube, Annamaria Spiegel, and Natascha Wanninger for help in collecting and digitizing the historical Caribbean and Indian newspapers and preparing the samples for analysis, to Bernd Ludwig for giving us access to computational superpower, to Tobias Bernaisch again for encouraging feedback on a first draft of this article, and to the two anonymous reviewers for their helpful comments and criticism. As always, all remaining errors are our own.

Notes

1. We also assessed the interrater reliability of our animacy classification; parallel coding of a random subset of 199 genitive possessors by two trained coders yielded a simple agreement rate of 86% and a Cohen's κ of 0.79.

2. In English, in general only indefinite NPs can fill the slot in existential constructions such as *There is/are* _____ *outside*. With definite NPs, special readings apply (Garretson 2004:19-20).

3. In order to substantiate this claim, we compared lexical density for the two text types in each of our samples. We opted for the most popular "minor variant" of calculating this parameter for texts of differing lengths, i.e., noun density (cf. Johansson 2008:65), and, in fact, the latter is consistently higher in news reports than in editorials. While LOB Press, for example, has 24% nouns in editorials,

in reportage this proportion rises to 28%. For the other corpora, we obtained the following (rounded) percentages: BE06 Press 23% vs. 27%, Brown Press 25% vs. 30%, AE06 Press 26.6% vs. 27.4%, BAH 1968 26% vs. 32%, BAH 2012 27% vs. 29%, JAM 1968 26% vs. 29%, JAM 2012 28% vs. 32%, IND 1968 26% vs. 36%, and IND SAVE 28% vs. 34%.

4. Z-scoring is a linear transformation which first centers values around the mean, which in our sample amounts to -0.25, and then divides them by the standard deviation, which in our case equaled 12. In other words, an absolute syntactic weight of 0 corresponds to -0.25 in Figure 4, whereas the z-scores must be multiplied by 12 in order to obtain the non-transformed values.

5. Contexts marked "n.a." refer to the very first genitive token in any individual corpus text.

6. The following pseudo R^2s were obtained: $RF_2(1)$ LOB Press: 0.92, $RF_2(2)$ Brown Press: 0.93, $RF_2(3)$ BE06 Press: 0.94, $RF_2(4)$ AE06 Press: 0.94.

7. We used centered means to account for deviations that occur in the reference models themselves. Even though we obtained very good *C* values for all RF_1s, it is clear that none of these analyses fully accounts for each and every genitive token in the underlying data set, which results in deviations from predicted outcomes. In a nutshell, since we did not want to compare our postcolonial varieties to perfect but non-existent yardsticks but to real ones, we calculated deviations not in absolute terms but by centering them around the mean deviations of each reference data set.

References

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

Appadurai, Arjun. 1990. Disjuncture and difference in the global cultural economy. *Public Culture* 2(2). 1-24.

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

Baker, Paul. 2009. The BE06 Corpus of British English and recent language change. International *Journal of Corpus Linguistics* 14. 312-337.

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

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

Biber, Douglas. 2003. Compressed noun-phrase structure in newspaper discourse: The competing demands of popularization vs. economy. In Jean Aitchison & Diana M. Lewis (eds.), *New media language,* 169-181. London: Longman.

Bruckmaier, Elisabeth & Stephanie Hackert. 2011. Bahamian Standard English: A first approach. *English World-Wide* 32. 174-205.

Census of India. 2011. Paper 1 of 2018: Language. New Delhi: Office of the Registrar General. https://www.censusindia.gov.in/2011Census/C-16_25062018_NEW.pdf (October 30, 2020).

Craton, Michael & Gail Saunders. 1998. *Islanders in the stream: A history of the Bahamian people*. Vol. 2: *From the ending of slavery to the twenty-first century*. Athens, GA: University of Georgia 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. 1-28.

Deuber, Dagmar, Stephanie Hackert, Eva Canan Hänsel, Alexander Laube, Mahyar Hejrani & Catherine Laliberté. Fc. The norm orientation of English in the Caribbean: A comparative study of newspaper writing from ten countries. *American Speech*.

Devonish, Hubert & Ewart A.C. Thomas. 2012. Standards of English in the Caribbean. In Raymond Hickey (ed.), *Standards of English: Codified varieties around the world*, 179-97. Cambridge: Cambridge University Press.

Feist, Jim. 2012. What controls the “genitive variation“ in present-day English? *Studies in Language*

36. 261-299.

Fuchs, Robert. 2020. The progressive in 19th and 20th century settler and indigenous English. *World Englishes* 39. 394-410. https://onlinelibrary.wiley.com/doi/epdf/10.1111/weng.12480 (December 9, 2020).

Gamer, Matthias, Jim Lemon, Ian Fellows & Puspendra Singh. 2019. irr: Various coefficients of interrater reliability and agreement. R package version 0.84.1. https://CRAN.R-project.org/package=irr (October 8, 2020).

Garretson, Gregory. 2004. Coding practices used in the project Optimal Typology of Determiner Phrases. Unpublished manuscript, Boston University, Boston, MA. http://npcorpus.bu.edu/html/documentation/index.html (July 27, 2018).

Gonçalves, Bruno, Lucía Loureiro-Porto, José J. Ramasco & David Sánchez. 2018. Mapping the Americanization of English in space and time. *PLoS ONE* 13.5: e0197741. https://doi.org/10.1371/journal.pone.0197741 (15 August 2019).

Görlach, Manfred. 2002. *Still more Englishes*. Amsterdam: Benjamins.

Grafmiller, Jason. 2014. Variation in English genitives across modality and genres. *English Language and Linguistics* 18. 471-496.

Gries, Stefan Th. & Tobias Bernaisch. 2016. Exploring epicenters empirically: Focus on South Asian Englishes. *English World-Wide* 37. 1-25.

Gries, Stefan Th., Tobias Bernaisch & Benedikt Heller. 2018. A corpus-linguistic account of the history of the genitive alternation in Singapore English. In Sandra C. Deshors (ed.), *Modeling world Englishes. Assessing the interplay of emancipation and globalization of ESL varieties*, 245-279. Amsterdam: Benjamins.

Gries, Stefan Th. & Sandra Deshors. 2014. Using regressions to explore deviations between corpus data and a standard/target: two suggestions. *Corpora* 9(1). 109-136.

Hackert, Stephanie. 2015. Pseudotitles in Bahamian English: A case of Americanization? *Journal of*

English Linguistics 43. 143-167.

