A multifactorial analysis of contact-induced change in speech reporting in written White South African English (WSAfE)

HAIDEE KRUGER
Macquarie University / North-West University

and

BERTUS VAN ROOY
North-West University
(Received 21 June 2017; revised 15 December 2018)

This article presents a corpus analysis of changes over a period of two centuries in speech-reporting constructions in written White South African English (WSAfE), a native variety of English that has been in contact with Afrikaans throughout its history. The analysis is based on register-differentiated comparable diachronic corpora of WSAfE, its parent variety, British English (BrE), and the contact language, Afrikaans. Three related reported-speech constructions are analysed, focusing on changes in the relative frequencies of variants of each construction. These constructions show ongoing change, with similar trajectories of change for WSAfE and BrE in some cases, but divergent trajectories in others. In the latter case, WSAfE and Afrikaans converge on similar frequency distributions, which follow from an accelerated rate of change or a slowing down of the rate of change for particular features in WSAfE in comparison to BrE. Descriptive findings are supported by conditional inference tree modelling. The effect of frequency on reinforcing similar patterns of change in WSAfE and Afrikaans, as well as simplification through the levelling of register differences in WSAfE and Afrikaans are proposed as explanations. The study highlights the importance of converging norms in a multilingual publication industry as a site of contact.

1 Introduction

That English is a contact language is widely acknowledged (e.g. Hundt & Schreier 2013). Through its history in England and the larger British Isles, speakers of Celtic and Scandinavian languages as well as French shifted to English, and in the course of doing so contributed to language change (Fischer 2013). The effect on English vocabulary is widely recognised, but the grammatical effects have only been acknowledged in more recent reappraisals of the evidence (Hundt & Schreier 2013: 3–4). Since the onset of the colonial diffusion of English, two specific kinds of contact-induced change have attracted attention: koineisation or convergence in contact between transplanted[1] dialects of English, for instance in New Zealand

---

[1] 'Colonial' varieties are known by a number of different names: Trudgill (2004) uses the term 'colonial Englishes'; Schneider (2007) denotes them as 'settler (STL) strands'; and Kachru (1985) calls them 'Inner

(Trudgill 2004), and transfer of first-language (L1) features to English as a second language (L2). The latter scenario includes English as used in the Outer Circle (Kachru 1985), or language-shift Englishes such as Irish English (Hickey 2013) and Indian South African English (Mesthrie 1992).

Notably absent from most discussions is the possibility of language change in modern and contemporary L1-varieties of English through contact with other languages in the same multilingual environment. In this article, we investigate a situation of sustained contact between two languages in a multilingual setting, to analyse how a less dominant language, Afrikaans, may have affected English over the course of two centuries in the South African context. Our focus is on the native variety of English (termed White South African English, or WSAfE), transplanted to South Africa with the arrival of the British settlers in 1820.

The particular nature of settings of language contact is important. Thomason & Kaufman (1988: 35) postulate as point of departure that the sociolinguistic history of speakers, rather than the structure of the languages, is the primary determinant of the outcome of language contact. Matras (2009: 47) points out that '[a]symmetry in the social roles of languages and in the directionality of bilingualism is a crucial factor in determining the impact that contact is likely to have on the structures of the relevant languages'. Generally, the assumption is that languages of high status, used across a wide range of domains by mainly monolingual speakers, are less likely to be affected by crosslinguistic influence (CLI) in contact settings than languages of lower status used by highly bilingual speakers in intensive contact with high-status languages (Thomason 2001: 66; Matras 2009: 47; Fischer 2013: 19–20). Under these conditions, it seems unlikely that Afrikaans would have considerable influence on WSAfE, given that speakers of WSAfE are not on average more bilingual than Afrikaans speakers, and English has enjoyed a higher status than Afrikaans for most of their shared history (Van Rooy 2017: 513–15).

The extent to which Afrikaans has influenced the grammatical features of English has, indeed, been a matter of debate. Lass & Wright (1986) argue that the case for Afrikaans influence is overstated for a number of features where they identify endogenous sources. Mesthrie (2017: 535-9) reviews previous work and concludes that a selection of informal features of Afrikaans (associative plurals, reduplication, diminutives and the invariant tag *nè*) made their way to English, but that not all cases of parallelism between the two languages have necessarily been shown to be contact-induced. Wasserman & Van Rooy (2014) argue that semantic and pragmatic uses of modal constructions have been transferred to WSAfE from Afrikaans, alongside increased frequency of use of some modals.

Studies of language change, including changes that are influenced by language contact, are underpinned by the assumption that face-to-face interaction is a required mediating factor for language change to spread (Sayers 2014). Written sources are typically analysed for the evidence they yield about the spoken language of earlier

Circle Englishes'. Relevant to the present purposes is that they are native varieties, spoken outside the ancestral home of the United Kingdom.

times (e.g. Mesthrie 1996). While we do not deny the importance of spoken inter-action, in this article we argue that it is crucial also to consider written communication as a possible site of language change (see also Biber & Gray 2011). In the usage-based paradigm (especially Bybee 2010) which we adopt in this article, written input, just like spoken input, serves to cognitively entrench grammatical constructions – which, in turn, raises the accessibility of constructions for users, who will then be more inclined to use these constructions in both spoken and written settings, where appro-priate, setting in motion a frequency-entrenchment cycle that cuts across written and spoken language. In highly literate communities, exposure to written texts (particularly of widely disseminated genres like newspapers and magazines, or, more recently, online communication) forms an important component of the language input that users receive alongside spoken interaction, as argued by Kruger & Van Rooy (2017).

In other words, in contexts where contact may introduce grammatical changes (in whichever dimension), such changes may be taken up and disseminated in written as much as in spoken language, setting in motion the dynamics of change. A further dimension to the process of dissemination of changes in writing, as Kruger & Van Rooy (2016) argue, can be identified in the case of edited published writing, where favourable conditions for language contact may exist within the publishing landscape of a society. This, certainly, has been the case in South Africa, where the local English and Afrikaans publishing industries have developed alongside each other for approximately two centuries – often from within the same publishing houses (Kantey 1990), which opens up the opportunity for shared text-production norms to emerge. However, there is scant historical research on publishing and editorial practices in South African news, magazine and book publishing.

The dynamics of contact in the written medium of course also need to be considered from the broader perspective of the changing language ecology of South Africa over the course of the two centuries in which English and Dutch/Afrikaans have co-existed. A complete survey of the historical development of English and Afrikaans in South Africa is not possible within the scope of this article (see Van Rooy 2017: 509–11 for the most recent survey). For the purposes of the argument here, it will suffice to point out that clear shifts in the nature of the contact can be identified, accompanied by shifts in power relationships and attitudes of animosity and reconciliation.

A Dutch colony at the Cape was established in 1652, and under the influence of various factors, the settlers' Dutch gradually diverged from metropolitan or continental Dutch. By the early nineteenth century, a local Cape Dutch vernacular was recognised as distinct from metropolitan Dutch. Afrikaans gained recognition as a language separate from Dutch in the late nineteenth and early twentieth century. The British captured the Cape Colony in 1795, and in 1814 the Cape became a British territory. As a result, the Cape Dutch settlers had to yield to the British in all domains of public life. In 1820 the first sizeable group of British settlers arrived at the Cape, and in 1822 English was declared the only official language of the Cape, accompanied by aggressive Anglicisation policies (Van Rooy 2017: 509–10). In 1910, the Union of South Africa was formed out of the various separately governed territories under British rule, adopting a bilingual policy with Dutch and English as official languages,

with the former replaced by Afrikaans in 1925. Afrikaans and English remained the official languages of South Africa (through the establishment of the Republic of South Africa in 1961) until 1994, when eleven languages (including Afrikaans, English and nine African languages) were recognised as official languages.

The linguistic features we investigate in this contact setting are three syntactic constructions associated with the representation of direct and indirect speech and thought: the order of the reporting and reported clause in direct speech reporting; inversion of subject and verb in reporting clauses that are not in sentence-initial position; and the omission of the complementiser *that* in indirect speech and thought representation. These are more formal syntactic features, and perform linguistic functions that are more commonly expressed in writing than in speech, which have particular partially similar and partially distinct forms as well as pragmatic associations and functions in Afrikaans and non-contact varieties of English. In a previous study of these features, Kruger & Van Rooy (2016) found that, synchronically, WSAfE and BrE differ from each other in respect of their preferences for variants of these constructions. In this article, we investigate these constructions in comparable diachronic corpora of WSAfE and BrE, with reference also to a diachronic corpus of Afrikaans, to determine whether CLI (of different kinds) from Afrikaans may account for divergences between WSAfE and its British parent variety. Before turning to these constructions, we first briefly outline our view of how contact potentially influences language change. Subsequent to this, we briefly discuss the diachronic corpora used in the study, and the quantitative analysis methods. This is followed by the presentation and discussion of findings, and a summary and concluding perspective on the results of the study.

2   Crosslinguistic influence

If a linguistic change occurs under the influence of a grammatical construction from another language, then the change is called contact-induced (Thomason & Kaufman 1988; Thomason 2001). This process of change has been referred to in various terms, including borrowing, interference, code-copying, and replication (see Weinreich 1979; Thomason & Kaufman 1988; Thomason 2001; Johanson 2002; Heine & Kuteva 2005), but in this article, we refer to the general process as CLI.[2] Our approach to CLI is functional in that it extends to change in use and not only change in form, and aligned with the usage-based view of grammar developed by Bybee (2010), in terms of which a change in the frequency of a construction or of its context of use is an important part of language change.

---

[2] We prefer the term 'crosslinguistic influence (CLI)' to 'contact-induced change'. The latter term evokes the sense of a change initiated and driven by contact, which is a valid account of some phenomena discussed in the literature, but we extend the concept to changes that occur under the influence of contact, and where contact is an enabler or mediator of the change, without necessarily being the driving force. This extended view is still consistent with the definition of 'contact-induced language change' proposed by Thomason (2001: 62), which specifically notes that 'any linguistic change that would have been less likely to occur outside a particular contact situation is due at least in part to language contact'.

Thomason & Kaufman ([1988](): 36-9) propose a sharp distinction between borrowing, which typically involves specific lexical elements, and interference, which typically affects grammatical patterns. The latter is characterised by imperfect acquisition of a second language, a factor that is absent from the first manifestation of contact-induced change.[3] They highlight the importance of social factors such as the intensity of contact and the asymmetry in the relation between the speakers in the contact setting. They also show that borrowing affects lexical items first, and only after much more intense contact are phonological or morphosyntactic forms borrowed. Where a non-dominant population approximates the target language of a dominant speech community, phonological and syntactic patterns of their native language may be transferred in the first instance, to the degree that they do not converge on the target language as spoken by the dominant population.

