

# English Today

<http://journals.cambridge.org/ENG>

Additional services for **English Today**:

Email alerts: [Click here](#)

Subscriptions: [Click here](#)

Commercial reprints: [Click here](#)

Terms of use : [Click here](#)



---

## Corpus linguistic work on Black South African English

Bertus van Rooy

English Today / Volume 29 / Issue 01 / March 2013, pp 10 - 15

DOI: 10.1017/S0266078412000466, Published online: 27 February 2013

**Link to this article:** [http://journals.cambridge.org/abstract\\_S0266078412000466](http://journals.cambridge.org/abstract_S0266078412000466)

### How to cite this article:

Bertus van Rooy (2013). Corpus linguistic work on Black South African English. English Today, 29, pp 10-15 doi:10.1017/S0266078412000466

**Request Permissions :** [Click here](#)

---

# Corpus linguistic work on Black South African English

BERTUS VAN ROOY

An overview of the corpus revolution and new directions in Black English syntax

---

## Corpora of Black South African English

Black South African English (henceforth BSAfE) has received more attention from corpus linguists than any other variety in the country. In the early 2000s, two corpus projects were initiated more or less simultaneously: the corpus of spoken Xhosa-English, compiled by Vivian de Klerk at Rhodes University, and the Tswana Learner English corpus that I compiled. A number of further corpora have also been compiled in the more recent past, many of which are available in the public domain for research purposes.<sup>1</sup>

Before examining the findings of the various studies, it is important to consider why corpus linguistics is a fruitful approach to the study of a New/Outer Circle Variety of English. When the late Sydney Greenbaum (1988) initiated the International Corpus of English (ICE) project, Schmied (1990) took up the challenge to compile an ICE-East Africa. He noted that there were a number of unique challenges to overcome in collecting all the data categories, and also to identify the types of speakers to include in the corpus, especially in terms of their language proficiency and language backgrounds. Since then, a number of ICE-corpora have been completed for non-native varieties. They enable researchers of varieties of English to look for patterns of similarities and differences across various countries, larger sub-continental groupings and, especially, across the native/non-native divide. A complete ICE-corpus has not yet been completed for BSAfE, but a wide enough range of data types is available to allow for a similar range of research possibilities.

The most important advantage of a corpus for research on non-native varieties is to ground statements about the incidence of a particular feature in proper data (see Minow, 2010: 1). While intuitive

judgements are useful and valuable as data for some linguistic research questions, they are much less reliable when it comes to matters of frequency. The reason for this is the human inclination to overestimate the frequency of rare events. Kahneman (2011: 322–33) explains why this happens: the mere fact that an event can be recalled from memory tends to bias the observer to overestimate its frequency, a process he calls the confirmation bias (2011: 81). Thus, grammatical descriptions of BSAfE may potentially claim that a particular linguistic form (e.g. the use of the unmarked verb form in past time contexts) is ‘characteristic’ of BSAfE, but corpus evidence allows Minow (2010: 111) to conclude that this phenomenon occurs in only 15% of possible past tense contexts, whereas the standard form of past tense marking is present in 85% of the possible contexts.

The other advantages are related to the advantages of corpus linguistics in general. By studying



*BERTUS VAN ROOY is professor in English Language Studies at the Vaal Triangle Campus of the North-West University in Vanderbijlpark, South Africa. He is a past president of the International Association for World Englishes. His current*

*interests include the grammatical features of non-native varieties of English, and the development of South African varieties of English in the nineteenth and twentieth centuries. He works on the compilation of a number of synchronic and diachronic corpora of varieties of South African English. Email: Bertus.VanRooy@nwu.ac.za*

corpora of naturally occurring data, the linguist is able to understand the functions of language in use much better, particularly since many instances of a particular linguistic feature can be retrieved to enable a comprehensive overview of the functional possibilities. Moreover, particularly within the more inductively-oriented corpus-driven approach, new and unexpected features of a language can be discovered, patterns that occur too far apart for the human analyst to spot the correspondences in active memory. A look at much of the published corpus work on BSAfE shows that this opportunity is not yet fully capitalised on. The ‘comparative fallacy’ of determining what is different from some standard variety remains a methodological characteristic of current research (see Van Rooy, 2008).

