This article analyzes the sociophonetic dynamics of the acquisition of a prestige variety of English in post-apartheid South Africa. New economic and educational opportunities have seen the rapid growth of a Black middle class, whose younger members have increasingly come into contact with what used to be ‘(L1) White South African English’. The article documents the differences among younger Black peoples’ English from the traditional L2 of their parents’ generation, via acoustic analyses of two sets of variables: schwa as a full vowel, and neutralizations of vowel length. At the same time there is considerable internal difference among younger Black speakers in relation to social class and gender. The finding that young Black women are in the lead in acquiring the prestige variety is triangulated with a perception experiment and citations from commentaries in the press and young peoples’ own attitudes.

Keywords: Black South African English, sociophonetics, schwa, vowel-length variation, race, class, and gender

1. Introduction. This article is concerned with the extent to which linguistics (and especially sociolinguistics) can illuminate at a microlevel the large-scale changes that continue to take place in societies outside the Western industrial and postindustrial mainstream. It focuses on the changing sociolinguistic dynamics in post-apartheid South Africa, where there are now robust generational differences in the new society, new social-class considerations, and salient gender differences within the younger generation. These considerations form the core of a larger project on English in a multilingual society, focusing on South Africans of different backgrounds in five cities. This article concentrates on the interplay between L2 BLACK SOUTH AFRICAN ENGLISH (henceforth BSAE) and the L1 English formerly associated with the White middle class. There is differential access to this prestige L1 variety in the new post-apartheid society, based largely on type of schooling, which has made possible varied positionings among individuals that reveal their personal background and their quest for new, meaningful social roles—to rephrase Le Page and Tabouret-Keller’s formulation (1985:14) somewhat. This article demonstrates the class and gender effects among young Black students that result in the use of crossover varieties of English, showing substrate erasure, with young middle-class Black women leading the way. The term ‘crossover varieties’ refers to a range of new varieties among younger people that eschew the main features of traditional BSAE and approximate to varying extents the use of the main features of what used to be thought of as WHITE SOUTH AFRICAN ENGLISH (henceforth WSAE). The term is thus

* Acknowledgments to people for a paper long in gestation are many. The principal funders are the National Research Foundation via a SARCHI chair (grant no. 64805) held by the author, with secondary funding from the University of Cape Town’s research committee. The editors of Language and anonymous referees provided acutely perceptive comments and suggestions that improved the original submission immensely. Earlier versions of the paper benefited from comments at the Linguistics Society of Southern Africa (Witwatersrand and Potchefstroom), NAW (Chicago), and at several university departments (UCT, UKZN, Queen Mary, Michigan, Duisburg-Essen). Erik Thomas gave generous advice on setting up the acoustic aspects of analysis, as did the researchers at the University of Pennsylvania who developed the software used in this analysis. Thanks are also due to my co-researchers at UCT: Alida Chevalier, Tracey Toefy, and Yolandi Klein.

1 The concept of a crossover variety applies to middle-class younger Colored and Indian speakers too, especially those who have been to the model-C schools described in this study (see below).
very different from the more familiar sociolinguistic term ‘crossing’ (Rampton 1995), which refers to a temporary, marked, stylized adoption of facets of a variety (usually not a prestige or dominant variety) traditionally associated with ‘other’ groups of speakers. Crossover varieties, by contrast, approximate to the prestige variety used in the educational systems. The deracialization of the title thus alludes to the fact that the prestige variety of English in the country, as favored in the spoken-cum-visual media, is no longer associated with one group alone. For many of the Black middle-class students of this study, the main dichotomy between Englishes of South Africa is not over race, but over an L2 variety marked by heavy substrate influence from Bantu languages versus their own (crossover) English. Substrate erasure in the title refers to the avoidance of all (or nearly all) markers of African language influence in crossover varieties and, by implication, avoidance of traditional forms of L2 English.

This article has three parts. The first is sociophonetic, showing the differential use of two sets of variables by Black speakers of different sexes and schooling backgrounds. These variables are (i) schwa, and secondarily (ii) vowel length. The first of these is analyzed in detail using the techniques of acoustically based sociophonetics. The variables were selected to represent the range of fine-grained phonetic possibilities between the two erstwhile poles: older L2 BSAE and middle-class L1 WSAE. As such they may be thought of as Crossover variables, which measure the degree of adherence to the more traditional variety as against crossover speech showing varying degrees of substrate erasure. This part of the article demonstrates that there are class effects correlating with schooling for some subsets of the schwa variable and for vowel length. However, equally strong (and perhaps more so) are gender considerations, which override the effects of schooling in some salient subsets of the schwa and vowel-length variables. The second part of the article strengthens the gender findings by summarizing a perception test involving accent identification of speakers, undertaken by the author and two colleagues. The concluding part of the article completes the triangulation with an analysis of attitudes expressed in the media and in some of the interviews themselves, showing the generalizability of the findings nationally. Overall, it is suggested that women do show a greater investment in new forms of status, whereas men’s investment in this regard is mediated by older considerations of solidarity and perhaps language loyalty. First, I provide the background to the recent sociopolitical and linguistic changes in South Africa.

2. Sociopolitical and linguistic change in South Africa. South Africa’s history of racial segregation, first in a de facto colonial form, and rigidified legally and in practice under apartheid (1948–1994), provides almost laboratory-like conditions for sociolinguistics. The explicit classification and labeling of people of different ‘races’ or ‘population groups’ were used to control and manipulate social groups and networks, with individuals sometimes being forcibly put into one or another box. The four main designated groups (White, Black, Colored, and Indian) were then separated from birth (segregated hospitals) and through early childhood (segregated living areas), education (segregated schools, colleges, and universities), religion (separate churches), work (job reservation, separated canteens), sexual relations and marriage (the ‘Immorality Act’), and finally death (segregated cemeteries). Contact across the ethnicities fostered by this history was largely limited to transactional and asymmetrical work situations, rather

speaker’ is sometimes used to avoid the more cumbersome ‘speaker of a crossover variety’. It should not be taken to imply that such a speaker necessarily switched from BSAE to a WSAE-like variety in his/her lifetime. Many speakers of crossover varieties had early exposure to WSAE from crèches.
than extended social interaction. The consequences for language learning are obvious, and as far as English is concerned, one outcome was the growth and maintenance of five distinct Englishes, each more or less associated with one group rather than the others: it is still customary to speak of Black, White, Colored, and Indian English in the country (and to further differentiate Afrikaans-based L2 English from L1 English among Whites).  

The collapse of apartheid in the early 1990s—spurred by radical internal opposition, especially from Black youth, as well as international political pressure—led to the first democratic elections of 1994 and the installation of a government dedicated to overcoming apartheid imbalances and to improving conditions for those previously oppressed.  

Perhaps the most striking change in the last twenty years is the transfer of power to a Black majority in Parliament, the state media, and the economic sector, via a program of Black economic empowerment. The economists Seekings and Nattrass (2005) pointed to the resulting rapid changes in social class: class was emerging out of race. Within each ‘population group’, class differences were being magnified to the extent that it was possible to imagine class as a major competing parameter to the old arrangement of ethnicity. The greatest changes have occurred within the Black population, with Seekings and Nattrass (2005:309) showing a rise of 78% in middle-class Black households between 1991 and 1996. They later comment on the continuing deracialization of the South African upper (or upper-middle) classes, with Blacks rising from proportionally low levels in the 1980s to just under half by 2008 (Seekings & Nattrass 2015:115). In terms of political and economic power, including consumption patterns, this is now the major group (Sunday Times, 28 April 2013).