Matras ([2009]()) distinguishes between the replication of linguistic matter and linguistic patterns, where the former is identifiable as discrete morphs, and the latter are more abstract structures for the organisation of linguistic matter, or grammatical constructions (Matras [2009](): 234). Heine & Kuteva ([2005]()) hold a similar view. Their point of departure is that the early stages of grammatical change through contact take place through the replication of use patterns drawn from the other language known by the speakers. While it is possible for a new use pattern to be introduced to a language through contact, Heine & Kuteva ([2005](): 45–50) argue that one possibility that is sometimes overlooked is where CLI enables a 'minor use pattern' in the replica (or target) language to develop into a 'major use pattern' under the influence of contact with a model (or source) language where a similar pattern exists. In other words, there is an increase in the frequency of an existing use pattern (or construction) under the influence of the frequency pull of a pattern in another language in the contact situation that is perceived to be similar. Such a change is therefore a quantitative development only, and corresponds to what Mougeon *et al.* ([2005](): 102–3) refer to as covert transfer: contact causes an increase in the frequency of a construction in a target language that shows similarity with one in the source language. Exemplars of a construction from the source language are categorised with exemplars of a similar construction in the target language (based on similarities in form and function) – compounding the entrenchment of the construction and leading to an increase in frequency of use. These frequency changes may at later stages lead to an extension to new contexts and the development of new functions (Heine & Kuteva [2005](): 45).

Overt transfer, in contrast, is a 'qualitative development' (Mougeon *et al.* [2005](): 102) in the target language, introducing a construction into the target language that did not exist prior to contact. Such a language change presupposes that the languages in question differ in some respect in the inventories of constructions available to users. A

---

[3] Thomason & Kaufman ([1988](): 39) explicitly offer the disclaimer that 'imperfect learning' does not pertain to lack of the ability to learn, and is only judged from the perspective of the pre-existing system of the target language. A number of terms used in the literature, often similar to terms used in the study of second language acquisition, may be contested in terms of the implicit view of speakers from already marginalised communities. We will therefore, without disrupting the flow of the argumentation unnecessarily, seek to use terms that are not obviously judgemental as far as possible.

construction from the source language must occur with such frequency of use that it becomes so cognitively salient that it transfers across the two languages in the form of an innovative structural pattern or innovative functional use in the target language.

There are a number of reasons why a change emanating from CLI may disseminate further. In the case of second-language speakers who transfer features to a target language, the degree of bilingualism and intensity of contact act as constraining forces, while limitations of access to the target language and attitudes towards the target language from the side of the shifting population may facilitate greater divergence from the pre-existing form of the target language (Thomason & Kaufman 1988: 43-8). When the speech of the target-language speech community itself is considered, the default option is the borrowing of elements from the speakers of the other language in the contact situation, which are usually lexical features first, with grammatical features only following after very intense and prolonged periods of contact (Thomason & Kaufman 1988: 41–3).

The grammatical constructions under investigation in this article are, however, features of the dominant language that appear to be influenced by the less dominant language in the contact situation, which raises questions about the kinds of mechanisms that may promote CLI in this case. Thomason (2001: 142) identifies the possibility of implicit 'negotiation' between speakers who aim to reproduce a pattern from another language. In more general terms, this may be framed as accommodation between speakers in a face-to-face contact situation, as proposed by Trudgill (2004) for dialect contact.

Alongside such social factors, two possible psycholinguistic factors may also play a role. The first of these is frequency of exposure, where even (extensive) passive exposure to a form may eventually lead to a speaker adopting it (Thomason 2001: 140). This ties in more generally with Bybee's (2010) view of the importance of frequency of exposure as part of the emergence of grammatical constructions beyond the context of language contact: the more a particular construction is encountered, the more it is reinforced in the mind of an individual user, which in turn leads to higher frequencies of use for that speaker. The likelihood of frequency entrenchment in contact situations is increased by the higher intensity of such contact.

Matras (2009: 151) identifies another psycholinguistic factor. He notes that the cognitive effort of keeping apart structures or preferences for use of constructions across two languages may be eased by simplifying the choices to be made. Bilingual individuals may opt to treat elements from different languages they perceive to be similar in the same way (e.g. use articles in the same way) even if monolingual speakers of the two languages use them in different ways. This could, in theory, extend to stylistic homogenisation of registers across two languages in contact, where similar constructions are selected in similar frequencies in the corresponding registers of two languages in contact as a way of simplifying different crosslinguistic register preferences.

In this article, we investigate the possibility that CLI effects from Afrikaans may account for some of the divergences of WSAfE from its parent variety, British English (BrE). We investigate three grammatical constructions associated with the representation of direct and indirect speech and thought in written texts, which demonstrate differential frequency distributions among variants in contemporary synchronic corpora of written BrE and WSAfE (Kruger & Van Rooy 2016). This analysis investigates the proposal made by Kruger & Van

Rooy (2016) that these frequency differences may be the result of prolonged contact with Afrikaans. It proceeds from the assumption that in considering contact explanations for change, it is necessary to consider the historical trajectories of change of all three language varieties in question: the model language (Afrikaans) and two variants of the target language: the one in contact with the model language (WSAfE) and one that is not (BrE).

3 Three reported-speech constructions in Afrikaans and non-contact varieties of English

Our selection of reported-speech constructions as an area of investigation is motivated not only by the synchronic differences observed by Kruger & Van Rooy (2016), but also by the fact that speech and thought reporting is an area in which not only lexical but also structural aspects have been 'fundamentally reorganized' over the course of the past 150 years (D'Arcy 2015: 56), which creates an opportunity to observe whether CLI plays a role in a change that is ongoing in any case. A further motivation is the fact that speech and thought reporting have very specific functions across different written registers, and therefore offer opportunities to investigate the effects of language contact in different registers – which may, in themselves, be more or less receptive to change (Hundt & Mair 1999; Biber & Gray 2013).

3.1 The position of the reporting clause in direct speech and thought reporting

In both English and Afrikaans, reporting clauses are syntagmatically variable. The reporting clause can occur before the reported clause, be inserted 'within' the reported clause, or occur after the reported clause, as illustrated in (1) to (3) (where the reporting clauses are underlined).

(1) <u>She then said</u>, 'I am going to set the house on fire tonight, at seven o'clock.' (WSAfE, Reportage, 19th-1)[4]

(2) 'Perhaps,' <u>I said</u>, 'she was anxious to be married, and he was her only suitor' (BrE, Fiction, 20th-2)

(3) 'The streets presented a very "sulphury" appearance,' <u>said our informant</u>. (WSAfE, Reportage, 19th-2)

Although these positional variants are formally identical in English and Afrikaans, they demonstrate distinct register preferences in the two languages, particularly in reportage and fiction. According to Biber *et al.* (1999: 923), the final position is preferred in British and American English news and fiction, but the initial choice is the preferred option in fiction. In Afrikaans, data from the *Taalkommissiekorpus*[5] reveal a consistent preference of slightly more than two-thirds for the final position in both reportage (68%) and fiction (70%).

---

[4] Examples are from the corpus compiled for this study (see section 4.1), with variety, register and period indicated. Periods are distinguished as the first and second half of each century (19th-1, 19th-2, 20th-1, 20th-2).

[5] The *Taalkommissiekorpus* [Corpus of the Language Commission] (Taalkommissie 2011) is a corpus of 57 million words containing published written Afrikaans from a wide range of registers. Data from those registers

Kruger & Van Rooy (2016) found that the same pattern reported by Biber *et al.* (1999) also holds for ICE-GB, but that ICE-SA, representing WSAfE, is different. That difference makes WSAfE very similar to Afrikaans: the final position for the reporting clause is the dominant pattern in reportage and fiction.

3.2   *Quotative inversion in reporting clauses in non-initial position in direct and indirect speech and thought*

Quotative inversion is one of the few vestiges of historical verb-second (V2) word order in English syntax (D'Arcy 2015: 47). In English, quotative inversion occurs freely in medial or final reporting clauses containing a noun phrase subject and a simple verb (Biber *et al.* 1999: 921), in both direct and indirect speech and thought. In the inverted form, the reporting clause has VS order, as in (4a) and (5a), with the corresponding non-inverted form with SV order in (4b) and (5b).

(4) (a) 'We feel a bit hard done by to have lost,' said Clark. (BrE, Reportage, 20th-2)
     (b) 'I've wanted you for a long time,' counsel added. (BrE, Reportage, 20th-2)
(5) (a) But these new mystery signals, said Mr. Garratt, seemed to be coming from nowhere… (WSAfE, Reportage, 20th-2)
     (b) High real interest rates, Manuel said, were harmful to economic growth and job creation… (WSAfE, Reportage, 20th-2)

D'Arcy (2015: 47) points out that these kinds of reporting structures are strongly associated with written rather than spoken language, and that inversion, in particular, 'is maintained in more formal, reflective usage'.

Biber *et al.* (1999: 922) note that where the subject is an unstressed pronoun, the non-inverted order is 'virtually the rule', except where used for stylistic effects. The questionable current acceptability of inversion with a pronoun-subject reporting clause is illustrated in (6b) (which was, however, acceptable in the nineteenth century).

(6) (a) 'We are not seeing any reduction in the amount of downgrading activity,' he said. (BrE, Reportage, 20th-2)
     (b) ?'Stay a minute', said he, as she was on the point of departure. (BrE, Fiction, 19th-2)

In non-initial reporting clauses in Afrikaans, however, no alternation is possible, and the order is obligatorily inverted VS, as in (7a). The SV construction in (7b) is therefore ungrammatical.

(7) (a) 'Ek is seker die vrou is nie van hier nie,' sê Ana-Linda. (TK)
         '"I am sure the woman is not from here," says Ana-Linda.'

corresponding to the ones analysed for English in this article were used for comparative purposes. Afrikaans examples taken from *Taalkommissiekorpus* are referenced TK in the text.

(b) *'Ek is seker die vrou is nie van hier nie,' <u>Ana-Linda sê.</u>
'"I am sure the woman is not from here," <u>Ana-Linda says.</u>'

Kruger & Van Rooy (2016) identify these formal asymmetries as presenting opportunities for both overt and covert transfer in contemporary usage. Overt CLI would occur if the less acceptable or archaic verb-pronoun structure occurred more frequently in contemporary WSAfE than in BrE, under influence of the obligatory inverted order in Afrikaans. Covert CLI would be evident in the higher frequency of the inverted VS order with noun subjects, and corresponding lower frequency of SV order in WSAfE. They find no evidence of overt CLI, but do find a stronger preference for the VS order in WSAfE compared to BrE with non-pronoun subjects. Given D'Arcy's (2012, 2015) comments on changes in the system of English speech and thought reporting, this raises questions about the extent to which, during an ongoing language change in English, there is evidence of a CLI effect from Afrikaans that slows down the change in WSAfE.