Hackert, Stephanie. 2016. Standards of English in the Caribbean: History, attitudes, functions, features. In Elena Seoane & Cristina Suárez-Gómez (eds.), *World Englishes: New theoretical and methodological considerations*, 85-112. Amsterdam: Benjamins.

Hackert, Stephanie & Dagmar Deuber. 2015. American influence on written Caribbean English: A diachronic analysis of newspaper reportage in the Bahamas and in Trinidad and Tobago. In Peter Collins (ed.), *Grammatical change in English world-wide*, 389-410. Amsterdam: Benjamins.

Hackert, Stephanie & Magnus Huber. 2007. Gullah in the diaspora. Historical and linguistic evidence from the Bahamas. *Diachronica* 24. 279-325.

Hänsel, Eva Canan & Dagmar Deuber. 2013. Globalization, postcolonial Englishes, and the English language press in Kenya, Singapore, and Trinidad and Tobago. *World Englishes* 32. 338-357.

Heller, Benedikt, Tobias Bernaisch & Stefan Th. Gries. 2017. Empirical perspectives on two potential epicenters: The genitive alternation in Asian Englishes. *ICAME Journal* 41. 111-144.

Heller, Benedikt, Benedikt Szmrecsanyi & Jason Grafmiller. 2017. Stability and fluidity in syntactic variation world-wide: The genitive alternation across varieties of English. *Journal of English Linguistics* 45. 3-27.

Heller, Benedikt & Benedikt Szmrecsanyi. 2019. Possessives world-wide: Genitive variation in varieties of English. In Nuria Yáñez-Bouza, Emma Moore, Linda van Bergen & Willem B. Hollmann (eds.), *Categories, constructions, and change in English syntax*, 315-335. Cambridge: Cambridge University Press.

Hinrichs, Lars & Benedikt Szmrecsanyi. 2007. Recent changes in the function and frequency of Standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language and Linguistics* 11. 437-474.

Hothorn, Torsten, Peter Buehlmann, Sandrine Dudoit, Annette Molinaro & Mark van der Laan. 2006.

Survival ensembles. *Biostatistics* 7(3). 355-373.

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

Jespersen, Otto. 1909-49. *A modern English grammar on historical principles*. 7 vols. Copenhagen: Munksgaard.

Johansson, Victoria. 2008. Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund University Department of Linguistics and Phonetics Working Papers* 53. 61-79.

Kachru, Braj B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In Randolph Quirk & H.G. Widdowson, (eds.), *English in the world. Teaching and learning the language and literatures*, 11-30. Cambridge: Cambridge University Press.

Kapur, Akash. 2012. How India became America. *The New York Times*. https://www.nytimes.com/2012/03/11/opinion/sunday/how-india-became-america.html (August 22, 2020).

Laliberté, Catherine. 2016. A diachronic study of modality in Indian English. Munich: LMU Munich MA thesis.

Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press.

MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4(226). 1-16.

Mair, Christian. 2002. Creolisms in an emerging standard: Written English in Jamaica. *English World-Wide* 23. 31-58.

Mair, Christian. 2006. *Twentieth-century English: History, variation and standardization*. Cambridge: Cambridge University Press.

Mair, Christian. 2013. The World System of Englishes. Accounting for the transnational importance

of mobile and mediated vernaculars. *English World-Wide* 34. 253-278.

Majumdar, Manabi & Jos Mooij. 2012. The mark's race: India's dominant education regime and new segmentation. In Christine Sleeter, Shashi Bhushan Upadhyay, Arvind K. Mishra & Sanjay Kumar (eds.), *School education, pluralism and marginality. Comparative perspectives*, 207-246. New Delhi: Orient Blackswan.

Merrill, John C. & Harold A. Fisher. 1980. *The world's great dailies. Profiles of fifty newspapers*. New York: Hastings House.

Mesthrie, Rajend & Rakesh M. Bhatt. 2008. *World Englishes: The study of new linguistic varieties*. Cambridge: Cambridge University Press.

Meyer, Charles F. 2002. Pseudo-titles in the press genre of various components of the International Corpus of English. In Randi Reppen, Susan M. Fitzmaurice & Douglas Biber (eds.), *Using corpora to explore linguistic variation*, 147-166. Amsterdam: Benjamins.

Nadeem, Shehzad. 2011. Accent neutralisation and a crisis of identity in India's call centres. *The Guardian*. https://www.theguardian.com/commentisfree/2011/feb/09/india-call-centres-accent-neutralisation (August 22, 2020).

Nongrum, Catherine Meryl. 2018. *Collins English grammar and composition 6*. Noida, Uttar Pradesh: Collins India.

Oenbring, Raymond. 2010. Corpus linguistic studies of standard Bahamian English: A comparative study of newspaper usage. *The International Journal of Bahamian Studies* 16. 51-62.

Oenbring, Raymond & William Fielding. 2014. Young adult's attitudes to standard and nonstandard English in an English-creole speaking country: The case of the Bahamas. *Language, Discourse & Society* 3. 28-51.

Paradis, Emmanuel & Klaus Schliep. 2019. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35. 526-528.

Rau, M. Chalapathi. 1958. India. In *The training of journalists. A world-wide survey on the training*

of personnel for the mass media. Paris: The United Nations Educational, Scientific and Cultural Organization, 144-9. https://unesdoc.unesco.org/ark:/48223/pf0000043231 (November 1, 2020).

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (October 8, 2020).

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

Rosenbach, Anette. 2008. Animacy and grammatical variation – Findings from English genitive variation. *Lingua* 118. 151-171.

Rosenbach, Anette. 2014. English genitive variation – the state of the art. *English Language and Linguistics* 18. 215-262.

Salwen, Michael B. & Bruce Garrison. 1991. *Latin American journalism*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schneider, Edgar W. 2003. The dynamics of New Englishes: From identity construction to dialect birth. *Language* 79. 233-281.

Schneider, Edgar W. 2006. English in North America. In Braj B. Kachru, Yamuna Kachru & Cecil L. Nelson (eds.), *The handbook of World Englishes*, 58-73. Oxford: Blackwell.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.

Scott, Mike. 2012. WordSmith Tools, version 6. Stroud: Lexical Analysis Software.

Storr, Juliette. 2000. *Changes and challenges: A history of the development of broadcasting in the Commonwealth of the Bahamas, 1930-1980*. Athens, OH: Ohio University dissertation.

Storr, Juliette. 2016. *Journalism in a small place: Making Caribbean news relevant, comprehensive, and independent*. Calgary, Alberta: University of Calgary Press.

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

Mouton de Gruyter.

Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide* 37. 109-137.