Within the Black population, young people of the new or emerging middle class are arguably the greatest beneficiaries of the political and socioeconomic changes, being afforded a quality education once reserved for Whites. That other young Black people—a majority—are still excluded from these benefits can be seen from the vociferous protests by university students in 2015 and 2016 against the country’s stark economic polarities. These events, which postdate the research and writing of the main draft of this article, are nevertheless relevant in my arguments in §3 regarding the stability and counter-prestige of traditional BSAE. In the 1990s (and even slightly before), young Black children formed a minority at elite, highly expensive private schools, and at the then ‘model-C’ White government schools, which charge a moderately high fee. The latter denotes a particular option within a reformist model offered to White schools before the collapse of apartheid. Model-C schools could receive a government subsidy (unlike the private schools) if they charged moderate fees and opened their doors to small numbers of Black pupils. For convenience, in the rest of this article I use the label ‘model-C’ for both types of prestigious schools—that is, private schools and what are now ex-model-C schools (and the abbreviation C in this article will always stand for this strand of relatively advantaged and multiracial schooling).

2 In South Africa, ‘Colored’ prototypically denotes people of multiple ancestries, one important strand of which includes the most indigenous KhoeSan cultures. The term is not synonymous with ‘Black’, which prototypically refers to speakers of Bantu languages. The term ‘black’ (in lower case), which denotes a broader grouping of Blacks, Coloreds, and Indians, while still valid politically is not needed in this article. It is acknowledged that all race terms are contested in South Africa, especially if used to stigmatize or marginalize people.

3 This bald summary of necessity glosses over certain complexities. White academics, social workers, and clergy often played a role that went contrary to what the apartheid ideologues intended, and the post-apartheid political record of some politicians leaves much room for skepticism about political claims and slogans, as evident in the current student protests at universities and outside Parliament.
From a sociolinguistic perspective, the change in admissions policy meant the rapid integration of initially small numbers of Black students within the prevailing Western ethos of the schools, and into peer groups in which White students were likely to predominate. For a critique of the class bias of this policy, see Soudien 2007. The 1990s thus saw the rapid emergence of a group of Black schoolchildren showing a degree of westernization and socialization into White society that was hard to envisage for most of the twentieth century. A report from the University of Cape Town Unilever Institute of Strategic Marketing (2012) indicated that 75% of the children of the new Black middle class (then numbering four million) were attending model-C schools. With this education comes exposure to a variety of English previously associated only with middle-class, English-speaking Whites, the elites of the old society. These changes were astonishing to older people of all backgrounds, and young Black people of more modest socioeconomic backgrounds were not slow in noting the differences between their own interests, values, language preferences, and accents as against those of the new middle classes. The term coconut was popularized by young people from the mid-1990s onward to characterize this sense of difference between Black children of a more traditional township background and those of the new elites. Rather like the dynamics of the term oreo in North America, coconuts are alleged to be ‘dark on the outside, white on the inside’. We might take this as an example of ‘re-racing’, discussed in a US context by Sweetland (2002) and Fought (2006:32–33). This very salient term among young people forms the title and theme of an award-winning novel Coconut (discussed further in §7) by Kopano Matlwa (2004), herself a young medical student at the time. The term forms the cover image of a humorous book of essays on modern social relations, Some of my best friends are White (Ngcobo 2007). It is known among young Indian and Colored South Africans, mainly among the middle classes. While some people use and/or accept the term in jest, others use it to offend and suggest a sense of betrayal of ‘traditional’ behavior patterns (see Bangeni & Kapp 2007, McKinney 2007, Rudwick 2008). A blog by Ndumiso Ngcobo (2008), the author cited above, shows the class attitudes and characteristics that are at stake:

What is a coconut? I think that it is sufficient to explain the term ‘coconut’ as a Black person with white values. Fair enough? OK, if you want to be pedantic […] I suppose you could throw in white mannerisms, speech, behaviour and perhaps even dress sense.
1. Coconuts speak English most of the time and they do so with an accent that is ‘White’.
2. Coconuts seem to be very comfortable in White people’s company and may have many White friends.
3. Coconuts, in their sickening coconuttiness, will often express views that seem at loggerheads with African values.

The prominence of Ngcobo’s first criterion has spawned a complaint tradition (see Milroy & Milroy 1999) in the media showing deep linguistic anxieties and polarities among the South African citizenry. A brief overview of comments in the newspapers in the last twenty years will show how South Africans are divided over the accents and behavioral changes of the new young elites. These attitudes are crucial to evaluating the sociolinguistic data in the main part of this article. Russell Kaschula, a professor of African languages, calls South Africa ‘a society obsessed with the acquisition of a “model-C” accent’ and one that disvalues other languages, styles, and accents (Cape
Some politicians link this directly to imminent language shift, as Aaron Motsoaledi, MEC for Education in Limpopo province, is cited as saying:

‘Many parents are stupidly excited about how well their children speak English. They proudly display them at funerals to read obituaries,’ he said … ‘When children from former model-C schools visit the rural villages where their parents were born you still find parents who are proud to tell the elders: “No you can’t talk to my child in Tswana, she doesn’t hear you. She only speaks English.”’ (Sunday Times, 9 November 2008, report by Prega Govender)

The regular surfacing of debate about language and accents is a symptom of deeper political anxieties, positions, and crises, arising in the interregnum between the passing of an old order and the successful establishment of a new one, as theorized by Antonio Gramsci (1971). One such crisis is the growing and almost unbridgeable gap between the very rich and very poor: ‘If you do not have the language skills to access the huge store of information available in English, then you are in a prison’ (J. Skiraya, letter to Cape Argus, 16 October 2008). Another is the feeling of cultural betrayal, that the new educated, wealthy youngsters are becoming ‘un-African’ compared to the less privileged, who ‘continue the vibrant tradition of a struggle culture that was at its height in the 1980s’ (S. Memela, letter to Sunday Times, 14 December 2008). More trenchantly, ‘this African elite with their split personalities and blurred vision of everything cannot be agents of development in their society’ (Mosibudi Mangena, Azapo leader and former Minister of Science and Technology, reported in the Cape Argus, 29 August 2006). Whereas the commentators cited thus far express their fears in strongly worded rhetoric, the targets of their criticism, who are comfortable with their African heritage and political ancestry as well as their Western-oriented academic education and tastes, often use humor or irony to puncture the political certitude of their critics (as best illustrated by the Ngcobo blog above). Most young people interviewed in this study did not show any of the alleged betrayals. Almost all middle-class students interviewed remain multilingual, with loyalty to one or more African languages, and affection for their grandparents’ generation and their language use (see also Morreira 2012). While language shift does not generally feature in their life histories, a consequence of their education is a shift in language dominance, with English becoming far more prominent in their lives compared to previous generations. This could lay the seeds of future middle-class shift among Black people living in elite multiracial suburbs, despite avowals among the interviewees that they would want their children to be proficient in English and an African language (see also de Klerk 2000).

The above discussion should make it clear that class cleavages are central in understanding the range of varieties between older BSAE and newer crossover varieties. The other salient consideration is gender, which I introduce here (as a prelude to the sociophonetic hypotheses of this article) before returning to it in detail in §9 (on ‘supportive media evidence’). In the early days of societal and university transformation of the 1990s, when university staff expressed surprise to me at the first Black students who sounded White, they were invariably referring to female speakers. This is paralleled by reports in the newspapers that explicitly mention women’s English, and the implied generation and gender gap it symbolizes. Thus Lindiwe Mazibuko ‘offers the soft cadences of a private school education’, whereas her political rival Julius Malema speaks

---

6 *Hear* is a common form in Africa for ‘understand’. It is probably relevant that the sex of the child of this anecdote is female.

7 The phonetic comparisons in this article thus involve varieties that are clearly L2, clearly L1, or in between. The uniting factor is not L1 or L2 status, but fluency in English amid a multilingual repertoire.

This sociological background, together with my own phonetic observations, led to this study, which sets out to demonstrate that (i) class differences are salient in young Black peoples’ English, and (ii) gender is equally and possibly more salient.