3.3 *The presence or absence of the complementiser* that *in indirect speech and thought*

In English and Afrikaans indirect speech and thought reporting, the complementiser introducing the complement clause (*that* or *dat*) is optional. This represents a recent language change in Afrikaans, since its Early Modern Dutch ancestor and Contemporary Dutch sibling do not allow complementiser omission freely (Van Bogaert & Colleman 2013: 496). Contact with English is identified as an important factor in the dissemination of this construction in nineteenth- and twentieth-century Afrikaans (Feinauer 1990: 119). There is a syntactic difference in that Afrikaans has verb-final subordinate clause word order, so that when the complementiser is present, all verbs in the complement clause are clustered together in the final position. Where the complementiser is omitted, the complement clause has main-clause (V2) word order. This alternation is illustrated in (8). In contrast, English has no word-order alternation to mark dependency, and word order is the same for cases where the complementiser is absent or present (see (9)).

(8) (a) <u>Hy het</u>   erken   ***dat*** *hy alkohol gedrink   het.* (TK)
        He have.AUX admit.PST.PTCP COMP he alcohol PST.PTCP-drink have.AUX
        '<u>He admitted</u> ***that*** *he had drunk alcohol.*'
    (b) <u>Hy het</u>   erken   ***ø*** *hy het   alkohol gedrink.*
        He have.AUX admit.PST.PTCP ø he have.AUX alcohol PST.PTCP-drink
        '<u>He admitted</u> ***ø*** *he had drunk alcohol.*'
(9) (a) … and <u>it is said</u> ***that*** *motions will be made to rescind these appointments.*
        (BrE, Reportage, 19th-1)
    (b) … and <u>it is said</u> ***ø*** *motions will be made to rescind these appointments.*

Afrikaans and English both demonstrate a high frequency of complementiser omission, with Afrikaans demonstrating a particularly strong preference for the form

without the complementiser even in written registers. Van Rooy & Kruger (2016) find an omission ratio of 67 per cent in contemporary written Afrikaans, based on an analysis of the *Taalkommissiekorpus*.

In British and American English, the form without the complementiser is the dominant option in speech, but in writing, the form with the complementiser is overall more frequent (Biber *et al.* 1999). In respect of preferences in specific registers, Biber *et al.* (1999: 680) find that in news reportage the form with the complementiser is strongly preferred, but in fiction, the form without the complementiser is slightly more frequent than the form with the complementiser. In Afrikaans, the form without the complementiser is the dominant choice in both fiction and reportage. Based on this difference, Kruger & Van Rooy (2016) predict a higher omission rate in WSAfE reportage in comparison to BrE as a consequence of covert CLI at the pragmatic (register) level, possibly supported by convergence in publishing norms. They do not find support for this specific effect, but they do find that WSAfE, across all registers except the instructional register, tends towards a higher omission rate than BrE. They propose that given the particularly high omission rate in Afrikaans across written registers, the higher rate of omission in WSAfE may be the consequence of contact with Afrikaans more generally (Kruger & Van Rooy 2016: 129), but this hypothesis needs to be evaluated against diachronic data.

4   Method

4.1   Corpora

The primary analysis for this study was carried out using two comparable diachronic corpora of BrE and WSAfE. For the WSAfE component, we used selected registers from the corpus compiled by Wasserman (2014), as well as the most recent version of the *International Corpus of English for South Africa* (ICE-SA). The registers selected were dependent on the availability of comparable registers in the diachronic BrE corpus, for which we used ARCHER-3.2 (*A Representative Corpus of Historical English Registers*, 2013) and the *International Corpus of English for Britain* (ICE-GB; Nelson *et al.* 2002). The registers included in the corpora are letters (social and business), reportage and fiction.

Speech-reporting constructions fulfil distinct functions in these registers. A full appraisal of these functions falls outside the scope of this article; for the purposes of the argument it will suffice to point out that while in letters and reportage a shared core function of reporting the words of others can be identified, the encoding of evaluative stance is also an important function in these registers (see Jullian 2011). In fiction, the function is quite different, as speech reporting forms part of the representation of the interaction of characters in a narrative world. The representation of speech and thought is widely researched in literary stylistics (see, for example, Jahn 1992; Short 2012).

An important distinction between these registers is that letters tend to be unedited, whereas reportage and fiction are subject to editorial processes, which may be affected by publishing conventions in particular contexts (in themselves subject to change).

Table 1. *Word counts for BrE corpus*

Period	Fiction	News	Letters	**Total**
1800–49 (ARCHER)	45,107	22,951	10,819	**78,877**
1850–99 (ARCHER)	43,427	23,072	12,529	**79,028**
1900–49 (ARCHER)	45,408	21,977	11,361	**78,746**
ICE-GB	58,275	50,329	60,979	**169,583**
Total	**192,217**	**118,329**	**95,688**	**406,234**

Table 2. *Word counts for WSAfE corpus*

Period	Fiction	News	Letters	**Total**
1820–49 (Wasserman)	36,674	41,976	39,639	**118,289**
1850–99 (Wasserman)	76,233	69,507	38,175	**183,915**
1900–59 (Wasserman)	37,852	35,134	53,222	**126,208**
ICE-SA	35,346	39,346	14,415	**89,107**
Total	**186,105**	**185,963**	**145,451**	**517,519**

Moreover, it is well known that different written registers are more or less receptive to linguistic change (Hundt & Mair 1999; Biber & Gray 2013), with reportage generally identified as a register that is particularly apt to take on innovative forms relatively quickly.

The period covered by the two corpora is roughly 1800–2000, divided into four fifty-year periods. The composition of the corpora used is reported in tables 1 and 2.

In the discussion, we also make reference to the *Taalkommissiekorpus* of contemporary written Afrikaans (see footnote 5), and a diachronic corpus of Afrikaans compiled by Kirsten (2016), where relevant, to compare diachronic changes in Afrikaans. This corpus contains 250,000 words of data for each of three decades: the 1910s, 1940s and 1970s.

4.2   Analysis

4.2.1   Data extraction

To identify cases of direct and indirect speech and thought reporting, we used the same verb list as in Kruger & Van Rooy (2016) as search items. This list includes all the public and suasive verbs listed by Quirk *et al.* (1985: 1180–3), as well as an additional selection of verbs identified as introducing speech and thought reporting, which correspond to the verbs used in the analysis of Afrikaans complementation patterns by Van Rooy & Kruger (2016): *think, moan, mumble, mutter, whisper, yell, admonish, joke, snarl, warn, respond, wish* and *guess*. Using the concordances of all these verb lemmas, we analysed the entire data set to identify and code the following grammatical variants:

1. Direct speech with the reporting clause in (a) initial, (b) medial and (c) final position.
2. In all instances of direct and indirect speech with the reporting clause in medial and final position, constructions with the following word orders in the reporting clause: (a) noun phrase + verb, (b) verb + noun phrase, (c) pronoun + verb, and (d) verb + pronoun.
3. Indirect speech with the reporting clause in initial position, with the complementiser (a) present or (b) absent.

A total of 13,951 English concordance lines were analysed, which yielded 4,870 relevant cases included in the analysis, with frequencies of all variants of the three constructions under investigation ranging between 357 and 771 per 100,000 words per period per variety. Since the focus in this study is on proportional changes, raw frequencies rather than normalised frequencies are reported in most cases. Data for Afrikaans are taken from ongoing work on *Taalportaal*, an online grammar of Afrikaans (www.taalportaal.org), as well as Van Rooy & Kruger (2016). The total sample of concordance lines on which the analysis of Afrikaans is based exceeds 60,000, but includes a wider range of written registers than the data for English.

4.2.2 Data analysis

For all three constructions, the analysis proceeds in two steps. The first step compares the diachronic changes in the relative (proportional) preferences of one option over the other(s) for each of the three constructions, across the two varieties, to identify whether there are overall patterns of divergence in the frequency with which certain options are selected. Only if such divergence is demonstrated, can we consider whether it reflects grammatical changes related to CLI through comparison with the Afrikaans data. Register differences are considered, where relevant to the construction in question. The Afrikaans data are not included in the statistical analysis, however, because of differences in the way they were extracted, and also because the Afrikaans historical data go back only to the beginning of the twentieth century and have a different register composition.[6]

Following this, a statistical evaluation of the differences in development across the two varieties and three registers in question is carried out. For this purpose, we use conditional inference trees, a method for regression and classification using binary recursive partitioning (see Tagliamonte & Baayen 2012), as implemented in the R-package 'partykit' (Hothorn & Zeileis 2016). Hothorn & Zeileis (2016: 7–8) explain that conditional inference trees

> estimate a regression relationship by binary recursive partitioning in a conditional inference framework. Roughly, the algorithm works as follows: 1) Test the global null hypothesis of independence between any of the input variables and the response (which

---

[6] Prior to the late nineteenth century, a diglossic situation obtained for Afrikaans speakers, who generally used Dutch rather than Afrikaans as the language of writing and formal communication (education, government, religion). It is therefore not possible to develop a corpus of nineteenth-century written Afrikaans; a representative corpus would mainly contain texts that were intentionally written in (the local variety of) Dutch.

may be multivariate as well). Stop if this hypothesis cannot be rejected. Otherwise select the input variable with strongest association to the response. This association is measured by a p-value corresponding to a test for the partial null hypothesis of a single input variable and the response. 2) Implement a binary split on the selected input variable. 3) Recursively repeat steps 1) and 2).

Conditional inference trees offer several advantages for linguistic analysis. They are suited to conditions of data sparseness, deal well with high-order interactions and highly correlated predictors, do not assume normal distribution of data, and are robust to outliers (Levshina 2015: 292). Importantly, they straightforwardly visualise how multiple predictors (or independent variables) operate together in conditioning the selection of a linguistic choice (Tagliamonte & Baayen 2012: 135).

We use conditional inference tree modelling to determine the relative importance of the three factors (or independent variables) of Variety, Register and Period in conditioning the choice between the options for the three constructions (the dependent variable). Given Thomason's (2001) recommendation that multiple causation should be considered when dealing with contact-induced language change, we formulate the following principles for interpreting the conditional inference trees. Evidence that Period significantly conditions the choice between alternatives indicates that there is change over time. A statistically significant contribution of Variety to the inference tree structure, coupled with correspondence of WSAfE to the Afrikaans frequency distributions, is accepted as evidence for language contact.

One limitation of the current research design should be highlighted. Other factors, especially grammatical ones, are also likely to play a role in the choice between variants. These factors are not included in the current analysis, but in section 6 we reflect on the implications that the inclusion of such factors may have for the current findings, and outline some future research possibilities.

5   Findings and discussion

5.1   The position of the reporting clause in direct speech and thought reporting

Positional variation for the reporting clause is rare in letters. The non-initial position is attested only thirteen times across the two corpora and four periods, and therefore the analysis in this section includes only fiction and news reporting. The two registers are analysed separately, because the distributions in fiction in contemporary BrE and SAfE are similar, while the distributions in reportage are quite different.

