

Prescriptivism, editing, and morphosyntactic variation in written South African Englishes: A case study of relative *who*, *that* and *which*



Haidee Kotze^{a,1,*}, Melanie Ann Law Favo^{b,2}

^a Department of Languages, Literature and Communication, Utrecht University, Trans 10, Utrecht, JK 3512, The Netherlands

^b School of Languages, North-West University, Hendrik van Eck Boulevard, Vanderbijlpark 1609, South Africa

Received 29 December 2023; revised 2 December 2024; accepted in revised form 2 December 2024;

Abstract

In variationist research there is little reflection on the multiple agents involved in written textual production, whose distinct linguistic and normative representations influence the linguistic composition of a text. In this study, we ask where the uniformity and prescription-oriented nature of published written language originate – directly from writers, or from those responsible for editorial intervention. Focusing on the alternation between *who*, *that* and *which* in subject restrictive relative clauses with animate antecedents, we analyse patterns of variation in unedited written texts and their edited counterparts, across three subvarieties of South African English. We find that, generally, in the unedited writing of the subvarieties the distribution of relativisers patterns similarly to other varieties of English, reflecting the stability of the English relativisation system. However, in line with previous findings on New Englishes, there are lower-level divergences in the subvarieties investigated – and, in general, higher proportional frequencies of *that* and *which* with animate antecedents in restrictive relative clauses in subject position than is generally the case in previous studies that have focused on written language. Where a variety tends towards less prototypical usage, editorial interventions at times mask this by amending usage towards more prototypical usage. However, the patterns of intervention are not always consistent. The linguistic background of the editor in relation to that of the author (i.e. whether they are users of the same subvariety of South African English or not) appears to be one of the factors conditioning editorial intervention.

© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Relativiser choice; Restrictive relative clauses; Animacy; South African English; Editing; Prescriptivism

1. INTRODUCTION

That language use varies, and is shaped by the broader context of culture, the immediate context of communication, and the individual characteristics and linguistic background of speakers and writers is a fundamental premise of

* Corresponding author.

E-mail addresses: h.kotze@uu.nl (H. Kotze), Melanie.Law@nwu.ac.za (M.A. Law Favo).

¹ Utrecht University / North-West University.

² North-West University.

sociolinguistic approaches to linguistic variation. The study reported on in this article is motivated by a lacuna in variationist research: a lack of attention to the multiple agents who may be involved in specifically published written textual production, and whose distinct linguistic representations (including representations about normative correctness) influence the ultimate linguistic composition of a text. As D'Arcy and Tagliamonte (2015: 255) point out, written and spoken language differ substantially, with "standard written language" (particularly in published texts) being "highly uniform and governed by prescription", while spoken language is "revealing of structured heterogeneity". Of course, distinctions between written and spoken language are not always clear cut (as is, for example, the case in computer-mediated communication, or in prepared speeches), and some forms of written language are less normatively constrained and influenced by prescription (e.g., personal emails and blogs). However, if one considers published written texts (of the kind typically included in linguistic corpora), a factor that has remained largely unaccounted for is how editorial interventions, of various kinds, reshape such texts. In other words, where precisely does the uniformity and prescription-oriented nature of published written language originate – directly from writers, or from those responsible for editorial tasks, like copy-editors and proofreaders?

Recent work on the role of editing in linguistic variation has focused particularly on how editorial processes may provide insight in normative orientation in contexts where linguistic norms are in flux or contested – as is the case for pluricentric languages like English, where multiple norm orientations exist and compete. Kruger and Van Rooy (2017) have argued that consideration of editorial practices is important not only because such practices provide insights into normative orientation, but also because editorial practices themselves form part of what they refer to as a conventionalisation–legitimation feedback loop: if editors accept innovative or unconventional linguistic features, thereby legitimising them, these pass into published media. This contributes to conventionalisation, since it increases the frequency of the feature in the input that users receive, which in turn makes selection of the feature more likely by writers. This, in turn, increases the frequency of exposure for editors, who are then more likely to endorse it – and so on. Kruger and Van Rooy (2017: 28) regard this feedback loop as an important mechanism of the establishment of a local norm, which functions even in the absence of codified norms.

This feedback loop integrates two dimensions of standardisation into a single process. Peters (2014: 582–583) refers to these two dimensions as standardisation viewed as "progressive adaptations to the common language that occur with its application to an ever-wider range of functions in a given speech community", versus standardisation viewed as "direct intervention in the forms of written language ... associated with the work of editors and publishers". The conventionalization–legitimation loop proposed by Kruger and Van Rooy (2017) offers a way of linking bottom-up and top-down, and unconscious cognitive and consciously engineered social factors that influence language variation and change (see Labov, 2007) – firmly grounded in language use, and language users. As Milroy and Milroy (2012: 48) emphasise, languages "do not exist independently of speakers, and if changes take place in them, such changes must be the reflexes of speaker-innovations, established as new norms by speaker acceptance". We argue that in this it is essential to broaden the scope not just to speakers, but also writers and editors, all of whom play a role in the iterative processes that diffuse, conventionalise and legitimise particular linguistic usages, in particular contexts of use.

Existing studies of this dynamic in postcolonial varieties of English (e.g., Van Rooy and Kruger, 2016; Kruger and Van Rooy, 2017; Kotze, 2019; Law, 2019; Law and Kotze, 2021) have demonstrated that editorial practice does provide an index of the endonormativity of a variety of English. When editors accept variety-specific usages this may be seen as a marker of the establishment of a local norm. However, editorial practice is also highly complex and dependent on various factors, such as the particular linguistic feature in question and the linguistic background of the editor in relation to the author. While the importance of this latter aspect is suggested by several of the existing studies, the role of the linguistic background of the editor (also in relation to that of the author) has hardly been the explicit focus of any empirical research. The potential role of this variable is raised in Kruger and Van Rooy (2017), who argue that acceptance of an innovative feature (such as extended uses of the progressive) by editors who are *not* users of the same (sub)variety of English as the authors who use these features is an especially stringent criterion for the establishment of a local norm. Such acceptance indexes acknowledgement of a norm even from 'outsiders' to the immediate linguistic community. While there is an extensive body of sociolinguistic research focusing on how speakers' and writers' language use varies, in relation to variables related to region, ethnic background, gender, class and age, existing research on *editors* as language users whose linguistic practices may be shaped by the same variables as those influencing writers, is scant.

The South African context is particularly useful for studying this dynamic. English in South Africa is used as an L1 and L2 variety in complex sociolinguistic configurations, arising from the sociopolitical history of the country. In this article, we focus on three subvarieties of South African English (SAfE): so-called White South African English (WSAfE), an L1 variety used by descendants of British colonial settlers; Afrikaans English (AfrE), an L2 variety used by Afrikaans L1 users; and Black South African English (BSAfE), a mostly L2 variety used by L1 users of an indigenous African language. Only approximately 10% of South Africans describe themselves as speaking English at home (or as an L1), and most of these are white South Africans (see Botha et al., 2021 for more detail). However, English is widely used

as a language of informal and formal spoken and written communication, and publication, in South Africa. It is a high-status language, associated with upward socio-economic mobility. In the educational environment, despite policies intended to promote mother-tongue education and additive bilingualism, English and Afrikaans continue to dominate. At the moment, the African languages are used as a medium of instruction to a significant degree only in the first three years of schooling (in 2007, 80% of learners in this age group were learning in their home language; see [Posel and Zeller, 2020](#): 291). However, from Grade 4 onwards, the majority of learners in South Africa have English as medium of learning and teaching. In higher education, too, English predominates (see [Mkhize and Balfour, 2017](#)), and academic writing is mostly in English, irrespective of the L1 of the author. English is the dominant language across all publishing sectors. Academic book publishing is near exclusively in English, while in the trade sector (fiction and non-fiction) Afrikaans and English have a strong presence – with hardly any publication in the African languages (see [Möller, 2013](#)). The only sector where the African languages are reasonably well represented is in educational publishing for the primary schools market, but here production is often in English (or Afrikaans) first, with books subsequently being translated to other languages (see [Kotze, 2012](#)). Magazine and newspaper publishing have a strong presence of both Afrikaans and English, and while there are some African-languages newspapers and magazines, these are few, often with a small circulation and directed to a regional rather than national readership.

The majority of textual production in South Africa, across most registers (see also [Section 3.1](#)), is thus in English – but English produced by users of L2 varieties of English, particularly BSAfE and AfrE. However, it is important to point out that such L2 English language production happens in the context of stable (and even growing) bilingualism rather than language shift, with African-language and Afrikaans speakers alike maintaining their L1 alongside English (albeit in often complex linguistic repertoires) (see [Coetzee-Van Rooy, 2013](#); [Posel and Zeller, 2016](#)).

While written text production in English in South Africa is done by both L1 (WSAfE) and L2 (AfrE and BSAfE) users, hardly any professional editing in South Africa is done by BSAfE users. For example, a quick overview of the editors offering services in English on the website of the South African Translators' Institute shows that just less than a fifth of these editors are BSAfE users ([SATI, 2023](#)). In [Law Favo's \(2024\)](#) data, none of the editors who responded to her questionnaire were users of BSAfE, and of those who did respond, 96.5% were multilingual, most of whom had English and Afrikaans as their strongest two languages. This is a distinctive feature of the South African context, with both authors and editors typically being bi- or multilingual, although in the case of editors, language profiles tend to be strongly weighted towards Afrikaans–English bilingualism.

In this article, we extend the current work on the role of editors by focusing on the alternation between *who*, *that* and *which* in restrictive relative clauses (RRCs) with animate antecedents in subject position. We consider patterns of variation in unedited original written texts and their edited counterparts, across different registers, and three subvarieties of South African English: AfrE, BSAfE and WSAfE. Our choice of this feature is motivated by two factors. First, in comparison to the choice of relativiser in RRCs with inanimate antecedents, which has been widely studied, the choice of relativiser in RRCs with animate antecedents has received far less attention. Second, this choice is both grammatically and prescriptively constrained – but there is evidence of variability, in both usage and prescription, across different varieties of English. This variability makes this particular feature especially suited to investigating the normative orientation of both writers and editors with different linguistic backgrounds.

[Section 2](#) provides an overview of the English relativiser system in general ([Section 2.1](#)), considers existing research on relativiser choice in RRCs with animate antecedents ([Section 2.2](#)), and the prescriptive advice related to this ([Section 2.3](#)). In [Section 3](#) we describe our corpus and method: we analyse the proportion of each relativiser (*who*, *that* and *which*) according to four animacy types (animals, collectives, humans and the word *people*), across the three varieties in question, in unedited and edited writing. We investigate the percentage change in frequency between the unedited and edited texts – while taking account of the linguistic background of editors in relation to those of authors, and other factors, like the particular antecedents. The findings of the analysis are presented in [Section 4](#), together with a discussion of the implications of the patterns we uncover. [Section 5](#) concludes the article, mapping out the limitations of the study, and looking ahead at future research.

2. ENGLISH RELATIVISERS: STRUCTURE, USAGE AND PRESCRIPTION

2.1. The English relativiser system: A brief overview

English relative clauses can be described and distinguished along three main axes: the degree of integration of the relative clause (usually captured in a distinction between restrictive and non-restrictive relative clauses (RRCs and NRRCs, respectively)); the animacy of the antecedent noun (usually captured on a scale of animacy ranging from human to inanimate nouns); and the syntactic function of the relative clause (whether in subject or non-subject position).