Given such advantages, this article sets out to synthesise the insights that have been gained from corpus linguistic investigations into the grammatical features of BSAfE. A number of features related to the noun phrase and the verb phrase have been identified by researchers, and supported by corpus data. However, corpus data have also indicated that some features are either quite rare, or clearly not stabilised features, but rather transitional phenomena that disappear when language proficiency increases. The established features will be presented first, before looking at those linguistic features that are shown not to be stable features of BSAfE.

### Established features

The *progressive aspect* has a long history of scholarly attention in BSAfE and many other New Englishes. The standard account has focused on the ‘extension of the progressive to stative verbs’ (Gough, 1996: 61; Mesthrie, 2008: 489), which corpus analysis confirms (De Klerk, 2006: 140). Siebers (2012: 149) and Van Rooy (2006) show that the progressive is still by far the most frequent with activity verbs, so the ‘extension’ does not alter the core possibilities of the construction. Furthermore, Minow (2010: 144) points out that the frequency of the progressive is inversely related to the proficiency levels of the speakers in her corpus: the more proficient a speaker, the less frequent the progressive is. These two observations may lead us to suspect that the extension of the progressive is a learner language phenomenon, which is bound to disappear as speakers adjust their grammatical usage with increased proficiency.

A functional approach, which goes beyond observing the presence of a particular feature, paints a different picture, though. Van Rooy

(2006) analyses a sample of 100 progressives from the TLE, and concludes that the underlying semantics of the construction is consistently different from the native speaker prototype of a dynamic event with limited duration. Rather, extended duration is profiled by the majority of progressive usages in the data. Once extended, rather than limited, duration is profiled, the construction is equally compatible with dynamic and stative predicates. Thus, Van Rooy (2006) argues against the interpretation that the progressive is ‘extended to stative verbs’. Siebers (2012: 150–3) likewise indicates that a large number of instances of the progressive in her corpus are compatible with extended, rather than limited, duration. In ongoing work that I am doing on the semantics of the progressive when used with stative verbs, it emerges that about half of all the examples are used with the semantics of extended duration (out of more than 500 drawn from all the corpora I had access to, including the XE1, VW, TLE and my own corpora). By contrast, about a quarter of the examples show the standard usage of states with limited duration, but just about the same number of instances denote states with unlimited duration, i.e. permanent states.<sup>2</sup> Thus, corpus linguistic research on the progressive refutes the simplistic view that the progressive is merely extended to stative contexts, as in the following example (from Siebers, 2012: 151):

- (1) So the way **we are thinking**, we’re thinking like like like whites we mustn’t have a – we mustn’t have this part of the body out on the open, you must cover it.

Siebers (2012: 151) points out that, in context, a general attitude is denoted rather than a temporary state, which is clearly a meaning that is not ascribed to the progressive in standard grammars. However, the temporal meaning in example (2), taken from Van Rooy (2006: 57), which is fully consistent with example (1), points to extended duration as well, even if the predicate denotes a dynamic event:

- (2) In prison is where they graduate in their criminal activities because all the criminals are infested there they become more wicked and dangerous because the society **is treating** them like outcasts or the worst sinners.

A second set of verb phrase features relate to the *expression of modality*. Older accounts have identified a number of unique expressions, especially the occurrence of the form ‘can be able to’ (Gough, 1996: 63). De Klerk (2006: 150) confirms

the presence of this form in her XE1 corpus. Using her data alongside the TLE, Van Rooy (2011) takes a closer look at the semantics of ‘can be able to’ in examples such as the following (from Van Rooy, 2011: 200):

- (3) People become sick for a long time and this caused Aids because this disease will kill all your immune system and the body **can’t be able to** defend itself against other diseases.

Semantically, ‘can’ conveys the sense of extrinsic possibility in combination with the expression ‘be able to’, which is a semi-modal expression synonymous with the intrinsic ability sense of ‘can’. Thus, unlike accounts that suggest redundancy or hyperclarity in the use of ‘can’ with ‘be able to’, a functional analysis of corpus data shows that the expression is not, logically speaking, problematic, but merely not conventionalised in present-day native varieties of English, unlike in the Early Modern English Period, when it was used more widely, and even made it into the King James Bible translation (Crystal, 2008).

