

*Gender, Writing and Editing in  
South African Englishes  
A Case Study of the Genitive Alternation*

*Melanie A. Law and Haidee Kotze*

### 9.1 Introduction

Of the social factors that are known to influence linguistic behaviour, gender is considered to be among the clearest and most consistent (see Labov 1990: 205).<sup>1</sup> A substantial body of cognitive linguistic, sociolinguistic and variationist research demonstrates that there are differences in the linguistic behaviour of men and women and that men and women play different roles in processes of language variation and change (e.g. Lakoff 1977; Milroy and Milroy 1993; Holmes 1997; Romaine 1999; Nevalainen 2000). These roles give rise to an apparent gender paradox: women are more likely than men to conform to variants that are overtly prescribed or standard and thus prestigious forms (Trudgill 1972, 1983), but women are also quicker than men to adopt innovations and diffuse forms that are not overtly prescribed (in other words, non-standard forms) within the speech community (see Labov 2001: 291; Tagliamonte 2012: 63). Women thus apparently orient their linguistic choices towards both linguistic features that reflect overt, prestige norms (thereby acting as a conservative force) and innovative usages that reflect emerging covert, communal norms (thereby acting as a progressive force). Other researchers, however, point out that it is not so much the case that women adhere to or favour standard or prestige forms – but rather that they are instrumental in creating and establishing these norms (see Milroy and Milroy 1993: 65). This view resolves the apparent gender paradox: women are ‘instrumental in setting new standards’ because of their ‘openness to innovations, stylistic

<sup>1</sup> The distinction between gender and sex offers two ways to approach differences in linguistic behaviour. *Sex* refers to the biological distinction between males and females, while *gender* is regarded as a socially constructed factor. In this study, we follow Labov’s (2001: 263) view that ‘language is not differentiated by the biological aspects of sex differences’ and regard gender-differentiated linguistic behaviour as a result of the causal factors involved in the social instantiation of gender roles.

flexibility, contextual sensitivity, and variety of domestic roles' (Nevalainen 2000: 53). Thus, innovation and standardisation are two sides of the same genderlectal coin as far as women's linguistic behaviour is concerned.

Differences in orientation towards normativity in male and female language use has been an important concern in studies of gender variation in language. Norm orientation is also crucial in understanding processes of ongoing language change generally and, within the context of the evolutionary development of World Englishes specifically, for understanding the progression of postcolonial varieties towards endonormativity. Normativity is a complex phenomenon in which multiple dimensions interact and form part of the socially motivated processes of conventionalisation, diffusion and the establishment of shared linguistic usage (Kruger and Van Rooy 2017; Kotze 2020). These dimensions include the distinction and interaction between individual cognitive and collective social norms, explicit and implicit norms, spoken and written norms and overtly codified and covertly emergent norms (Kotze 2020). Given what is known about genderlectal variation and the dual role of women as both conservative and progressive forces in processes of variation and change, it may be argued that an understanding of genderlectal variation in the evolutionary development of different varieties of English is essential to gain a deeper understanding of the development of varieties towards endonormativity.

The investigation in this chapter focuses on the role of gender variation in processes of ongoing change in World Englishes, casting this through the lens of editorial practices in one variety of English, namely South African English (SAfE). SAfE consists of various strands, including two indigenous strands, Afrikaans English (AfrE) and Black South African English (BSAfE), and the settler strand, White South African English (WSAfE). Within this context, the normative environment is heterogeneous and competition arises between developing local norms and the overt norms of an extraterritorial variety, British English (BrE), adding another dimension to the already composite notion of normativity. Normativity is central to Schneider's (2007) Dynamic Model, which proposes a uniform underlying developmental process in which postcolonial varieties of English move through five stages, progressing from an exonormative orientation in the early stages of the development of the variety (Stages 1 through 3) to an endonormative orientation in the later stages (Stages 4 and 5). A number of social, political and linguistic criteria are set for the classification of a variety as having attained endonormativity, one of which is the acceptance within the language community that the

local linguistic norm is adequate in formal usage which ‘can be imposed only for formal written domains’ (Schneider 2007: 50).

There is limited corpus-based research on gender-differentiated linguistic behaviour in SAfE. Existing research (e.g. Mesthrie 2017) focuses on sociophonetics and genderlectal variation in writing has hardly been considered. Furthermore, in the World Englishes paradigm, the influence of editorial practices on innovations or shifting usage patterns in written registers has only recently started to receive attention (Van Rooy and Kruger 2016; Kruger and Van Rooy 2017; Law 2019; Kotze 2020). Much of this work argues that editors’ assessments of these features serve as a particularly stringent point for their acceptance into published written texts. The analyses of the interventions of editors might therefore provide information on the conventionalisation of these features and might also provide some measure of the endonormative stabilisation of a variety. The role of women in this process is important: industry-based surveys of editors reveal that the overwhelming majority are female (e.g. IPEd 2016; Milliot 2018). Thus, the editorial intervention that characterises published written texts provides a unique opportunity to investigate the intersection of gender and normativity in World Englishes and the potential role of editorial intervention in the processes through which varieties attain endonormativity.

In this chapter, we use random forests and generalised linear mixed-effects decision-tree modelling to analyse the effects of gender amongst a set of linguistic and extralinguistic variables conditioning the genitive alternation (i.e. the choice between the *s*-genitive, as in *the table’s legs* and the *of*-genitive, as in *the legs of the table*) in three subvarieties of SAfE (AfrE, BSAfE and WSAfE). Section 9.2 presents an overview of the genitive alternation and the various linguistic and extralinguistic factors known to influence this alternation. Section 9.3 discusses writing and editing in the three subvarieties in South Africa and situates this within the World Englishes context. The data and method used in the study are presented in Section 9.4, while Section 9.5 presents the results and analysis. Section 9.6 offers some conclusions and final remarks.

## 9.2 The Genitive Alternation

The genitive alternation, or the choice between different constructions to express similar genitive relations, is probably the best researched of all syntactic alternations in the English language (Rosenbach 2014). The consensus in most research into this alternation in English is that two

constructions, the *s*-genitive and the *of*-genitive, are the two major adnominal constructions available to express possessive relations and, in line with the variationist tradition (Labov 1972), that there is potential for either construction to be used in the same context to express approximately the same meaning. The two constructions each comprise two noun phrases, the possessor and the possessum, which are linked via a relational marker. In the *s*-genitive, the possessor occurs before the possessum (in prenominal position) and is linked to it via the 's (see [1]). In the *of*-genitive construction, the possessor occurs after the possessum (in postnominal position) and is linked to it via the preposition *of* (see [2]) (Rosenbach 2002, 2017).

- (1) The table<sub>[POSSESSOR]</sub> 's legs<sub>[POSSESSUM]</sub>.  
 (2) The legs<sub>[POSSESSUM]</sub> of the table<sub>[POSSESSOR]</sub>.

Historically, the alternation between the two constructions has been in a constant state of fluctuation (see Rosenbach [2014] for an overview and also Chapter 6). In Present-Day English, the *s*-genitive represents the incoming form and has slowly been increasing in frequency at the expense of the *of*-genitive. Most work on genitive alternation in Present-Day English has focused on either American English (AmE) or BrE and has shown that AmE appears to be leading this change, while BrE lags behind (Rosenbach 2002; Hinrichs and Szmrecsanyi 2007; Szmrecsanyi and Hinrichs 2008).