3. Older BSAE, the new generation, and the sample.

3.1. The sample and subgroupings. This study draws upon seventy-seven young people interviewed under conditions roughly in accord with what I call the Labovian interview, following guidelines devised for variationist sociolinguistics by Labov (1972, 1984). They were all university students at the time of the interviews, between the ages of eighteen and twenty-six. The majority (fifty-three) or core group of the study were chosen to represent the current generation of educated young Black South Africans experiencing the effects of rapid social change involving desegregation, the possibility of schooling along nonracial lines, and greater economic opportunities for Black people. These are sometimes called the ‘born frees’, since they had been born after the announcement of the scrapping of apartheid and the release of Nelson Mandela after twenty-seven years in prison. These young students and graduates-to-be are on an upward class trajectory, by virtue of their education and socialization at a prestigious tertiary institution.

The fifty-three Black students were divided into four main groups according to gender and schooling.

(1) **Cf**: Black females who had some model-C schooling 

(2) **Cm**: Black males who had some model-C schooling

(3) **Nf**: Black females who had no model-C schooling

(4) **Nm**: Black males who had no model-C schooling

Boldface type is used to motivate the C, N, f, and m labels used throughout this article. Lowercase symbols for gender were used, since upper case F and M are later used for phonetic environments. C, as stated earlier, stands for ‘model-C and/or private school background’, N for its opposite (viz. ‘not of model-C background’). An international readership should note further that ‘schooling’ in the British/South African sense refers to junior and high school, and not college or university.

Groups **Nf** and **Nm** comprised ten speakers each, whose schooling began after 1994, the year in which South Africa became a new democracy. These children did not, however, enjoy the benefits of nonracial schooling, being constrained by what is sometimes termed ‘economic apartheid’, that is, not being able to afford an education away from the traditionally underresourced Black townships and their schools. The education sys-

---

8 Comments in our interviews from female interviewees also mention being alleged to ‘speak through the nose’ and to eat salads, especially rocket. Both charges are more common in the neighboring Zimbabwean context, where the equivalent term to ‘coconut’ is ‘salads’.

9 To circumvent inelegant phraseology I sometimes refer to students in the **Cf** and **Cm** groups (see 1 below) as middle class, even though they are not yet adults and therefore are open to other trajectories. While it is also the case that some students from disadvantaged, non-middle-class backgrounds are at model-C schools (through parental sacrifice or scholarships), none fitting this designation turned up in our interviews. Conversely, no students of middle-class family background turned up in our **Nf** and **Nm** school samples.
term in these schools has not improved much in the twenty years after apartheid. **Nf** and **Nm** students speak in the interviews of the excitement at nevertheless being admitted against these odds to former White universities. They also speak of the initial culture shock of being in an environment where English predominates, but on the whole claim to have adapted reasonably well to the institution. It is clear from the interviews that home languages like Xhosa or Tswana are highly valued by the students for social purposes, and that even as they become more proficient in academic English, some students embark on learning new African languages via friendships with other Black students. Students of this group now have multiracial social networks at university, but by and large have a majority of Black friends.

Groups **Cf** and **Cm** form the main focus of this study and represent the most upwardly mobile young Black people of this survey. **Cf** comprised twenty female speakers, and **Cm** thirteen speakers (it being difficult to locate or attract more than thirteen from the **Cm** group for interviews to match the intended total of twenty—see §7). Students categorized in these two groups all had one common characteristic: they had had some schooling in a multiracial school formerly reserved for Whites. The range of years of such schooling was considerable and would have been hard to control for beyond this stipulation.

In addition there were two main control groups of older Black and young White speakers, chosen here as a baseline against which the variation and change reported among the young post-apartheid generation can be gauged. Taking gender into account resulted in four control subsets.

(2) **Of**: older Black females with no model-C schooling by default ($n = 6$)
**Om**: older Black males with no model-C schooling by default ($n = 6$)
**Wf**: White females with model-C or private schooling by default ($n = 6$)
**Wm**: White males with model-C or private schooling by default ($n = 6$)

Groups **Of** and **Om** comprised twelve speakers in all, equally divided among males and females, interviewed while at the University of Cape Town in the early 1990s, prior to the collapse of apartheid. They formed part of a larger real-time sample of L2 Black South African speakers from different parts of the country reported on in Mesthrie 1997. All twelve speakers were subject to the constraints of the times: segregated schooling, economic hardships, and turbulent political conditions that affected their schooling and personal lives. But they had been accepted at the University of Cape Town, as the first trickle of Black students admitted to formerly White universities in the 1980s. All speakers were highly multilingual, speaking anything from three to six languages (typically Zulu, Xhosa, Sotho, Tswana, Afrikaans, and English). Although English was not a language of the home or neighborhood, it was their main educational language, and all speakers of this group used what may now be termed ‘older’ or ‘traditional’ BSAE.

Groups **Wf** and **Wm** comprised twelve White university students in all, divided equally for gender. They had studied at desegregating model-C schools and were chosen as a control group to confirm what the norms of WSAE currently are.

All interviews (ranging from forty to sixty minutes in duration) explored students’ backgrounds, primary and high school life, experiences at university, interests, best and worst moments, and the danger of death via recounting narratives of being mugged.

---

10 Universities were compelled by apartheid law from the 1950s to the 1980s not to accept Black students, except under certain very specific conditions. Prior to this time there had been a limited degree of multiracialism in the universities.
3.2. Older BSAE. The linguistic variables of vowel length and schwa are crucial elements that differentiate older BSAE from newer crossover varieties. As Hundleby (1963:62) put it, ‘there can be no doubt that it is the vowel allophones which are of greatest significance in determining the quite distinctive pronunciation pattern of Xhosa English’. I first provide a thumbnail sketch of the phonological system in which these features are embedded. Like many varieties of English in sub-Saharan Africa, traditional BSAE is notable for the following features that draw on the Southern Bantu substrates.

(3) Traditional BSAE features
   a. Use of a vowel system with five monophthongs and at least four diphthongs
   b. Rarity of schwa
   c. A tendency toward syllable timing, rather than stress timing
   d. Vowel lengthening being phonologically predictable
   e. Tone or pitch accent being employed in addition to, or sometimes instead of, stress

![Figure 1. An idealized vowel system of Older BSAE monophthongs, using Wells’s (1982) lexical sets as illustration.](image)

Regarding vowel length, older BSAE utilizes a largely five-vowel system, /a e i o u/ as charted in Figure 1, with some phonologically conditioned variation for /e/ and /o/ (Hundleby 1963:66–73). In Fig. 1, TRAP₁ denotes [e] in monosyllabic words like trap, have, had, back, while TRAP₂ denotes [a] in mostly polysyllabic words, for example, standard, adamant, manage. As far as diphthongs are concerned, most fluent traditional speakers have at least /ei/ for face, /at/ for price, /au/ for mouth, and /oi/ for choice, while goat is variably realized as [o] or [uu] (van Rooy & van Huyssteen 2000:29).

Hundleby (1963:63–66) and van Rooy (2004:945–46) note that vowel length is variable, with no consistent distinction between short and long pairs (i.e. tense and lax vowels). Length, which ‘may take place as a cue for stress placement’, thus does not have a phonemic function (van Rooy 2004:945). If the final syllable is heavy, its vowel is lengthened. If not, then it is the vowel of the penultimate syllable that is lengthened. Thus speakers of older BSAE lengthen the vowel in the final (superheavy, CVCC) syllable in contact [kʰontəkʦt] and parliament [pʰaləmənt], but in farmer [faːma] and seventy [sevənti], it is the vowel of the penultimate syllable that attracts length.

11 Regarding tense-lax distinctions, van Rooy finds intermediate realizations between, say, /i/ and /i/. I therefore treat tense-lax distinctions as equivalent to vowel-length differences (or lack thereof) in this article.