Different relative pronouns, such as *that*, *which*, *who*, *whom*, *whose* or zero, are available for use in relative clauses, and "... for every combination of antecedent type and grammatical role, there are very different frequencies of *that*, *zero*, and *wh*-relativisers" (D'Arcy and Tagliamonte, 2015: 266). Table 1, adapted from Quirk et al. (1985: 366), who base their findings on British English (BrE) and American English (AmE), summarises the choice of relative pronouns according to antecedent type and grammatical function.

The distinction between RRCs and NRRCs reflects the degree to which the content of the relative clause is "integral to the meaning of the clause" (Collins et al., 2014: 129). RRCs, as in Example (1) and Example (2)³ modify their antecedent noun typically by restricting the denotation of the antecedent (ibid.), though, less prototypically they may also provide essential information about the antecedent without fulfilling this restrictive function. NRRCs, in contrast, provide supplementary information, as in Example (3). In written, edited texts, they are generally set off by commas (or other forms of parenthesis) – though this is not reliably the case, as is illustrated in Example (4).

-
- (1) To the east and west of Kabwe are some natural *attractions* **that** have not yet been developed for tourism, but road access is very poor. (SA_I-084-E)
- (2) Implied consent: *consent* **which** is not spoken or given in writing, but rather indicated by the actions of the parent or guardian and the facts and circumstances of a particular situation. (SA_I-074-E)
- (3) The *trainers*, **who** held at least Master's Degrees, were considered knowledgeable ... (SA_A-124-E)
- (4) A family needs help finding *David Mpotu* **who** was last seen when he went to work last year. (SA_R-102-E)
-

Table 1 shows that there is an asymmetry in relativiser choice based on restrictiveness: *which* can be used for both RRCs and NRRCs, but *that* only for RRCs. This asymmetry has formed the foundation of prescriptive advice to restrict relativiser *that* to RRCs and *which* to NRRCs, in a quest for a more elegant and orderly grammar. Curzan (2014: 24) refers to this as a form of stylistic prescriptivism. The prescriptive advice, which originates from Fowler (1926/1994: 635), was subsequently taken over in Strunk and White's (1959) influential US English usage guide, and Hinrichs, Szmeccsanyi and Bohmann (2015) have argued that the prescriptive *that*-rule may be seen as a case in which prescriptivist advice aligned with other factors to contribute to the decline of *which* and the rise of *that* in AmE.

The animacy of the antecedent noun is a second constraint that affects the choice of relativiser. According to Collins et al. (2014: 132), numerous studies, such as Quirk et al. (1985) and Huddleston and Pullum (2002) have observed that *who* and *whom* are restricted to human antecedents (as in Example (3) and (4)), while *whose* shows more variability. *That* and *which*, in contrast, are in principle available to both human and non-human antecedents, although, as argued by D'Arcy and Tagliamonte (2015: 266), who focus on vernacular spoken Toronto English, the "foundational association of *who* with human antecedents effectively blocks its use with nonhuman ones". There thus appears to be a clear division of labour: *who* is reserved for human antecedents, while *that* and *which* tend to take on the role of relativiser for other types of antecedents (as in Example (1) and (2)).

The syntactic function of the relativiser also governs relativiser choice, and the distinction centres on whether the relativiser functions as a subject (as in Example (1)) or object (as in Example (5)) of the relative clause. According to Quirk et al. (1985: 366), *who* and *that* are the available relativisers for subject RRCs with personal antecedents, and *which* and *that* for subject RRCs with non-personal antecedents, while *whom*, *that* and zero-relativiser are available for object RRCs with personal antecedents, and *which*, *that* and zero-relativiser for object RRCs with non-personal antecedents. Zero-relativisers are therefore not (technically) available in subject RRCs,⁴ and *whom* is limited to object RRCs.

-
- (5) Stevens (2010) discusses four different *expectations* **that** employers need to match up to in order to keep their employees engaged. (SA_A-006-E)
-

³ Unless otherwise noted, examples are taken from the corpus of unedited and edited texts used in this study. The filenames indicate the variety of English, the register, the file number, and whether the text in question is unedited or edited. So, the filename SA_I-084-E in Example (1) indicates that it is from the SAfE subcorpus (SA) (see Section 3.1 for details of the full corpus from which the data are taken), in the instructional register (I), text 084, in its edited version (E). Where relevant to the analysis, relativisers are in bold, antecedents in italics, and relative clauses are underlined.

⁴ There is, however, some evidence that zero-relativisers do occur in vernacular varieties, as demonstrated in, for example, Tagliamonte et al. (2005).

Table 1
Relative pronouns available in different contexts, adapted from Quirk et al. (1985: 366).

	Restrictive		Non-restrictive	
	Personal/human	Non-personal	Personal/human	Non-personal
Subject	<i>who</i> <i>that</i>	<i>which</i> <i>that</i>	<i>who</i>	<i>which</i>
Object	<i>whom</i> <i>that</i> zero	<i>which</i> <i>that</i> zero	<i>whom</i>	
Genitive	<i>whose</i>			

Relativisation in English has been extensively researched, from synchronic and diachronic perspectives, and in spoken as well as written language across various registers. The substantial body of research on the English relativiser system has shown that apart from the three factors discussed, other language-external and –internal factors also play a role in the choice of relativiser. For example, formality has been shown to condition this alternation and the *wh*-relativisers are strongly associated with formal written registers. D’Arcy and Tagliamonte (2010: 385) argue that these relativisers entered the English linguistic system as a “change from above”, which continues to be evident in the fact that their use is associated not only with more formal and written registers, but is also correlated with user variables such as high social status and level of education. Clear register differences are also attested across numerous studies (e.g., Biber et al., 1999; Collins et al., 2014; Grafmiller et al., 2018), and there is a substantial body of research demonstrating that there are also cross-varietal differences. However, the focus of such research has mostly been on northern hemisphere native varieties of English like BrE, AmE, and Canadian English and there is limited research on other varieties. Exceptions include Sigley (1997) on New Zealand English; Gut and Coronel (2012) on Nigerian English, Jamaican English, Philippine English (PhilE) and Singapore English (SingE); Collins et al. (2014) on PhilE; and Lee (2020) on SingE. Such studies find that while there is much stability in the relative system across different varieties of English, differences can be found in terms of the choice of relative marker used, with some studies suggesting that lower-level divergences in some varieties may be attributed to the endonormative status of the variety (see, e.g., Gut and Coronel, 2012; Collins et al., 2014; Lee, 2020).

However, the role of language-internal factors has received significantly more research attention than external factors like register and variety (D’Arcy and Tagliamonte, 2010; Lee, 2020). Apart from the factors already discussed (restrictiveness, animacy of the antecedent and syntactic function of the relative clause), a number of other language-internal factors are known to influence relativiser choice. Complexity (measured by, e.g., the length of the antecedent noun or the length of the relative clause) is a much researched factor. There is also evidence of the priming effect of the preceding relativiser, as shown by, for example, Hinrichs et al. (2015); and Grafmiller et al. (2018), who draw on the Brown family of corpora. The antecedent part of speech (e.g., whether it is a noun, or some other part of speech, such as a pronoun or numeral), and the definiteness of the antecedent noun also play a role, as shown, for example, by Tagliamonte et al. (2005); Hinrichs et al. (2015); and Grafmiller et al. (2018).

2.2. Relativiser *who*, *that* and *which* with subject RRCs: Frequency and functional differentiation

Studies of the English relativiser system have focused principally on the choice between *that* and *which* (and zero) in RRCs with inanimate antecedents (see Hinrichs et al., 2015; Grafmiller et al., 2018; Labat et al., 2023), and there have been fewer studies of relativiser choice in RRCs with animate antecedents, where a different choice is possible: animate antecedents tend to take *who*, *whom*, and *whose*, while *that* and (to a limited degree) *which* is also available for animate antecedents (in addition to being available for inanimate antecedents). With regard to subject RRCs specifically, Jankowski (2013: 57) notes that “while the locus of variability in the vernacular system of subject relatives lies with the competing use of *who* and *that* with human antecedents, in the written registers, the variability lies between *that* and *which*”. As a result, the focus of most research on subject RRCs has either been on the alternation between *that* and *which* with inanimate antecedents in written texts, or on alternation between *that* and *who* with animate antecedents in spoken language. Far less research has been done on the alternation between *that* and *who* with subject RRCs with animate antecedents in written language, despite the fact that variation does exist: *that* is used with some frequency with animate nouns and there is also some variation in prescriptive advice on the use of *that* with animates (see Section 2.3). For this reason, in this study, we focus only on RRCs in subject position, and more specifically, those with animate antecedents, for which, in principle, *who*, *which* and *that* are available as relativisers.

During the course of the seventeenth century, *who*, the last to develop a relativising function (Romaine, 1980), gradually came to mark human antecedents, with *that* more commonly used for non-human antecedents (D'Arcy and Tagliamonte, 2015). *Who* is therefore highly restricted to human antecedents only, and was specialised for human subjects from the beginning (Tagliamonte et al., 2005) but *that*, the most common relativiser (Biber et al., 1999), has more flexibility, and is operational across all contexts of RRCs (subject and non-subject). Several recent studies have demonstrated the variability of *that* with animate antecedents in subject RRCs. For example, investigating three registers of Canadian English (face-to-face speech, press writing and Hansard) over a 100-year period (1906, 1956 and 2006 as sampling years), Jankowski (2013) finds that *that* and *who* parallel each other in proportion over time in subject RRCs in press writing until 2006, when the proportions begin to diverge – *who* starts to decrease in frequency, while *that* increases so that *who* is used in 41.5% of cases and *that* in 58%. She also finds visible cross-register differences in 2006 between press writing and Hansard: “the gap between the rates of use of [*that* and *who*] is considerably more pronounced in *Hansard*, with a noticeably lower proportion of use of *who* (41.5 percent in 2006 *Maclean's* [press writing] vs. 30% in 2006 *Hansard*)” (Jankowski, 2013: 50). In terms of animacy, she finds that in her Hansard and newswriting corpora, human antecedents in RRCs in subject position near categorically take *who*, but that there is more variability in the spoken data, where *that* is used in subject RRCs with human antecedents in 45.2% of cases. She further finds that it is in the category of collectives referring to humans where the use of *that* and *who* is similar across the written and spoken registers she investigates: *who* is used in around a quarter of RRCs in subject position. The variability evident in categories like collectives suggests the fuzzy nature of animacy distinctions (Jankowski, 2013).

The fuzziness of animacy distinctions is also noted in research on genitive alternation, a feature that is known to be strongly conditioned by animacy constraints. Some researchers have shown that animacy constraints operate less strongly in L2 varieties of English (see, e.g., Heller et al., 2017) with others arguing that this might be as a result of contact with substrate languages (Rosenbach, 2017). The findings for genitive alternation open up the possibility that because animacy constraints may operate less strongly in L2 varieties, likely under the influence of substrate languages, different distributional patterns of other linguistic features known to be conditioned by animacy constraints (such as relativiser choice) may be observed in these varieties too.