Szmrecsanyi, Benedikt, Jason Grafmiller, Joan Bresnan, Anette Rosenbach, Sali Tagliamonte & Simon Todd. 2017. Spoken syntax in a comparative perspective: The dative and genitive alternation in varieties of English. *Glossa* 2(1). 86. 1-27.

Tagliamonte, Sali A. 2013. Comparative sociolinguistics. In J.K. Chambers & Natalie Schilling (eds.), *The handbook of language variation and change*. 2nd edn., 128-156. Malden, MA: Wiley-Blackwell.

Wolk, Christoph, Joan Bresnan, Anette Rosenbach & Benedikt Szmrecsanyi. 2013. Dative and genitive variability in Late Modern English. *Diachronica* 30. 382-419.

Wulff, Stefanie & Stefan Th. Gries. 2015. Prenominal adjective order preferences in Chinese and German L2 English: A multifactorial corpus study. *Linguistic Approaches to Bilingualism* 5(1). 122-150.

TABLE 1

Corpora Used for the Present Study

Corpus	Variety	Year(s) covered	Word count	Composition
LOB Press	BrE	1961	143,022	LOB sections A (reportage: ca. 88,700 words), B (editorials: ca. 54,300 words)
BE06 Press	BrE	2003-08 (2005-07)	160,803	BE06 sections A (reportage: ca. 100,000 words), B (editorials: ca. 61,000 words)
Brown Press	AmE	1961	143,158	Brown sections A, B
AE06 Press	AmE	mostly 2006	163,696	AE06 sections A, B
BAH 1968	BahE	1968	181,823	reportage (ca. 120,000 words) + editorials (ca. 60,000 words) from the *Nassau Guardian and Bahamas Observer* (*NGBO*)
BAH 2012	BahE	2011/12	180,017	reportage + editorials from the *Nassau Guardian* and the *Freeport News*
JAM 1968	JamE	1968	180,769	reportage + editorials from the *Kingston Gleaner*
JAM 2012	JamE	2012	180,336	reportage + editorials from the *Jamaica Gleaner*, the *Western Mirror*, and the *Jamaica Observer*
IND 1968	IndE	1968	181,108	reportage + editorials from the *Times of India*
IND SAVE	IndE	2002-05	180,077	reportage + editorials from the *Times of India* and the *Statesman*

TABLE 2

Animacy Categories with Examples from Our Data

Category	Examples
Animate	*the Prime Minister's special committee* (JAM 1968 KG_Ed018.txt) *the office of the Prime Minister* (JAM 1968 KG_Ed062.txt)
Collective	*the government's Department of Food and Rural Affairs* (BE06_A34) *the responsible agencies of her Government* (JAM 2012 JamEd100)
Inanimate	*the system's antenna* (BAH 2012 Pr_138) *the weakness of our system* (IND SAVE TI_37763)
Locative	*Iraq's borders* (AE06_B22) *the borders of Iraq* (BE06_A35)
Temporal	*last year's general election campaign* (BE06_A08) *the Manipur Assembly arson case of last year* (IND SAVE SM_2002-08-13)

TABLE 3

Predictive Accuracies (in Percent) for Relevant Pairs of Metropolitan and Postcolonial Englishes

	LOB Press	Brown Press	BE06 Press	AE06 Press
BAH 1968	**75.4**	74.4		
JAM 1968	**80**	78.9		
IND 1968	**77.5**	75.1		
BAH 2012			**74.2**	72.8
JAM 2012			**77.6**	74
IND SAVE			**74.2**	71.1