The choice between the two constructions is not random but is influenced by the interaction of a number of linguistic and extralinguistic factors, with considerably more work having been done on the conditioning effects of linguistic factors (Rosenbach 2014). The most important linguistic factors known to condition the genitive alternation include the animacy of the possessor, the definiteness of the possessor, the length of the possessor and possessum, final sibilancy of the possessor and the semantic relation between the possessor and possessum. The *s*-genitive is generally favoured with possessors that are animate, definite, shorter in length than possessums, that do not have a final sibilant or which express prototypical semantic relations (Rosenbach 2014). The *of*-genitive is favoured for possessors that are inanimate, indefinite, longer in length than possessums, that have a final sibilant or which express non-prototypical semantic relations.

Recently the focus has shifted to the influence of extralinguistic factors in conditioning the alternation. For example, Grafmiller (2014) investigates the interaction of several linguistic factors across spoken and written language and, within written language, across five different published registers in AmE. He finds that modality is an important factor in

conditioning genitive alternation, but also that there is variability across different written registers with respect to the importance of certain linguistic variables (which, he argues, is moderated by the stylistic norms of the different registers). In a similar vein, some recent work on cross-varietal differences in varieties of English other than BrE or AmE reveals that while the influence of the different linguistic factors is relatively stable in different varieties of English, the strength of these factors differs across varieties (e.g. Szmrecsanyi et al. 2016; Heller et al. 2017a, 2017b; Rosenbach 2017; Szmrecsanyi et al. 2017). Heller et al. (2017b) hypothesise that the strength of the linguistic factors may be proportional to how advanced a variety is in terms of the Dynamic Model (i.e. the further along the stages of the Dynamic Model a variety is, the stronger the effect of the factor is), while Rosenbach (2017) argues that in second-language varieties of English contact with substrate languages might account for the differences in the strength of the linguistic factors.

Despite the substantial body of research on this alternation and the growing interest in (and evidence to support) the subtle variability introduced by the influence of extralinguistic factors, hardly any investigations of this alternation have included gender as a conditioning factor and, to the best of our knowledge, only two studies have investigated this alternation within SAfE, Rosenbach (2017) and Law (2019).<sup>2</sup> SAfE therefore offers a unique opportunity to investigate the interaction between gender, writing and editing, and linguistic and extralinguistic factors in conditioning the genitive alternation.

### 9.3 Writing and Editing in South African Englishes

The linguistic diversity and widespread multilingualism of South Africa are well known and while English is the most widely used language in formal written communication, it is used as a home language by only 8.1% of the population (see StatsSA 2019: 8). The use of English in South Africa presents a unique challenge in models of World Englishes: it is difficult to map to Kachru's Three Circles Model because it qualifies as both an Inner-Circle (in terms of first-language users) and an Outer-Circle (in terms of second-language users) variety. Furthermore, the complex conditions

<sup>2</sup> Ford and Bresnan (2015) include gender as a conditioning variable in their analysis of AmE and Australian English (AusE) and find that men show a slightly higher preference for *s*-genitives and women a slightly higher preference for *of*-genitives. The difference between the two genders is not significant.

under which English in South Africa has evolved and continues to evolve have given rise to a variety that is unique in its internal heterogeneity: several distinct subvarieties are developing within the same context and are in close contact with one another and while there is some evidence of linguistic convergence in terms of a handful of features at the levels of pronunciation and lexis and (to a lesser extent) grammar, different ethnic and social identities persist (Van Rooy 2014).

Three prominent subvarieties of English in South Africa are the settler strand WSAfE, as well as two indigenous strands, namely, AfrE (a second-language variety used by first-language users of Afrikaans) and BSAfE (a mostly second-language variety used by black South Africans). WSAfE and BSAfE hold special positions within the complex South African sociolinguistic landscape: WSAfE is a visible and influential subvariety and is usually regarded as the standard prestige form towards which many users of the indigenous strands aspire, even if many WSAfE users regard BrE as the standard prestige form (Van der Walt and Van Rooy 2002; De Klerk 2003; Mesthrie 2017). BSAfE, on the other hand, accounts for the overwhelming majority of English users in South Africa, with some scholars noting the emergence of norms in BSAfE and the growing status of the subvariety (De Klerk 2003). WSAfE and AfrE have been in continual contact for an extended period of time and some evidence of linguistic convergence between the two subvarieties can be found at the levels of lexis and grammar. There has been less contact between BSAfE and the other two subvarieties as a result of apartheid, which has had significant effects on the development of this subvariety.

This complexity has led some scholars, like Van Rooy (2014), to argue that in countries like South Africa, where multiple subvarieties of English develop in complex contact settings, the analysis of the endonormative status of the variety should be studied relative to the subvarieties present in the contact situation, rather than from an overarching approach in which the different subvarieties are absorbed into an all-encompassing assessment of endonormativity. Several studies have found support for this shift in perspective, with a handful framing the editorial acceptance of the patterns of use in the written forms of such subvarieties as an indication of the endonormative status at the local, subvarietal level (e.g. Kruger and Van Rooy 2017).

An important point in respect of this study is that there is potential for the substrate languages to influence the linguistic choices of users of the two second-language subvarieties in South Africa. In terms of the genitive alternation, Rosenbach (2017) shows this to be the case in AfrE.

Table 9.1 *Genitive constructions available in English and three substrates of second-language English users in South Africa (Law 2019)*

	Prenominal genitive			Postnominal genitive		
	Possessor	Relational marker	Possessum	Possessum	Relational marker	Possessor
<b>English</b>	<i>Harry</i>	’s	<i>heart</i>	<i>the corner</i>	<i>of</i>	<i>the street</i>
<b>Afrikaans</b>	<i>Harry</i>	<i>se</i>	<i>hart</i>	<i>die boek</i>	<i>van</i>	<i>die straat</i>
<b>Southern Sotho</b>		∅		<i>pelo</i>	<i>ya</i>	<i>Harry</i>
				[heart]	[of]	[Harry]
				<i>khona</i>	<i>ya</i>	<i>seterata</i>
				[corner]	[of]	[street]
<b>Zulu</b>		∅		<i>inhliziyo</i>	<i>ka</i>	<i>Harry</i>
				[heart]	[of]	[Harry]
				<i>ikhona</i>	<i>ye</i>	<i>stratu</i>
				[corner]	[of]	[street]

In Afrikaans, as in English, two constructions are available to express genitive relations: one in which the possessor is realised before the possessum and is linked to it via the *se* relational marker and one in which the possessor occurs after the possessum and is linked to it via the *van* relational marker (see Table 9.1 from Law [2019], which is adapted from Rosenbach [2017]). There is therefore strong overlap between English and Afrikaans in the structure and semantics of the genitive constructions (Rosenbach 2017). However, in Afrikaans, the animacy of the possessor is a less strong factor in conditioning the genitive alternation than it is in English and Rosenbach (2017) finds that this carries over to the genitive choices of AfrE users, who use *s*-genitives in contexts where native English users would be inclined to use the *of*-genitive. In Southern Sotho and Zulu (two of the most widely spoken indigenous languages among black South Africans), only one construction is available to express genitive relations (in which the possessor is realised after the possessum) (see Table 9.1). It therefore stands to reason that if contact with substrate languages influences the genitive alternation, AfrE users would use *s*-genitives in contexts where WSAfE users would not and BSAfE users would show a stronger preference for *of*-genitive constructions.

As already discussed in Section 9.1, all written published texts undergo some form of editorial intervention and in South Africa these editors are usually either WSAfE or AfrE users (Law 2019). Thus, while multiple subvarieties are represented in the writing of published English texts in

South Africa, the language of these texts is subjected to the acceptability judgements of a corps of editors who may not be users of the subvariety being edited. Most notably, this also means that the writing of BSAfE authors is subjected to the acceptability judgements of either WSAfE or AfrE editors. Furthermore, the editors of published written English texts in South Africa are, as in other countries, overwhelmingly older females (Law 2019), opening up the possibility for gender to play an important role in the kinds of acceptability judgements that these editors make.