5.1.1   Fiction
The assumption in the analysis that follows is that the initial position is the most conservative option that resembles the syntax of complementation most, while the alternative positions are regarded as parenthetical, but are interpretable in terms of their correspondence, or, in a loose sense of the word, derivational relationship, with the initial variant (see Huddleston 2002: 1024–9). Houston (2013: 203) argues that the non-initial variants developed during the course of the eighteenth century in tandem
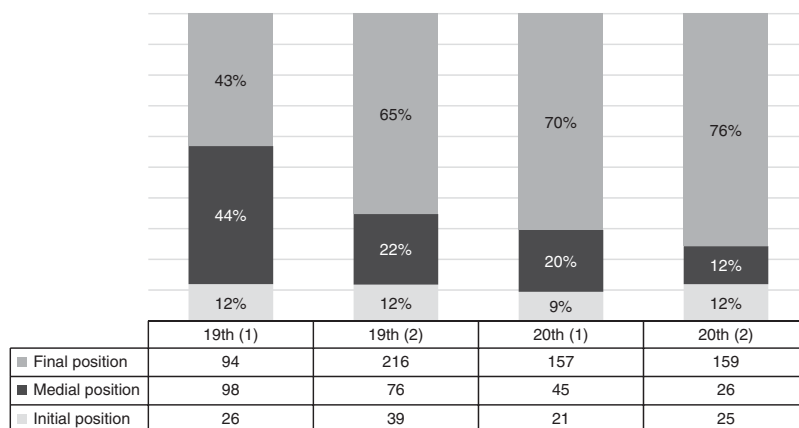
Figure 1. Proportional frequency of positional variants for the reporting clause across four periods: BrE fiction
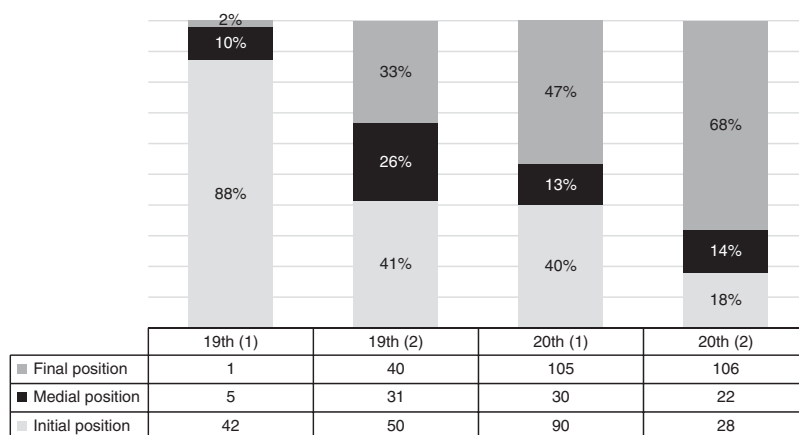
	19th (1)	19th (2)	20th (1)	20th (2)
▨ Final position	94	216	157	159
■ Medial position	98	76	45	26
▨ Initial position	26	39	21	25

Figure 2. Proportional frequency of positional variants for the reporting clause across four periods: WSAfE fiction

	19th (1)	19th (2)	20th (1)	20th (2)
▨ Final position	1	40	105	106
■ Medial position	5	31	30	22
▨ Initial position	42	50	90	28

with the development of the conventions of the novel, specifically its need for direct speech as a more dramatic way of presenting the words of characters. Figures 1 and 2 compare the proportional (relative) frequency of the three positional variants (initial, medial and final) across the four periods in BrE and WSAfE respectively, with raw frequencies also provided in data tables.

In BrE, the initial position is a marginal choice in fiction throughout the period under investigation, which means that it had already receded quite dramatically by the end of the eighteenth century as the novel became an established genre. The medial position shows very strong representation in the first half of the nineteenth century, but thereafter
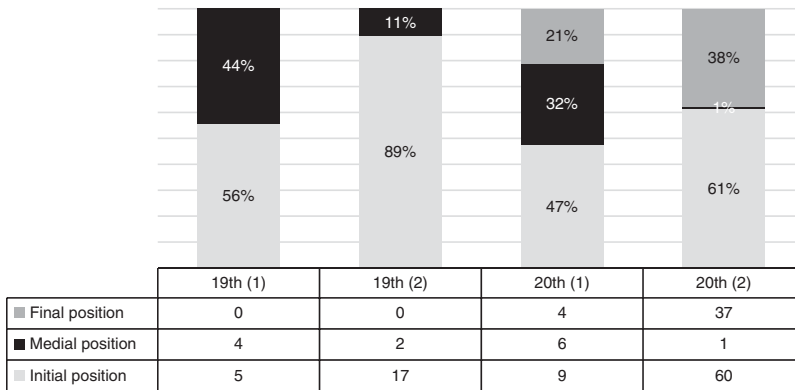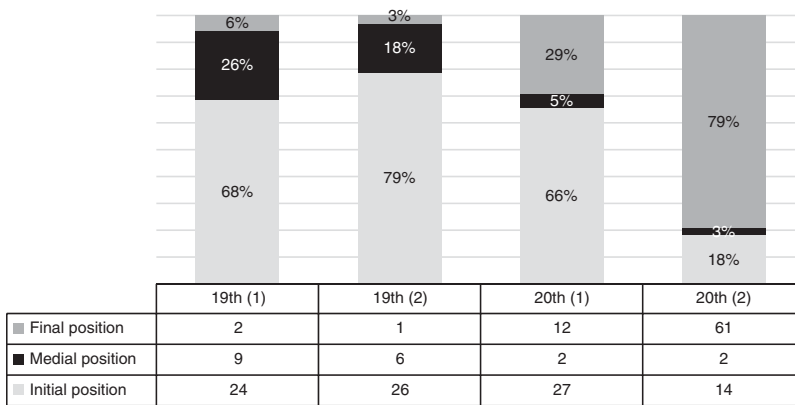
	19th (1)	19th (2)	20th (1)	20th (2)
▦ Final position	0	0	4	37
■ Medial position	4	2	6	1
▫ Initial position	5	17	9	60

Figure 3. Proportional frequency of positional variants for the reporting clause across four periods: BrE reportage

	19th (1)	19th (2)	20th (1)	20th (2)
▦ Final position	2	1	12	61
■ Medial position	9	6	2	2
▫ Initial position	24	26	27	14

Figure 4. Proportional frequency of positional variants for the reporting clause across four periods: WSAfE reportage

gradually declines. The final position vies with the medial position as the dominant pattern in the first half of the nineteenth century, after which it gradually increases to the major use pattern in this register in the second half of the twentieth century.

In comparison, the conservative initial position is very dominant in the first half of the nineteenth century in WSAfE fiction, with just one case of the final position. However, a note of caution should be sounded: in this period, there is, strictly speaking, no published fiction in WSAfE, and this register consists mostly of auto-biography and journals, where the established conventions of fiction did not play such an important role. Nevertheless, the initial position remains a comparatively strongly preferred option across at least the first three periods compared to BrE, converging with BrE norms only in the second half of the twentieth century. The medial position

is never as strongly represented in WSAfE fiction as it is in BrE fiction, although it also converges with the decline in BrE in the second half of the twentieth century. The increase in preference for the final position evident in BrE is also evident in WSAfE, but the rate of increase is much steeper, and by the late twentieth century, what was a decidedly minor use pattern has become strongly dominant.

5.1.2   Reportage

What emerges clearly from a comparison of relative frequency of the three positional variants of the reporting clause in reportage (figures 3 and 4) and fiction (figures 1 and 2) is that across both varieties the initial position is much more strongly preferred in reportage than in fiction, and compared to fiction, the medial variant is very marginal in reportage by the end of the twentieth century.

There is a gradual increase in the prevalence of the final position in reportage in both WSAfE and BrE, starting in the first half of the twentieth century in both varieties. However, in BrE this increase does not continue to the second half of the twentieth century, and the initial position remains the major use pattern in this register in this variety. By contrast, in WSAfE the change is quite extensive. As pointed out in section 3.1, in contemporary Afrikaans, the preference in news reportage is strongly in favour of the final position (in contrast with BrE, which prefers the initial position), and Kruger & Van Rooy (2016) propose that the differences in preference in contemporary BrE and WSAfE may be the consequence of contact with Afrikaans. The diachronic data allow us to refine this analysis: the proportional frequency of the final position increases much more in WSAfE, making it very similar to contemporary Afrikaans, where the final position is the major use pattern in this register.

The conditional inference tree analysis (see figure 5) both supports this overall analysis and allows us to further refine it.[7]

The highest-level predictor identified is Register, confirming that fiction and reportage demonstrate different preferences across time and variety. When we consider news reporting (the right main branch of node 1), it is clear that the divergence between BrE and WSAfE in positional preferences emerges only in the second half of the twentieth century (node 17): WSAfE strongly prefers the final position, and BrE retains a comparatively stronger preference for the more conservative initial position. Prior to this, the two varieties share the same diachronic developments (see nodes 13, 14), and Period is a higher-level predictor of the choice than Variety. An external factor, contact with Afrikaans in this case, is likely, given the quite dramatic divergence in the rate of change between WSAfE and BrE. The timeline is consistent with the development of joint publication houses for Afrikaans and English primarily in the second half of the twentieth century (see Kantey 1990).

---

[7] A crosstabulation of predicted probabilities with observed outcomes demonstrates that the model correctly classifies 62 per cent of the total of 1,863 observations, which offers some improvement over a baseline of random classification. However, this does suggest that other factors may play a role in conditioning the choice (see section 6).
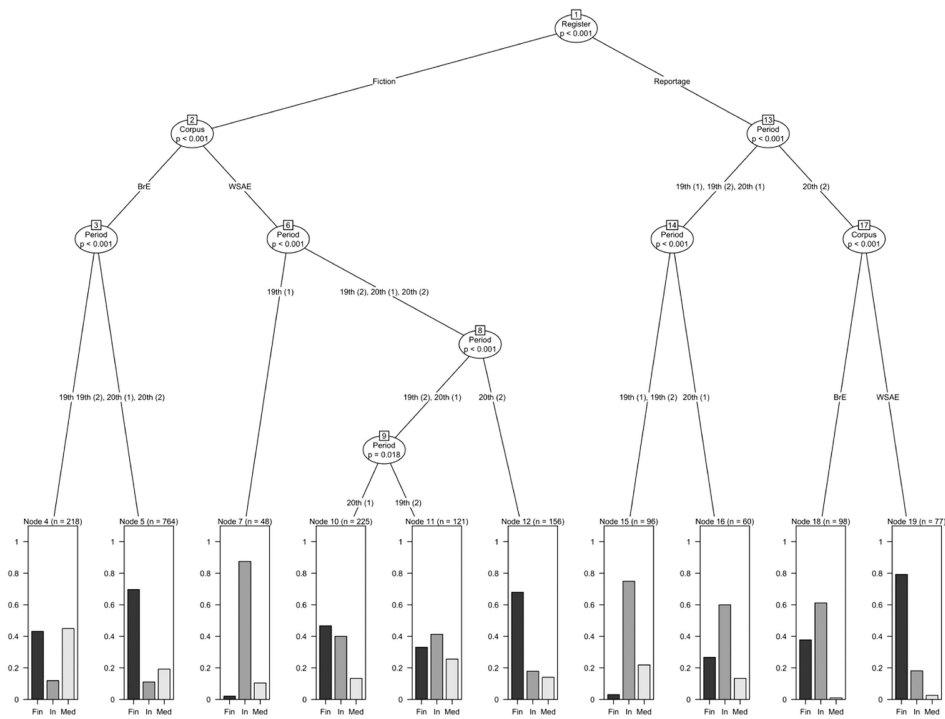
Figure 5. Conditional inference tree for the choice of positional variant (Fin = final, In = initial, Med = medial), with Register, Period and Variety as predictors

In fiction, the subsequent split is for corpus. British English (see node 3) follows a relatively straightforward diachronic development, in which the final position gradually increases at the expense of the medial position, with the nineteenth and twentieth centuries distinct in this respect. WSAfE shows a more complex, but still coherent development, with a slower and more gradual change in the shift in preference from initial to final position, without going through a phase where the medial position is particularly important.