Figure 1: Genitive Types by Data Set

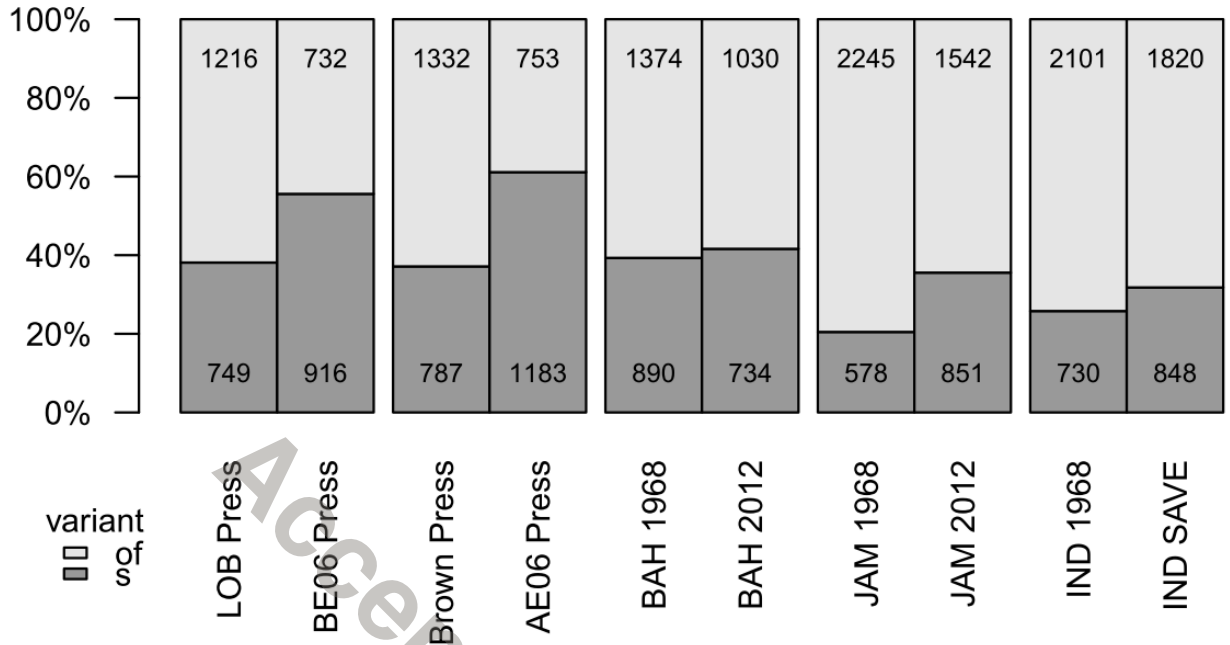


Figure 2: *S*-Genitive Probabilities by Animacy and Data Set

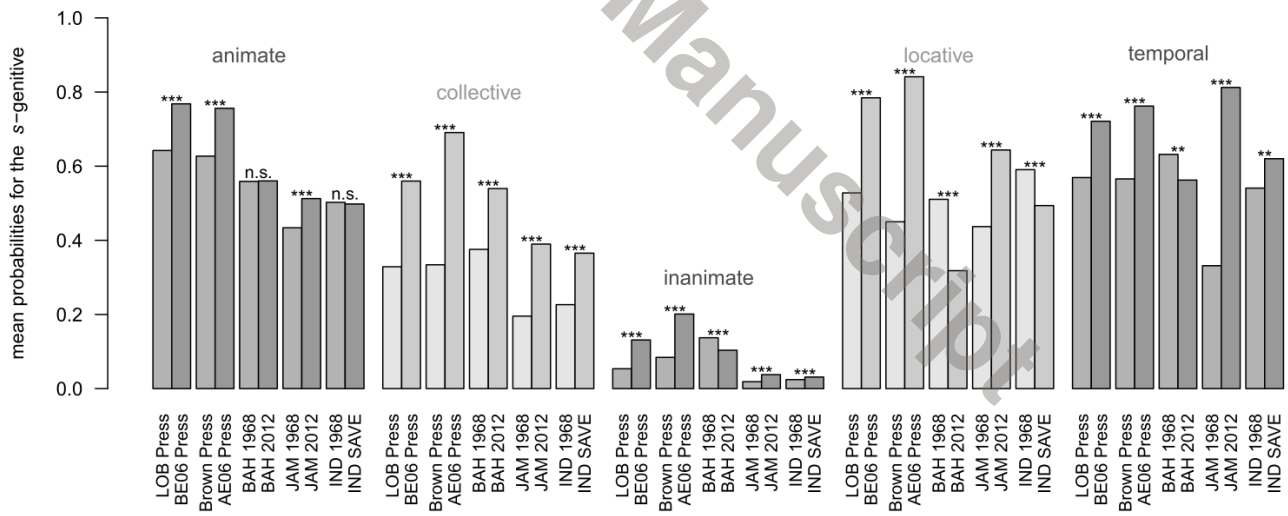


Figure 3: *S*-Genitive Probabilities by Semantic Relation and Data Set.