Studies of relativiser choice in new Englishes have tended to focus on the normative orientation of these varieties, exploring how they pattern in relation to their parent varieties, and many include animate antecedents in subject RRCs in their analyses. For example, Collins et al. (2014) investigate relative clauses in PhilE, compared to BrE, in corpora representing two time periods (the 1960s and 1990s), and four genres (press, general prose, learned writing and fiction). They find that over the three decades *that* has become the most frequent relativiser in PhilE, and ask whether an increasing acceptance for the use of *that* with human antecedents might account for this increase, so that *that* takes over some of the domain of *who*. Their findings suggest that this is not the case, and they conclude that “while acceptance of human *that* may be increasing, it represents only a very small contributing factor to the rising popularity of *that*” (Collins et al., 2014: 137). Looking at SingE in newswriting, over three time periods (1993, 2005 and 2016), Lee (2020) finds that *that* is very rarely used with human antecedents, and that relativisation patterns in SingE are very similar to BrE. Gut and Coronel (2012) investigate relativisation strategies in Nigerian, Jamaican, PhilE and SingE. Their emphasis is on normativity, motivated by the fact that in many new Englishes the “major contact with English is through literacy” (Gut and Coronel, 2012: 216). They argue that this leads to reduced variability in style, and a convergence between spoken and written styles (adjusted to normative expectations for written language). They hypothesise that “the greater the influence of external linguistic norms, the smaller the stylistic variability will be” (Gut and Coronel, 2012: 218), but find little evidence of monostylistic usage as far as relativisation is concerned (Gut and Coronel, 2012: 237). Nevertheless, in the two Englishes with the clearest exonormative orientation, Nigerian and Jamaican English, stylistic variability is lower than in PhilE and SingE (Gut and Coronel, 2012: 238). They also find that in written language, the relative marker *that* with human antecedents is hardly used in PhilE and SingE, and that the highest percentage of *that* with human antecedents is exhibited by Nigerian English, at a rate of 7.5% (Gut and Coronel, 2012: 226). They further find that, as is the case more generally, *that* with human antecedents occurs more frequently in spoken than written language for the four new Englishes that they investigate, but the proportional frequency of *that* in this context does not reach 40% in any of the four varieties (unlike in AmE and New Zealand English), where frequencies are higher; Gut and Coronel, 2012: 228).

2.3. Depriving them of their humanity: Prescriptive advice and the alternation between *who* and *that* for animate antecedents

As noted in Section 1, there is a long history of prescriptive advice targeting the use of relativisers. With regard to the choice between *that* and *who* in RRCs with animate antecedents, the differing prescriptive advice in normative sources

originating from the US, the UK, and South Africa is noteworthy, and there is great variation in the advice, across both time and regions. For example, in the first and second editions⁵ of the influential BrE usage guide, *Fowler's Modern English Usage*, Fowler expresses the ideal for a clear distinction between the functions of *that* and *who*: "It would be satisfactory if the same clear division of functions that can be confidently recommended for *that* and *which*, viz. between defining and non-defining clauses, could be established also for *that* and *who*; this would give us *that* for all defining clauses whether qualifying persons or things, and *who* for persons but *which* for things in all non-defining" (Fowler, 1926/1994: 716, Fowler, 1965/1968: 701). However, Fowler notes that there appears to be a reluctance to use *that* for persons, which he suggests is motivated by politeness: it "is perhaps felt to be a sort of slight, depriving them of their humanity as one deprives a man of his gentility by writing him Mr instead of Esq." (ibid.). As a result, in the first two editions of the usage guide, the prescriptive advice offered is summarised as follows: "(A) of *which* and *that*, *which* is appropriate of non-defining and *that* to defining clauses; (B) of *which* and *who*, *which* belongs to *things* and *who* to persons; (C) of *who* and *that*, *who* suits particular persons and *that* generic persons" (Fowler, 1926/1994: 713, Fowler, 1965/1968: 699). This advice, however, changes in the third and fourth editions of the usage guide,⁶ where the ideal expressed in the first two editions to limit *that* to RRCs only, irrespective of the animacy of the antecedent, is removed entirely and readers are instead advised that *who* is the relative pronoun following a human antecedent and *that* (or *which*) the relative pronoun following an inanimate antecedent (Burchfield, 1996/1998: 773; Butterfield, 2015: 808). However, the following additional refinements regarding the use of *who* or *that* in relation to animate antecedents is provided: "Either *who* or *that* may be used when the antecedent is animate but not human, or when the antecedent is human but representative of a class [...]. In contexts containing double antecedents, of which the first is human and the second inanimate, *that* is naturally required" (Burchfield, 1996/1998: 773; Butterfield, 2015: 808).

In the AmE usage guides by Garner (1998, 2003, 2009, 2016), there is initially a very clear distinction in functions between *that* and *who*, with the use of *that* with human antecedents (indirectly) proscribed: "*who* is the relative pronoun for human beings: *that* & *which* are relative pronouns for anything other than humans, including entities created by humans" (Garner, 1998: 694–695). This strict separation of roles does not carry through to subsequent editions. In the second edition, Garner (2003: 836) points to the acceptability of *that* with human beings through the addition of the following parentheses: "*who* is the relative pronoun for human beings (although *that* is also acceptable)". In the third and fourth editions, Garner (2009: 808, 2016: 903) becomes more emphatic in the acceptability of *that* with human antecedents, stating that "*people that* has always been good English, and it's a silly fetish to insist that *who* is the only relative pronoun that can refer to humans". Overall, in both the BrE and AmE usage guides, normative advice becomes more diffuse and less prescriptive over time.

A stricter distinction emerges and is maintained in SAfE usage guides, where there is a very clear division of roles between *that* and *who* in RRCs. Linnegar (2009: 146) advises readers to "[u]se 'who' to refer to people [...] and] 'that' is more usually used for animals and things". He further notes that *that* is frequently used incorrectly in formal writing, citing the following examples, all of which refer to human antecedents, as cases of incorrect usage:

- The person *that* took my laptop must please return it. (use *who*)
- People *that* drive around while on the cellphones are a menace. (use *who*)
- 'It can't be that it is white students *that* are still dominating – there is no reason for that 15 years into our democracy,' Blade Nzimande said. (use *who*)
- 'It's difficult to understand the behaviour of a partner towards someone *that* has been convicted and surrounded by allegations of inappropriate behaviour,' said Shelver. (use *who*)
- 'The feedback from 53rd Venice Biennale has been positive for the South African arts and artists *that* made it there, despite Arts and Culture's complete lack of support for a venture showcasing out artists internationally,' wrote Minnaar. (use 'art that and artists who') (Linnegar, 2009: 146–147)

The strict proscription against the use of *that* with human antecedents is upheld in a style guide published by the communications department of the South African government (GCIS, 2013) as well as in a guide to grammar and punctuation published by the Professional Editors Guild (PEG) in South Africa, some 10 years after Linnegar (2009). Steynberg (2018: 21–22) notes that "'[w]ho' (subject of the verb) and 'whom' (object of the verb) are used when referring to people [...] '[t]hat' or 'which' is used when referring to things". In Steynberg (2018: 22), however, further guidance on

⁵ The first and second editions are substantially the work of Henry Watson Fowler, although the second edition was revised by Ernest Gowers.

⁶ The third and fourth editions, while still carrying the name of Fowler in the title, were substantially revised and edited by R.W. Burchfield and Jeremy Butterfield, respectively.

regional differences are noted when readers are alerted that “American English frequently uses ‘that’ to refer to people, whereas British and South African English use ‘who’”. However, no evidence is provided in support of this claim of regional differences. A similar claim is made in two articles in the organisation’s newsletter, *PEGBoard*, where [Aecer \(2019, 2022\)](#) states that the use of *that* with human antecedents is acceptable in AmE, but is not preferred in BrE or SAfE. Within the context of SAfE, these prescriptive sources frame the use of *that* with human antecedents as an error typically committed by ESL or EFL users ([Aecer, 2022: 4](#)) or else as a “common confusable” among writers and editors who “aren’t aware of the difference in meaning or function” ([Linnegar, 2009: xiv](#)). These assertions, however, are not without opposition and the view emerges as a point of contention among South African editors, with arguments arising both in discussion forums for editors and in articles in *PEGBoard*. For example, [Blight \(2014\)](#) argues against the prescription to use *that* with human antecedents that has emerged in the discourse among South African editors, citing the two editions of the *South African Concise Oxford Dictionary* as evidence that such usage is acceptable and questioning: “Why, then, are so many competent speakers of English, including highly qualified editors with such dictionary evidence available to them, so set against using *that* with human antecedents?” ([Blight, 2014: 19](#)).

There is therefore considerable variation in the prescriptive advice over time and across varieties. Contemporary advice, at least since the start of the twenty-first century can be described on a cline of openness to the use of *that* with human antecedents in RRCs, with AmE usage guides providing no circumstances in which such usage is unacceptable, BrE usage guides specifying specific contexts in which such usage is acceptable, and SAfE guides proscribing any such usage. [Law Favō \(2024\)](#) finds that while editors of English in South Africa make use of a small selection of South African usage guides, in general, there is a much stronger preference for BrE and, to a lesser extent, AmE sources. Overtly, editors therefore appear to be oriented towards an external norm – but a norm that is, in itself variable and pluralistic. However, as argued by [Kruger and Van Rooy \(2017\)](#), editors are not ‘norm machines’, but are themselves language users whose own linguistic background and experiences influence their editorial work. While prescriptive or normative advice may form a more salient part of editors’ linguistic representations compared to other language users, it is, ultimately, just one part. The fact that South African editors are themselves embedded in the complex sociolinguistic and diffuse normative space of SAfE, with its multiple subvarieties embedded into a broader multilingual context, and that they regularly encounter texts produced by users from these subvarieties, is likely to affect their editorial work as much as their own linguistic representations.

What, then, are the usage preferences for relativisers with animate antecedents in RRCs in subject position, by writers of the different South African varieties investigated in this study? And how do editors respond to such preferences? Does the linguistic background of the editor (in relation to that of the author) matter? And what does this tell us about the normative orientation of South African editors? In the next section we set out the method used to answer these questions.

3. METHODOLOGY

3.1. Corpus

The corpus used in this article is taken from a larger, register-differentiated, aligned corpus of unedited English texts and their edited counterparts from Australia and South Africa (see [Law, 2019](#)). The construction of the South African component of the corpus was initially commenced under the directorship of Haidee Kotze and was extended (also to include texts from Australia) as part of the PhD study by [Law \(2019\)](#). To obtain unedited texts as well their edited counterparts, various publishing organisations (such as publishers and news agencies), organisations representing editors and authors, and language services agencies in the two countries were contacted via email with a request to contribute texts. In each case, the contributing publisher, editor or author obtained the appropriate permission from the relevant parties and supplied the texts. The corpus, which is currently housed in Sketch Engine, has been automatically tagged using the English TreeTagger PoS tagger with Sketch Engine modifications (<https://www.sketchengine.eu/english-tree-tagger-pipeline-2/#toggle-id-2>).

For this study, we made use of texts from the South African component of the corpus representing five registers: academic, creative, instructional, popular and reportage. Academic writing is composed of academic articles, dissertations and theses, while creative writing consists of narrative fiction. The instructional register includes instruction manuals and administrative documents. The popular register is composed of non-fiction texts published for a general readership, while reportage consists of newswriting published in several regional newspapers.