12 Since stress timing is not a characteristic of BSAE, stress is not generally marked in data from this variety.
4. On schwa. The BSAE restructuring of the standard vowel system(s) of English is completed by the infrequent occurrence of schwa and syllabic nasals and laterals.\textsuperscript{13} Schwa of standard English is usually replaced by a ‘full’ vowel: van Rooy and van Huyssteen (2000:21) give the following phonetic equivalents of schwa in Tswana-English: [æ ɛ ι ə ɑ ɔ] plus ‘other’.\textsuperscript{14} Moreover, this equivalent of schwa follows the same lengthening rules for ‘full’ vowels: it attracts length in a final heavy syllable, or otherwise in penultimate position. Thus examples like [miljɒn] ‘millions’, [kʰaːriəz] ‘careers’, and [sirjəs] ‘serious’ are traditional BSAE forms, with a long vowel in the final syllable, corresponding to schwa in WSAE and international standard English forms like [ˈmɪljənz], [kʰaːriəz], and [ˈsɪrjəs]. Examples from penultimate syllables include WSAE [ɒd(i)ˈdʒɪk] ‘audiology’, [saːˈkɒlədʒɪ] ‘psychology’, and [dɪkʃəˈnəri] ‘dictionary’ having BSAE equivalents [od(i)jolədʒi], [saːkolədʒi], and [dikʃənəri], with variable high tone and lengthening of the penultimate vowel, where WSAE and international standard varieties have schwa.

Schwa may not in and of itself seem a promising candidate vowel to illuminate the large-scale sociolinguistic and sociopolitical change discussed above. After all, as Silverman (2011:628) observes of standard varieties, schwa is short in duration, open to significant coarticulatory effects as the tongue moves from preceding to following contexts, and gravitates toward the center of the vowel space. Hence it is considered a ‘featureless’ vowel by some phonologists and as ‘weightless’ by others (Silverman 2011:629). Furthermore, the acoustically based program FAVE that we use (discussed further below) recommends discarding vowels in a data set that are less than 50 ms in duration. But for BSAE, schwa is often well beyond this length desideratum and is socially salient. Thus some ‘super-long’ tokens of schwa in words cited above include: dictionary (117 ms), seriously (304 ms), audiology (203 ms), and, as illustrated in Fig. 2 below, academically (123 ms). Such tokens contrast with typical realizations in the range of 10 to 50 ms among the WSAE control speakers in the database.\textsuperscript{15} Where social salience is concerned, there are indications from the media that some schwa effects are now being ‘enregistered’ (in the sense of Agha 2003) to portray a broad speaker of BSAE. One popular series of advertisements for Lunchbar chocolate uses the word obvious in its by-line (Speaker A: ‘Lunchbar!’; Speaker B: ‘It’s obvyas!’), with the traditional pronunciation [ɒbˈvjas], to sell this popular, inexpensive chocolate. It is unlikely that this pronunciation would be used in more upmarket ads, say, a big bank ‘serious’ about its wealthy clientele.\textsuperscript{16}

Flemming (2009) argues further that there are two different subsets of schwa: (i) medial schwa, which accommodates to its environments and lacks contrasts with other vowels, and (ii) final schwa, which contrasts with other vowels (actually with the high

\textsuperscript{13} Since the relevant historical input for South African English was mainly southern British English (Lass 1995), this is the phonological model intended by the label ‘standard English’ in this article.

\textsuperscript{14} As I show below, I have found it more convenient to work with the following equivalents of schwa: [æ ɛ i o u] and schwa itself. I include [o] since it is a common variant of schwa in my database.

\textsuperscript{15} Syllabic nasals and liquids are excluded from the study, though it is clear that these are much more common in WSAE and Cf speech in the database. This observation is also true of short ‘excluded’ schwas of less than 25 ms.

\textsuperscript{16} In another ad, the word picture is used to broaden the appeal of the ad for a television channel promising the ‘whole picture’ [hoʊl ˈpɪktə] to its sporting viewers. The final [ə] in place of schwa was a deliberate ‘downshifting’ on the part of the Black female uttering the phrase, intended to incorporate speakers beyond the hitherto middle-class females of the advertisement. Thus medial and final schwa are being enregistered, but not, to my knowledge, initial schwa.
vowels /iː/ and /uː/; cf. *bitty* versus *bitter*). On the basis of these arguments as well as experience with traditional BSAE, it was an easy decision to keep these two subtypes of schwa separate. And although Flemming is silent on initial schwa in words like *ago* and *ahead,* this was a clear third subset of schwa in the data. The analyses of initial and final schwa are relatively straightforward, since they are prototypically realized as [a] in older BSAE and as [ə] in WSAE. The ensuing analysis measures the extent to which individuals within the four younger Black core groups cohere with one or the other of these poles. Medial schwa is more complex, since it has any of the full realizations of [a e i o u] according to phonological environment, or by convention in certain suffixes often—but not always—cohering with the orthographical form. Thus -al, -ed, -ion, -ology, and -ful have the five realizations, respectively. The medial schwa subset therefore has to be broken down into five further divisions to allow for these subtypes (see Table 1 below).

4.1. Hypotheses. The hypotheses concerning the sociolinguistics of schwa for the core groups of young Black speakers of today can now be stated. These hypotheses concern vowel quality, rather than quantity, which is treated in §5.

(4) a. **Hypothesis 1:** The biggest difference occurs between Cf and Nm norms. Schwa is essentially a mid-central vowel in Cf speech. In Nm speech it is a peripheral ‘full’ vowel, which ranges well away from mid-central toward [a] for initial and final positions and toward any of [a e i o u] for medial schwa, depending on environment.

b. **Hypothesis 2:** This difference between Cf and Nm is strongest for final schwa, discernible for initial schwa, but less salient for medial schwa, where more purely linguistic constraints are expected to be stronger.

c. **Hypothesis 3:** Between the two poles identified in hypothesis 1 are the norms of Cm and Nf. In some instances each of these groups shows the same norms as their counterparts of the opposite sex by type of schooling (a class effect via schooling). In other instances, notably final and medial schwa, they differ from those counterparts (thus indexing gender effects that override class/schooling).

These hypotheses together reflect the following holistic expectation: overall, both class and gender considerations are involved in a move away from the norms of traditional BSAE. Middle-class females lead the way, middle-class male speakers are more ambivalent in this shift, and working-class females show closer affinity to middle-class norms than do working-class males. This generalization may be formalized as follows: $\text{Cf} > \text{Cm} = \text{Nf} > \text{Nm}$. Here ‘$>$’ means ‘is significantly different from’ and in fact denotes increasing distance from older traditional BSAE norms, while ‘$=$’ means ‘is not significantly different from’. This symbolism forms a useful shorthand heuristic for evaluating the degree to which the hypotheses hold in each of fourteen different statistical runs (one each for final and initial schwa, five runs for medial schwa, multiplied by two overall to account for vowel height and backness independently).

4.2. Acoustic and statistical methods. The analysis of schwa was undertaken acoustically, using the forced alignment program FAVE. Developed at the University of

---

17 Lass (2007) identifies seven different schwas in his own speech, arguing that they should be transcribed differently.

18 See Mesthrie 2005 and n. 25 below for a refutation of the belief that ‘spelling pronunciations’ are a driving force in the phonology of traditional BSAE.
Pennsylvania, this suite (Rosenfelder et al. 2011) is an adaptation of P2FA (developed by Yuan and Lieberman (2008)), written in the programming language Python. It consists of two toolkits: (i) FAVE-align, written by Rosenfelder, which is geared toward aligning sociolinguistic transcripts with the acoustic input, and (ii) FAVE-extract, an adaptation of ‘extractFormants’ (Evanini 2009), written by Rosenfelder and Fruehwald, which provides automatic formant readings of vowels. FAVE was first used to analyze data from the Philadelphia Neighborhood Corpus (PNC) at the University of Pennsylvania, and so was trained on American English data. Within FAVE-extract the Mahalanobis distance option was used for formant prediction, based on an algorithm developed by Evanini (2009) and discussed further by Labov and colleagues (2013:36). Essentially, this method simulates the continual adjustments that a researcher would carry out in a manual analysis in PRAAT (Boersma & Weenink 2016) to provide more accurate formant readings. We follow Evanini’s recommendation of using the ‘faav’ setting option that measures formants one third of the way into a vowel’s duration. We also used the vowel normalization procedure offered by FAVE, taking the ‘Lobanov’ (1971) option. As a first step in exploring the data statistically, linear mixed-effects regression ‘lmerTest’ was used in R (Johnson 2009, Kuznetsova et al. 2014, R Core Team 2014). The fixed-effect factors (predictors) for this study were ‘gender’, ‘phonetic environment’, and ‘type of school’. The random-effect predictors were ‘word’ and ‘individual’ (word-list style words were excluded and analyzed separately in the section concerning word length). The results are presented as random forests and conditional inference trees processed in R using the ‘partykit’ package via the ‘ctree’ function (Hothorn et al. 2006), which is discussed further in §4.5.