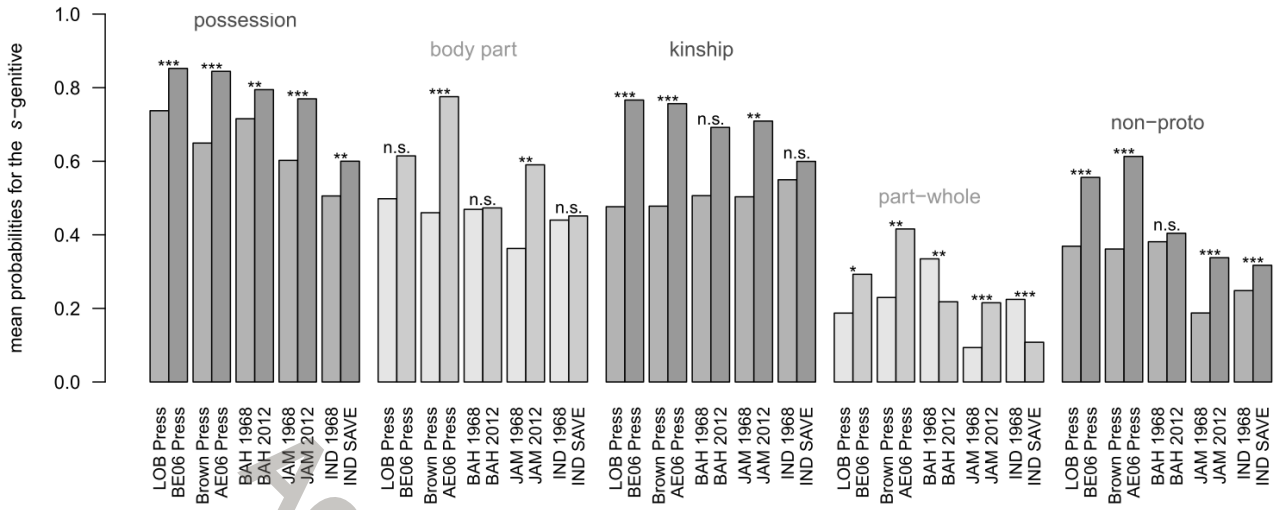


Figure 4: *S*-Genitive Probabilities by Syntactic Weight and Data Set

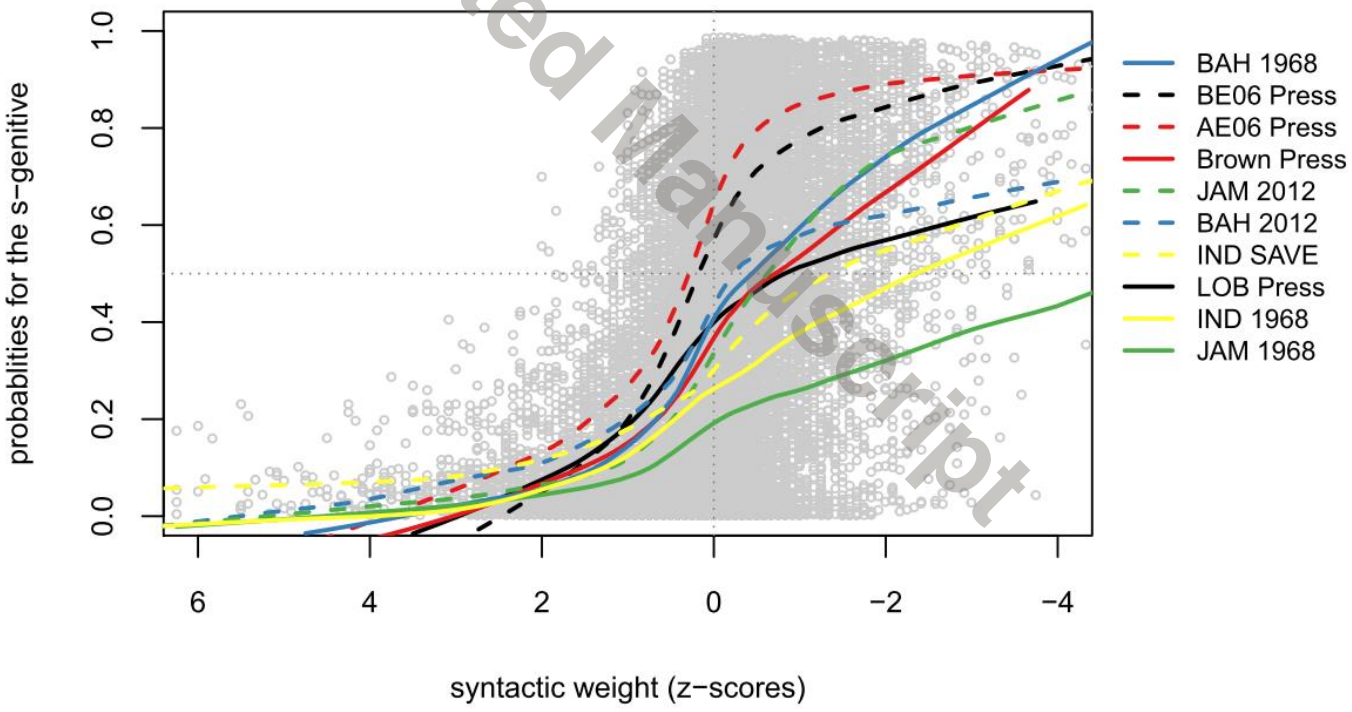


Figure 5: *S*-Genitive Probabilities by 1. Definiteness, 2. Givenness, 3. Persistence, 4. Possessor Sibilancy, 5. Embedding, 6. Text Type, and Data Set

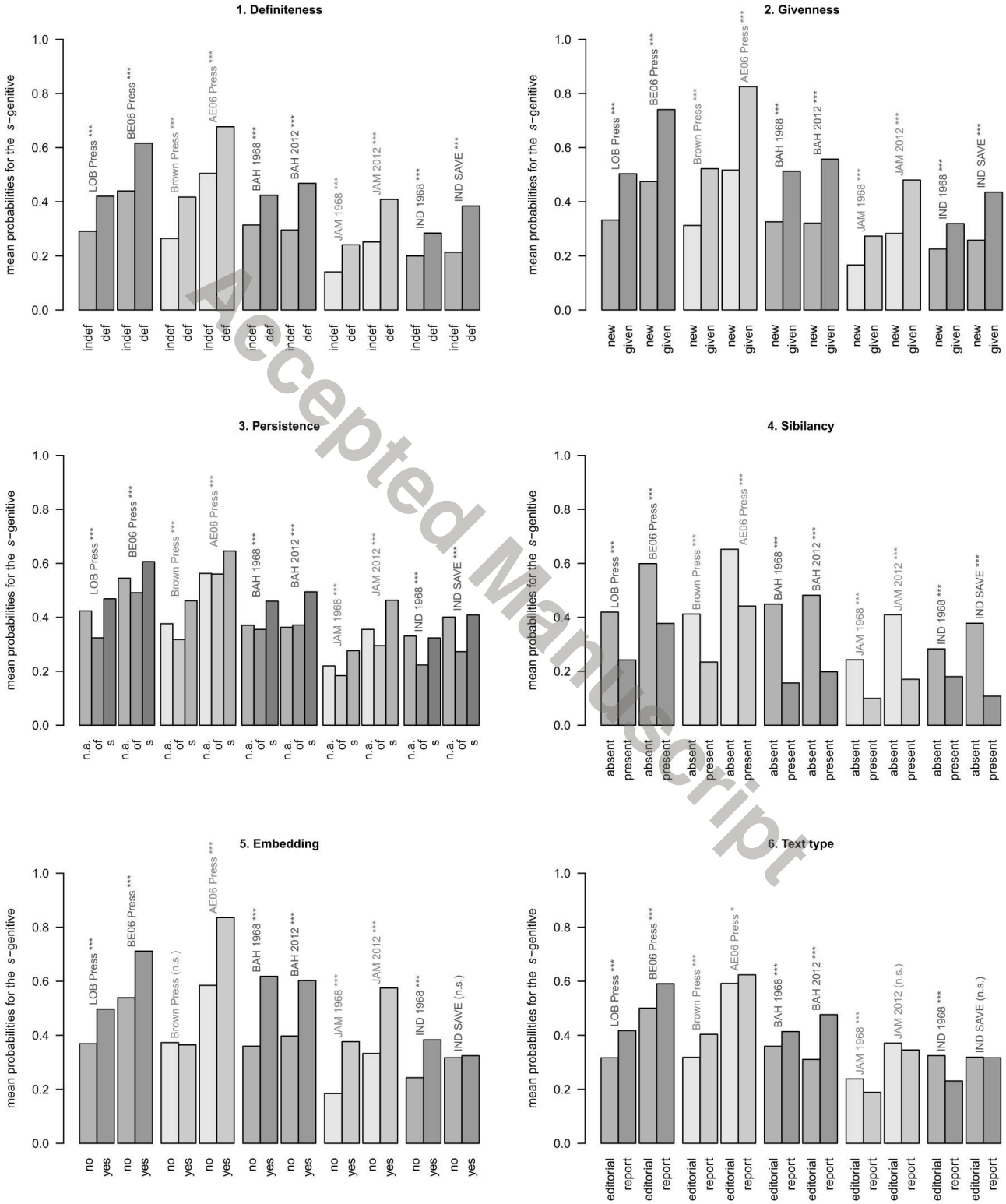


Figure 6: Predictor Importance Ranking for 1960s and Contemporary Data Sets by Variety