All texts included were professionally edited in South Africa and metadata are available for the different subvarieties of the authors and editors. We thus include texts authored by AfrE, BSAfE and WSAfE authors,⁷ and edited by either WSAfE or AfrE editors. While most texts are authored by a single author, some are written by multiple authors representing the three different subvarieties investigated. We include these multi-authored texts in this study, labelling them as of mixed provenance. Tables 2 and 3 show the structure and size of the subcorpus used, with Table 2 reflecting word counts across each register, per author subvariety for the unedited and edited components of the corpus and Table 3 reflecting the number of texts per subvariety and register alongside mean text length and standard deviation.

There are several difficulties associated with the construction of this type of corpus, not least of which is obtaining the unedited versions of edited, published texts. As is clear from Table 2, texts are not available in all of the registers for all of the subvarieties investigated, and the corpus is not balanced across registers and subvarieties. In part, this simply reflects the realities of writing and editorial practices in South Africa. For example, in the South African context, Afrikaans authors are less likely to write creative texts in English than they are in Afrikaans, thereby accounting for the absence of AfrE texts in the creative register. In other cases, the lack of texts in some registers in the corpus reflects the difficulty of accessing texts; for example, in news production contexts original unedited texts are often not retained, and in collecting corpus texts we were restricted to newspapers where we could make personal contact with journalists to supply original texts. This is part of the reason why there are (for example) no news texts authored by WSAfE journalists. At most of the local newspapers that were willing to contribute texts, only L2 English journalists were employed (see, Law 2019: 113–117 for more detail).

A further challenge with a corpus of this nature is whether to include equal length samples of texts or whole texts. Since the corpus has been constructed to investigate the interventions of editors, and since these interventions occur sporadically only, the corpus includes the whole version of texts submitted by publishers, authors, and editors to ensure that as many instances of intervention are captured as possible. As a consequence of this, and as is clear in Table 3, there is a great deal of variation, not only in terms of text length (as evidenced by the standard deviations reported) but also in terms of the number of texts represented in each register, by variety.

Although our analysis focuses on the impact of subvariety on unedited and editing writing, and does not focus on distributional patterns across registers, the lack of texts in certain registers for some of the subvarieties, and the uneven distribution of texts (and word counts) across the subvarieties investigated should be considered when interpreting the findings.

3.2. Data extraction, coding and analysis

As already indicated, our main interest is RRCs in subject position with animate antecedents, which can be introduced by either *that* or *who*, and less frequently, *which*. Since existing studies have found low frequencies of *which* in these contexts (see, e.g., Jankowski, 2013), and in order to include the full envelope of variation, we include all three relativisers. First, we extracted all instances of *that*, *who* and *which* from the corpus and inspected a random sample of 300 instances for each of the five PoS tags that occurred with each search term in the data frame. This initial inspection showed that three of the five tags in the data frame, ‘*that*-subordinator’, ‘*determiner*’ and ‘*adverb*’, yielded cases in which *that*, *which* and *who* were not used as relative pronouns. We therefore excluded these tags and only extracted instances where *that*, *who* and *which* had been tagged as *wh*-pronouns and *wh*-determiners – an extraction strategy also used by Lee (2020). In order to ensure that we isolated cases in which we would be dealing with RRCs in subject position, we followed the principles outlined in Labat (2018) and Labat et al. (2023), and excluded instances where the relativiser is followed by a personal pronoun. Following Labat (2018) and Labat et al. (2023), a similar restriction was placed on possessive pronouns, adjectives, nouns and determiners. Unlike most research on RRCs, we did not make use of commas to distinguish between RRCs and NRRCs when extracting the data frame, since we are dealing with unedited written texts where authors might erroneously use commas in RRC contexts.

Once the data frame had been extracted from Sketch Engine, we manually inspected each line of data and discarded those that were not cases of RRCs in subject position. From those that remained, we extracted the antecedent head noun of the NP, which did not necessarily have to come immediately before the relativiser, and coded each for animacy, using the animacy coding scale in D’Arcy and Tagliamonte (2010) and Brook and Tagliamonte (2023), who take it from Hinrichs and Szmrecsanyi (2007: 449), who in turn get it from Zaenen et al. (2004). On a scale from most inanimate to most animate, these categories are inanimates or things (coded as I), animals (coded as A), collectives, including

⁷ Although multiple varieties of English are used in South Africa, including South African Indian English, which is regarded as a major variety of English in the country (Mesthrie, 2010; Van Rooy, 2014), the corpus used in this study does not include enough instances of texts written by authors of South African Indian English, and we therefore limit our analysis to the three other major varieties, AfrE, BSAfE and WSAfE. The selection of these three varieties is aligned with previous studies of editorial intervention in SAfE.

Table 2

Corpus composition: Word count per author subvariety by register for unedited and edited texts.

	Academic	Creative	Instructional	Popular	Reportage	Total
Unedited						
AfrE	571 450	0	239 136	41 977	12 591	865 154
BSaFE	191 958	0	20 151	20 749	27 875	260 733
Mixed	97 853	0	30 295	0	0	128 148
WSaFE	38 346	159 308	307 971	142 145	0	647 770
Total	899 607	159 308	597 553	204 871	40 466	1 901 805
Edited						
AfrE	566 807	0	242 739	43 290	12 455	865 291
BSaFE	194 102	0	19 671	20 772	25 435	259 980
Mixed	100 778	0	33 913	0	0	134 691
WSaFE	39 161	154 152	295 619	170 362	0	659 294
Total	900 848	154 152	591 942	234 424	37 890	1 919 256

Table 3

Corpus description: Number of texts per subvariety and register alongside mean text length and standard deviation.

	Texts	Unedited subcorpus		Edited subcorpus	
		M	SD	M	SD
AfrE	94	9 203.8	20 872.2	9 205.2	20 856.8
Academic	39	14 652.6	27 434.0	14 533.5	27 314.4
Instructional	19	12 586.1	20 915.3	12 775.7	21 147.6
Popular	11	3 816.1	6 693.8	3 935.5	7 054.4
Reportage	25	503.6	523.5	498.2	532.7
BSaFE	115	2 267.2	6 226.3	2 260.7	6 281.5
Academic	19	10 103.1	11 571.3	10 215.9	11 660.2
Instructional	2	10 075.5	4 278.7	9 835.5	4 782.2
Popular	1	20 749.0	n/a	20 772.0	n/a
Reportage	93	299.7	96.8	273.5	92.8
Mixed	7	18 306.9	17 755.1	19 241.6	18 284.8
Academic	5	19 570.6	20 659.4	20 155.6	21 731.7
Instructional	2	15 147.5	12 500.9	16 956.5	10 115.2
WSaFE	44	14 722.0	25 936.4	14 984.0	25 201.3
Academic	11	3486.0	1 730.2	3560.1	1 690.3
Creative	3	53 102.7	33 624.5	51 384.0	31 665.8
Instructional	27	11 406.3	25 101.7	10 948.9	22 163.3
Popular	3	47 381.7	21 766.9	56 787.3	26 564.2
Total	260	7 314.6	17 823.4	7381.8	17 698.1

groups and indefinite pronouns (coded as C), humans (coded as H), and the word *people* (coded as P). We included a sixth category, mixed (coded as M), for coordinated antecedents where the items belong to different animacy categories. Since our interest is in subject RRCs with animate antecedent nouns, we exclude inanimate antecedents from the further analysis. This process yielded 5 920 cases that were used for the analysis in this article. The distribution of cases across the animacy categories used is shown in Table 4.

Following previous work on RRCs in subject position with animate antecedents (see, e.g., Jankowski, 2013; Collins, 2014; D'Arcy and Tagliamonte, 2015), for our analysis, we start with a basic distributional analysis of the proportion of the three relatives per antecedent type by author subvariety in the unedited and edited texts (we exclude the 'mixed' category from further analysis, due to the low frequency of observations in this category). We use this to calculate the frequency change (in percentage points) in the relativisers across the unedited and edited texts, in order to reflect on how editorial intervention (re)shapes relativiser distribution. Subsequently we investigate the link between these changes and the backgrounds of the editors (in relation to that of the authors), while also considering other factors, such as the influence of particular lexical items. The cross-tabulation of the proportional distribution exposes the intersection of the factors we are most concerned with in this study (namely animacy, subvariety, editorial intervention) and allows us to investigate (1) how the relativisers pattern in terms of the different animacy categories we coded for, as well as the

Table 4
Raw count of cases per animacy category.

Category	Count
Animals	208
Collectives	2 169
Humans	2 913
Mixed	86
<i>People</i>	544
Total	5 920

different subvarieties investigated, in original unedited writing, and (2) how editorial intervention contributes to the distributional proportions reported for written, edited texts in current studies.

4. FINDINGS

In this section, we present the findings of our analysis partitioned according to the four animacy categories investigated: animals, collectives, humans, and *people* (ordered from 'least' to 'most' prototypically animate).

4.1. Animals

The results for animal antecedents in subject RRCs, across varieties and unedited texts, are shown in Table 5. As is evident, *that* is preferred across the three subvarieties (though frequencies in the texts by BSAfE authors are too low to provide a reliable analysis). In AfrE, however, *that* is far more strongly preferred (95.7%) than in WSAfE (75%), where more variability is evident – also because *which* is used in the WSAfE data as relativiser with animal antecedents in RRCs. The stronger preference for *that* in the writing of AfrE authors may be accounted for by the fact that in Afrikaans (the L1 of the authors), relative clauses are only formed by one relativiser, *wat* 'that', which is used irrespective of the animacy of the antecedent (see Ponelis, 1979; Kirsten and Breed, 2020). There are two exceptions: *wie* (the translation equivalent of *who* in Afrikaans) is used with animate antecedents in the genitive construction with *se* (the equivalent of the English s-genitive; in English, *whose* would be used in this context), and in constructions with prepositions. Kapp (2018) points out that the use of *wie* 'who' with animate antecedents beyond these restricted contexts occurs under the influence of English. As a result of this, while WSAfE writers use *who* with animals around a quarter of the time, AfrE writers, most likely under the influence of Afrikaans, opt for *that* with animal antecedents almost without exception. Of course, *that* is also the most common relativiser, and as such may represent a neutral 'default' option.

In editing, the proportional preference for *that* remains very similar; however, the somewhat stronger preference for *who* in WSAfE is toned down by editors, and some cases of *which* are removed and replaced with *that*, resulting in an increased preference for *that* (an increase of 16.7 percentage points). This serves to adjust the proportional distribution of *that* and *who* in edited WSAfE so that it is similar to that of unedited and edited AfrE writing.

The texts in this subset have mostly been edited by editors with a WSAfE background; of the total of 101 edited observations, 97 were edited by editors with a WSAfE background. Closer inspection of the data yields some evidence that these WSAfE editors occasionally directly target perceived inappropriate uses of *who* with animal antecedents, as in Example (6). They also replace *which* as a relativiser with animal antecedents in RRCs with *that*, as in Example (7) – though this is not done consistently, and some cases of *which* remain in the edited texts, as in Example (8).

Table 5
Frequency of *that*, *which* and *who* with animal antecedents in unedited and edited writing, across author varieties.