FAVE’s dictionary reflects US English forms, and has accordingly been adjusted for British English varieties (Robinson 1994). We undertook further adjustments for South African English departures from British English variants: chiefly schwa rather than [ɪ] in unstressed syllables like roses, wanted, and exactly, and [ɒ] as a covariant of schwa in initial prefixes like com-, con-, col- (e.g. computers, concern, collect). FAVE works well for middle-class L1 South African English (SAE) accents (Chevalier 2016), but does less well with other SAE accents. This is partly due to the enormous amount of substrate influence in L2 speech, including the use of schwa as a full long vowel. Getting FAVE to produce fully reliable textgrids (i.e. alignments of vowel and consonant segments in the continuous acoustic signal) requires speech technology and engineering expertise beyond what we currently have available, but is desirable for long-term diaclectology in South Africa. Figure 2 doubles as an example of a textgrid and of lengthened penultimate schwa in the word academically, as uttered in conversational style by an Nm speaker. The manual checking of the textgrids of full transcripts produced by FAVE proved enormously time consuming—initially in the region of thirty hours per one-hour transcript. For this reason subsequent analysis had to be limited to five minutes of speech per subject, excerpted from a point of the interview when the speaker was in full flow in casual speech mode. For the study of individual stressed vowels a five-minute excerpt is unlikely to produce a sufficient yield of tokens for sociolinguistic analysis, but for schwa this proved adequate, given its frequency of occurrence. This study reports on 4,746 tokens of schwa: 3,353 in the main group and 1,393 in the control groups (an average of sixty-two tokens per speaker).

19 To simplify matters we left out all tokens with these prefixes.
20 Important work on speech technology in South Africa occurs at North-West University and Stellenbosch University. FAVE is, however, only starting to be used at these universities.
For the traditional BSAE speakers of this study the excerpted textgrid still had to be checked carefully (about six hours per five-minute excerpt), with particular attention to ‘AH0’, the FAVE code for schwa. Checking of alignment and redrawing of boundaries were undertaken by a team of assistants, then I double-checked every textgrid myself, and especially triple-checked AH0. We were thus confidently able to run outputs for analysis of all tokens of schwa of 25 ms duration or more. FAVE produces normalized Lobanov (1971) outputs of all vowels, showing formant readings and duration. At the stage of analysis we selected just the AH0s and classified them according to environment. Yet one additional checking stage involved examining all apparent outliers to ensure that they were not due to misalignment, mistranscription, misreading, or interference of the speech signal missed in the earlier rounds of checking.

4.3. PHONETIC ENVIRONMENTS. The environments were arrived at by a mixture of inspection of the data, expectations from phonological trends in English, and acoustic caveats. Easily discernible from inspection of the output data are the fronting effects of a preceding /j/ and the backing effects of a following /l/ for all speakers. Two environments were generated internally from our data, both from traditional BSAE effects. The first involved /w/ before a final schwa, which induced backing (often with syllable restructuring, as in [flaːˈwə] ‘flower’ where WSAE has [flaʊə]). The second included a similar effect of vowel harmony with clear [i]-like realizations of schwa after an /i/ in the preceding syllable, hence [iːvɪn] ‘even’, [rizɪn] ‘reason’, and [iziːli] ‘easily’. To keep the data analysis at a manageable level in terms of the number of working environments, preceding environments (essentially /j/ and /i/) took precedence over following ones. This decision was based on inspection of the formant readings for words like *provincial*, *onion*, and *gracious*, which have both preceding (/j/) and following contexts of relevance (/l/, /n/, and /s/). Other effects were less clear-cut on first inspection, but in line with the

---

21 The following were excluded from our outputs: (i) short grammatical words like *was*, *would*, *just*, which are likely to be elided, (ii) schwa in the square and cure diphthongs, which have widely divergent realizations in SAE compared to their international counterparts (i.e. [wei] and [jo], respectively). 

22 On inspection, the fronting effects of /j/ are indeed more apparent than, say, the backing effects of /l/.
acoustic literature (notably Thomas 2011), the following environments were settled on, with labels in capitals indicating the mnemonic used in the statistical runs.23

(5) Final schwa

J: after /j/ or /ʒ/ in /ɔː/, for example idea, here, peer
W: after /w/ or /au/, for example hour, flower, power
G: ‘other’, for example butter, feather, gonna

(6) Initial schwa: Same following environments as for medial schwa below.

(7) Medial schwa: As noted above, medial schwa has to be divided into five subsets, which I call ‘a-schwa’, ‘e-schwa’, ‘i-schwa’, and so forth, depending on the typical older BSAE realizations.24 Analyzing these subsets separately is necessary to avoid a false averaging out of these peripheral realizations to a mid-central value. For each subset of schwa it was still necessary to factor in the influence of phonetic environments. The examples in Table 1 follow the sequence [a e i o u] of potential schwa realizations, with em-dashes denoting lack of tokens in that subset of schwa.25

<table>
<thead>
<tr>
<th>ENVIRONMENT</th>
<th>‘a-schwa’</th>
<th>‘e-schwa’</th>
<th>‘i-schwa’</th>
<th>‘o-schwa’</th>
<th>‘u-schwa’</th>
</tr>
</thead>
<tbody>
<tr>
<td>J: after /j/</td>
<td>serious</td>
<td>requires</td>
<td>—</td>
<td>canyon</td>
<td>—</td>
</tr>
<tr>
<td>L: before /l/</td>
<td>normal</td>
<td>novel</td>
<td>privilege</td>
<td>ecological</td>
<td>people</td>
</tr>
<tr>
<td>N: before /n/</td>
<td>balance</td>
<td>stolen</td>
<td>tension</td>
<td>second</td>
<td>unfortunate</td>
</tr>
<tr>
<td>M: before /m/</td>
<td>tantrum</td>
<td>cinema</td>
<td>—</td>
<td>freedom</td>
<td>—</td>
</tr>
<tr>
<td>P: before labials (including /w/, excluding /m/)</td>
<td>—</td>
<td>supposed</td>
<td>celebratory</td>
<td>responsible</td>
<td>bishop</td>
</tr>
<tr>
<td>F: before /f/ and /v/</td>
<td>paraffin</td>
<td>benefit</td>
<td>beautifully</td>
<td>renovation</td>
<td>—</td>
</tr>
<tr>
<td>H: before /h/ in following syllable</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>alcohol</td>
<td>—</td>
</tr>
<tr>
<td>R: before /r/ in following syllable</td>
<td>primary</td>
<td>funeral</td>
<td>inspiration</td>
<td>memory</td>
<td>—</td>
</tr>
<tr>
<td>K: before velars</td>
<td>stomach</td>
<td>subjects (n)</td>
<td>certificate</td>
<td>allocated</td>
<td>—</td>
</tr>
<tr>
<td>G: ‘other’ (before [θ ð t d s ʃ z ʒ ʒ ʃ dʒ], that is, before coronals, excluding [n l r], plus alveopalatals)</td>
<td>brothers</td>
<td>biggest</td>
<td>stupid</td>
<td>carrots</td>
<td>—</td>
</tr>
<tr>
<td>I: after /iː/ in preceding syllable</td>
<td>—</td>
<td>—</td>
<td>even</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 1. Example tokens of five types of medial schwa by structural environment.26

In most cases this choice of environments is well supported. The analysis below gives no motivation to further collapse environments, since they do not cluster in exactly the same ways within the three schwas. Unfortunately, this means that the acoustic analysis does not always give us the neat environmental effects via natural classes that phonologists might hope for. However, in all applicable cases J is clearly a fronting en-

23 I thank Erik Thomas (p.c., 11 August 2015) for vetting these choices and for detailed advice regarding grouping. Note that SAE varieties are generally nonrhotic, though some speakers are becoming responsive to postvocalic /r/ from American English.