As far as the noun phrase is concerned, corpus research on BSAfE also yields confirmation of some existing research, but elaborates on existing insights at the same time. The *resumptive pronoun* (or left dislocation) strategy has been reported for BSAfE since the earliest descriptions. Mesthrie (1997), in a pre-corpus study that draws on an extensive ‘corpus’ of sociolinguistic interviews, offers a substantial body of insight into the pragmatic functions and syntactic environments in which various topicalisation phenomena occur in spoken BSAfE. He notes that the construction is much more frequent in BSAfE than in other varieties, and while it shares some of the functions with other varieties, such as the reintroduction of given information, it sometimes functions in contexts where no prominent pragmatic function is prevalent. Two syntactic environments, with partitive-of-constructions and relative clauses, are identified, as well as the high frequency of the combination ‘people they’, as illustrated by the following example from Mesthrie (1997):

- (4) The people, they got nothing to eat.

Corpus studies by De Klerk (2006: 140), Minow (2010: 193) and Siebers (2012: 204–6) all confirm Mesthrie’s account, using three different spoken corpora. Botha (2012: 176–88) likewise confirms the existing accounts, but what is different is that her data are drawn from the written student work in the TLE. She finds that a much bigger proportion

of instances can be attributed to referent tracking in especially relative clause and partitive constructions (2012: 187–8), and also explains that the high frequency of the ‘people they’ combination should not be overinterpreted: ‘people’ is simply by far the most frequent lexical noun in the entire TLE, and it is thus entirely expected that the most frequent noun should be the one that enters into the most frequent combination with a resumptive pronoun as well (2012: 176).

An area in which pre-corpus research is considerably less precise is the *use of articles*. There are three logically possible ways in which BSAfE articles may differ from native varieties, all of which have been reported in the literature: articles are omitted (Gough, 1996: 61), or inserted (Mesthrie, 2008: 496), or substituted for each other. Greenbaum and Mbali (2002: 241–3) mention all three possibilities in the same article. De Klerk’s corpus analysis confirms the use of articles with non-count nouns (2006: 146), and she furthermore identifies a range of usage that will be unacceptable in native varieties, thereby largely confirming pre-corpus accounts of the unsystematic use of articles.

Minow (2010) and Siebers (2012) add to our understanding by showing that the omission and substitution of articles, compared to native norms, occurs with a rather low frequency – between them they find article occurrence rates of between 87% and 97%, depending on corpus and whether indefinite or definite. The insertion of articles in positions where no overt article would be used in native varieties is the most frequent deviation from the norm, and both find that native-like usage increases with proficiency levels. Siebers (2012: 120–1) identifies one idiomatic expression that is characteristic of BSAfE usage, the form ‘kind of a NOUN’, exemplified by the following example:

- (5) Because if you go and look for a job, you must be doing **some kind of a research** in order to know what kind of company or what kind of institution is that (M1).

Botha (2012) identifies a few more systematic patterns of different usage in BSAfE. Firstly, BSAfE seems to use articles more widely than native varieties before human institutions: besides ‘go to the bank/shop’, as in native varieties, BSAfE speakers also prefer the formulation ‘go to the school/university/hospital/jail’ (Botha, 2012: 257). Secondly, indefinite articles are used more widely in noun phrases with non-particular interpretations where such nouns are