The corpus-based analysis of the unedited writing of AfrE, BSAfE and WSAfE authors and the editorial choices of the editors of these texts presented in this study thus provides a unique opportunity to investigate the dual role of women in the evolutionary development of different (sub) varieties of English (in South Africa) and to gauge whether there are gender differences in the propagation of new standards in the different subvarieties, by both authors and editors.

## 9.4 Data and Method

### 9.4.1 Corpus

The corpus used in this study is a register-differentiated corpus representing unedited SAfE texts and their edited counterparts, which were written and edited in South Africa since approximately 2000. The construction of the corpus is ongoing and, at the time of writing, comprises just over 6 million words spread over 380 texts and representing 5 written, published registers (see Table 9.2).

Table 9.2 *Token count per register, subvariety and unedited and edited texts in SAfE corpus*

	AfrE		BSAfE		WSAfE	
	Unedited	Edited	Unedited	Edited	Unedited	Edited
<b>Academic</b>	926,328	914,753	302,589	305,091	45,604	46,536
<b>Creative</b>	0	0	0	0	405,994	365,912
<b>Instructional</b>	275,058	281,767	58,013	63,519	367,430	356,002
<b>Popular</b>	55,148	57,021	21,471	21,351	145,977	175,609
<b>Reportage</b>	14,158	14,032	31,020	27,961	0	0
<b>Total</b>	1,270,692	1,267,573	413,093	417,922	965,005	944,059



The corpus is clearly not balanced across subvarieties and registers and there are far fewer tokens for BSAfE than there are for the other two subvarieties. In addition, there are no data for the creative register in AfrE and BSAfE, nor for WSAfE in reportage. The causes of this imbalance are twofold. First, in some registers for some subvarieties texts are not originally produced in English (e.g. Afrikaans authors do not typically write creative texts in English). Second, there are considerable challenges associated with obtaining access to unedited written texts.

#### 9.4.2 Data Extraction and Coding

The data were extracted using the concordance tool in *Sketch Engine*. Because this corpus comprises unedited texts aligned to their edited counterparts, a two-pronged approach to the extraction of the data was used. In the first step, a random sample of 1,000 instances of each variant (*'s* and *s'* and *of*) was extracted from each unedited register (together with its aligned segment from the edited registers). Where an unedited register did not contain more than 1,000 instances of *'s* and *s'* and *of*, the entire set of instances was extracted. This process yielded a sample of 9,230 tokens. The second step was to extract a sample of instances of each variable that occurred in the edited sub-corpus but which did not occur in the unedited sub-corpus (together with its aligned segment from the unedited sub-corpus) – in other words, genitives that editors *added* to the texts and therefore would not appear in the sample extracted from the unedited texts. This process yielded a further 2,740 cases, resulting in a total sample of 11,970 aligned pairs.

Many of the texts in the corpus include metadata on the subvariety and gender of the authors and editors; however, for some texts this information is not yet available (or is still being added). Following the extraction of the data, it was necessary to identify cases where this metadata was not available and then to either add this by consulting the files or exclude the observation if such data could not be obtained. Each of the remaining cases was then manually inspected to determine if it constituted a choice context using the criteria set out in Bresnan et al. (2017). Following this, we arrived at a total set of 4,001 observations for the unedited texts and 3,237 observations for the edited texts, spread over 282 text pairs.

Each observation was annotated to indicate whether it had occurred in an unedited or edited text: coded as *uned* for unedited cases and *ed* for edited cases. We also included an annotation for author ID as a way to account for multiple observations contributed by the same author(s).

Table 9.3 *Unique authors by gender, per subvariety and register*

	AfrE		BSAfrE		WSAfrE		Overall		Total
	Female	Male	Female	Male	Female	Male	Female	Male	
<b>Academic</b>	25	20	8	8	5	4	38	32	70
<b>Creative</b>	0	0	0	0	5	1	5	1	6
<b>Instructional</b>	5	6	1	1	4	4	10	11	21
<b>Popular</b>	4	3	0	2	3	0	7	5	12
<b>Reportage</b>	12	4	0	13	0	0	12	17	29
<b>Total</b>	46	33	9	24	17	9	72	66	138

This was not possible for the editors, since in many cases identifying information for the editors was not included in the metadata or had been removed from the files prior to their inclusion in the corpus. Table 9.3 reflects the number of unique authors, by gender, per subvariety and register and shows that 138 unique authors are represented in the dataset. Table 9.3 also shows that the overall representation of unique authors by gender is relatively even (with slightly more female authors represented than males), even if there are slight differences across subvarieties within registers.

Table 9.4 reflects the proportional distribution of authors and editors by gender per subvariety and register, across all texts. Although it is clear from the table that there is a somewhat uneven distribution of editor and author genders across the texts at the subvarietal and register level, the overall total proportions for the two genders are similarly represented across the data, with more female editors than males.

Each token was manually coded for a set of linguistic and extralinguistic factors. Five linguistic factors known to play an important role in genitive alternation were included:

- The ANIMACY of the possessor was coded using five levels: *A* for animates such as humans (dead or alive) and higher animals; *C* for collectives referring to organisations as well as nouns not referring to organisations but with potentially variable verbal and pronominal concord (such as *society*, *National Party*); *T* for temporal possessors with a time reference (such as *the afternoon*, *the twentieth century*, *today*); *L* for locatives referring to places (such as *the Missouri River*, *Africa*); and *I* for inanimates, such as inanimate concrete and abstract nouns, as well as gerunds/infinitives (such as *HbS*, *pain*, *vessels*, *test*, *multiculturalism*).

Table 9.4 Proportional distribution of male and female authors and editors by text, per subvariety and register

Authors								
	AfrE		BSAfE		WSAfE		Total per register	
	%	%	%	%	%	%	%	%
	Female	Male	Female	Male	Female	Male	Female	Male
<b>Academic</b>	48	21	9	10	10	4	66	34
<b>Creative</b>	0	0	0	0	83	17	83	17
<b>Instructional</b>	20	13	2	7	9	50	30	70
<b>Popular</b>	33	25	0	17	25	0	58	42
<b>Reportage</b>	16	6	0	78	0	0	16	84
<b>Overall</b>							51	49

Editors								
	AfrE		BrE		WSAfE		Total per register	
	%	%	%	%	%	%	%	%
	Female	Male	Female	Male	Female	Male	Female	Male
<b>Academic</b>	30	4	0	0	38	28	68	32
<b>Creative</b>	0	0	33	0	67	0	100	0
<b>Instructional</b>	54	0	0	0	17	29	71	29
<b>Popular</b>	0	0	0	0	38	63	38	63
<b>Reportage</b>	3	5	0	0	19	72	22	78
<b>Overall</b>							60	40

- The definiteness (DEFNESS) of the possessor was coded using two levels: *Def* for all possessors that were definite (such as proper nouns and possessors that start with a definite article) and *Indef* for all indefinite possessors.
- The presence or absence of a final sibilant (FIN\_SIB) in the possessor was coded. If the possessor ended in an /s/, /z/, /ʃ/, /ʒ/, /tʃ/ or /dʒ/, then it was coded as *Yes*, otherwise, the observation was coded as *No*.
- The semantic relation (SEM\_REL) between the possessor and possessum was coded based on Rosenbach (2002), who distinguishes between prototypical semantic relations (such as legal ownership, kinship, part-whole and body part relations) and non-prototypical semantic relations (all other relations that do not fall into the prototype category).