24 Note that the differentiation of medial schwa into five subtypes is most obvious in traditional or older BSAE. For comparative purposes the data set had to consider these subtypes even for the White control group and crossover speakers (CF and CM). As Figs. 5a and 5b below show, this differentiation turns out to apply reasonably well to these groups too (especially Fig. 5d in backing environments for medial schwa).

25 The examples also serve the purpose of illustrating that the choice of schwa is by convention, rather than being dictated by the spellings. More trenchantly, button has [e], not [o], donor [a], not [o], and bigger has [a], but biggest has [e] in traditional BSAE.

26 M and N are separated because although nasality is relevant for F1, place of articulation is more relevant for F2. This consideration does not arise for /r/, since there were no relevant tokens with schwa + /r/. /w/ was treated as a separate environment for final schwa, but was otherwise subsumed under P (labials).
environment, a backing environment, and G—as expected from its definition—a neutral environment regarding advancement. The other environments are not as clear-cut in the coarticulatory influence on schwa.

4.4. Statistical results for the control groups. The existing literature leads us to expect unstressed, short, mid-central schwa in WSAE (Lanham & Macdonald 1979, Lass 1995, Bekker 2009) and peripheral full values for BSAE (as cited above). Figures 3–5 give the medians for final, initial, and medial schwas and the distribution of 50% of the data around each of these medians for the White and older Black control groups, using the methods (linear mixed-effects regression runs in R) outlined above. The full set of statistical results are provided in the online supplementary materials. The figures indicate any significant differences by gender and environment within each control group. The figures also show the overlaps in the set of tokens around the medians. Figure 3a shows a narrow range of variation in the ‘mid’ region for height for final schwa of the White control group and a much wider region for ‘advancement’ according to environment. The general ‘G’ environment, which is arguably the most neutral, shows what might be considered a ‘classic’ schwa as a mid-central realization, with no gender difference. The medians are (624 Hz, 1556 Hz) for White females and (647 Hz, 1595 Hz) for White males. For J there is a difference by gender, with males being fronter. W shows a great deal of backing. In contrast, the Older BSAE control group in Figure 3b shows a broader range for height, with the lowest realizations near 800 Hz for F1, indicating an [a] realization, with fronting after J. There are no gender differences in this control group.

Figure 4a shows a bifurcation by environment, not gender, for initial schwa in the White control group, which has a mid-central schwa, with backed realizations before L, P, and R. By contrast, Figure 4b shows no significant differentiation by environment or gender in the older Black control group and a range of variation from mid-central to notably low realizations.

Matters are more complex for the five subtypes of medial schwa, indicated in the figures by ‘a’, ‘e’, ‘i’, ‘O’, ‘u’. To add to the complexity, three of these (‘a’, ‘e’, and ‘O’) show differentiation by environment. In the interests of space and clarity the information is presented in four figures (Figures 5a to 5d), rather than ten separate ones. Figures 5a and 5b give the medians for five subtypes of medial schwa for the White control group, showing that environment, but not gender, has an effect on each subtype. Figures 5c and 5d give the medians for the older Black control group, showing ‘a-schwa’ with its more complex patterns separately in Fig. 5d, and the rest in Fig. 5c (where structural environment is not significant). Despite these complications, Figs. 5a to 5d follow the broad trends of Figs. 3 and 4. Thus there is a spread for F2 (front-backness) in the White group, but not for F1 (height). For the Black group there is considerable spread in both dimensions, resembling the distribution we would expect for ‘full’ vowels [a e i o u].

27 The online supplementary materials can be accessed at http://rajendmesthrie.com/ or http://muse.jhu.edu/resolve/14.
28 The medians are per environment and gender if significant, and strictly speaking each box illustrates the intersection of the range of 50% of the data around the median for F1 and 50% of the data around the median for F2 for the relevant gender and/or environment.
29 In this conventional (x, y) notation, ‘x’ gives the first formant reading in Hertz, and ‘y’ the second formant reading. For US English, Flemming gives an average schwa reading of (665 Hz, 1772 Hz). A comparable average for the WSAE control group in the neutral ‘G’ environment is (575 Hz, 1685 Hz), with no differentiation for gender.
Summing up, Figs. 3a to 5d show that for the young White control group schwa is indeed in the mid-central region, except in fronting environments (typically after J) or backing environments (typically before L). In contrast, schwa is a low central vowel in initial and final positions in the older Black control group, as well as for medial ‘a-schwa’. The one exception is that even in this group final schwa is fronted in the J environment. Medial schwa indeed shows five different realizations in this control group, unlike the White control group, which shows a mid-central norm, but with backing in certain environments.

A further function of the control groups is to ascertain whether gender is significant in them. The linear mixed-effects regression runs in R indicate that there are no significant gender differences for F1 or F2 in the White control group for initial schwa (Fig. 4a) or any of the five medial subsets (Figs. 5a and 5b). This is also true of final schwa (Fig. 3a), with one exception: the J environment for F2, where males show fronter realizations (me-
In the Older Black (Of and Om) control groups there are no gender differences for F1 or F2 of final or initial schwa (Figs. 3b and 4b, respectively). For medial schwa (Figs. 5c and 5d) there are four significant differences in gender—in F1 and F2 of medial ‘a-schwa’, F2 of medial ‘e-schwa’, and F2 of medial ‘i-schwa’, for all of which

30 The conditional inference trees for this set are too many to reproduce here and, together with the linear mixed-effects regression in R (lmer), can be accessed in the online supplementary materials at http://rajendmesthrie.com/ or http://muse.jhu.edu/resolve/14.
males produce more peripheral tokens. Thus ten of the fourteen tests for F1 and F2 in the Older Black control group show no significant gender differences, including the arguably more salient initial and final positions. Figures 3 to 5 thus set the scene for evaluating changes for the core group of young Black students studied.

Figure 3a. White control group medians and distribution of 50% of data around them for five subtypes of medial schwa, showing environments that inhibit backing per vowel and absence of gender effects.

Figure 3b. White control group medians and distribution of 50% of data around them for five subtypes of medial schwa, showing environments that induce backing per vowel and absence of gender effects.

‘Peripheral’ is used in the standard sense of ‘non-mid-central realizations’, largely [æ e i o u] in this article (which differs from Labov’s concept of ‘peripheral track’). The medians are for medial ‘a-schwa’, F1: males 770 Hz, females 655 Hz ($p = 0.003$); F2: males 1441 Hz, females 1623 Hz ($p = 0.003$). For medial ‘e-schwa’, F2: males 1766 Hz, females 1592 Hz ($p = 0.005$). For medial ‘i-schwa’, F2: males 1832 Hz, females 1609 Hz ($p = 0.049$).
4.5. Results for the core groups of the study. We now analyze the social differences in the younger core groups, via the conditional inference trees in R. The trees visually represent the recursive binary partition of a large data set using regression analysis. They show the subtle interactions in the data using a hierarchical display, reporting accessibly on the p-values upon which the branching is based. The trees are read from top to bottom in ways familiar from diagrams in linguistics, each branch showing significant partitions and subpartitions. The box-and-whiskers plots at the lowest tier give the following information: (i) the main box colored in gray shows the distribution of 50% of the data, with the median given as a dark line; (ii) the broken lines (whiskers) above and below the box denote the upper quartile and lower quartile of the normally distributed data, respectively; and (iii) the small circles above and below the whiskers denote the outliers.