	Unedited				Edited				Percentage point change in preference for <i>that</i>
	<i>that</i>	<i>who</i>	<i>which</i>	% <i>that</i>	<i>that</i>	<i>who</i>	<i>which</i>	% <i>that</i>	
AfrE	45	1	1	95.7	47	1	1	95.9	+0.2
BSAfE	2	1	0	66.7	2	1	0	66.7	0
WSAfE	42	4	10	75.0	44	1	3	91.7	+16.7
Mixed authors	1	0	0	100	1	0	0	100	0
	90	6	11	84.1	94	3	4	93.1	+9.0

-
- (6a) The Jane Goodall Institute provides rescue and care for orphaned chimps **who** have been misplaced from the natural habitat around Africa. (SA_P-018-O)
- (6b) The Jane Goodall Institute provides rescue and care for orphaned chimps **that** have been misplaced from their natural habitat in Africa. (SA_P-018-E)
-

- (7a) Other marine life that can be viewed, not hunted, in the Mafia Island Marine Park, are the gentle dugongs, and endangered green turtles **which** lay their eggs here and on the smaller islands. (SA_I-084-O)
- (7b) Other marine life that can be viewed (not hunted) in the Mafia Island Marine Park are dugongs, as well as the endangered green turtles **that** lay their eggs here and on the smaller islands. (SA_I-084-E)
-

- (8a) Savuti is known for its elephants **which** wander through the campsite as it is unfenced. (SA_I-084-O)
- (8b) Savuti is known for its elephant **which** wanders through the campsite, as it is unfenced. (SA_I-084-E)
-

However, the shift in preference towards *that* with animal antecedents, most evident for WSAfE writing edited by WSAfE speakers, is accounted for also by other more comprehensive changes (additions and deletions) that alter the proportional frequencies of the three relativisers.

4.2. Collectives

As far as collectives that can be construed as animate are concerned, there is considerable variability in the choice between *that* and *who*, though *who* is generally somewhat more strongly preferred (see Table 6). Collectives, then, are clearly considered more prototypically animate than animals, as reflected in the stronger preference for *who* (compared to the stronger preference for *that* with animals). Writing in the two L2 varieties (AfrE and BSAfE) is more similar, with a stronger preference for *who* with collectives that may be construed as animate (58.1 and 59.5% of the time). In WSAfE writing there is a balance in the preference for *who* and *that*, with *who* preferred only 50% of the time. Collective antecedents, then, are perceived as more ambiguous in terms of animacy, across all three subvarieties. Writing produced by groups with a mixed linguistic background demonstrates by far the strongest preference for *who*: 67.2% of the time (though it should be noted that this is a small subset of the data). In all subsets of the data *which* is sometimes used as relativiser with animate collectives: 1.8% of the time in AfrE writing, and at somewhat higher frequencies in BSAfE and WSAfE (3.4 and 3.3% of the time). The frequency of *which* with animate collective antecedents is highest for groups of authors with mixed linguistic backgrounds, at 9.8%.

Editing makes hardly any difference to AfrE and BSAfE writing, but in WSAfE writing there is a slight increase in the preference for *who* with animate collective antecedents (5.5 percentage points), which is at least in part accomplished both by a reduction in the use of *that* (from 170 to 158 cases) and removal of *which* (from 12 to 4 cases). For groups of authors with mixed linguistic backgrounds, the adjustment is also notable (4.2 percentage points). In this case, the stronger preference for *who* is minimised, in favour of a stronger preference for *that*. Editors therefore appear to endorse usage in the two L2 varieties (without adjusting to another, external norm). Instead, writing in WSAfE, the native variety, is more targeted for changes, along with texts written by mixed groups of authors.

A closer look at the animate collective antecedents in this dataset reveals that of the five most frequently occurring antecedents in this subset, four are pronouns. By far the most frequent is the demonstrative pronoun *those* (311 in the

Table 6
Frequency of *that*, *which* and *who* with collective antecedents in unedited and edited writing, across author varieties.

	Unedited				Edited				Percentage point change in preference for <i>who</i>
	<i>that</i>	<i>who</i>	<i>which</i>	% <i>who</i>	<i>that</i>	<i>who</i>	<i>which</i>	% <i>who</i>	
AfrE	155	225	7	58.1	162	234	6	58.2	+0.1
BSAfE	98	157	9	59.5	94	152	8	59.8	+0.4
WSAfE	170	182	12	50.0	158	202	4	55.5	+5.5
Mixed authors	14	41	6	67.2	23	46	4	63.0	-4.2
	437	605	34	56.2	437	634	22	58.0	+1.8

unedited corpus; 322 in the edited corpus), followed by the indefinite pronouns *one* (143 in the unedited corpus; 147 in the edited corpus), *someone* (54 in the unedited corpus; 61 in the edited corpus), and *anyone* (28 in the unedited corpus; 33 in the edited corpus). The only collective noun in the top-5 is *party* (37 in the unedited corpus; 39 in the edited corpus). These four pronouns all occur only with *who* or *that*; *party* occurs once with *which*, in the unedited corpus.

Close reading of the data suggests that the increase in the proportional frequency for *who* in WSAfE writing can at least in part be ascribed to targeted changes to the use of *that* with these and also other indefinite pronouns (like *everyone*), which is replaced with *who*. Table 7 shows the raw frequencies for the four most frequent pronoun antecedents with *that* and *who*, in the unedited and edited corpus produced by WSAfE writers, respectively.

In WSAfE writing, the preference for *who* with *someone* and *anyone* is near categorical, in both original and edited writing. For *one(s)* there is a significant amount of variability, and this variability is hardly altered by the editors (though some changes are made; see Example (9)). For *those*, there is a reduction in the frequency of *that*, and an increase in *who*, as illustrated in Example (10).

-
- (9a) “I’m the only one **that** got a job and a free vetkoek.” (SA_C-005-O)
 (9b) “I’m the only one **who** got a job and a free vetkoek.” (SA_C-005-E)
-

-
- (10a) Over a three year period the number of children being cared for significantly reduced and those **that** were there did not remain in the orphanage for longer than three months. (SA_I-074-O)
 (10b) Over a three year period the number of children being cared for significantly reduced and those **who** were there did not remain in the orphanage for longer than three months. (SA_I-074-E)
-

It is striking, however, that most of the cases of *those* in texts by WSAfE writers occur in two texts (SA_I-074 and SA_P-017), both edited by WSAfE editors. The former case accounts for almost all of the cases of a shift from *that* to *who* with *those*: in the original version of the text, the ratio of *that* to *who* with *those* is 17:19, in the edited version it is 8:27 – with a much stronger preference for *who*. In the case of the latter text, the ratio in the original text (3:12) already favours *who*, and is not altered in editing (3:13). Clearly, there is a degree of individuality in authors’ and editors’ usage patterns. It should also be noted that even in a single text, neither editors nor authors are consistent in their usage; in the case of the text discussed above, there are also instances where *that* is retained with *those*, as in Example (11).

-
- (11) A church in Khayletisha decided to set up an outreach to support the orphans in their community, especially those **that** lived in child-headed families. (SA_I-074-O/E)
-

What precisely prompts an editor to make a change or not, is not self-evident from the corpus data available.

Which occurs with animate collectives at non-negligible frequencies, but localised in two texts (SA_I-084 and SA_A-049), and with a small set of collective nouns: *church(es)*, *communities*, *companies*, *country*, *group*, *species*, *state*, *couples* and *families*. It is significantly reduced in frequency. While this is occasionally the result of *which* specifically being targeted, as in Example (12), it is also the consequence of other changes, such as deletion or rewriting of text.

-
- (12a) There are a number of companies in Malawi **which** organise tours and/or safaris for travellers. (SA_I-084-O)
 (12b) There are a number of companies in Malawi **that** organise tours and/or safaris for travellers. (SA_I-084-E)
-

Table 7
 Raw frequencies of *that* vs *who* for the four most frequent collective antecedents in WSAfE writing, across edited and unedited writing.

	<i>that</i>		<i>who</i>	
	Unedited	Edited	Unedited	Edited
<i>those</i>	29	22	56	68
<i>one</i>	23	24	28	29
<i>someone</i>	1	0	30	33
<i>anyone</i>	2	2	20	21

4.3. Humans

Which occurs at negligible frequencies as relativiser with human antecedents in RRCs: only 3 times in the unedited corpus, and 2 times in the edited corpus (in texts written by AfrE writers). Given this very low frequency, we omit it from the further analysis in this section, only focusing on *that* and *who* (see Table 8).

As expected, RRCs with human antecedents, prototypically animate, have a strong preference for *who* as a relativiser, with WSAfE writers opting for *who* rather than *that* 94.1% of the time with human antecedents. Writers of the L2 varieties, and the writer groups with mixed linguistic backgrounds, however, demonstrate a less strong preference, with the AfrE writers using *that* with human antecedents 20% of the time. The higher frequency of *that* among AfrE writers even when *who* is strongly prototypical with human antecedents can likely be ascribed to the influence from the L1: since Afrikaans does not generally use *wie* 'who' as a relativiser for human antecedents (having only *wat* 'dat' as an option), Afrikaans writers might be prone to avoid or underuse *who* (see Section 4.1). BSAfE writers and groups of writers with mixed linguistic backgrounds tend to approximate usage in WSAfE more closely, though here, too, *who* is used less commonly (89.4% and 87.2% of the time) compared to the L1 variety.

In editing, there is a consistent pattern of adjustment towards an increase in proportional frequency of *who*, particularly strongly evident for AfrE writing, adjusted upwards in editing by almost 10 percentage points. In the following sections, we briefly consider each of the four groups of writers in turn, considering how the use of *that* or *who* is associated with different antecedents in both unedited and editing writing, and whether the linguistic background of the editor in relation to the author's background appears to influence the choice to replace *that* with *who* in RRC with human antecedents.

4.3.1. AfrE

As shown in Table 8, in 20% of cases (117 out of 584 cases) human antecedents in original AfrE writing collocate with *that*. The human antecedents that occur with *that* in AfrE original writing are shown in Table 9, ordered from most to least frequently occurring items, and within this, from highest to lowest proportion of *that*.

The most frequent human antecedents where *that* is used by AfrE writers are *learners*, *person*, *students*, *participants*, *individuals*, *respondents*, *roleplayers*, *researchers*, *consumers*, *employee* and *officials* (occurring 10 or more times in this dataset). These antecedents (while often specific to the academic and instructional text types that dominate the AfrE writing in this corpus) have in common that they tend to refer to generic groups (most are plural forms). Among this set of antecedents, the use of *that* varies from 6.9% to 59%.

Most targeted in editing are the antecedents *students*, *participants* and *respondents* (occurring mostly in academic texts) (see Table 10), where the relatively stronger preference for *that* by AfrE authors is toned down in favour of the use of *who*, as shown in Example (13). For other antecedents (with the exception of *consumers*, where there is an increase as a result of an addition of *that* by an editor) the same pattern of correction is observed, if more modestly.

(13a) Most of the students **that** registered for the degree also did their undergraduate studies at the [redacted]. (SA_A-009-O)

(13b) Most of the students **who** registered for the degree also did their undergraduate studies at the [redacted]. (SA_A-009-E)

In the AfrE dataset of relative clauses with human antecedents, the vast majority of cases were edited by editors with a WSAfE background (519 edited cases), with the remainder of cases (59) edited by editors who are themselves AfrE users. An analysis of the cases where editors have specifically targeted *that* as relativiser with human antecedents

Table 8

Frequency of *that* and *who* with human antecedents in unedited and edited writing, across author varieties.