Two levels were thus coded: *Pro* for prototypical semantic relations and *Non-pro* for non-prototypical semantic relations.

- The relationship between the length of the possessor and possessum was operationalised as the length difference between the possessor and possessum (ORUM\_LENDIFF) in terms of the number of characters (excluding whitespaces). Determiners in the possessum of *of*-genitives were not included in the length measurement. Thus, in observations where the possessor is longer than the possessum  $\text{ORUM\_LENDIFF} > 0$ , while in observations where the possessum is longer than the possessor  $\text{ORUM\_LENDIFF} < 0$ .

All observations were coded for three extralinguistic factors:

- The REGISTER from which each observation was drawn included five levels: *Aca*, *Cr*, *Ins*, *Pop* and *Rep* for the academic, creative, instructional, popular and reportage registers, respectively.
- The subvariety best reflecting the author's language background (A\_VARIETY) included three levels: *AfrE*, *BSAfE* and *WSAfE*.
- The author's gender (A\_GENDER) was coded using two levels: *fem* to indicate females and *male* for males.

One of the challenges with corpus-based investigations of unedited and edited texts is that the observations drawn from edited texts cannot be unambiguously attributed to the linguistic choices of the author alone, but they also reflect the acceptability judgements of the editors who edited those texts. It is thus necessary to also account for the editors' linguistic background and gender. Edited observations therefore include two further extralinguistic factors:

- E\_VARIETY (with three levels: *AfrE*, *BrE* and *WSAfE*)
- E\_GENDER (with two levels: *fem* and *male*).<sup>3</sup>

We also categorised each observation based on what the editor did when encountering it in the text. This categorisation was not included as a factor in the modelling in the study, but served to facilitate the qualitative

<sup>3</sup> These factors cannot be included in the coding of the unedited observations. However, the inclusion of these two variables for the edited observations results in a series of empty cells for these two factors for all unedited observations, which presents challenges for the multifactorial modelling technique used in this study. To deal with this, we considered various methods to impute the missing data and ultimately opted to replace all NAs in the unedited observations with the mode of the factor (*WSAfE* for E\_VARIETY and *fem* for E\_GENDER). Trialling various models without the imputed values (e.g. separate models for the unedited and edited data, or a combined model with E\_VARIETY and E\_GENDER excluded) yielded very similar results.

analysis of the complex mix of potential editorial interventions. Four levels were included to distinguish the different types of editorial intervention: (1) the editor **retained** the genitive construction as is (and it thus occurred unchanged in the unedited and edited texts); (2) the editor **replaced** the genitive construction with the alternative construction (and there is thus alternation across the unedited and edited texts); (3) the editor **removed** the genitive construction from the edited text (as a consequence of either removing sentences or paragraphs from the text or removing only the genitive construction or part of it) and (4) editors **added** a genitive construction to the edited text where one does not exist in the unedited text (through either the addition of sentences and paragraphs to text or the addition of the genitive construction or part of it to the text). We therefore included observations where editorial intervention extends beyond the retention or replacement of the two constructions to include interventions that led to the addition and omission of the genitive constructions. This was done to highlight how multifaceted and complex editorial intervention is and to investigate how the many dimensions of editorial intervention lead to shifts in the overall configuration of genitive constructions across unedited and edited texts.

### 9.5 Analysis and Results

The analysis is developed in three steps. First, as background in Section 9.5.1, we sketch the overall proportions of the two constructions across the three subvarieties and five registers, for the unedited and edited texts, in order to explore if editorial intervention shifts the proportions of the two constructions. In Section 9.5.2, we use random forests to investigate the explanatory importance of the individual independent variables (the linguistic and extralinguistic factors) in conditioning the outcome of the response variable (the genitive construction). Random forests rank the individual independent variables according to their explanatory importance in conditioning the response variable given all other independent variables: the higher a variable's importance score, the stronger its impact on the response variable (see Levshina 2015: 292). In Section 9.5.3, we zoom in on the contribution of the various linguistic and extralinguistic factors in conditioning the genitive alternation. We make use of a generalised linear mixed-effects model tree (GLMM tree), which uses model-based recursive partitioning to detect subgroup interactions and a generalised linear mixed-effects model (GLMM) to estimate the random-effects parameters (Fokkema et al. 2018). Gries (2015) advises that the advantages of GLMMs for corpus-based investigations include that they

easily integrate numeric predictors (such as `ORUM_LENDIFF`), that they are good at handling unbalanced designs and that they address the variability introduced by the same speaker contributing multiple data points in the dataset. GLMMs thus allow for the inclusion of multiple fixed effects while simultaneously controlling for random (speaker-specific) effects. By using GLMM trees, we are able to construct a model in which it is possible to investigate the influence of the various linguistic and extralinguistic factors in conditioning genitive choices across unedited and edited texts, while controlling for the random effect of authors contributing large numbers of observations to the dataset.

### 9.5.1 Proportions of Genitive Variants across Varieties, Registers and Version

Table 9.5 shows the proportions of the two constructions across the unedited and edited texts in each subvariety and register. BSAfE authors favour the *s*-genitive least (as might be predicted from a substrate transfer effect), followed by AfrE authors and WSAfE authors (who appear to be particularly receptive to the *s*-genitive in the creative, instructional and popular registers). Editorial intervention leads to substantial shifts of the proportions in WSAfE creative, instructional and popular writing, in AfrE academic writing and reportage and in BSAfE reportage.

A qualitative analysis of the types of editorial interventions shows that while editors overwhelmingly endorse existing usages in unedited texts (by retaining them), they make a fair number of interventions that add and remove the two constructions, with only a handful of changes targeting the replacement of the two constructions. The shifts in the proportions of the

Table 9.5 Proportions of *s*- and *of*-genitives across variety, register and unedited and edited texts

	AfrE		BSAfE		WSAfE	
	Unedited ( <i>s</i> % : <i>of</i> %)	Edited ( <i>s</i> % : <i>of</i> %)	Unedited ( <i>s</i> % : <i>of</i> %)	Edited ( <i>s</i> % : <i>of</i> %)	Unedited ( <i>s</i> % : <i>of</i> %)	Edited ( <i>s</i> % : <i>of</i> %)
<b>Academic</b>	14 : 86	18 : 82	13 : 87	13 : 87	19 : 81	19 : 81
<b>Creative</b>					61 : 39	40 : 60
<b>Instructional</b>	18 : 82	20 : 80	5 : 95	5 : 95	50 : 50	18 : 82
<b>Popular</b>	27 : 73	26 : 74	18 : 82	18 : 82	38 : 62	35 : 65
<b>Reportage</b>	22 : 78	18 : 82	14 : 86	23 : 77		

two constructions are therefore largely due to more substantive editorial interventions, rather than to the targeted replacement of authors' usages.

Editing is most interventionist in WSAfE and in particular in the creative and instructional registers, where editors drastically reduce the proportion of *s*-genitives in these two registers. In the creative register, this shift appears to be due to the removal of sections of text (with such sections containing many more *s*-genitives than *of*-genitives) and in the instructional register to the substantive addition of sections of text, reflecting editors' own preferences (which appear to favour the *of*-genitive).<sup>4</sup> These shifts bring the proportions much closer to those of AfrE, especially in instructional writing.

Editorial intervention is less interventionist in AfrE, by comparison. The most substantive shifts in AfrE occur in the academic and reportage registers, but the effect is in opposite directions: in academic writing, editing increases the proportion of *s*-genitives, but lowers this in reportage. The net effect of this appears to smooth out register differences in AfrE writing, especially in the academic, instructional and reportage registers.