N	Concordance	Be	Word #	File
1	government the people must look at that how they they can be able to assist the government in ensuring that black		526	#dpc234.txt
2	money like so that we have a capital so that the bank can be able to assist us to borrow the money to start the		1,152	#dpc129.txt
3	BY THEIR HANDS SEE BY EYES SO THAT THEY CAN BE ABLE TO BE OPEN MINDED AND DISCOVER		152	TSNO1427.txt
4	that we need to rewrite our history and people who can be able to bring new cultural developments for our society		914	#dpc230.txt
5	HIV/AIDS in Africa people could get work, so that they can be able to buy food and everything they run in short of.		292	TSNO1079.txt
6	their bodies only to get that little money so that they can be able to buy something to eat -And some students the		110	TSNO1131.txt
7	in this way in this way in this way the the newspaper can be able to come up with the right sources right uh place		737	#dpc251.txt
8	you have to have an tertiary qualification so that you can be able to compete so i think so it's a better idea just like		1,530	#dpc073.txt
9	who've got the the potential in sport or the people who can be able to do their best in sport i'm no uh uh my man		490	#dpc066.txt
10	able to ensure that it after matric that person passes can be able to ensure that that person is going through a		990	#dpc230.txt
11	we need a well balanced secondary education that can be able to ensure that it develops an individual that		967	#dpc230.txt
12	that visitors when they come around grahamstown we can be able to give them accommodation at the moment there		1,935	#dpc024.txt
13	[it's true.] [Only in] winter when it's little bit quiet, you can be able to go away. Mh. [mh.] [Yes.] Mh. (...) Ja uhm (1.9		2,610	Wellington Gardens2.txt
14	, you need to do a lot of practical work so that you can be able to help patients immediately they arrive at the		223	TSNO1460.txt
15	to win what? we want to win the public so that they can be able to mean we can be able to serve them okay? so that		828	#dpc005.txt
16	an internal source of data or of information that you can be able to obtain whereas for instance in a business in a		1,318	#dpc011.txt
17	source of data that means that information that you can be able to obtain it inside the business that it information		532	#dpc011.txt
18	. And people should all be empowered so that people can be able to participate in the positive growth of the		644	TSNO1516.txt
19	like before. When players play in their countries, they can be able to plough back to their communities, they can be		310	TSNO1509.txt
20	involved with a man owning a posh car one who can be able to provide hard cash that is someone who is a		258	TSNO1237.txt
21	language where everybody who can be able to read can be able to read it then second reproduction reproduction		1,743	#dpc011.txt
22	it has been processed into something whereby you can be able to read and understand it whereas when it was in a		1,169	#dpc011.txt
23	understand? because n- now we can be able to read it and understand it whereas before it was		424	#dpc011.txt
24	format that means everybody who is able to read there can be able to read that because it has been processed into		1,721	#dpc011.txt
25	processed into the language where everybody who can be able to read can be able to read it then second		1,738	#dpc011.txt
26	that have been processed the information so that we can be able to read it- are you happy? now that means you are		1,015	#dpc011.txt
27	data now it has been processed in a text whereby you can be able to read it and understand for instance u-grocotts		1,579	#dpc011.txt
28	been processed whereby each and every individual can be able to read that information and understand it then you		1,142	#dpc011.txt

Figure 1. Screenshot of WordSmith concordances of 'can be able to'

conventionally construed as uncountable; thus as well as the native-like 'to have a better life', BSAfE speakers also select 'to have a time to relax' (Botha, 2012: 265). Lastly, Botha (2012: 253–4) also identifies the use of the definite articles with ascriptive nominals in BSAfE, for instance:

(6) It's the question of loyalty.

Botha (2012: 278) concurs with Siebers (2012: 131) that the basic underlying system of article usage is the same in BSAfE as in native varieties. To the one idiomatic alternative in BSAfE that Siebers identifies, Botha adds three more. In her conclusion, Botha (2012: 78) notes that the small differences between BSAfE and native varieties are due to alternative constructions that are conventionalised in BSAfE. These alternatives develop in the leaky edges where native speaker grammar is also less regular.

The distinctive use of *quantifiers* in BSAfE has been noted in pre-corpus accounts as well. Gough (1996: 62–3) lists a number of features that can be grouped together under the broad umbrella of quantification (the use of 'too much', 'very' and 'very much', 'some few', the 'other...other' construction, 'the most thing' and 'X's first time'), whereas Mesthrie (2008: 495–6) includes the 'other...other' construction in his discussion of subordination and coordination. De Klerk

(2006: 143) identifies instances of almost all the constructions listed by Gough in her XE1 corpus, and thus confirms their presence in the data.

Botha (2012: 309–14) points out that most of the exceptional usages noted by De Klerk and pre-corpus accounts of BSAfE are due to systematic extensions of other uses of quantifiers that do resemble native varieties more closely. The construction 'most of the NOUNS' occurs almost ten times as often in the TLE compared to the native speaker control corpus, and while this construction is perfectly acceptable in native speaker writing, it is quite rare. A blended construction, not used by native speakers, is the construction 'most of NOUNS', which appears to take features of the 'most of the NOUNS' and 'most NOUNS' constructions. In fact, Botha (2012: 312) concurs with Mesthrie's (2006: 139–40) undeletion account where the 'of' is inserted, rather than an article omitted, in the construction. Thus, the following example (from Botha, 2012: 312) shows that the 'of' will probably be omitted in native varieties, since the context does not require the quantified head noun to be definite:

(7) **Most of club owners** complain about the standard of soccer in our country.

Botha (2012: 302–7) also offers an account of the 'some few' construction. She firstly points

out that the form ‘some’ is used more extensively in BSAfE than in native varieties, due to its function as overt marker of indefiniteness in contexts where the indefinite article is not typically found (like indefinite plurals), as illustrated by the following example (2012: 305):

(8) So you cannot play soccer each and every week and after that you’ve been paid **some peanuts**.