	Unedited			Edited			Percentage point change in preference for <i>who</i>
	<i>that</i>	<i>who</i>	% <i>who</i>	<i>that</i>	<i>who</i>	% <i>who</i>	
AfrE	117	467	80	59	519	89.8	+9.8
BSAfE	23	194	89.4	12	200	94.3	+4.9
WSAfE	33	531	94.1	18	552	96.8	+2.7
Mixed authors	11	75	87.2	9	88	90.7	+3.5
	184	1267	87.3	98	1359	93.1	+6

Table 9

Human antecedents that occur with *that* in AfrE original writing: Frequency in unedited corpus, and proportional preference for *that*.

Antecedent	Raw frequency of antecedent in unedited corpus	Proportional preference for <i>that</i> (vs <i>who</i>) in unedited corpus (%)
<i>learners</i>	115	7
<i>person</i>	35	14.3
<i>students</i>	29	34.5
<i>participants</i>	22	59.1
<i>trainers</i>	18	11.1
<i>role players/role-players/roleplayers</i>	16	56.2
<i>individuals</i>	16	37.5
<i>respondents</i>	14	28.6
<i>researchers</i>	12	16.7
<i>consumers</i>	10	30
<i>employee</i>	10	20
<i>officials</i>	10	10
<i>persons</i>	8	37.5
<i>individual</i>	7	42.9
<i>clients</i>	7	14.3
<i>children</i>	6	33.3
<i>officers</i>	5	20
<i>workers</i>	5	20
<i>managers</i>	4	100
<i>assistants</i>	3	100
<i>old-timers</i>	3	100
<i>customers</i>	3	66.7
<i>guests</i>	3	33.3
<i>leader</i>	3	33.3
<i>man</i>	3	33.3
<i>suspects</i>	3	33.3
<i>intruder</i>	2	100
<i>thinkers</i>	2	100
<i>assessors</i>	2	50
<i>infants</i>	2	50
<i>speakers</i>	2	50
<i>user</i>	2	50
<i>academics</i>	1	100
<i>candidates</i>	1	100
<i>child</i>	1	100
<i>clerk</i>	1	100
<i>entrepreneur</i>	1	100
<i>foreigners</i>	1	100
<i>friends</i>	1	100
<i>graduates</i>	1	100
<i>human being</i>	1	100
<i>lady</i>	1	100
<i>leaders</i>	1	100
<i>lecturer</i>	1	100
<i>ministers</i>	1	100
<i>pioneer</i>	1	100
<i>problem-solvers</i>	1	100
<i>solvers</i>	1	100
<i>tourists</i>	1	100

for replacement with *who*, compared with cases where *that* was left unaltered (see Table 11) suggests that AfrE editors tend to be more prone to hypercorrection (implementing a correction 72% of the time) – perhaps because they are particularly aware of the overuse of *that* with human antecedents under the influence of Afrikaans, or because they follow particularly closely the advice in SAfE usage guides (see Section 2.3). In contrast, WSAfE editors are more likely to

Table 10

Percentage point changes across unedited and edited texts for the most frequent human antecedents occurring with *that* in AfrE.

Antecedent	Proportional preference for <i>that</i> (vs <i>who</i>) in unedited corpus (%)	Proportional preference for <i>that</i> (vs <i>who</i>) in edited corpus (%)	Percentage point change in preference for <i>that</i> (vs <i>who</i>)
<i>learners</i>	6.9	3.6	-3.3
<i>person</i>	14	10	-4
<i>students</i>	34.5	3	-31.5
<i>participants</i>	59	19	-40
<i>trainers</i>	11.1	10.5	-0.6
<i>individuals</i>	37.5	26.1	-11.4
<i>respondents</i>	28.6	0	-28.6
<i>role players/role-players/roleplayers</i>	56.2	42.9	-13.3
<i>researchers</i>	16.7	8.3	-8.4
<i>consumers</i>	30	44.4	+14.4
<i>employee</i>	20	10	-10
<i>officials</i>	10	10	0

Table 11

Frequency of AfrE and WSAfE editors targeting *that* with human antecedents in AfrE.

	Changed <i>that</i> to <i>who</i>	Retained <i>that</i>
AfrE editors	13	5
WSAfE editors	35	51

retain *that* with human antecedents (implementing a change 40.7% of the time), potentially reflecting their alignment with British or American usage guides.⁸ Of course, the much smaller number of cases for AfrE editors means that generalisations can only be made cautiously. Nevertheless, for WSAfE editors practices are more variable, and it is more likely that WSAfE editors editing the work of AfrE writers will leave relativiser *that* with an animate antecedent intact than when an AfrE editor edits the work of an AfrE writer.

4.3.2. BSAfE

As shown in Table 8, in 11% of cases (23 out of 218 cases) human antecedents in original BSAfE writing collocate with *that*. The human antecedents that occur with *that* in BSAfE original writing (mostly from academic writing, and to a lesser degree from popular writing and reportage) are shown in Table 12, ordered from most to least frequently occurring items, and within this, from highest to lowest proportion of *that*.

Among the five most frequent human antecedents in RRCs in BSAfE original writing occurring with *that*, the proportional preference is shifted to *who* for *person*, *leaders* and *employees* (see Table 13), in amendments such as those shown in Example (14).

(14a) Employees **that** feel appreciated are more likely to be engaged in their jobs. (SA_A-006-O)

(14b) Employees **who** feel appreciated are more likely to be engaged in their jobs. (SA_A-006-E).

Overall, then, there is a move towards increasing the preference for *who* in edited BSAfE (academic) writing, though this is a relatively small dataset, and caution should be exercised in generalising the findings.

In the subset of cases of RRCs with animate antecedents in texts written by BSAfE authors, 70 of the cases were edited by editors with an AfrE background, and 143 by editors with a WSAfE background. An analysis of cases where *that* is used with RRCs with animate antecedents shows that WSAfE editors are more inclined to change *that* to *who* in

⁸ There are 3 cases of *that* in the edited corpus without a matching text fragment in the unedited corpus, suggesting that editors added these cases of *that*, and 13 cases of *that* in the unedited corpus without a matching text fragment in the edited corpus, suggesting that these sentences were removed or radically rewritten during editing.

Table 12

Human antecedents that occur with *that* in BSAfE original writing: Frequency in unedited corpus, and proportional preference for *that*.

Antecedent	Raw frequency of antecedent in unedited corpus	Proportional preference for <i>that</i> (vs <i>who</i>) in unedited corpus (%)
<i>person</i>	24	4.2
<i>members</i>	10	10
<i>leaders</i>	9	22.2
<i>God</i>	7	14.3
<i>employees</i>	6	75
<i>member</i>	4	25
<i>pastors</i>	4	25
<i>heroes</i>	3	33.3
<i>Christians</i>	3	33.3
<i>respondents</i>	3	33.3
<i>student</i>	3	33.3
<i>boyfriend</i>	1	100
<i>caregivers</i>	1	100
<i>character</i>	1	100
<i>citizen</i>	1	100
<i>councilors</i>	1	100
<i>FC</i>	1	100
<i>strangers or men</i>	1	100
<i>theologians</i>	1	100

Table 13

Percentage point changes across unedited and edited texts for the most frequent human antecedents occurring with *that* in BSAfE.

Antecedent	Proportional preference for <i>that</i> (vs <i>who</i>) in unedited corpus (%)	Proportional preference for <i>that</i> (vs <i>who</i>) in edited corpus (%)	Percentage point change in preference for <i>that</i> (vs <i>who</i>)
<i>person</i>	4.2	0	-4.2
<i>members</i>	10	9.09	-0.91
<i>leaders</i>	22.2	11.1	-11.1
<i>God</i>	14.3	14.3	0
<i>employees</i>	75	16.7	-50

these cases (9 cases out of 14, or 64.3%), than to leave *that* as it is. This appears to be in contrast with the approach when editing the texts of AfrE users discussed in Section 4.3.1 (although the lower number of cases and editors involved here means that caution is warranted in making generalisations). In contrast, AfrE editors never change *that* to *who* under these circumstances, retaining *that* in all 5 cases (as shown in Example (15)).⁹

(15) There have been many theologians **that** have used the covenant concept. . . (SA_A-047O/E)

4.3.3. WSAfE

Only a very small set of human antecedents in original WSAfE English writing occur with *that*, as shown in Table 14. These cases occur mostly in instructional writing, and to a lesser degree in creative, popular and academic writing.

For the five most frequent antecedents, the proportional frequency of *that* is reduced in favour of *who* (see Table 15), with the exception of *adults*, where *that* is still used more than 20% of the time in edited texts.

In original WSAfE writing, RRCs with human antecedents are strongly associated with *who*, used 93.8% of the time, and editors further strengthen this preference. In this subset of the data, most cases (502) were edited by WSAfE edi-

⁹ In this subset of the data there are 2 cases of *that* with human antecedents in RRCs in the edited corpus that do not have corresponding text in the unedited corpus (both added by an editor with an AfrE background), and there are 4 cases of *that* in this construction in the unedited corpus, with no corresponding text in the edited subcorpus.

Table 14

Human antecedents that occur with *that* in WSAfE original writing: Frequency in unedited corpus, and proportional preference for *that*.

Antecedent	Raw frequency of antecedent in unedited corpus	Proportional preference for <i>that</i> (vs <i>who</i>) in unedited corpus (%)
<i>person</i>	115	4.3
<i>children</i>	71	11.2
<i>child</i>	51	7.8
<i>parents</i>	21	4.8
<i>adults</i>	9	22.2
<i>patients</i>	3	100
<i>persons</i>	3	33.3
<i>teachers</i>	3	33.3
<i>teenagers</i>	3	33.3
<i>you</i>	3	66.7
<i>refugees</i>	2	50
<i>flossie</i>	1	100
<i>panel beater</i>	1	100
<i>professionals</i>	1	100
<i>Ross Macdonald</i>	1	100
<i>therapists</i>	1	100

Table 15

Percentage point changes across unedited and edited texts for the most frequent human antecedents occurring with *that* in WSAfE.

Antecedent	Proportional preference for <i>that</i> (vs <i>who</i>) in unedited corpus (%)	Proportional preference for <i>that</i> (vs <i>who</i>) in edited corpus (%)	Percentage point change in preference for <i>that</i> (vs <i>who</i>)
<i>person</i>	4.3	1.8	-2.5
<i>children</i>	11.2	4.1	-7.2
<i>child</i>	7.8	6.1	-1.7
<i>parents</i>	4.8	0	-4.7
<i>adults</i>	22.2	22.2	0

tors, and only 68 cases by AfrE editors. Inspection of the data shows that only one case of *that* in an original text was edited by an AfrE editor, who retained it. WSAfE editors edited the remainder of the cases, changing *that* to *who* in 14 cases, and retaining *that* in 15 cases.¹⁰

4.4. People

In our dataset, *which* never occurs as relativiser with the word *people*. As shown in Table 16, the word *people* used as antecedent in RRCs is strongly associated with relativiser *who* in BSAfE and WSAfE original writing (96.8 and 96.6% of the time), and this strong association is endorsed (and slightly amplified) in editing. AfrE writers, however, opt for *that* with *people* comparatively more frequently – *who* is selected only 69.6% of the time. Once again, this comparative under-representation of *who* may be accounted for by influence from the L1, Afrikaans (see Section 4.1). Editors reduce the frequency of *that*, and increase the frequency of *who*, by 12.8 percentage points. The subset of cases produced by authors with a mixed linguistic background is small, but here, too, the comparatively weaker preference for *who* is strengthened in editing.