Such strategies do not appear to inform editorial intervention in BSAfE: the strong preference for the *of*-genitive in unedited BSAfE writing is maintained by editors (with the exception of reportage) and the divergent usage patterns across the registers are thus preserved. In terms of reportage, editors appear to add *s*-genitives due to either substantive rewriting or addition of text (thereby imposing their own preferences), while the removal of *of*-genitives appears to be motivated by factors of economy (a key feature of this register [e.g. Grafmiller 2014]).

The proportional analysis suggests several points: there appears to be divergence in the usage practices of AfrE, BSAfE and WSAfE authors, but this is smoothed over in AfrE and WSAfE and maintained in BSAfE. Thus, editors bring WSAfE and AfrE closer together, while simultaneously accepting and reinforcing the divergent usage practices of BSAfE. Shifts, however, are not due to editors specifically targeting the genitive constructions, but appear to be the by-product of more substantive editing practices, such as the removal of sections of text or the

<sup>4</sup> The instructional register includes a number of schoolbooks and teacher's guides targeted at school-level learners and since the *of*-genitive is considered the more analytic, periphrastic and therefore more transparent of the two constructions (see Heller et al. 2017b: 20), it is likely that the stronger preference for the *of*-genitive in editors' additions is motivated by the need to facilitate readers' ease of access to the content of these texts.

addition of sections of text (where editors employ their own preferences for the two constructions).

### 9.5.2 *Random-Forest Analysis: Determining the Overall Importance of Predictor Variables*

Although the analysis in Section 9.5.1 provides some information on how editorial intervention shifts usage preferences, it does not provide information on the role of the linguistic and extralinguistic factors in conditioning the genitive alternation. To explore the overall importance of the linguistic and extralinguistic factors in conditioning the genitive alternation, we ran a random forest utilising the `PARTYKIT` package in R (Hothorn and Zeileis 2015). All linguistic and extralinguistic factors described in Section 9.4.2 were included in the random forest, with the genitive construction set as the outcome variable. The accuracy of the random forest is determined using the `confusionMatrix()` function in the package `CARET` (Kuhn 2017).

The classification accuracy of the random forest is 91.79%, which is highly significantly better ( $p < 0.001$ ) than the baseline model predicting the most frequent choice (i.e. the *of*-genitive) all the time, which has an accuracy of 59.84%. The variable importance plot in Figure 9.1 shows that, as is to be expected, linguistic factors play the most important role in conditioning the genitive alternation in *SAfE*. With the exception of author subvariety and register, none of the extralinguistic factors, including gender, plays a notable role in conditioning the alternation. The importance of subvariety and register seems to confirm the findings of recent investigations into the importance of varietal and genre differences in genitive alternation (see Section 9.2).

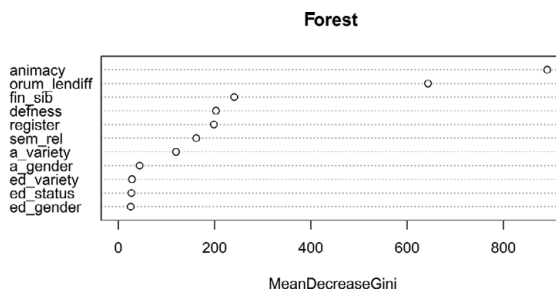


Figure 9.1 Variable importance plot for the linguistic and extralinguistic factors



The ANIMACY of the possessor is by far the most important predictor for which genitive construction is used, followed by the length difference between the possessor and the possessum. The presence of a final sibilant in the possessor plays a somewhat less important role, by comparison, but is still clearly an important predictor in genitive alternation. Definiteness of the possessor and register play similarly important roles, more so than semantic relation (which is the least important of the linguistic factors coded for in the dataset).

### 9.5.3 Generalised Linear Mixed-Model Tree: Analysis of the Effects of Predictor Variables

In this section, we analyse the interaction of the linguistic and extralinguistic factors outlined in Section 9.4.2 in conditioning the genitive alternation across unedited and edited texts. We fit a GLMM tree model using the R package GLMERTREE (Fokkema et al. 2018) in which GENITIVE is set as the binary response variable, ED\_STATUS is set as the subgroup-specific fixed effect, AUTHOR\_ID is set as the random effect and ANIMACY, ORUM\_LENDIFF, DEFNESS, FIN\_SIB, SEM\_REL, REGISTER, A\_GENDER, A\_VARIETY, E\_GENDER and E\_VARIETY are set as potential partitioning variables. We specified a maxdepth of 6 for the tree. The accuracy of the GLMM tree is determined using the confusionMatrix() function in the package CARET (Kuhn 2017). The resulting tree is plotted in Figure 9.2.

The classification accuracy of the GLMM tree is 85.56%, which is highly significantly better ( $p < 0.001$ ) than the baseline model predicting the most frequent choice (i.e. the *of*-genitive) all the time, which has an accuracy of 59.84%. As anticipated from the random-forest analysis, gender (of authors and editors) and editors' language background do not play a visible role. Varietal differences occur relatively low down in the tree for the more prototypically animate possessors only, signalling that choices across the three subvarieties are conditioned in very similar ways. As evidenced by the terminal nodes, editorial changes introduced are not dramatic (again, as expected), but, nevertheless, there are notable changes under some conditions.

The left main branch of the tree, split on ANIMACY, contains the more prototypically animate categories (animates [A], collectives [C] and locatives [L]). For cases where the possessor is shorter than the possessum, or the possessum is less than nine characters longer than the possessor (ORUM\_LENDIFF < 9) (node 2), the next split is on ANIMACY. In the most

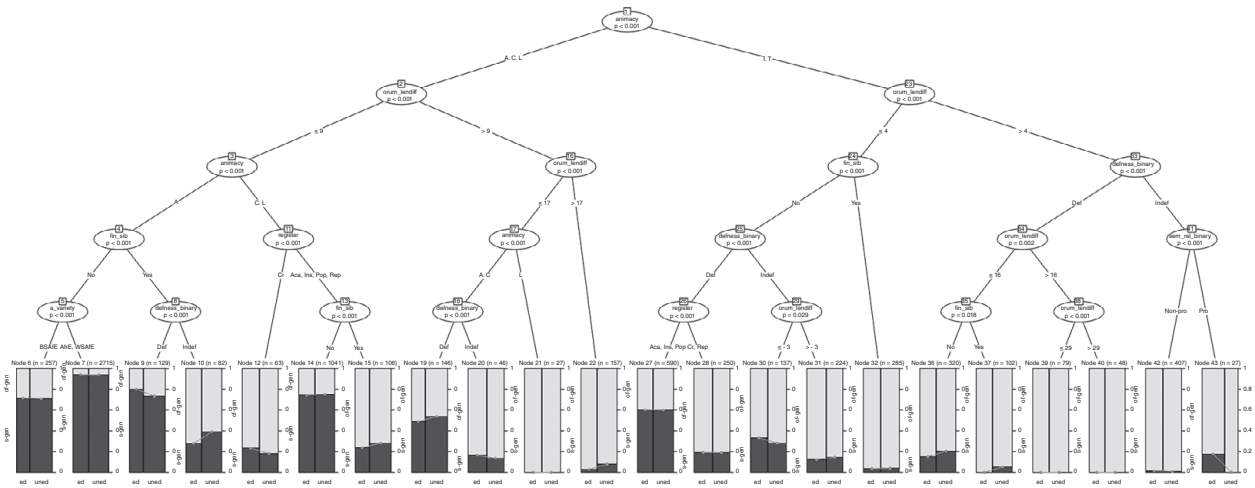


Figure 9.2 GLMM tree

prototypically animate category (A), FIN\_SIB is the next distinction. In cases where a final sibilant is absent, the final split is for A\_VARIETY and terminal nodes 6 and 7 show that whereas AfrE and WSAfE here choose almost without exception the *s*-genitive, this preference is far less strong for BSAfE. In other words (possibly due to substrate influence), BSAfE users do not use the *s*-genitive in the most prototypical circumstances where it would typically be used by the other two varieties (as exemplified in [3] and [4]). Also notable is that editors do not alter this preference – the ratios of *s*-genitives and *of*-genitives in node 6 and node 7 are almost unaltered in the edited and unedited texts. This provides some evidence of endonormativity for BSAfE in this respect and aligns with the findings of the proportional analysis (Section 9.5.1).