She proceeds to show that in contexts where ‘some’ combines with ‘few’, ‘some’ is used in its determinative role, and not its quantifier role. Hence, no conflict should arise, within the BSAfE system, in the combination ‘some few’. Native varieties would permit ‘a few + NOUNS’ in this context, exceptionally combining the indefinite article with a plural noun. The two instances of the expression in the TLE can both be paraphrased as ‘a few’ in native varieties: ‘After **some few** days’ and ‘In **some few** years ago’. Thus, the ‘some few’ construction is not so much a case of unusual quantification, but follows from the functional extension of ‘some’ to complement ‘a/an’ as marker of indefiniteness.

One morphological feature of the noun that has received attention in research on BSAfE is the distinction between mass and count nouns. In pre-corpus overviews (Mesthrie, 2008: 497; Gough, 1996: 61), the claim is made that non-count nouns are used as if they are count nouns. De Klerk (2006: 146) reports data from her XE1 corpus that show that certain mass nouns are used with plural suffixes (*homeworks*, *equipments*, *moneys* and *advices*). However, apart from ‘home-works’, which occurs in the plural 13 times while the singular is used only 7 times, all other items are used in the singular form the majority of the time. Siebers (2012: 134) furthermore finds that such usage occurs mainly with the least proficient speakers in the XE2 corpus.

Botha (2012: 318) finds that the form ‘equipments’ is the most frequent pluralised mass noun in the TLE corpus, and it is the only form that has more plural usages than singular usages. She argues, however, that it is not so much that the contrast between mass and count nouns is violated, but rather that a number of nouns are re-analysed as count nouns, where native varieties of English typically use such nouns as mass nouns. The nouns that are most likely to undergo such reanalysis are nouns that refer to countable objects, such as ‘equipment’ or ‘furniture’. Abstract mass nouns, such as ‘information’ or ‘advice’, are the other type that undergo such reanalysis. These are

variously construable as unbounded or bounded objects, as is shown by the contrast between German and English as far as ‘information’ is concerned: German ‘die Information’ can be pluralised to ‘die Informationen’.

### Non-features

In the absence of a systematic data base of language, from a range of speakers, it is difficult to determine which observations in a small sample constitute a ‘pattern’ of BSAfE usage, and which are individual errors or transitional phenomena. Both Gough (1996) and Buthelezi (1995) concede that they mainly drew on student writing as the source of evidence for their discussion of BSAfE features, while other writers, such as Greenbaum and Mbali (2002), did not intend to identify properties of a variety, but specifically intended to pinpoint recurring errors in student writing. The availability of bigger corpora enables researchers to quantify the extent to which a particular feature occurs, and if the frequency is negligible in the face of an overwhelming trend to select the native-like variant (or some other variant), then there is little reason to regard such a feature as an established feature of BSAfE. Minow (2010: 3) argues that a stable feature of BSAfE is one that is used to some degree by all speakers, regardless of proficiency level, while a feature that is restricted to the least proficient speakers in the data should not be regarded as a stable feature. Siebers (2012: 187) refers to the view of Romaine (2005: 427) that a feature which appears 80–90% of the time should be regarded as acquired by a speaker, and the remainder should be regarded as performance errors.

Based on the analyses of various researchers, the following features are not regarded as stable features of BSAfE, but are occasional performance errors instead:

- The omission of the suffix *-ly* on adverbs such as ‘quickly’ (De Klerk, 2006: 153);
- Subject–verb concord errors (Siebers, 2012: 187);
- Overgeneralisation of the past tense suffix *-ed* on irregular verbs (De Klerk, 2006: 145);
- Use of ‘does + VERB’ to express present tense (De Klerk, 2006: 153);
- The neutralisation of the contrast between *themself/themselves* (De Klerk, 2006: 154);
- Conflation of pronoun gender by using ‘he’ and ‘she’ for the same referent (from my own unpublished analyses, incidence below 5% even in the writing of school pupils, and below 2% in the writing of adults).

## Conclusion

Corpus linguistic research into BSAfE has enriched our understanding of the grammatical features of this variety in three ways. On a purely quantitative level, it has provided support for claims made in research based on less extensive data sets or less formalised analyses, to put the discussion of BSAfE on a surer footing. It has also shown that some features that are attributed to the grammar of BSAfE should rather be regarded as performance errors, since they represent a small minority of variants that differ from the majority variants in the data. Finally, for at least some features, such as the use of the progressive aspect (Van Rooy, 2006; Siebers, 2012), ‘can be able to’ as modal expression (Van Rooy, 2011), and articles and quantifiers (Botha, 2012), functionally oriented corpus analyses have added to our understanding of features that were previously noticed mainly for their deviance, without a thorough grasp of ways in which they are consistent with the grammar of BSAfE.