In the case of the AfrE dataset, where editing increases the lower proportional frequency of *who* by 12.8 percentage points, most of the editing is done by WSAfE editors (40 cases are edited by WSAfE editors, and 11 by AfrE editors). Cases where *that* occurs in the unedited subcorpus are mostly edited by WSAfE editors (all but one case). The WSAfE

¹⁰ In this subset of the data there are 2 cases of *that* with human antecedents in RRCs in the edited corpus that do not have corresponding text in the unedited corpus, added by editors with a WSAfE background; there are 3 cases where this construction occurs in the unedited corpus with no equivalent in the edited corpus.

Table 16

Frequency of *that* and *who* with the word *people* as antecedent in unedited and edited writing, across author varieties.

	Unedited			Edited			Percentage point change in preference for <i>who</i>
	<i>that</i>	<i>who</i>	% <i>who</i>	<i>that</i>	<i>who</i>	% <i>who</i>	
AfrE	14	32	69.6	9	42	82.4	+12.8
BSAfE	2	60	96.8	2	64	97.0	+0.2
WSAfE	5	141	96.6	4	155	97.5	+0.9
Mixed authors	1	5	83.3	0	8	100	+16.7
	22	238	91.5	15	269	94.7	+3.2

editors retain 9 cases of *that* with *people* (see Example (16)), and alter only 1, whereas the AfrE editor changes one case of *that* to *who*. Here, too, as in the case of RRCs with human antecedents, there is a tendency for WSAfE editors who encounter *that* with prototypically animate antecedents in AfrE writing to endorse this usage, rather than changing the relativiser to *who* (see Section 4.3.1).¹¹

(16) Even though diabetes is a chronic life-long condition that needs daily management, people **that** manage their blood glucose levels effectively, can live normal, active and high-quality lives. . . (SA_A-144-O/E)

What, then, is the cause of the adjustment to an increased proportional preference for *who* with *people* in edited AfrE writing? Clearly, it is not so much the case that (WSAfE) editors specifically target the use of *that* with *people*; this occurs rarely. Instead, in editing, more relative clauses with *who* are simply added by the editors, which accounts for the shift in proportions. Most of these additions are made by WSAfE editors (a total of 9 relative clauses with *people* as antecedent and *who* as relativiser), while some are also added by AfrE editors (4 relative clauses). The same applies to the proportional shift evident in the edited writing of groups of authors with mixed backgrounds. Where one case of *that* with *people* is replaced with *who*, the increase in proportional frequency of *who* is brought about by the addition of relative clauses with *who*. In such additions, editors are acting as writers, and their 'own' linguistic constructions therefore come into play. When writing, editors are inclined to use *who* as relativiser with *people* – even if they are otherwise inclined to leave *that* with *people* when editing.

4.5. Discussion

The analysis presented in Sections 4.1 to 4.4 provides information on how the relativisers pattern in terms of the different animacy categories we coded for, as well as the different subvarieties of English investigated. In broad strokes, in the unedited writing of the three subvarieties of SAfE, the distribution of relativisers patterns similarly to other varieties of English, reflecting the stability of the relativisation system in English (see discussion in Sections 2.1 and 2.2). We find that the scale of prototypicality for animacy plays an important role in conditioning the variation between *that* and *who*, with a stronger association of *who* with antecedents higher on the animacy scale (humans, *people* and collectives). However, in line with the findings of Gut and Coronel (2012) and Lee (2020) we find evidence for lower-level divergences in the subvarieties investigated, especially in the case of AfrE – and, in general, higher proportional frequencies of *that* and *which* with animate antecedents in RRCs in subject position are found than is mostly the case in previous studies that have focused on written language. In the first instance, this suggests that editing is a factor to be considered in studies of written language.

In AfrE, although there is an overall stronger preference for *who* with human and collective antecedents and the word *people*, AfrE authors use *that* in these contexts more frequently than the BSAfE and WSAfE authors do (as they also do for animal antecedents). There may be two reasons for this; one related to language-internal factors specific to AfrE users, the other to perceptions about normative orientation. On the one hand, there appears to be an influence of the L1, Afrikaans, on the English usage of these authors. As noted in the analysis, in Afrikaans, the only available relativiser for relative clauses is *wat* 'that', irrespective of the animacy of the antecedent, and this might account for the generally more frequent use of this relativiser among AfrE authors. The fact that relativiser *that* is the most frequent and most prototypical relativiser in all likelihood strengthens this tendency. A further effect of the L1 might be the weaker

¹¹ There are 5 cases of unmatched *that* in the unedited subcorpus, which means that this material was removed altogether in the edited version. There are no cases of unmatched *that* in the edited subcorpus; therefore no additions of *that* were made by the editors.

predictive power of animacy in AfrE, as argued by [Rosenbach \(2017\)](#) in relation to genitive alternation. This weaker effect of animacy might also be evident in relativiser variation. In other words, the sensitivity to animacy as a conditioning variable that is widely attested in relativiser choice has a less strong effect on AfrE writers' choices. This is least evident in the category of collectives, where the fuzzy nature of animacy distinctions is most strongly felt by all three groups of writers, and the choice of relativiser is most clearly defined.

[Rosenbach \(2017\)](#) and [Rosenbach and Kirsten \(2024\)](#) explicitly link the reduced sensitivity to animacy as a conditioning variable in the genitive alternation in Afrikaans, AfrE and WSAfE to cross-linguistic transfer effects, viewed both at the individual psycholinguistic level and at the varietal (and historical) level. However, how exactly such transfer effects occur at these various levels in populations with a high degree of stable bilingualism remains vastly under-researched, and is an important area of further research.

Normative orientation, and more specifically prescriptive advice, might also play an important role conditioning the AfrE users' preferences. As noted in [Section 2.3](#), the prescriptive advice in the BrE sources allows for alternation between *that* and *who* with human antecedents when the antecedent refers to a generic group of humans or when it is human but refers to a class. A similar distinction is not made in SAfE usage guides, which enforce the division of labour between *who* (for humans) and *that* (for animals and things) more unequivocally and consistently. Since the AfrE users are found to use *that* at higher frequencies with generic groups of humans (for example, *people, learners, person, students, participants, individuals, respondents, roleplayers, researchers, consumers, employee and officials*), it may be that the writing of these authors is influenced by the prescriptive recommendations of sources oriented towards the external, BrE norm, possibly also as captured in English L2 school curricula.

With regard to the impact of editorial intervention on the relativiser patterns in the South African data, the most striking finding is that where a variety tends towards less prototypical usage, the interventions of editors might mask this by amending usage more towards prototypical usage – in most cases, replacing *that* with *who*. However, although there is evidence that editing shifts patterns of relativiser usage, these shifts are not always consistent, and different trends are evident for the different subvarieties. The most striking editorial changes happen in the writing of AfrE authors for antecedents higher up on the animacy scale, humans and *people*, which are two of the categories where AfrE authors' preferences diverge from prototypical usage. In the case of human antecedents, AfrE editors are the ones mostly responsible for the replacement of *that* with the more prototypical *who* (which may reflect a hypercorrection, based on AfrE editors' awareness of this particular usage in AfrE writing), while WSAfE editors leave the greater portion of the less prototypical usages intact. However, for BSAfE writing, the two groups of editors do the opposite: WSAfE editors replace the less prototypical *that* with *who*, while AfrE authors never change these instances of *that* to *who*. This finding suggests that the tendency of editors to intervene in a text depends (also) on the relation between the language background of the author and editor.

Individual variability in editorial practice clearly plays a role, and, together with the uneven representation of different registers across the different varieties in the corpus, complicates the identification of broader trends, for which an even larger dataset would be needed. It should, however, be noted that in the data analysed, there is a very strong representation of WSAfE editors, and most changes appear to be due to their interventions to AfrE or WSAfE, rather than BSAfE. This finding is in line with previous findings in [Kotze \(2019\)](#) and [Law and Kotze \(2021\)](#): editorial intervention in the South African context tends to be focused on WSAfE and AfrE, with BSAfE comparatively untouched (by the predominantly WSAfE and AfrE editors). This may suggest an endorsement of the endonormativity of BSAfE, but not WSAfE (and AfrE), as argued in [Kotze \(2019\)](#).

However, a more careful and more fine-grained consideration of the sociolinguistic profiles of the editors raise other possibilities. As already discussed in [Section 1](#), South African editors tend to be bi- or multilingual, with the languages in their profiles mostly English and Afrikaans, combined with a few other European languages. [Law \(2019: 142\)](#) finds that 41% of the L1 Afrikaans editors view Afrikaans as their strongest language, with 45% of the L1 English editors likewise see English as their strongest language. However, 10% of the L1 Afrikaans editors view English as their strongest language. These findings point to the prevalence of Afrikaans–English bilingualism among South African editors, and particularly L1 Afrikaans editors, which echoes the more general findings of [Coetzee-Van Rooy \(2013\)](#), whose research reveals large-scale stable Afrikaans–English bilingualism and bi-literacy.

This raises the question: What are the consequences of bilingualism and biliteracy for editorial practice? On the one hand, it seems self-evident that, psycholinguistically speaking, bilingualism must open the door to greater tolerance for (or less sensitivity to) linguistic variation associated with language-contact phenomena, or differences in usage preferences across different languages. Afrikaans–English bilingual editors are likely to be susceptible to cross-linguistic influences of the kind outlined by [Rosenbach \(2017\)](#), [Rosenbach and Kirsten \(2024\)](#), and [Van Rooy \(2020\)](#). On the other hand, bilingual editors may precisely be more consciously aware of particular pitfalls associated with L2 writing (being L2 users themselves, at times), and may take a hypercorrective approach, as a consequence. In addition, where changes are made, it is not only targeted changes of phenomena, but also where editors rewrite, and particularly where

they add material, where their own linguistic preferences come into play. Edited texts are therefore at least to some degree ‘hybrid’ in their linguistic preferences, consisting of the linguistic usage patterns of at least two users – both of whom may, in this instance, be influenced to some degree by the sociocognitive interplay of different languages. However, while the sociolinguistic and historical aspects of Afrikaans–English bilingualism has been fairly well described (see, e.g., Coetzee-Van Rooy, 2013; Van Rooy, 2020), there is far less research on the actual language use and processing of this group of bilinguals, and none on editors, specifically. Our study suggests that editorial intervention, in the South African context, is shaped by the bilingualism of writers and editors – but more work needs to be done to determine how, exactly. In existing studies, the editorial dynamic appears to occur mostly in the interaction between white authors and editors, with a generally less interventionist approach taken to the writing of black authors (by white editors). It may be that (at a subconscious level) the shared Afrikaans–English bilingualism licences editorial intervention – there is a perception of shared linguistic ground which warrants or authorises normative tweaking, which tends to occur in the direction of normative prototypicality. However, this tentative proposal requires more research to explore.

Lastly, there is the question of how usage guides and other normative sources ‘feed into’ the linguistic representations of South African editors (and authors). The overall increase in frequency of *who* at the cost of *that* during editing suggests that there is an alignment with the South African usage guides, which are most unequivocal in insisting on *who* as relativiser for animate antecedents (see Steynberg, 2018; Linnegar, 2009). However, once again there is evidence of lower-level variation: the fact that WSAfE editors are sometimes more inclined to retain *that* (e.g., in the case of human antecedents and *people* as an antecedent, in AfrE writing) suggests a potential orientation of this group to BrE or AmE usage norms.