- (3) ... the Christ-centred teaching of the Chief Apostle Mkhonza... <SA-P-007-O><sup>5</sup>  
 (4) ... the unconventional behaviour of the offender... <SA-A-139-O>

Where a final sibilant is present in cases with animate possessors where ORUM\_LENDIFF < 9, definites show a much stronger preference for the *s*-genitives than indefinites (node 9 and node 10). However, there are some shifts introduced by editing, such that for the definites in this set the proportion of *s*-genitives is slightly increased and for indefinites it is decreased (somewhat more). An inspection of the data shows that the increase in *s*-genitives is largely due to the addition of the construction by editors, either as a way of clarifying the meaning in a sentence (as exemplified in [5]) or as part of the substantive editing process typical of certain registers (as exemplified in [6]).

- (5) a. In most cases shock, sorrow, frustration, guilt, blame, insecurity and anxiety were present. <SA-A-044-O>  
 b. In most cases, shock, sorrow, frustration, guilt, blame, insecurity and anxiety were present as part of the parents' initial reactions ... <SA-A-044-E>

<sup>5</sup> All exemplifications are referenced to their identifying label in the corpus. This abbreviated label contains three sets of information: a code for the register in which the genitive occurred, the number of the text pair in the corpus and whether the example is taken from the unedited text (coded as O) or the edited text (coded as E). Thus, example (3) is from the unedited version of text pair 007 in the popular register.

- (6) a. ... At the thought of his mother, his eyes welled up...  
<SA-C-006-O>  
b. ... The mention of Morag changed Ross's mood and his eyes  
became damp. <SA-C-006-E>

In other words, the increase in *s*-genitives seems to reflect the editors' own writing preferences for these types of constructions. Similarly, the decrease in *s*-genitives with indefinites is linked, on the one hand, to the addition of *of*-genitives by the editor as a consequence of improving formulation in the text (as exemplified in [7]) but also, in a few instances, to the replacement of *s*-genitives with *of*-genitives (see [8]). In the case of the latter, the head noun in the possessor phrase is plural and contains a word-internal sibilant and it therefore seems likely that the replacement is triggered by the repetition of sibilants in the construction.<sup>6</sup>

- (7) a. These children's safety lean on us as the landlords and it's their  
responsibility to protect themselves again. <SA-R-121-O>  
b. In my business the safety of these children is paramount. They  
depend on us for safety and it's the responsibility of landlords  
to protect them. <SA-R-121-E>  
(8) a. ... finance officers' workloads... <SA-A-119-O>  
b. ... the workload of finance officers ... <SA-A-119-E>

For collective and locative possessors (less strongly prototypically attracted to *s*-genitives) (right branch of node 3), REGISTER is the following split, with the creative register showing a stronger preference for the *of*-genitive than the other four registers.<sup>7</sup> This preference is slightly reduced in editing, with a small shift in the direction of more *s*-genitives in the edited texts. This shift appears either to be due to the removal of sentences containing *of*-genitives in the edited texts (as exemplified in [9], which was removed from the edited text) or through the replacement of (a non-alternating)

<sup>6</sup> Although the *s*-genitive in (7a) contains a definite possessor that does not end in a sibilant, the replacement of this construction by the editor may also be motivated by the presence of multiple sibilants in the possessor. It is, of course, also likely that this replacement is triggered by the length difference between the possessor and possessum.

<sup>7</sup> The proportions for the two constructions reported in Section 9.5.1 suggest a much stronger preference for *s*-genitives in the creative register. An inspection of the observations in this register shows that more than half of the possessors are animate. Thus, the higher proportion of *s*-genitives reported in Section 9.5.1 does not reflect a particular openness to *s*-genitives in this register or by WSAfE authors. Rather, the higher number of animate possessors (a feature typical of creative writing), together with the stronger preference for *of*-genitives with inanimate and temporal possessors reported here, suggests a stronger effect for animacy on genitive choices in this register (where only WSAfE is represented).

construction with an *s*-genitive in an attempt by the editor to amend awkward formulation (as in [10]).

- (9) Tears pricked at her eyelids as she sank into her mother's armchair in the bay window and looked out over the treetops of Johannesburg. <SA-C-006-O>
- (10) a. ... the difficulty of being a professional Muslim woman in a changing history of South Africa. <SA-C-003-O>  
 b. ...the difficulty of being a professional Muslim woman as South Africa's history changed. <SA-C-003-E>

In the other four registers, a further split occurs based on final sibilance (node 13), with the *of*-genitive preferred in cases where a final sibilant is present. Editorial intervention hardly makes any difference to preferences in terminal nodes 14 and 15 (apart from a slight reduction in the frequency of *s*-genitives with final sibilants).

Returning to node 2, in the right branch, reflecting cases where the possessum is relatively longer than the possessor (ORUM\_LENDIFF>9), the next split is also for ORUM\_LENDIFF (node 16). Where the possessum is substantially longer (ORUM\_LENDIFF>17), the preference is strongly for the *of*-genitive and editors facilitate a small increase in the strength of this preference (node 22) by adding sentences that contain *of*-genitive constructions, as in (11), or by adding *of*-genitives to help clarify meaning, as in (12). Editors therefore do not appear to replace *s*-genitives with substantially longer possessums with *of*-genitives: instead the shifts in the strength of this preference are due to editors adding *of*-genitives under these conditions and consequently reinforcing this preference.

- (11) The young men recently finished shooting a documentary which covers the lives of these young, talented artists who embrace 'Pantsula' dancing as a career and a lifestyle. <SA-R-086-E>
- (12) a. Learner can identify rural and urban needs. <SA-I-016-O>  
 b. The learner can identify the needs of people in urban and rural areas. <SA-I-016-E>

Where ORUM\_LENDIFF<17, the following split is for ANIMACY, with locatives preferring the *of*-genitive almost without exception. For animates and collectives, a following conditioning factor is definiteness, with definites again preferring the *s*-genitive more strongly than indefinites (node 19 and 20). Editorial intervention slightly reduces the preference for *s*-genitives with definites, but increases the preference for *of*-genitives with indefinites. Editors do not appear to remove *s*-genitives

with long definite possessors (as in [13]), but when they add material containing these constructions, their preference is for the *of*-genitive (as in [14]). A similar pattern emerges with indefinite animate and collective possessors: editors add *of*-genitive constructions, rather than remove *s*-genitives.

- (13) The most valuable form of positive discipline is self-discipline, which is derived from the individual official's integrity. <SA-I-080-O/E>
- (14) ...choose an ethical operator who contributes funds to the protection of this magnificent species. <SA-P-018-E>

This means that while editors accept the use of *s*-genitives for animate and collective possessors with longer possessums by authors, when they apply their own writing preferences this is more prototypically in line with the expected influence of `ORUM_LENDIFF`.