Corpus linguistic research shows us that, at this point, the grammar of BSAfE is to a large extent similar to that of native varieties of English. Differences mainly reside in a small number of new grammatical constructions of fairly restricted scope. In the leaky corners of grammar, a few constructions that are unique to BSAfE have attained (or are close to attaining) stability, and should therefore be regarded as characteristic features of this variety (even if they are shared by other New Varieties of English elsewhere on the African continent or further afield). ■

## Notes

1 The most important corpora that this paper will refer to are the following: Xhosa-English Conversation (XE1), of approximately 550,000 words – De Klerk (2006), Tswana Learner English student writing (TLE) of approximately 200,000 words – Van Rooy (2006), Volkswagen conversation corpus of approximately 85,000 words (VW), directed by Christiane Meierkord – (Minow, 2010), Siebers’ (2012) PhD corpus of Xhosa-English interviews (XE2) of approximately 100,000 words, and then as yet undistributed corpora in my own collection covering the written registers of fiction, journalism, and academic writing, each about 50,000 words in size.

2 Van Rooy (2012).

## References

Botha, Y. V. 2012. ‘Specification in the English nominal group with reference to student writing.’

- Unpublished PhD dissertation, North-West University.
- Buthelezi, Q. 1995. ‘South African Black English.’ In R. Mesthrie (ed.), *Language and Social History: Studies in South African Sociolinguistics*. Cape Town: David Philip, pp. 242–50.
- Crystal, D. 2008. ‘On “can be able to”.’ Online at <<http://david-crystal.blogspot.com/2008/08/on-can-be-able-to.html>> (Accessed January 19, 2010).
- De Klerk, V. 2006. *Corpus Linguistics and World Englishes*. London: Continuum.
- Gough, D. 1996. ‘Black English in South Africa.’ In V. de Klerk (ed.), *Focus on South Africa*. Amsterdam: John Benjamins, pp. 53–77.
- Greenbaum, S. 1988. ‘Proposal for an international computerized corpus of English.’ *World Englishes* 7, 3–15.
- Greenbaum, L. & Mbali, C. 2002. ‘An analysis of language problems identified in writing by low-achieving first-year students, with some suggestions for remediation.’ *Southern African Linguistics and Applied Language Studies* 20, 233–44.
- Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Mesthrie, R. 1997. ‘A sociolinguistic study of topicalisation phenomena in South African Black English.’ In E. Schneider (ed.), *Englishes around the World: studies in honour of Manfred Görlach, Volume 2*. Amsterdam: Benjamins, pp. 119–141.
- . 2006. ‘Anti-deletions in an L2 grammar: a study of Black South African English mesolect.’ *English World-Wide* 27, 111–45.
- . 2008. ‘Black South African English: morphology and syntax.’ In R. Mesthrie (ed.), *Varieties of English 4: Africa, South and South East Asia*. Berlin: Mouton de Gruyter, pp. 488–500.
- Minow, V. 2010. *Variation in the Grammar of Black South African English*. Frankfurt: Peter Lang.
- Romaine, S. 2005. ‘Variation.’ In C. J. Doughty & M. H. Long (eds), *The Handbook of Second Language Acquisition*. Malden/Oxford: Blackwell, 409–35.
- Schmied, J. 1990. ‘Corpus linguistics and non-native varieties of English.’ *World Englishes* 9, 255–68.
- Siebers, L. 2012. *Morphosyntax in Black South African English: a case study of Xhosa-English*. Tübingen: Gunther Narr.
- Van Rooy, B. 2006. ‘The extension of the progressive aspect in Black South African English.’ *World Englishes* 25, 37–64.
- . 2008. ‘An alternative analysis of tense and aspect in Black South African English.’ *World Englishes* 27, 335–58.
- . 2011. ‘A principled distinction between error and conventionalised innovation in African Englishes.’ In J. Mukherjee & M. Hundt (eds), *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam: Benjamins, pp. 191–209.
- . 2012. ‘Overextending the progressive aspect in the Outer Circle?’ Paper presented at IAWWE 2012.