5. CONCLUSION

The study reported on in this article demonstrates that editorial intervention is an important consideration in investigations of written language and that the linguistic composition of published written texts cannot be unambiguously attributed to the work of authors alone. Specifically, our findings, alongside those of previous studies that investigate whether editing matters in studies of language variation, demonstrate that editorial practices are variable, and that endorsement (or otherwise) of original writing depends on the complex interaction between the specific linguistic feature being targeted, the available normative advice, and the interaction of the author’s and editor’s sociolinguistic profiles. This latter factor emerges as potentially important in the context of South African Englishes, where editorial interventions appear to be shaped by the bilingualism of writers and editors, and the shared linguistic background (or not) among them.

Our findings should, however, be interpreted with caution: our study is based on only one feature, and further work would need to be done on a broader range of features. Furthermore, the limited data and unbalanced corpus used in this study do not allow for analyses across registers, and future work should include a more balanced corpus.

Although they are underrepresented in the industry, the interventions of BSAfE editors also play an important role in shaping the language of texts in South Africa. Future studies should include texts edited by these editors to determine whether the patterns of intervention identified for AfrE and WSAfE editors, are true for BSAfE editors too. Including more texts edited by BSAfE editors would also allow for a more comprehensive investigation of the ways in which the bilingualism of writers and editors shapes editorial interventions.

DECLARATION OF INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Haidee Kotze: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Melanie Ann Law Favo:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization.

Data availability

Data will be made available on request.

References

- Aecer, L., 2019. Relative pronouns and prepositions. *Pegboard*. 26 (3), 6–8.
- Aecer, L., 2022. A potpourri of common errors committed by EFL and ESL academic writers (part 2). *Pegboard*. 29 (3), 2–4.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., 1999. *Longman Grammar of Spoken and Written English*. Pearson, Harlow.
- Blight, N., 2014. Damsels **who** or **that**. *Pegboard*. 21 (2), 17–19.
- Botha, W., Van Rooy, B., Coetzee-Van Rooy, S., 2021. Researching South African Englishes. *World Englishes*. 40 (1), 2–11.
- Brook, M., Tagliamonte, S.A., 2023. Subject relative *who* in Ontario, Canada: Change from above in a transplanted ecology. *Journal of Linguistic Geography*. 11, 25–37.
- Burchfield, R.W. (Ed.), 1996/1998. *The New Fowler's Modern English Usage*, revised third ed. Oxford University Press, London.
- Butterfield, J. (Ed.), 2015. *Fowler's Dictionary of Modern English Usage*. 4th ed. Oxford University Press, Oxford.
- Coetzee-Van Rooy, S., 2013. Afrikaans in contact with English: Endangered language or case of exceptional bilingualism?. *International Journal of the Sociology of Language*. 224 179–207.
- Collins, P., Yao, X., Borlongan, A., 2014. Relative clauses in Philippine English: A diachronic perspective. *Language and Computers*. 78 (1), 125–146.
- Curzan, A., 2014. *Fixing English: Prescriptivism and Language History*. Cambridge University Press, Cambridge.
- D'Arcy, A., Tagliamonte, S.A., 2010. Prestige, accommodation and the legacy of relative *who*. *Language in Society*. 39, 389–410.
- D'Arcy, A., Tagliamonte, S.A., 2015. Not always variable: Probing the vernacular grammar. *Language Variation and Change*. 27, 255–285.
- Fowler, H.W., 1926/1994. *A Dictionary of Modern English Usage*. Wordsworth Editions, Ware.
- Fowler, H.W., 1965/1968. *A Dictionary of Modern English Usage*, second ed. Revised by Ernest Gowers. Oxford University Press, London.
- Garner, B.A., 1998. *A Dictionary of Modern American Usage*. Oxford University Press, New York.
- Garner, B.A., 2003. *Garner's Modern American Usage*. Oxford University Press, New York.
- Garner, B.A., 2009. *Garner's Modern American Usage*. Oxford University Press, New York.
- Garner, B.A., 2016. *Garner's Modern English Usage*. Oxford University Press, New York.
- GCSIS, 2013. Editorial Style Guide. https://www.gcis.gov.za/sites/default/files/docs/resourcecentre/guidelines/Editorial_Style_Guide.pdf.
- Grafmiller, J., Szmrecsanyi, B., Hinrichs, L., 2018. Restricting the restrictive relativizer: Constraints on subject and non-subject English relative clauses. *Corpus Linguistics and Linguistic Theory*. 14 (2), 309–355.
- Gut, U., Coronel, L., 2012. Relatives worldwide. In: Hundt, M., Gut, U. (Eds.), *Mapping Unity and Diversity World-Wide: Corpus-Based Studies of New Englishes*. John Benjamins, Amsterdam, pp. 216–241.
- Heller, B., Szmrecsanyi, B., Grafmiller, J., 2017. Stability and fluidity in syntactic variation world-wide: The genitive alternation across varieties of English. *Journal of English Linguistics*. 45 (1), 3–27.
- Hinrichs, L., Szmrecsanyi, B., 2007. Recent changes in the function and frequency of Standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language and Linguistics*. 11 (3), 437–474.
- Hinrichs, L., Szmrecsanyi, B., Bohmann, A., 2015. *Which*-hunting and the standard English relative clause. *Language*. 91 (4), 806–836.
- Huddleston, R., Pullum, G., 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press.
- Jankowski, B.L., 2013. *A Variationist Approach to Cross-register Language Variation and Change*. University of Toronto, Doctoral dissertation.
- Kapp, S., 2018. Die hoe en hoekom van “wie” en “wat”. <https://viva-afrikaans.org/lees-luister/blog/item/381-die-hoe-en-hoekom-van-wie-en-wat>.
- Kirsten, J., Breed, A., 2020. Relative pronouns. <https://taalportaal.org/taalportaal/topic/pid/topic-20190924184807143>.
- Kotze, H., 2012. Postcolonial Polysystems: The Production and Reception of Translated Children's Literature in South Africa. John Benjamins, Amsterdam.
- Kotze, H., 2019. Does editing matter? Editorial work, endonormativity and convergence in written Englishes in South Africa. In: Hickey, R. (Ed.), *English in Multilingual South Africa*. Cambridge University Press, pp. 101–126.
- Kruger, H., Van Rooy, B., 2017. Editorial practice and the progressive in Black South African English. *World Englishes*. 36, 20–41.
- Labat, S., Kotze, H., Szmrecsanyi, B., 2023. Processing and prescriptivism as constraints on language variation and change: Relative clauses in British and Australian English parliamentary debates. In: Korhonen, M., Kotze, H., Tyrkkö, J. (Eds.), *Exploring Language and Society with Big Data: Parliamentary Discourse across Time and Space*. John Benjamins, Amsterdam, pp. 250–276.
- Labat, S., 2018. *Relative Markers in Parliamentary Speech: A Real-Time Study of Australian and British English*. Master thesis. KU Leuven.
- Labov, W., 2007. Transmission and diffusion. *Language*. 83 (2), 344–387.
- Law, M.A., 2019. *The Role of Editorial Intervention in Ongoing Language Variation and Change in South African and Australian English*. Macquarie University/North-West University, Doctoral diss.
- Law Favo, M.A., 2024. Editors and world Englishes: Exploring (the interaction between) editors' sociolinguistic profiles and the norm-providing sources they use for editing work. *English World-Wide*. 45 (1), 61–93.

- Law, M.A., Kotze, H., 2021. Gender, writing and editing in South African Englishes: Genitive alternation in multifactorial perspective. In: Bernaisch, T. (Ed.), *Gender in World Englishes*. Cambridge University Press, pp. 205–232.
- Lee, K.M., 2020. Relative clauses in a modern diachronic corpus of Singapore English. *Asia Pacific Journal of Corpus Research*. 1 (3), 31–60.
- Linnegar, J., 2009. *English, Our English: Common Errors in South African English and How to Resolve Them*. Pharos, Cape Town.
- Mesthrie, R., 2010. *A Dictionary of South African Indian English*. UCT Press, Cape Town.
- Milroy, J., Milroy, L., 2012. *Authority in Language: Investigating Standard English*. Routledge, Abingdon.
- Mkhize, D., Balfour, R., 2017. Language rights in education in South Africa. *South African Journal of Higher Education*. 31 (6), 133–150.
- Möller, J., 2013. The state of multilingual publishing in South Africa. *e-Rea*. 11 (1) <http://journals.openedition.org/erea/3507>.
- Peters, P., 2014. Usage guides and trends in Australian and British English. *Australian Journal of Linguistics*. 34 (4), 581–598.
- Ponelis, F.A., 1979. *Afrikaanse sintaksis*. JL van Schaik, Pretoria.
- Posel, D., Zeller, J., 2016. Language shift or increased bilingualism in South Africa: evidence from census data. *Journal of Multilingual and Multicultural Development*. 37 (4), 357–370.
- Posel, D., Zeller, J., 2020. Language use and language shift in post-apartheid South Africa. In: Hickey, R. (Ed.), *English in Multilingual South Africa: the Linguistics of Contact and Change*. Cambridge University Press, pp. 288–309.
- Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., 1985. *A Comprehensive Grammar of the English Language*. Longman, London.
- Romaine, S., 1980. The relative clause marker in Scots English: Diffusion, complexity, and style as dimensions of syntactic change. *Language in Society*. 9 (2), 221–247.
- Rosenbach, A., 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.
- Rosenbach, A., Kirsten, J., 2024. Afrikaans influence on genitive variation in South African English? A comparative diachronic study of Afrikaans and White South African English. In: Van Rooy, B., Kotze, H. (Eds.), *Constraints on Language Variation and Change in Complex Multilingual Contact Settings*. John Benjamins, Amsterdam, pp. 29–57.
- SATI, 2023. Find a language practitioner. <https://www.translators.org.za/members-2/> Accessed 28 December 2023.
- Sigley, R., 1997. The influence of formality and channel on relative pronoun choice in New Zealand English. *English Language and Linguistics*. 1 (2), 207–232.
- Steynberg, M., 2018. *The PEG Guide to Grammar and Punctuation*. Professional Editors' Guild, Johannesburg.
- Strunk, W., White, E.B., 1959. *The Elements of Style*. Macmillan, New York.
- Tagliamonte, S., Smith, J., Lawrence, H., 2005. No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change*. 17 (1), 75–112.
- Van Rooy, B., 2014. English in South Africa. In: Filppula, M., Klemola, J., Sharma, D. (Eds.), *The Oxford Handbook of World Englishes*. Oxford University Press, pp. 508–530.
- Van Rooy, B., 2020. Present-day Afrikaans in contact with English. In: Hickey, R. (Ed.), *English in Multilingual South Africa: the Linguistics of Contact and Change*. Cambridge University Press, pp. 241–264.
- Van Rooy, B., Kruger, H., 2016. The innovative progressive aspect of Black South African English: The role of language proficiency and normative processes. *International Journal of Learner Corpus Research*. 2 (2), 205–228.
- Zaenen, A., Carlette, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., O'Connor, M.C., Wasow, T., 2004. Animacy encoding in English: Why and how. In: Byron, D., Webber, B. (Eds.), *Proceedings of the 2004 ACL Workshop on Discourse Annotation, Barcelona, July 2004*. Association for Computational Linguistics, Stroudsburg, pp. 118–125.