Written Afrikaans (see figure 6) shows a gradual increase in the choice of the initial variant, at the expense of the final variant for most of the twentieth century, but this trend is reversed sharply after the 1970s. Contemporary Afrikaans uses the initial variant slightly less than 30 per cent of the time, and the final variant around 60 per cent of the time. Despite the direction of change, though, the final variant has remained dominant throughout the twentieth century in written Afrikaans, and thus potentially exerted covert CLI on WSAfE to follow suit. A further similarity is that in Afrikaans and WSAfE, the register differentiation between fiction and reportage is quite small (see section 3.1 above), whereas BrE maintains a considerable differentiation between the two registers.
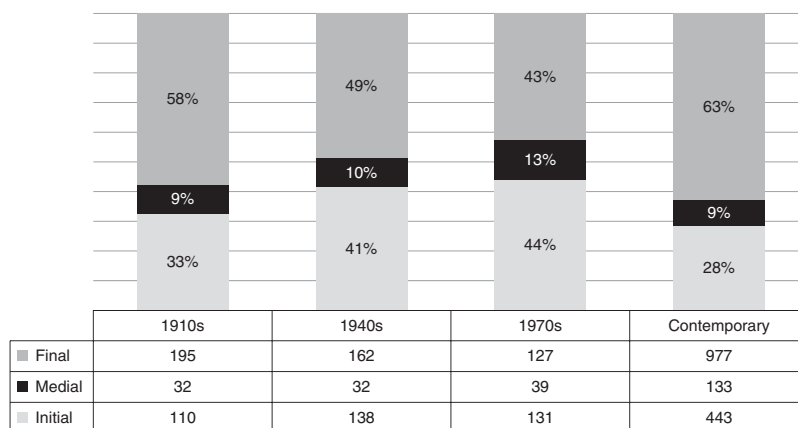
	1910s	1940s	1970s	Contemporary
■ Final	195	162	127	977
■ Medial	32	32	39	133
■ Initial	110	138	131	443

Figure 6. Proportional frequency of positional variants for the reporting clause: Afrikaans

5.2   Quotative inversion in reporting clauses in non-initial position in direct and indirect speech and thought

5.2.1   Non-initial reporting clauses with pronoun subjects

As discussed in section 3.2, quotative inversion is the only feature where the possibility of overt transfer exists, if the archaic V-Pronoun construction is retained in WSAfE under the influence of the obligatory V-Pronoun order of Afrikaans. The relative frequencies of the V-Pronoun versus Pronoun-V order over the four periods in the two varieties (for all three registers) are shown in figures 7 and 8.

There is a similar pattern across the two varieties of a feature receding in the nineteenth century, and all but gone by the twentieth century, although the change in WSAfE lagged behind in the second half of the nineteenth century, as illustrated in (10).

(10)   'Hadn't we better change the name, Bain,' said I when he had gone, 'and call this "Orthis Kloof?"' (WSAfE, Fiction, 19th-2)

The conditional inference tree (see figure 9)[8] shows that the effect of Period is the highest-level predictor, distinguishing the nineteenth and twentieth centuries (node 1). The left branch of the tree (the nineteenth-century data) further shows that BrE and WSAfE underwent somewhat distinct developments in this time period. BrE was advanced in the change relative to WSAfE, whereas in WSAfE this pattern was distinguished by register. Contact with Afrikaans is unlikely to explain the delay in the disappearance of the V-pronoun inversion, since the contact would have been at its lowest in the second half of the nineteenth century,

---

[8] A crosstabulation of predicted probabilities with observed outcomes demonstrates that the conditional inference tree has a very high classification accuracy, with 90 per cent of the 815 cases correctly classified.
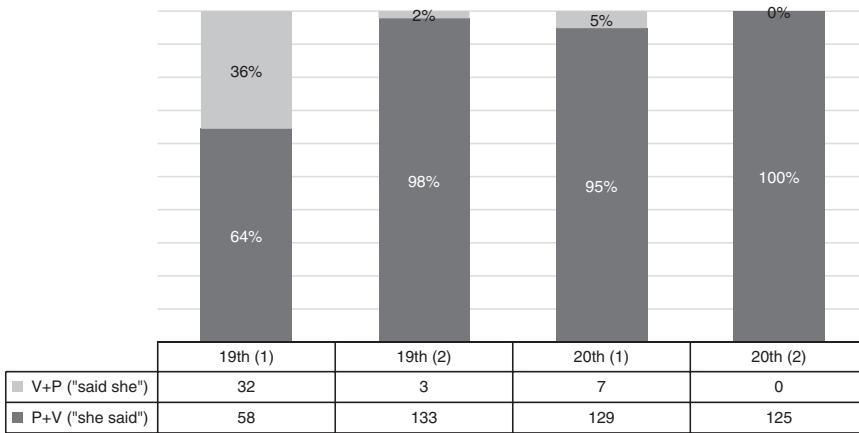
	19th (1)	19th (2)	20th (1)	20th (2)
▨ V+P ("said she")	32	3	7	0
■ P+V ("she said")	58	133	129	125

Figure 7. Proportional frequency of inverted vs non-inverted order in reporting clauses with pronoun subjects across four periods: BrE

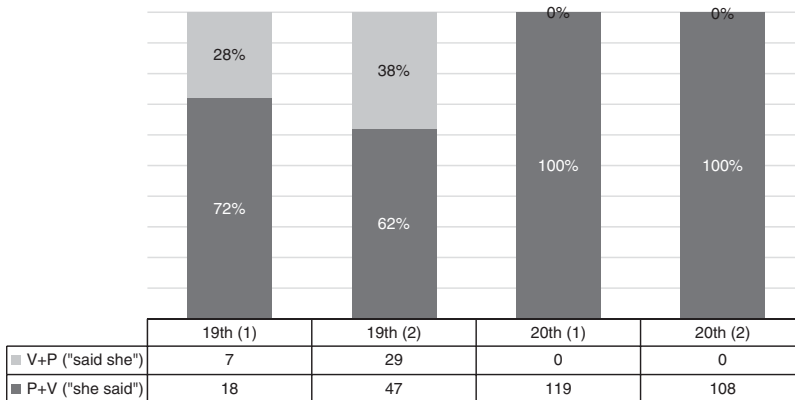|	19th (1)	19th (2)	20th (1)	20th (2)
▨ V+P ("said she")	7	29	0	0
■ P+V ("she said")	18	47	119	108

Figure 8. Proportional frequency of inverted vs non-inverted order in reporting clauses with pronoun subjects across four periods: WSAfE

and publishing in Afrikaans (and even Dutch) was very limited in the nineteenth century.

The register distinction in the nineteenth century in WSAfE (though with relatively small numbers) provides some support for the notion that some registers are more 'conservative' and others more receptive to ongoing change. Letters and reportage had already been closer to selecting only the non-inverted word order than fiction as far as pronoun subjects were concerned. As noted earlier, fiction established itself later in WSAfE, and register conventions might still have been in flux in this period.
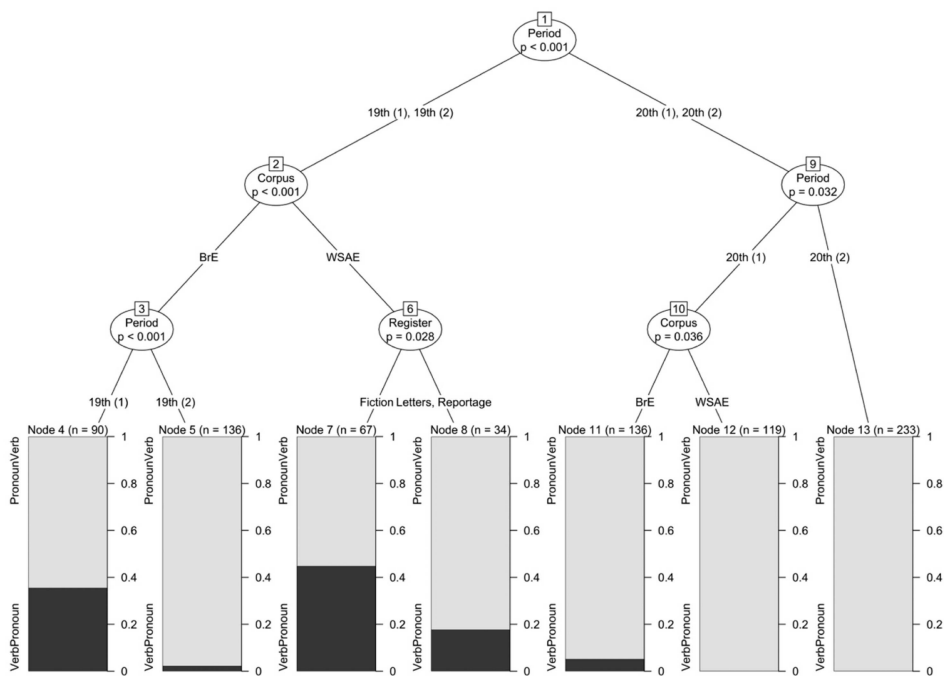
Figure 9. Conditional inference tree for quotative inversion with pronoun subjects, with Register, Period and Variety as predictors
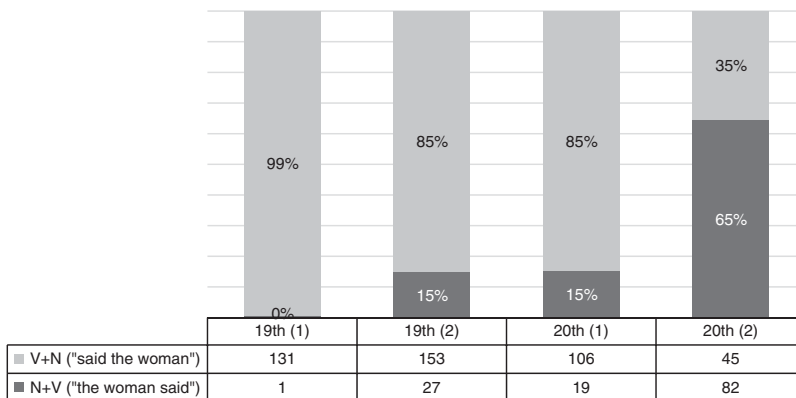

|	19th (1)	19th (2)	20th (1)	20th (2)
▨ V+N ("said the woman")	131	153	106	45
■ N+V ("the woman said")	1	27	19	82