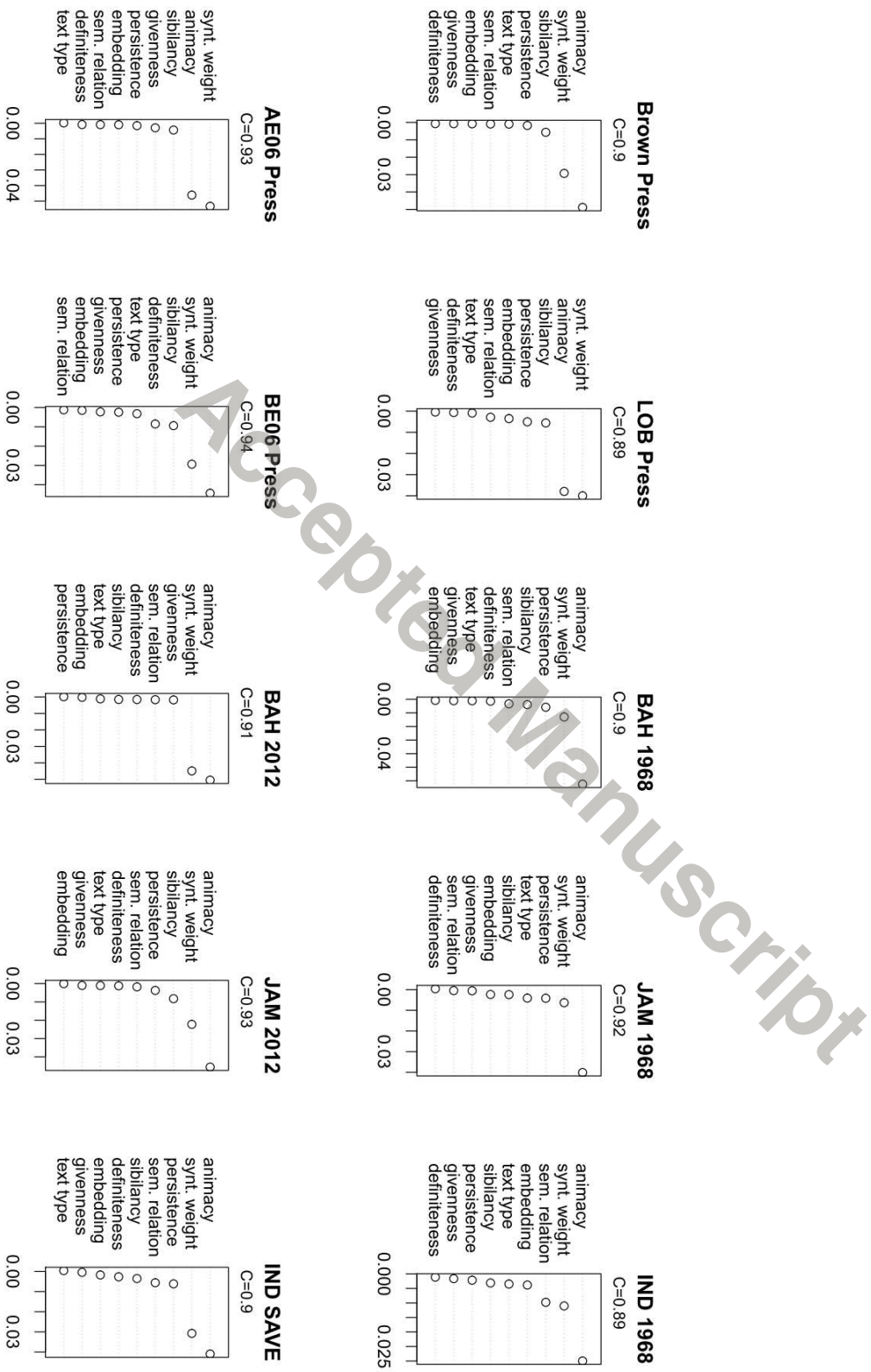


Figure 7: Steps Involved in the Present MuPDAR-F Analysis

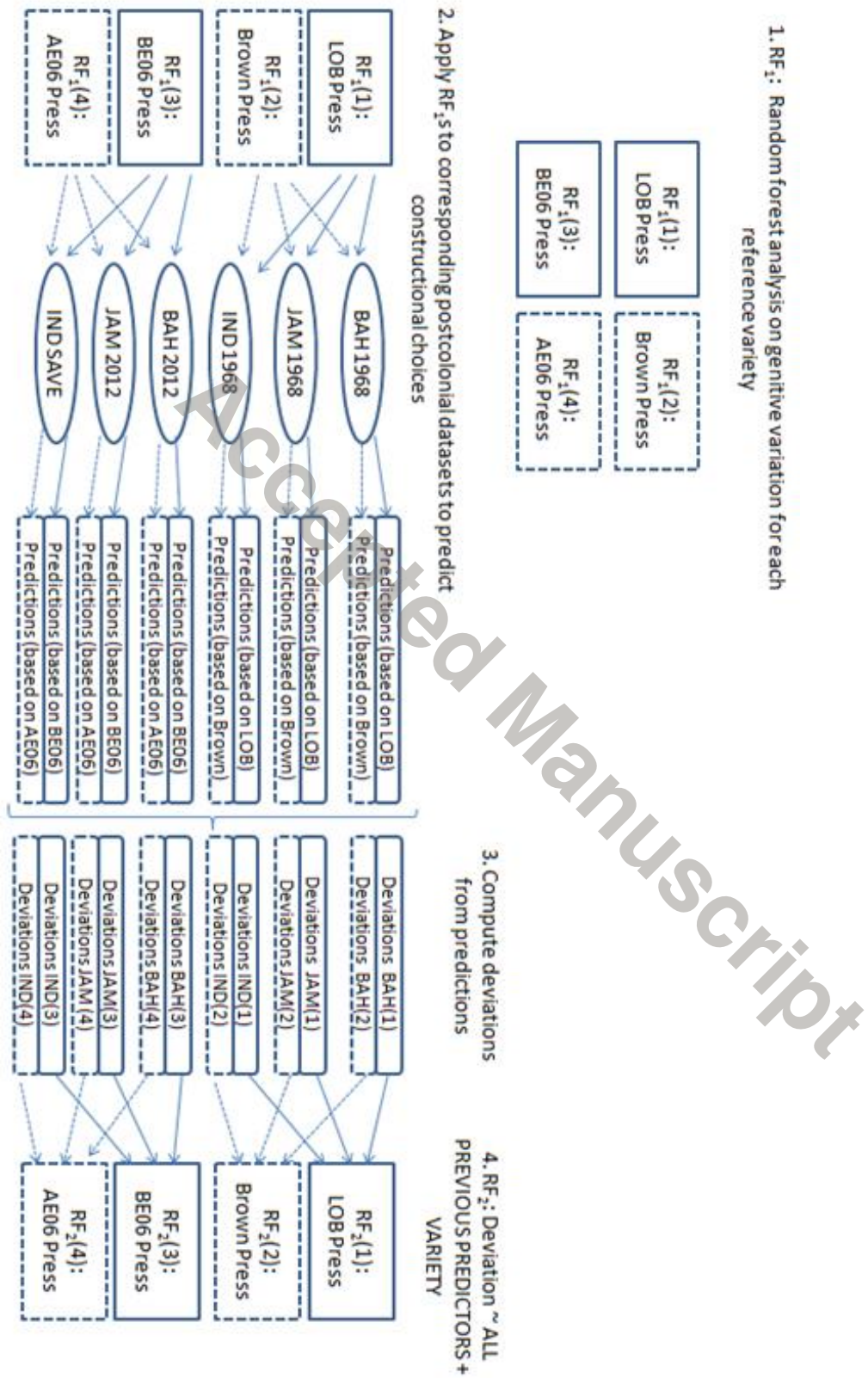


Figure 8: Centered Predicted Deviation Scores by Animacy for Postcolonial Data Sets, Historical (Left) and Contemporary (Right)