The right main branch of the tree (containing inanimate [*I*] and temporal [*T*] possessors) demonstrates a generally stronger preference for the *of*-genitive. The following split is for `ORUM_LENDIFF` (node 23): where possessums are more than four characters longer than inanimate or temporal possessors (`ORUM_LENDIFF`>4), there is a stronger preference for the *of*-genitive. There are only two conditions in this set where a slightly stronger preference for the *s*-genitive is evident. The first is with definite possessors where `ORUM_LENDIFF`>4 and where no final sibilant is present (terminal node 36). In this case, editing works to slightly reduce the frequency of the *s*-genitive, as it does when a final sibilant is present, but to a smaller degree (terminal node 37). These changes again appear to reflect the writing preferences of the editors: in most cases the increase in the *of*-genitive is due to the addition of these genitives to texts as a way to clarify meaning (see [15] and [16]).

- (15) a. If learners do not have a dice they can make a dice using a pencil. <SA-I-054-O>
- b. If learners do not have a dice they can make a dice using a pencil. To do this they scratch or write the numbers from 1 to 6 on the six sides of an ordinary pencil. <SA-I-054-E>
- (16) a. The main goal was to encourage Muslims to enhance their religious knowledge and religious commitments. <SA-A-030-O>
- b. The main goal of the call for Islamisation was to encourage Muslims to enhance their religious knowledge and religious commitments. <SA-A-030-E>

The second condition with a slight *s*-genitive preference is with indefinite inanimate and temporal possessors in prototypical semantic relations where  $ORUM\_LENDIFF > 4$  (terminal node 43) – but here the *s*-genitive is introduced by editing (it is almost completely absent in unedited texts). These changes appear to be due either to the replacement of *of*-genitives (as in [17]) or, in most cases, to the addition of an 's to a possessor where one does not occur in the unedited text (either as a straightforward copyediting correction or in an attempt to alter a noun–noun construction to an *s*-genitive, as in [18]).

- (17) a. The case of a tape measure is usually a fixed size. <SA-I-082-O>  
 b. A tape measure's case is usually a fixed size. <SA-I-082-E>
- (18) a. ... the depth which the drill bit penetrates into the material during one revolution of the drilling machine spindle. <SA-I-082-O>  
 b. ... the depth that a drill bit penetrates into the material during one revolution of a drilling machine's spindle. <SA-I-082-E>

Lastly, with inanimate and temporal possessors where the possessor is shorter or only marginally longer than the possessum ( $ORUM\_LENDIFF < 4$ ) (left branch of node 23), *FINAL SIBILANCE* is the next conditioning variable. When a final sibilant is present, the preference is almost exclusively for the *of*-genitive (terminal node 32). Where there is no final sibilant, definiteness is the next conditioning variable. Definite possessors are more likely to take an *s*-genitive under these conditions than an indefinite possessor, but for definite possessors *REGISTER* plays a further conditioning role (such that the *of*-genitive is preferred in creative writing and reportage and the *s*-genitive in the academic, instructional and popular registers [see terminal nodes 27 and 28]). Editing effects little change across any of these registers, however. For indefinites, the further conditioning variable is length difference (node 29). The *s*-genitive is more strongly preferred where the possessor is shorter than the possessum ( $ORUM\_LENDIFF < -3$ ) (terminal node 30) and editing further shifts this preference in the direction of the *s*-genitive (which again often appears to be due to the addition of an 's where editors deem this missing, as in [19]).

- (19) a. The importance of culture studies is therefore imperative to the understanding of tourism socio-cultural impacts. <SA-A-049-O>

- b. Cultural studies are therefore imperative to the understanding of tourism's socio-cultural impacts. <SA-A-049-E>

Where `ORUM_LENDIFF`>-3 (terminal node 31), the *of*-genitive is more strongly preferred and editing makes only a very small difference in further reinforcing this preference. Such reinforcements are due to the addition of a small number of *of*-genitives (mostly for clarification purposes, as in [20]).

- (20) a. Four frame picture showing a gold mine over 100 years. <SA-I-037-O>  
 b. A four-frame picture showing the development of a gold mine over 100 years. <SA-I-037-E>

## 9.6 Discussion and Conclusion

The findings presented in this chapter all consistently demonstrate that in the three SAfE subvarieties investigated, gender plays almost no visible role in conditioning the genitive alternation in writing or in editing. Thus, at least for the feature investigated in this study, there appears to be no difference in how males or females orient their choices, either in original writing (which echoes the findings in Ford and Bresnan [2015] for AmE and AusE) or in editing. However, this does not mean that genderlectal variation does not arise in the different varieties of SAfE (as evidenced by the work of Mesthrie [2017]), but it might suggest that gender-differentiated linguistic behaviour is less evident at the level of grammar than it is in lexis and pronunciation. Further investigations that include other grammatical features will need to be conducted in order to explore genderlectal differences at the level of grammar.

The analyses presented in Section 9.5.2 reveal that, as expected, animacy, length difference between the possessor and possessum, final sibilance and definiteness of the possessor play important roles in conditioning the genitive alternation in SAfE, with subvariety and register introducing subtle shifts only lower down in the tree. Semantic relation plays a role only with temporal and inanimate possessors. As expected, animacy and length difference are the strongest conditioning factors in genitive alternation. However, definiteness, final sibilance, author variety and register interact with animacy and length difference to temper this.

Most notable is the less strong preference for *s*-genitives in prototypical contexts by BSAfE authors – a preference that editors leave unchanged.



Given that only a postnominally realised construction is available in Southern Sotho and Zulu (see Section 9.3), it seems plausible that the substrate influences the genitive choices of BSAfE authors, which are accepted by editors (who are users of different varieties). Register differences emerge only for possessors lower down on the animacy scale (collective, locative, temporal and inanimate possessors), where the creative register (with collective and locative possessors) and the creative and reportage registers (with inanimate and temporal possessors) show a much stronger preference for the *of*-genitive. However, this finding should be interpreted with caution, particularly given the unbalanced design of the corpus and the exclusion of some subvarieties from these registers.

Definiteness emerges as a strong factor conditioning the genitive alternation, particularly with animate, collective and locative possessors, which favour the *s*-genitive, except when a final sibilant is present with animate possessors, or when animate and collective possessors are relatively longer than their possessums ( $ORUM\_LENDIFF > 9$  but  $\leq 17$ ).

Editorial intervention introduces subtle, non-significant shifts in the use of the constructions, which appear to reinforce the choices of the authors. It thus seems that editors reinforce the usage patterns of the authors they edit, something that is most striking in the writing of BSAfE authors (whose work is edited by WSAfE or AfrE editors). The qualitative analyses of the types of editorial interventions (in Sections 9.5.1 and 9.5.3) reveal that the subtle shifts in editing are not generally the result of editors specifically targeting the replacement of one construction over the other, but rather that the shifts in the proportions of the two constructions are mostly due to substantive editing practices or to changes related to improving formulation and ensuring clarity of expression. In other words, shifts in the proportions of the two constructions have less to do with a specific editorial focus on these constructions than they do with more general editing practices where the opportunity arises for editors' own writing preferences to come into play.

The findings regarding genderlectal variation in this study should be interpreted with caution, especially given the sparsity of data and the complexity of the model used to analyse this data. In addition, the corpus used in this study is unbalanced and therefore a more balanced corpus that more evenly represents the various subvarieties and genders might have provided slightly different results. A further limitation is the fact that we were unable to control for the random effects of the same editors represented in multiple observations and there is the chance that the patterns

revealed in editing reflect the preferences of the individual editors rather than generalisable patterns. In addition, further work using other (grammatical) features will need to be conducted in order to gain a more complete picture of gender-differentiated linguistic behaviour in published written texts.