Figure 10. Proportional frequency of inverted vs non-inverted order in reporting clauses with noun subjects across four periods: BrE

5.2.2 Non-initial reporting clauses with noun subjects

Figures 10 and 11 summarise the relative frequency of quotative inversion in reporting clauses with noun subjects in BrE and WSAfE respectively.
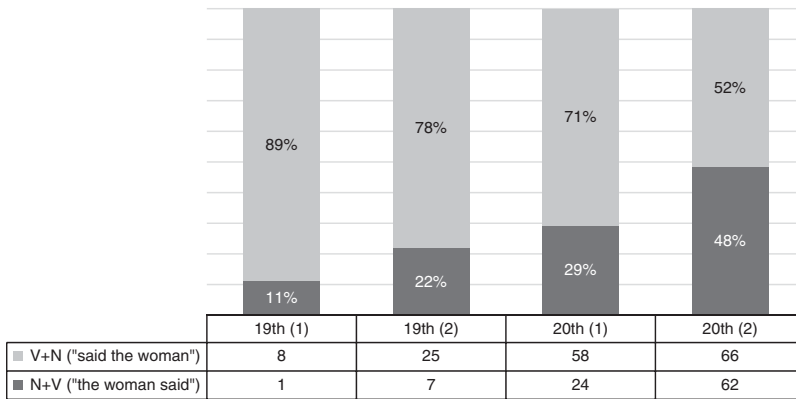
Figure 11. Proportional frequency of inverted vs non-inverted order in reporting clauses with noun subjects across four periods: WSAfE
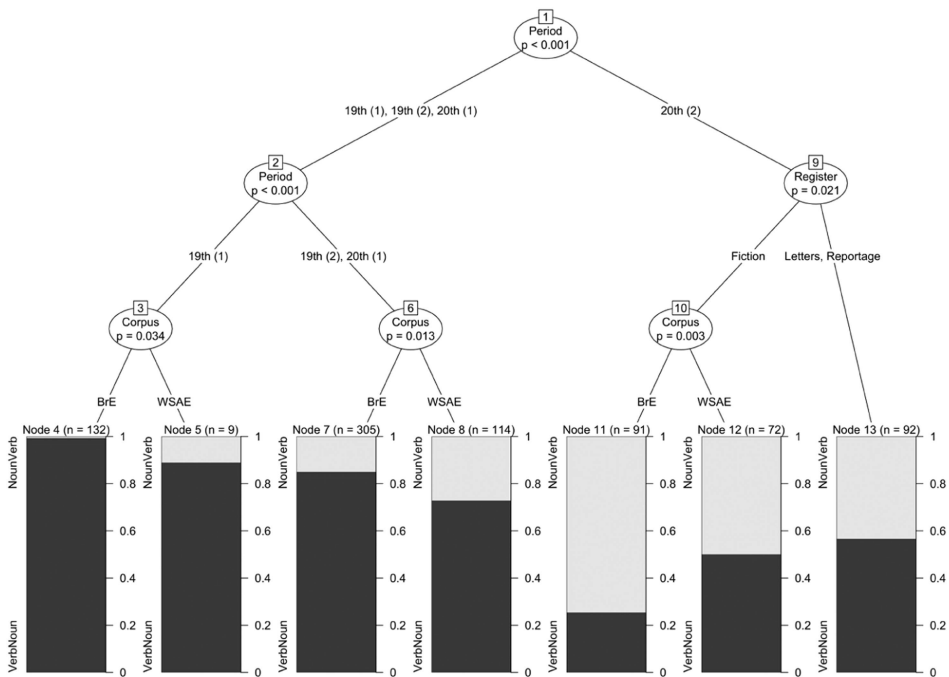

Figure 12. Conditional inference tree for quotative inversion with noun subjects, with Register, Period and Variety as predictors

In the case of this feature, there is a clear shared historical diffusion of minor to major use pattern across the two varieties. The language change is in the opposite direction to the pattern that would be influenced by the obligatory Afrikaans word
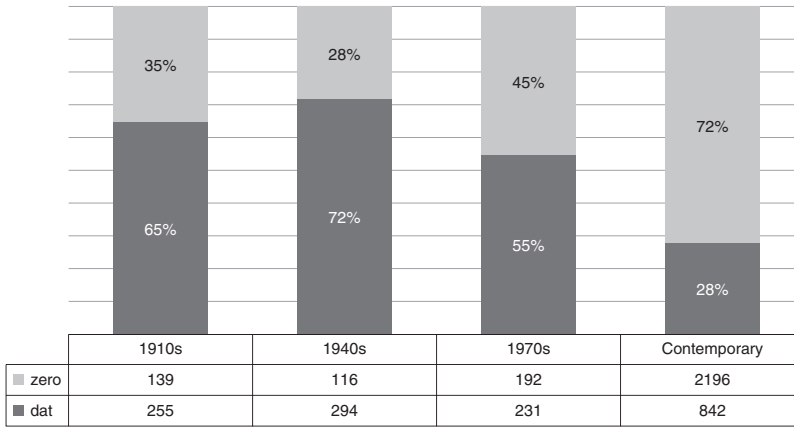
Figure 13. Proportional frequency of *dat* present and absent: Afrikaans

order. The non-inverted N + V construction is all but absent in the earlier nineteenth century (a definite minor use pattern), but is adopted at reasonably similar rates by both BrE and SAfE. By the second half of the twentieth century, the N + V pattern has developed to a major use pattern in BrE – but not quite in WSAfE, where the inverted V + N pattern is still more strongly represented. The stronger representation of the V + N pattern can perhaps be attributed to covert, rather than overt, CLI from Afrikaans, in that for contemporary written English by Afrikaans speakers, as reported by Kruger & Van Rooy (2016: 127), the inverted form is favoured strongly in reportage, and thus may exert a pull in the opposite direction from the general change in English on WSAfE (no results were reported for fiction, since there is very little fiction produced originally in English by Afrikaans speakers).

The conditional inference tree (figure 12)[9] confirms that this feature is most strongly conditioned by Period, with a split on node 1 setting the second half of the twentieth century apart from the earlier periods, where the V-N construction dominates. However, WSAfE seems to be somewhat ahead of BrE in adopting the innovative N-V construction in the second half of the nineteenth and early twentieth century. In the second half of the twentieth century, reportage and letters (on the one hand) and fiction (on the other) behave somewhat differently across the two varieties: reportage and letters, across the two varieties, have a relatively high frequency of the innovative N-V pattern. In fiction (the left branch of node 9), BrE and WSAfE behave differently: BrE is more advanced in the change towards the N-V pattern, whereas in WSAfE, the change lags somewhat behind, with a frequency similar to that in letters and reportage. The relative similarity of registers in WSAfE, compared to a larger degree of divergence across registers in BrE, again emerges as a finding from the diachronic data.

[9] Crosstabulation of predicted probabilities with observed outcomes demonstrates that the model has good classification accuracy, correctly classifying 78 per cent of the 815 cases in the analysis.
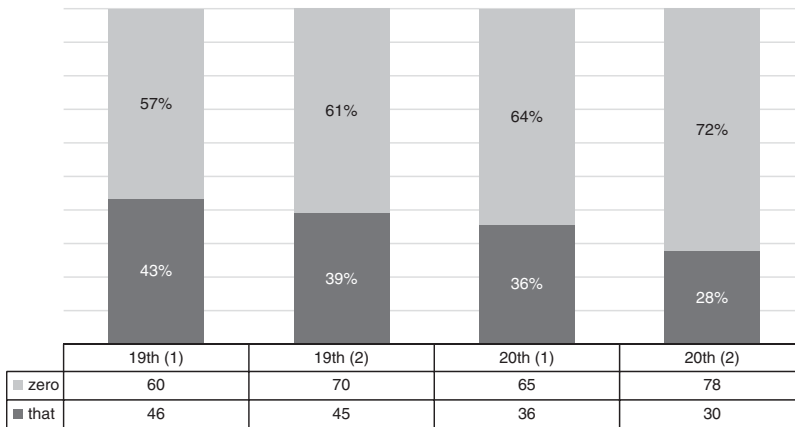
	19th (1)	19th (2)	20th (1)	20th (2)
▪ zero	60	70	65	78
▪ that	46	45	36	30

Figure 14.   Proportional frequency of *that* present and absent across four periods: BrE fiction

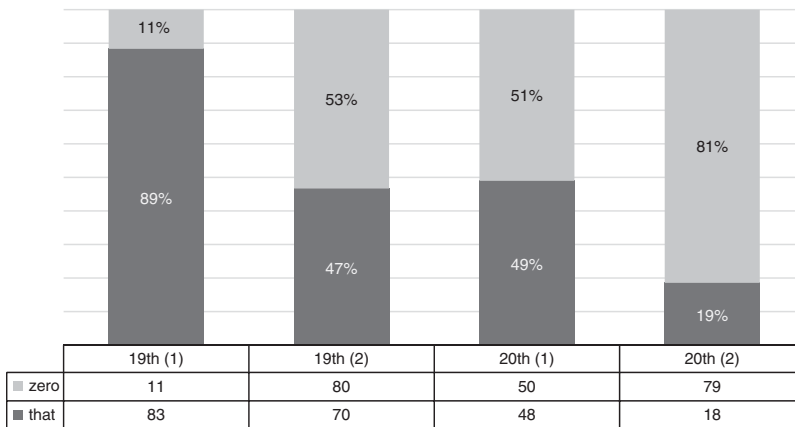|	19th (1)	19th (2)	20th (1)	20th (2)
▪ zero	11	80	50	79
▪ that	83	70	48	18

Figure 15.   Proportional frequency of *that* present and absent across four periods: WSAfE fiction

5.3   The presence or absence of the complementiser *that* in indirect speech and thought

For this analysis, only complement clauses with the reporting clause in initial position were analysed, since for the medial and final positions the version without the complementiser seems to be required almost categorically.[10] Summarising across the

---

[10] The values reported for the contemporary data in this section are in some cases slightly different from the values reported by Kruger & Van Rooy (2016). In that article, complement clauses with non-initial reporting clauses were included in the calculations. There was also a copying error in the text, which meant that the
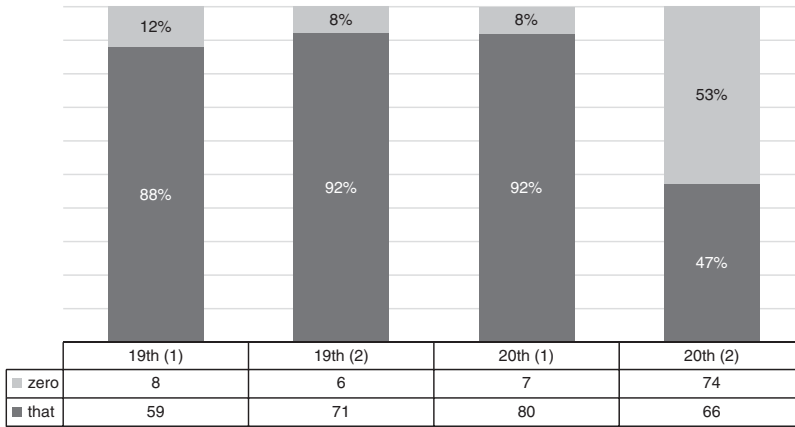
	19th (1)	19th (2)	20th (1)	20th (2)
zero	8	6	7	74
that	59	71	80	66