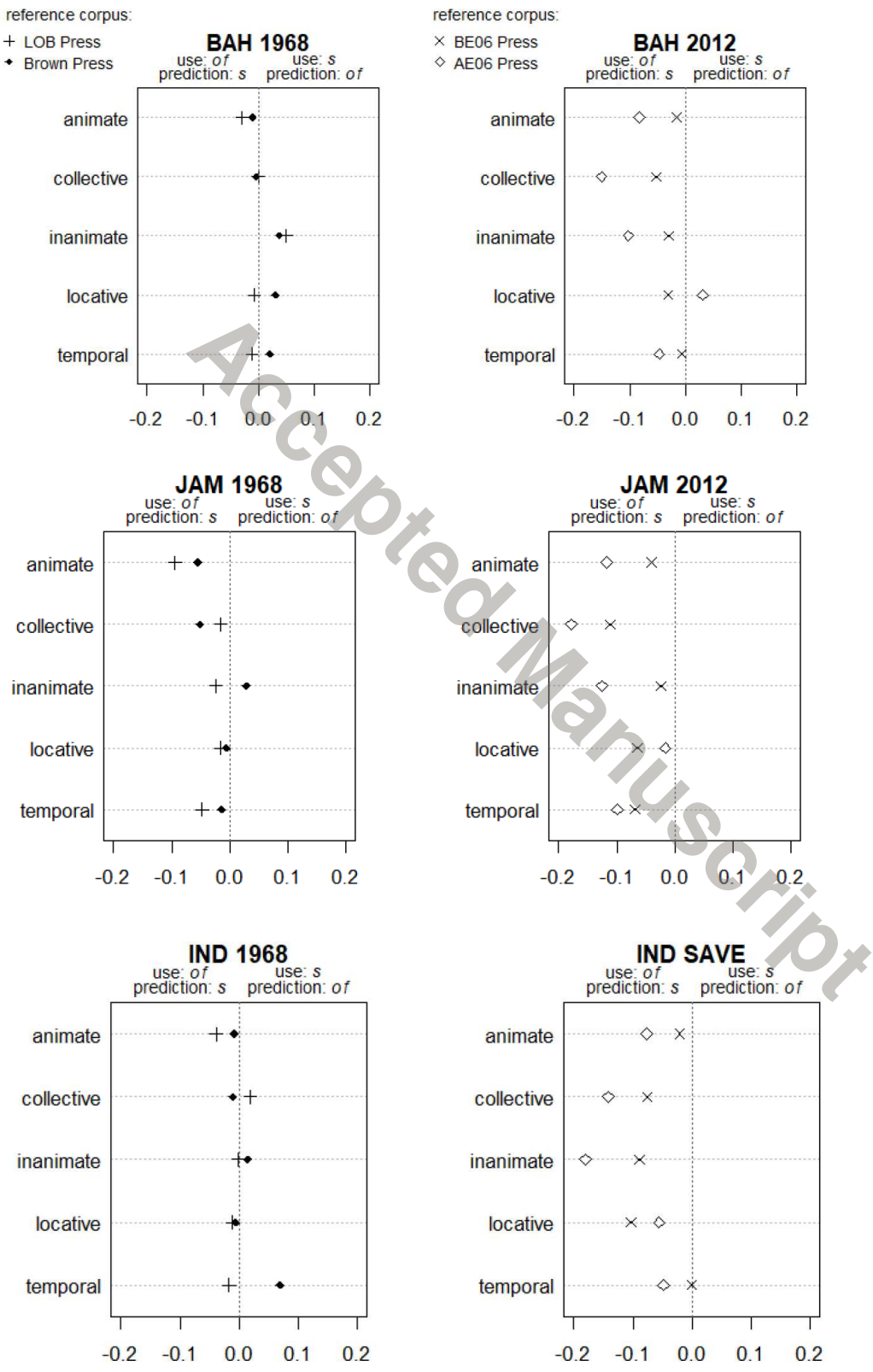
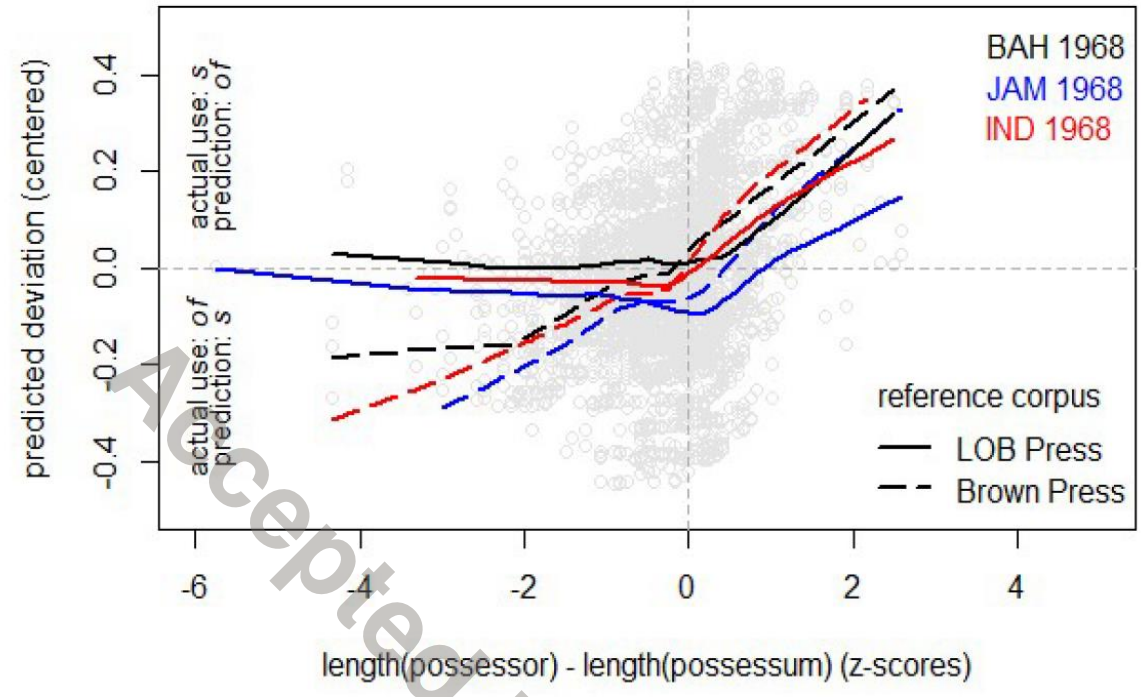