Figure 5c. Older Black control group medians and distribution of 50% of data around them for medial ‘e’, ‘i’, ‘o’, and ‘u-schwa’, showing gender effects but absence of influence by structural environments.

Figure 5d. Older Black control group medians and distribution of 50% of data around them for medial ‘a-schwa’, showing significant gender and structural effects.
Final schwa. The conditional inference trees for F1 and F2 of final schwa for the core group of young Black speakers (groups N and C) are given in Figures 6 and 7.\(^{32}\)

The branches in Fig. 6 show the following for F1 (height).

(8) Results for F1 of final schwa
   a. The primary difference is by gender \((p < 0.001)\) rather than school.
   b. For males there is further differentiation by school \((p < 0.001)\); for females there is none.
   c. There is no differentiation by environment for height of final schwa.

Thus the social hierarchy for F1 of final schwa may be formulated as \(Cf = Nf > Cm > Nm\), with females \((Cf\) and \(Nf)\) having the most ‘mid’ realizations (median of 630 Hz), \(Nm\) the lowest realizations (761 Hz), and \(Cm\) in between (668 Hz). Qualitatively, this is arguably the most important hierarchy since it deals with the most salient environment—schwa in final or unchecked position.

For F2 (backness) the branching in the conditional inference tree in Fig. 7 shows that the chosen environments are indeed significant, with J (fronting after /j/) significantly different from W and G (backing after /w/ and the ‘other’ or ‘general’ environment; \(p < 0.001\)). Furthermore, W is significantly different from G \((p < 0.001)\). The social correlates for these environments are as follows.

\(^{32}\) Figure 6 and all subsequent graphs for F1 are to be read as inversely representing vowel height (the lower the ‘box’, the higher the vowel). Similarly, Fig. 7 and all subsequent graphs for F2 inversely represent vowel backness (the lower the box, the backer the vowel).
A sociophonetic study of schwa in deracializing South African English

(9) a. There is no social differentiation in the W environment (all groups show some backing after /w/).

b. For G (general environment) there is a differentiation by gender, but not school ($p < 0.001$). Males (1504 Hz) are slightly backer than females (1638 Hz).

c. For J (fronting after /j/) the opposite holds: there is differentiation by school, not gender, with model-C students being slightly fronter ($p = 0.023$). The social hierarchy for F2 of final schwa for the G environment is thus: $C_f = N_f > C_m = N_m$. For the J environment the social hierarchy is: $C_f = C_m > N_f = N_m$. For the W environment there is no social hierarchy.

Comparing the medians of Figs. 6 and 7 for final schwa with Figs. 3a and 3b for the control group shows that $N_m$ is most similar to older BSAE norms (showing substrate fidelity), and $C_f$ the least (showing substrate erasure). The $N_f$ females, as we have seen, also side with $C_f$, except in the J environment (where they do not have medians as front as those of $C_f$ or $C_m$).

**Initial schwa.** The conditional inference trees for F1 and F2 of initial schwa are given in Figures 8 and 9, respectively. Figure 8 shows the following for vowel height.

(10) Results for F1 of initial schwa

a. The primary differentiation is by school ($p < 0.001$).

b. For ‘C’ (the ex-model-C) students there are no gender differences.

c. For the ‘N’ (or non-model-C) students there is a gender differentiation ($p < 0.001$).

d. There is no differentiation by environment for height of initial schwa.

The social hierarchy is thus: $C_f = C_m > N_f > N_m$. Model-C students have the most mid realizations (642 Hz); $N_m$ males have the lowest (806 Hz) realizations; $N_f$ are in between (680 Hz).

Turning to Fig. 9, the patterns for F2 of initial schwa are as follows.
Results for F2 of initial schwa

a. The primary differentiation is by backing (P the most, then L, R, F) versus other environments (G, K, M, and N).

b. There is no evidence of social differentiation, and hence no social hierarchy.
Summing up overall for initial schwa, we once again see \textsl{Nm} showing substrate fidelity (with their low-central medians not very different from the older BSAE control in Fig. 4b). This time \textsl{Cf} and \textsl{Cm} are united in their substrate erasure (with mid-central realizations of initial schwa not dissimilar from the White control group means in Fig. 4a). Figure 8 shows that \textsl{Nf} are intermediate, but statistically closer to \textsl{Nm}.

Medial schwa. It is not practical to reproduce the ten conditional inference trees for medial schwa (see the online supplementary materials). Instead, a summary is provided of groupings of the hierarchies produced by the trees.

(12) \textbf{\textsl{Cf} = \textsl{Cm} > \textsl{Nf} = \textsl{Nm}}: This pattern, with school being significant but not gender, applies to the following four cases for medial schwa:
   a. F1 of ‘e-schwa’ in the environments of L, R, N, F, J, K (model-C students are more mid/less peripheral; \(p = 0.014\))
   b. F2 of ‘e-schwa’ in the environment of L (model-C students are backer/more peripheral; \(p = 0.042\))
   c. F2 of ‘i-schwa’ in all environments (model-C students are more central/less peripheral; \(p < 0.001\))
   d. F2 of ‘o-schwa’ in all environments (model-C students are less back/less peripheral; \(p < 0.001\))

Related to the above hierarchy are the following patterns.

(13) \textbf{\textsl{Cf} = \textsl{Cm} > \textsl{Nf} > \textsl{Nm}}: This pattern, with school being significant overall, and gender for the non-model-C group only, applies to the following three cases:
   a. F1 of ‘a-schwa’ in all environments (\textsl{Nm} have the lowest realizations, \(p < 0.001\); \textsl{Cf} and \textsl{Cm} have more ‘mid’ realizations, \(p < 0.001\))
   b. F1 of ‘i-schwa’ in all environments (\textsl{Nm} have the highest realizations, \(p = 0.009\); \textsl{Cf} and \textsl{Cm} have more ‘mid’ realizations, \(p = 0.006\))
   c. F2 of ‘e-schwa’ in all environments, except (backing) L and (fronting) J (\textsl{Nm} have the frontest/most peripheral realizations, \(p = 0.001\); \textsl{Cf} and \textsl{Cm} have the most central/least peripheral, \(p = 0.002\))

(14) \textbf{\textsl{Cf} = \textsl{Nf} > \textsl{Cm} = \textsl{Nm}}: This pattern, with gender but not school being significant, applies to the following two cases:
   a. F2 of ‘a-schwa’ in the environments of J, N (males have fronter/more peripheral realizations, \(p = 0.029\))
   b. F2 of ‘u-schwa’ in the L environment (females have backer/more peripheral realizations, \(p = 0.001\))

(15) \textbf{\textsl{Cf} = \textsl{Nf} > \textsl{Cm} > \textsl{Nm}}: This pattern, with school being significant for males but not females, applies to the following case: F2 of ‘a-schwa’ in the environments of G and K—that is, all environments except (backing) L and (fronting) J, N (\textsl{Nm} have backer/more peripheral realizations, \(p = 0.047\))

To summarize these complexities concerning medial schwa, note that:

(i) The medians for model-C students (\textsl{Cf} and \textsl{Cm}) are mostly less peripheral than non-model-C students (\textsl{Nf} and \textsl{Nm}) (six cases).

(ii) In two cases gender reveals closer similarities than school (with males having more peripheral values of schwa than females).