Figure 16. Proportional frequency of *that* present and absent across four periods: BrE reportage

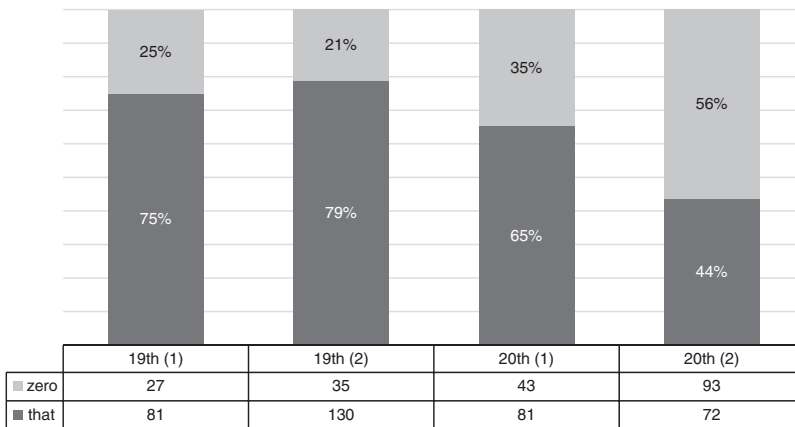|	19th (1)	19th (2)	20th (1)	20th (2)
zero	27	35	43	93
that	81	130	81	72

Figure 17. Proportional frequency of *that* present and absent across four periods: WSAfE reportage

registers, the proportion of complement clauses with the overt complementiser decreases in favour of the option with zero complementiser in both BrE and WSAfE. The same pattern is evident in Afrikaans (see figure 13).

However, the patterns in English are quite different across registers, and therefore each register is considered separately in the following discussion.

values for the fiction register were copied incorrectly from the spreadsheet to the article. The interpretation of results is unaffected by this, but the correct values are reported in the present article.
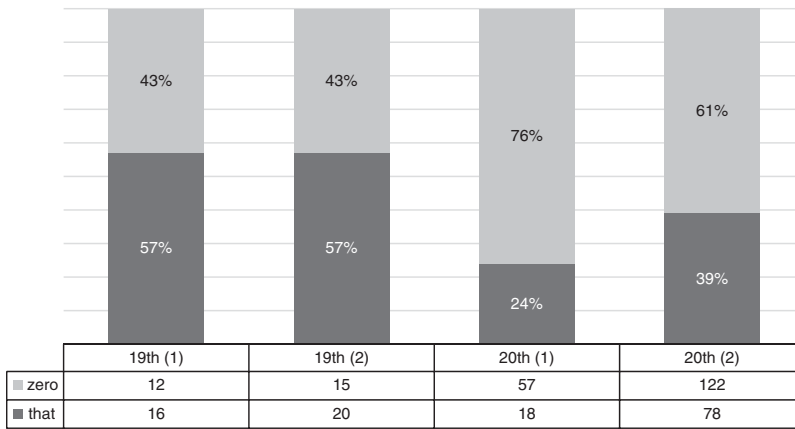
	19th (1)	19th (2)	20th (1)	20th (2)
▢ zero	12	15	57	122
▪ that	16	20	18	78

Figure 18. Proportional frequency of *that* present and absent across four periods: BrE letters

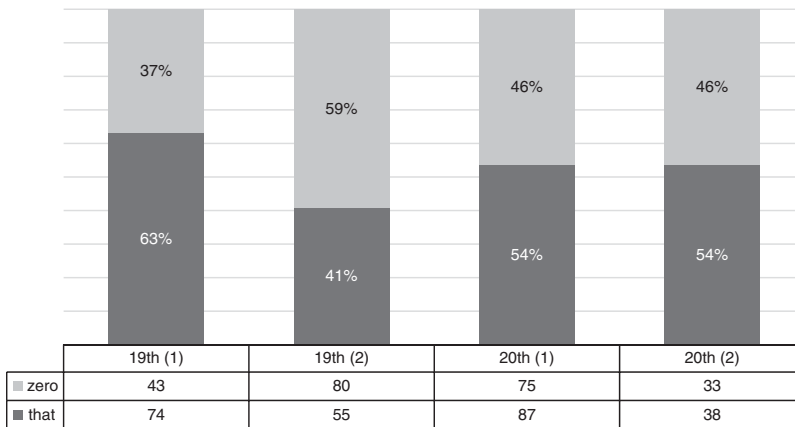|	19th (1)	19th (2)	20th (1)	20th (2)
▢ zero	43	80	75	33
▪ that	74	55	87	38

Figure 19. Proportional frequency of *that* present and absent across four periods: WSAfE letters

WSAfE fiction (figure 15) shows a very steep change in frequency, with an almost inverted preference for *that*/zero for the early nineteenth century compared to the late twentieth century. For BrE fiction (figure 14) the pattern is the same, but far more moderate – starting out with a preference for zero, and slowly increasing this preference.

For reportage, by the late twentieth century, very similar omission ratios are observed for the two varieties, with WSAfE slightly higher (as shown by figures 16 and 17), perhaps reflecting the development of a more general global shift to increased colloquialisation in reportage in native varieties of English. Prior to this, however, BrE demonstrated a strong preference for *that*, throughout the nineteenth and earlier twentieth century, whereas for WSAfE the increase has been more gradual in the
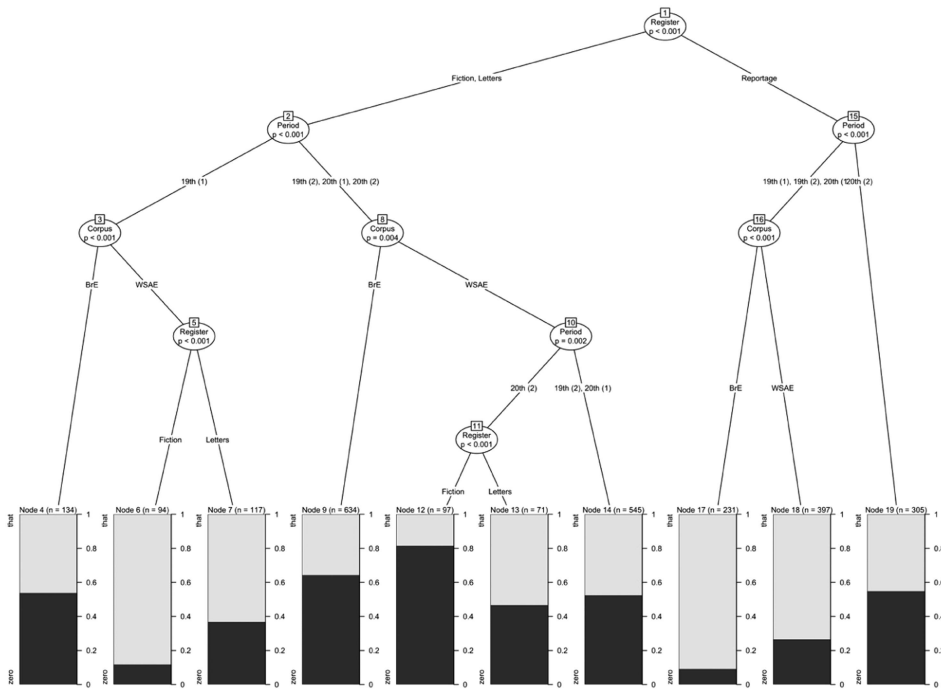
Figure 20. Conditional inference tree for complementiser omission, with Register, Period and Variety as predictors

direction of favouring the zero form – and with a stronger preference for zero from the beginning.

As far as letters are concerned, the data show that BrE participates in the overall change towards the zero complementiser (figure 18), but this is not the case for WSAfE, which continues to prefer the *that* form (figure 19). Considering the differences in register distribution, there is some evidence that in edited written registers, WSAfE and Afrikaans converge on similar omission preferences, whereas the unedited letters maintain a separate identity, further affirming our contention that the publishing environment is an important site of contact between English and Afrikaans. The evidence points to the likelihood of mutual reinforcement between Afrikaans and English; both varieties undergo the same change, but Afrikaans changes more extensively than WSAfE.

The conditional inference tree (figure 20)[11] demonstrates that the highest-level predictor of the choice between *that*/zero is register (node 1). For reportage (the right branch of node 1), there is a distinct diachronic development (node 15), which sets the latter half of the twentieth century apart from the preceding periods. In the earlier periods, BrE and WSAfE are significantly different (see node 16),

---

[11] A crosstabulation of predicted probabilities with observed outcomes demonstrates that the conditional inference tree has reasonably good classification accuracy, correctly classifying 65 per cent of 2,625 observations.

with WSAfE demonstrating a stronger preference for zero, but both varieties converge on the zero variant by the latter half of the twentieth century.

For fiction and letters, the following split is on period (node 2), isolating the first half of the nineteenth century (left branch of node 2). In this period, BrE and WSAfE also demonstrate distinct preferences, with WSAfE overall more conservative in selecting the form with *that*, but more so in fiction than in letters (see node 5). In the subsequent periods, BrE and WSAfE are also significantly different (see node 8). BrE, for the whole duration of this period, demonstrates a strong preference for zero, whereas in WSAfE, there are further distinctions: in the latter part of the nineteenth century and the early twentieth century (node 10), the preference is mostly for zero, whereas in the second part of the twentieth century, there is a register differentiation (node 11), with fiction demonstrating a far higher omission ratio than letters. In this case, it appears as if the role of CLI in the development of WSAfE is quite marginal, and it is more likely that Afrikaans has been influenced by English. Both varieties are also likely to be influenced by other, presumably grammatical, factors – a point taken up in the following section.

6 Summary and conclusion

A set of related language changes in WSAfE and BrE were investigated in this article to determine the degree to which contact with Afrikaans is required to explain the changes in WSAfE. In the case of the position of the reporting clause in direct reported speech and thought, there is a move from minor to major use pattern (in the increasing preference for the final position of the reporting clause) across both BrE and WSAfE. The pattern of change in relative frequency is more extensive in WSAfE reportage, which strongly prefers this position in comparison to BrE reportage. The same pattern of change is evident in Afrikaans, with a strong increase in preference for the final position after the 1970s, which suggests mutual reinforcement between English and Afrikaans in the South African publishing industry, leading to a similar frequency change in both languages, and at the same time, divergence between WSAfE and BrE. In this case, an existing change in progress from minor to major use pattern in English is amplified by contact with Afrikaans. Covert CLI, due to reinforcement of the frequencies with which the construction is encountered in texts in Afrikaans and/or the second-language English of Afrikaans speakers, is the main reason for the difference in the rate of change and the more advanced endpoint of the change in WSAfE. Afrikaans and WSAfE also turn out to have very similar frequencies in both reportage and fiction, which points to a simplification of the register differentiation within and between the two languages in the contact situation.