## REFERENCES

- Bresnan, Joan, Rosenbach, Anette, Szmrecsanyi, Benedikt, Tagliamonte, Sali A. and Simon, Todd. 2017. 'Syntactic alternations data: datives and genitives in four varieties of English' (Manuscript). Retrieved from <http://purl.stanford.edu/qj187zs3852>.
- De Klerk, Vivian. 2003. 'Towards a norm in South African Englishes: the case for Xhosa English', *World Englishes* 22(4): 463–81.
- Fokkema, Marjolein, Smits, Niels, Zeileis, Achim, Hothorn, Torsten and Kelderman, Henk. 2018. 'Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees', *Behavior Research Methods* (50): 2016–34.
- Ford, Marilyn and Bresnan, Joan. 2015. 'Generating data as a proxy for unavailable corpus data: the contextualized sentence completion task', *Corpus Linguistics and Linguistic Theory* 11(1): 187–224.
- Grafmiller, Jason. 2014. 'Variation in English genitives across modality and genres', *English Language and Linguistics* 18(3): 471–96.
- Gries, Stefan Th. 2015. 'The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models', *Corpora* 10: 95–125.
- Heller, Benedikt, Bernaisch, Tobias and Gries, Stefan Th. 2017a. 'Empirical perspectives on two potential epicentres: the genitive alternation in Asian Englishes', *ICAME Journal* 41: 111–44.
- Heller, Benedikt, Szmrecsanyi, Benedikt and Grafmiller, Jason. 2017b. 'Stability and fluidity in syntactic variation world-wide: the genitive alternation across varieties of English', *Journal of English Linguistics* 45(1): 3–27.
- Hinrichs, Lars and Szmrecsanyi, Benedikt. 2007. 'Recent changes in the function and frequency of Standard English genitive constructions: a multivariate analysis of tagged corpora', *English Language and Linguistics* 11(3): 437–74.
- Holmes, Janet. 1997. 'Setting new standards: sound changes and gender in New Zealand English', *English World-Wide* 18(1): 107–42.
- Hothorn, Torsten and Zeileis, Achim. 2015. 'partykit: a modular toolkit for recursive partitioning in R', *Journal of Machine Learning Research* 16: 3905–9.
- IPed. 2016. 'Report of the IPed National Survey of Editors' (Manuscript). Retrieved from <http://iped-editors.org/site/DefaultSite/filesystem/documents/IPed%20National%20Survey%20of%20Editors%202016%20%20FINAL%2020072017.pdf>.

- Kotze, Haidee. 2020. 'Does editing matter? Editorial work, endonormativity and convergence in written Englishes in South Africa'. In Raymond Hickey, ed. *English in Multilingual South Africa: The Linguistics of Contact and Change*. Cambridge University Press, 101–28.
- Kruger, Haidee and Van Rooy, Bertus. 2017. 'Editorial practice and the progressive in Black South African English', *World Englishes* 36: 20–41.
- Kuhn, Max. 2017. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan and Tyler Hunt. *caret: classification and regression training* [R package version 6.0-81]. <https://CRAN.R-project.org/package=caret>.
- Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
1990. 'The intersection of sex and social class in the course of linguistic change', *Language Variation and Change* 2: 205–54.
2001. *Principles of Linguistic Change, Vol. 2: Social Factors*. Oxford: Blackwell.
- Lakoff, Robin T. 1977. 'You say what you are: acceptability and gender-related language'. In Sidney Greenbaum, ed. *Acceptability in Language*. The Hague: Mouton, 73–87.
- Law, Melanie A. 2019. *The Role of Editorial Intervention in Ongoing Language Variation and Change in South African and Australian English* (Unpublished doctoral dissertation). Macquarie University, Sydney, Australia and North-West University, Vanderbijlpark, South Africa.
- Levshina, Natalia. 2015. *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins.
- Mesthrie, Rajend. 2017. 'Class, gender, and substrate erasure in sociolinguistic change: a sociophonetic study of schwa in deracializing South African English', *Language* 93(2): 314–46.
- Milliot, Jim. 2018. 'The PW Publishing Industry Salary Survey, 2018' (Manuscript). Retrieved from [www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/78554-the-pw-publishing-industry-salary-survey-2018.html](http://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/78554-the-pw-publishing-industry-salary-survey-2018.html).
- Milroy, James and Milroy, Lesley. 1993. 'Mechanisms of change in urban dialects: the role of class, social network and gender', *International Journal of Applied Linguistics* 3(1): 57–77.
- Nevalainen, Terttu. 2000. 'Gender differences in the evolution of Standard English: evidence from the Corpus of Early English Correspondence', *Journal of English Linguistics* 28(1): 38–59.
- Romaine, Suzanne. 1999. *Communicating Gender*. London: Lawrence Erlbaum.
- Rosenbach, Anette. 2002. *Genitive Variation in English*. Berlin: Mouton de Gruyter.
2014. 'English genitive variation – the state of the art', *English Language and Linguistics* 18(2): 215–62.

2017. 'Constraints in contact: animacy in English and Afrikaans genitive variation – a cross-linguistic perspective', *Glossa: A Journal of General Linguistics* 2(1): 1–21.
- Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the World*. Cambridge University Press.
- StatsSA. 2019. 'General Household Survey: 2018 (Po318)' (Manuscript). Retrieved from [www.statssa.gov.za/publications/Po318/Po3182018.pdf](http://www.statssa.gov.za/publications/Po318/Po3182018.pdf).
- Szmrecsanyi, Benedikt and Hinrichs, Lars. 2008. 'Probabilistic determinants of genitive variation in spoken and written English: a multivariate comparison across time, space, and genres'. In Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta and Minna Korhonen, eds. *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*. Amsterdam: John Benjamins, 291–309.
- Szmrecsanyi, Benedikt, Grafmiller, Jason, Bresnan, Joan, Rosenbach, Anette, Tagliamonte, Sali A. and Todd, Simon. 2017. 'Spoken syntax in a comparative perspective: the dative and genitive alternation in varieties of English', *Glossa* 2(1): 86: 1–27.
- Szmrecsanyi, Benedikt, Grafmiller, Jason, Heller, Benedikt and Röthlisberger, Melanie. 2016. 'Around the world in three alternations: modelling syntactic variation in varieties of English', *English World Wide* 37(2): 109–37.
- Tagliamonte, Sali A. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Oxford: Wiley-Blackwell.
- Trudgill, Peter. 1972. 'Sex, covert prestige and linguistic change in the urban British English of Norwich', *Language in Society* 1: 179–95.
1983. *On Dialect*. Oxford: Blackwell.
- Van der Walt, Johann L. and Van Rooy, Bertus. 2002. 'Towards a norm in South African Englishes', *World Englishes* 21(1): 113–28.
- Van Rooy, Bertus. 2014. 'Convergence and endonormativity at phase 4 of the Dynamic Model'. In Buschfeld, Sarah, Hoffmann, Thomas, Huber, Magnus and Kautzsch, Alexander, eds. *The Evolution of Englishes: The Dynamic Model and Beyond*. Amsterdam: John Benjamins, 21–38.
- Van Rooy, Bertus and Kruger, Haidee. 2016. 'The innovative progressive aspect of Black South African English: the role of language proficiency and normative pressures', *International Journal of Learner Corpus Research* 2: 205–28.