Figure 9: Centered Predicted Deviation Scores by Syntactic Weight for Postcolonial Data Sets, Historical (Top) and Contemporary (Bottom)



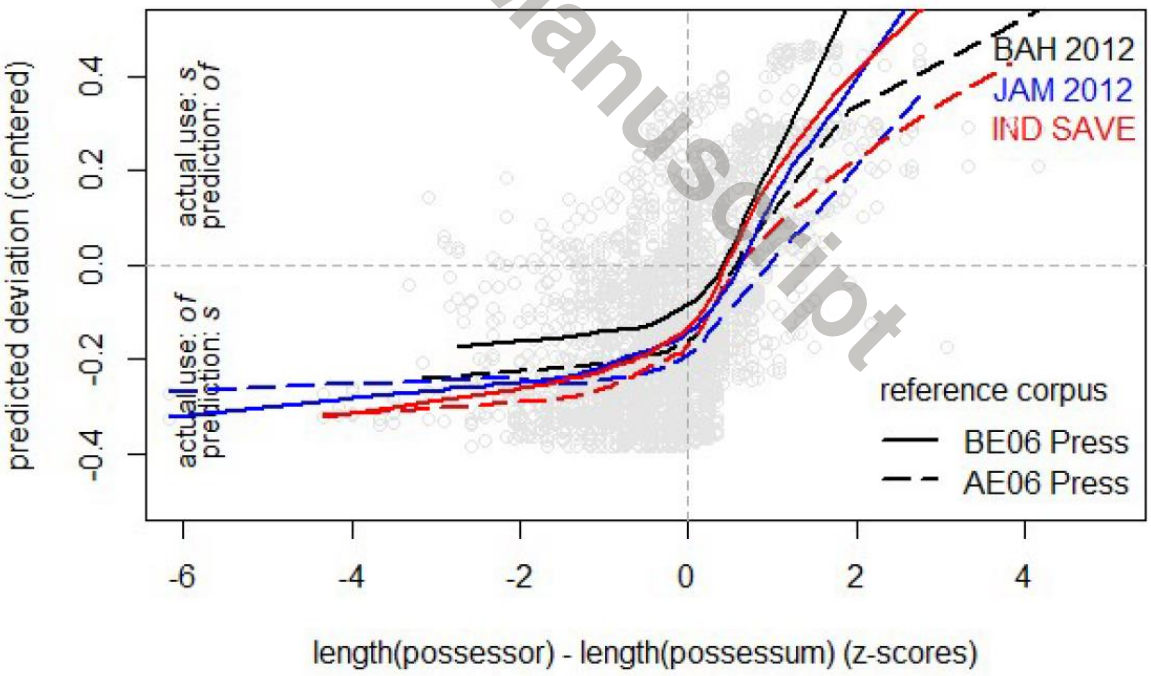
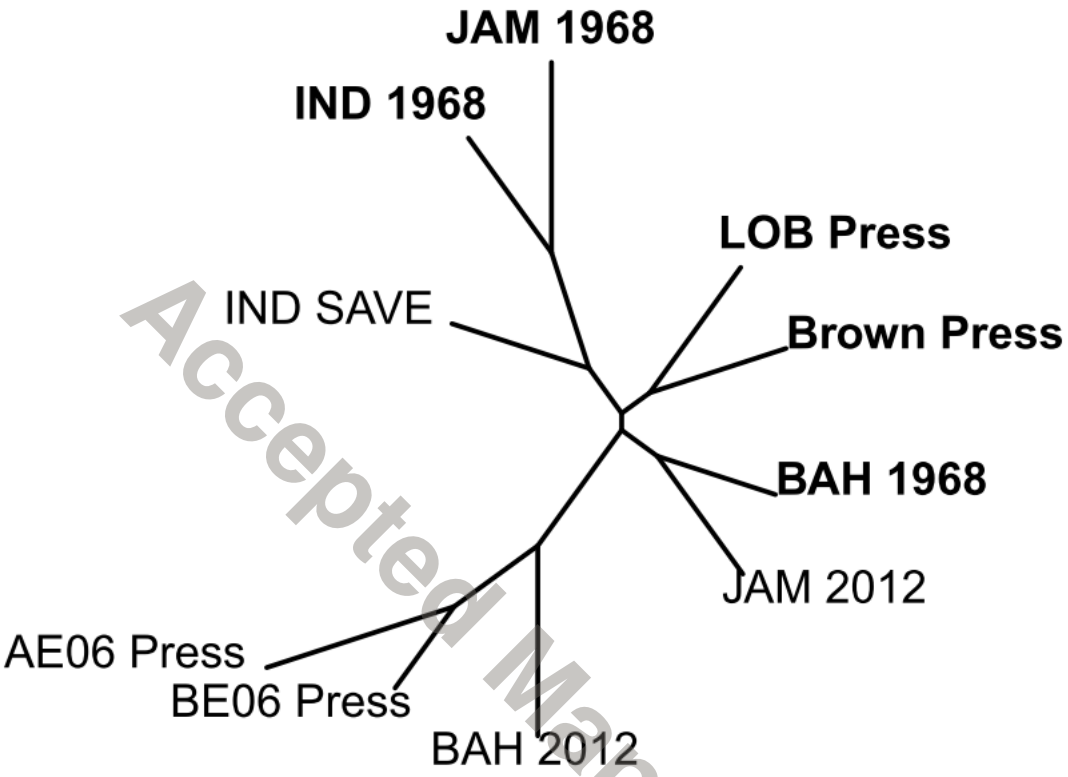


Figure 10: Phylogenetic Cluster Analysis of Varieties' Predicted Deviations in Terms of *S*-Genitive Probabilities



Author Biographies

Stephanie Hackert obtained her Ph.D. from the University of Heidelberg and has taught at various German universities. She now holds a Chair of English Linguistics at the University of Munich (LMU). Her research interests center around pidgins and creoles, variationist sociolinguistics, corpus linguistics, recent grammatical change, and historical discourse analysis. Her publications include monographs on urban Bahamian Creole English (Benjamins, 2004) and the history of the English native speaker (de Gruyter, 2012).

Email: st.hackert@lmu.de

Diana Wengler is a Ph.D. candidate and research assistant in English linguistics at LMU Munich. She is currently completing her dissertation on historical audio recordings from the Bahamas, looking into the recent history of Bahamian Creole as well as issues of dialect representation and the phonology-grammar interface in language change.

Email: diana.wengler@lmu.de