As far as quotative inversion is concerned, for pronoun subjects, inversion disappears in both varieties, and contact with Afrikaans in the form of overt CLI does not appear to play an important role. The ongoing change continues until its end in the categorical loss of a grammatical variant in both BrE and WSAfE.

For quotative inversion with noun subjects, there is a clear diffusion from minor to major use pattern in both varieties, with the inverted V + N form initially dominant and then gradually supplanted by the non-inverted N + V form. In BrE the latter develops as clear major use pattern by the 1990s, but the change in WSAfE is slower, and the inverted pattern with V + N is still strongly represented. Contact with Afrikaans (and its obligatory V + N pattern) therefore appears to slow the rate of change in WSAfE in the late twentieth century (when contact between the two languages increased), but the effect of Afrikaans is, given the current data, not strong enough to arrest the ongoing change in English.

For the third construction, the alternation between *that*/zero, a similar change from minor to major use pattern in the preference for zero is evident across the English varieties, as well as in Afrikaans. There appears to be mutual reinforcement in the rate of change between Afrikaans and WSAfE, and similar proportional frequencies emerge in fiction and reportage. Differences in published and unpublished registers are also suggestive of a convergence in publishing norms, with published registers demonstrating a higher omission rate, as is also the case for Afrikaans. The influential Afrikaans style guide by Müller (2003: 664), which is based on the in-house guide of the largest multilingual publishing house in the country, Media24, explicitly advocates the omission of the complementiser in certain contexts.

In sum, the analysis demonstrates that sustained contact can lead to linguistic change even with a less dominant language affecting the shape of a more dominant language. The data in this study show that such change happens at a covert rather than overt level: the frequencies with which variants are selected converge below the level of conscious awareness, which points to a cycle of reinforcement between the exposure to texts in WSAfE, Afrikaans English, and perhaps to a lesser degree, Afrikaans; and the production of such texts. In the case of the contact between Afrikaans and WSAfE, it is clear that the site of contact is in the publishing industry.

Besides the role of the cycle of frequency reinforcement, the data also show evidence of simplification of register differences of those registers that are in contact in the publication environment, which ties in with the second of the covert forces identified in section 2. It is likely that the processing strain that competing preferences can put on text producers is mitigated by choosing constructions in similar ways across registers and across the two languages.

As pointed out in section 4.2.2, a limitation of the current study is that it does not consider additional variables, particularly grammatical variables, that may condition the choice between variants of a construction. The classification accuracy of the statistical models for complementiser omission and the relative position of the reporting and reported clause are above chance, but not as high as the models for quotative inversion, pointing to other factors that also play an important role in conditioning the ongoing change. The role of grammatical variables in conditioning, for example, the presence or absence of the complementiser is well researched (see, for example, Torres-Cacoullos & Walker 2009). Investigating the role of these grammatical factors over time is an important next step in this research. Moreover, as Szmrecsanyi (2016) argues, it is important to consider that the frequencies of linguistic

features may be conditioned by changes to the textual environment – and thus may be a collateral effect of other changes in discourse patterns in different registers.

Authors' addresses:

Department of Linguistics
Macquarie University
Macquarie Walk
North Ryde NSW 2109
Australia
haidee.kruger@mq.edu.au

UPSET Research Focus Area
North-West University
Hendrik van Eck Blvd
Vanderbijlpark 1911
South Africa
bertus.vanrooy@nwu.ac.za

References

ARCHER-3.2 = *A Representative Corpus of Historical English Registers* version 3.2. 1990–2013. Originally compiled under the supervision of Douglas Biber and Edward Finegan at Northern Arizona University and University of Southern California; modified and expanded by subsequent members of a consortium of universities. Current member universities are Bamberg, Freiburg, Heidelberg, Helsinki, Lancaster, Leicester, Manchester, Michigan, Northern Arizona, Santiago de Compostela, Southern California, Trier, Uppsala, Zurich.

Biber, Douglas & Bethany Gray. 2011. Grammar emerging in the noun phrase: The influence of written language use. *English Language and Linguistics* 15(2), 223–50.

Biber, Douglas & Bethany Gray. 2013. Being specific about historical change: The influence of sub-register. *Journal of English Linguistics* 41(2), 104–34.

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

Bybee, Joan. 2010. *Language, cognition and usage*. Cambridge: Cambridge University Press.

D'Arcy, Alexandra. 2012. The diachrony of quotation: Evidence from New Zealand English. *Language Variation and Change* 24, 343–69.

D'Arcy, Alexandra. 2015. Quotation and advances in understanding syntactic systems. *Annual Review of Linguistics* 1, 43–61.

Feinauer, Ilse. 1990. Skoon afhanklike sinne in Afrikaanse spreektaal. *Suid-Afrikaanse Tydskrif vir Taalkunde* 8(3), 116–20.

Filppula, Markku, Juhani Klemola & Devyani Sharma (eds.). 2017. *The Oxford handbook of World Englishes*. Oxford: Oxford University Press.

Fischer, Olga. 2013. The role of contact in English syntactic change in the Old and Middle English periods. In Schreier & Hundt (eds.), 18–40.

Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. New York: Cambridge University Press.

Hickey, Raymond. 2013. English as a contact language in Ireland and Scotland. In Schreier & Hundt (eds.), 88–105.

Hickey, Raymond (ed.). 2017. *The Cambridge handbook of areal linguistics*. Cambridge: Cambridge University Press.

Hothorn, Torsten & Achim Zeileis. 2016. Package 'partykit': CRAN, https://cran.r-project.org/web/packages/partykit/partykit.pdf (20 June 2017)

Houston, Keith. 2013. *Shady characters: The secret life of punctuation, symbols & other typographical marks*. New York: W. W. Norton.

Huddleston, Rodney. 2002. Content clauses and reported speech. In Huddleston & Pullum (eds.), 947–1030.

Huddleston, Rodney & Geoffrey K. Pullum (eds.). 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Hundt, Marianne & Christian Mair. 1999. 'Agile' and 'uptight' genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4(2), 221–42.

Hundt, Marianne & Daniel Schreier. 2013. Introduction: Nothing but a contact language… In Schreier & Hundt (eds.), 1–17.

Jahn, Manfred. 1992. Contextualising represented speech and thought. *Journal of Pragmatics* 17, 347–67.

Johanson, Lars. 2002. *Structural factors in Turkic language contacts*. London: Curzon.

Jullian, Paula M. 2011. Appraising through someone else's words: The evaluative power of quotations in news reports. *Discourse and Society* 22(6), 766–80.

Kachru, Braj B. 1985. Standards, codification, and sociolinguistic realism: The English language in the Outer Circle. In Webster (ed.), 153–76.

Kantey, Mike. 1990. Publishing in South Africa. *Africa Bibliography 1989*, vi–xx. doi:10.1017/S0266673100005183 (20 June 2017).

Kirsten, Johanita. 2016. Grammatikale verandering in Afrikaans van 1911–2010. PhD dissertation, North-West University.

Kruger, Haidee & Bertus van Rooy. 2016. Syntactic and pragmatic transfer effects in reported-speech constructions in three contact varieties of English influenced by Afrikaans. *Language Sciences* 56, 118–31.

Kruger, Haidee & Bertus van Rooy. 2017. Editorial practice and the progressive in Black South African English. *World Englishes* 36(1), 20–41.

Lass, Roger & Susan Wright. 1986. Endogeny vs contact: 'Afrikaans influence' on South African English. *English World-Wide* 7(2), 201–23.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.

Mesthrie, Rajend. 1992. *English in language shift: The history, structure and sociolinguistics of South African Indian English*. Cambridge: Cambridge University Press.

Mesthrie, Rajend. 1996. Imagint excusations: Missionary English in the nineteenth century Cape Colony, South Africa. *World Englishes* 15(2), 139–57.

Mesthrie, Rajend. 2017. South Africa and areal linguistics. In Hickey (ed.), 527–50.

Mougeon, Raymond, Terry Nadasdi & Katherine Rehner. 2005. Contact-induced linguistic innovations on the continuum of language use: The case of French in Ontario. *Bilingualism: Language and Cognition* 8(2), 99–115.

Müller, Dalene. 2003. *Skryf Afrikaans van A tot Z*. Cape Town: Pharos.

Nelson, Gerald, Sean Wallis & Bas Aarts. 2002. *Exploring natural language: Working with the British component of the International Corpus of English*. Amsterdam: John Benjamins.

Quirk, Randolph, Sydney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Sayers, Dave. 2014. The mediated innovation model: A framework for researching media influence in language change. *Journal of Sociolinguistics* 18(2), 185–212.

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

Schreier, Daniel & Marianne Hundt (eds.). 2013. *English as a contact language*. Cambridge: Cambridge University Press.

Short, Mick. 2012. Discourse presentation and speech (and writing, but not thought) summary. *Language and Literature* 21(1), 18–32.

Szmrecsanyi, Benedikt. 2016. About text frequencies in historical linguistics: Disentangling environmental and grammatical change. *Corpus Linguistics and Linguistic Theory* 12, 1153–71.

Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns. 2011. *Taalkommissiekorpus* 1.1. North-West University: CTexT.

Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2), 135–78.

Thomason, Sarah Grey. 2001. *Language contact: An introduction*. Edinburgh: Edinburgh University Press.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley and Los Angeles: University of California Press.

Torres Cacoullos, Rena & James A. Walker. 2009. On the persistence of grammar in discourse formulas: A variationist study of *that*. *Linguistics* 47(1), 1–43.

Trudgill, Peter. 2004. *Dialect contact and new-dialect formation: The inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.

Van Bogaert, Julie & Timothy Colleman. 2013. On the grammaticalization of *('t) schijnt* 'it seems' as an evidential particle in colloquial Belgian Dutch. *Folia Linguistica* 47(2), 481–520.

Van Rooy, Bertus. 2017. South African English. In Filppula, Klemola & Sharma (eds.), 508–30.

Van Rooy, Bertus & Haidee Kruger. 2016. Faktore wat die weglating van die Afrikaanse onderskikker *dat* bepaal. *Tydskrif vir Geesteswetenskappe / Journal of Humanities* 56(1), 102–6.

Wasserman, Ronel. 2014. Modality on trek: Diachronic changes in written South African English across text and context. PhD dissertation, North-West University.

Wasserman, Ronel & Bertus van Rooy. 2014. The development of modals of obligation and necessity in White South African English through contact with Afrikaans. *Journal of English Linguistics* 42(1), 31–50.

Webster, Jonathan J. (ed.). 2015. *Collected works of Braj B. Kachru*, vol. 3. London: Bloomsbury.

Weinreich, Uriel. 1979. *Languages in contact: Findings and problems*. The Hague: Mouton.