33 Peripherality here (and for F2 of ‘u-schwa’) is an artifact of retraction before L, which gives the only two instances where model-C students are more peripheral than non-model-C students. The former appear to be following White norms where retraction before /l/ is common with many vowels (Lanham & Macdonald 1979:43, Bekker 2009:424). Figures 5a and 5b show that the most retracted tokens belong to these two subsets of schwa before L in the White control group. Note that almost all tokens (283 of 287) of ‘u-schwa’ in the database were followed by /l/.
(iii) Retraction of medial schwa before L produces two exceptional cases, where females are more peripheral than males (for ‘u-schwa’) and model-C students are more peripheral than non-model-C students (for ‘e-schwa’). These two exceptional cases, in fact, confirm the close and subtle similarities between Cf especially and the White control group (see Fig. 5b where these two types of schwa show the most backing among medial schwas in the L environment for the White control group).

(iv) There is no social hierarchy in two cases (F1 of ‘o-schwa’ and ‘u-schwa’).

More generally, we may conclude from these summaries that Nm veer toward substrate-influenced [a e i o u] for medial schwa (in ways still similar to Figs. 5c and 5d for the older BSAE control group). And again, Cf females show the most substrate erasure in generally having mid-central medians for the subtypes of medial schwa (similar to Figs. 5a and 5b for the White control group).

4.6. Confirming the hypotheses for initial, medial, and final schwa. Hypothesis 1 concerns the maximal difference between the Cf and Nm groups. This can be seen in the conditional inference trees and the social hierarchies derived from them. In all fourteen cases listed above that show statistically significant social differentiation, Cf and Nm are apart. Hypothesis 1 is thus on the whole confirmed. (For the sake of completeness, note that Cf and Nm do share similarities in (another) four cases; in these cases there was no social differentiation at all between the four core groups: F1 for ‘o-schwa’ and ‘u-schwa’; F2 for initial schwa; F2 for final schwa in the W environment only.)

Hypothesis 2 is that the difference between these same two social groups would be significant for final schwa, less so for initial schwa, and least for medial schwa (where more purely coarticulatory constraints were likely to operate). The hypothesis is not upheld, since there is a significant difference between Cf and Nm in all fourteen cases of initial, final, and medial schwa, showing statistically significant social differentiation.

Hypothesis 3 concerns the intermediate status of Cm and Nf. In five cases Cf and Nf cluster together, but not Cm (‘gender overrides class’). In another nine cases Cf and Cm cluster together, but not Nf (‘class overrides gender’). Hypothesis 3 is thus upheld.34

A general integration of the three hypotheses (and for all types of schwa) is therefore viable. Cf speakers show the furthest move away (‘substrate erasure’) from the older traditional norms (charted in Figs. 3b, 4b, 5c, and 5d), Nm speakers the least (‘substrate fidelity’). In five cases Nf speakers remain on the traditional side of the scale (having similar norms to Nm). In four cases (including the most socially and linguistically salient F1 and F2 of final schwa, and F2 but not F1 of the equally salient ‘a-schwa’) they move ‘upward’ to resemble the Cf group.35 Cm speakers are similar to their middle-class female counterparts in nine cases, but strategically deploy schwa ‘downward’ (socially) in three cases (including the most socially salient final schwa). They do this possibly to unconsciously signal a degree of covert solidarity with Black males of a more traditional schooling. This accords well with the broader adult society in which similar ambivalences are salient—see Modisha 2007 on the contradictory class location of the African corporate middle class in South Africa.36 Cm speakers would not generally be mistaken for Nm, but appear to strategically keep something traditional so as not

34 For the sake of completeness I note that in three of these cases Nf clusters with Nm, and in another three cases Cm clusters with Nm.
35 See §4 on the salience of final schwa and medial ‘a-schwa’ in relation to stereotypes and ‘enregisterment’.
36 Although Modisha does not say so explicitly, his generalizations apply mostly to males.
to cross over completely. Nf speakers, somewhat similarly, would not be mistaken for Cf speakers overall; but they appear to do their best to match the schwa variable of Cf speakers in the ‘easier’ initial and final environments (but not much in the more complex case of medial schwa). I now turn briefly to the related area of vowel length to ascertain whether the same social hierarchies apply.

5. A PRELIMINARY ANALYSIS OF VOWEL LENGTH. The current sociophonetic literature is pessimistic about the chances of studying vowel-length variation in natural speech, outside the word-list or carrier-phrase laboratory. Di Paolo, Yaeger-Dror, and Wassink (2011:98) put it thus:

there is a general tendency in actual conversations for people to speed up or slow down over the course of a sentence, to emphasise one word in the sentence, as a preface to a word search or other hesitation phenomenon. As a result, the measurement of duration from small quantities of actual conversations may not be very profitable.

Lengthening in my BSAE database may depend on a host of factors: the phonological rule identified earlier, minor lengthening of final vowels compared to other vowels, prepausal lengthening, possible minor variation according to vowel height, focus information effects, sentence-final position, and so forth. There is the added complexity of the coexistence of traditional BSAE and ‘target language’ L1 patterns that speakers might be gradually adjusting to at university. As an example of this complexity, I cite two sentences in casual style from the database, uttered by Dora M, a speaker of traditional BSAE.

(16) a. I don’t know what I’m going to eat.
   b. I didn’t really see the beauty of it.

The lengths of the vowels are almost the same in eat (160 ms) in 16a and it (186 ms) in 16b, both counting as [iː] phonetically, with the vowel in it in fact slightly longer than that in eat. This looks like sentence-final lengthening, but there are other examples in the database where sentence-final it or eat are not lengthened. At this stage we do not know enough about the pragmatics and suprasegmental phonology of BSAE to do justice to vowel-length variation in natural speech. For this reason, only a preliminary account is given via a comparison of vowel-length ratios of selected pairs from the word list read out at the end of the interviews. The reason I pursue this preliminary analysis at all is that vowel neutralizations and ‘full’ schwa often go together in traditional BSAE, with its penchant for syllable timing. It is important to note that unlike variation in schwa, length differences of other vowels have semantic repercussions. A neutralization in length can affect meaning, as some very salient examples from the traditional BSAE of radio journalists show:

(17) South Africa is a hard country to [liːv]. [Radio journalist intended live, but open to misinterpretation as leave in the context of large numbers of expatriates abroad.]

(18) With all this rain people are [slipiːŋ] on the pavement. [Radio journalist intended slipping, but open to being misinterpreted as sleeping in the context of urban poverty.]

Written English, even by graduates, has occasional misspellings showing the influence of neutralizations: feast for ‘fist’, seat for ‘sit’, living for ‘leaving’, and so forth.

37 These examples are not meant facetiously, and a reminder is in order that the burden is on the L1 listener, not just the L2 speaker of English in a highly multilingual and still unequal society.
Cross-dialectal misunderstandings are evident in the occasional mistranscriptions of BSAE in ELAN by junior research assistants of varying backgrounds in the early stages of this project (before a word list was hastily circulated). Examples include puck for ‘park’, but for ‘bath’, nest and ness for ‘nurse’, and teat for ‘kit’, while bit and beat were frequently interchanged. Cf and Cm students of this study show virtually no such variation in their casual speech; the others do. Therefore a working hypothesis this time was that while most neutralization occurs among Nm speakers, Cf and Cm norms coincide, and Nf are intermediate between these two poles (i.e. Cf = Cm > Nf > Nm). The word list of this study did not specifically test for length via minimal pairs, as it is part of a larger general study of SAE varieties. Rather, it used Wells’s (1982) well-known lexical sets kit, dress, trap, and so forth, together with a longer list of tokens deemed of interest to SAE studies, including the minimal pair beat–bit. I report briefly on /iː/ versus /ɪ/ in three pairs from the word list in Table 2. Citing a ratio rather than an absolute measure per vowel gives us the advantage of overcoming the need for normalization of length by speaker, since speech rates may differ (see Wassink 2006, Thomas 2011:144). Thomas notes that ‘ordinarily long vowels show durations only about 50% longer than those of short vowels, and often the difference is even less than that’ (2011: 143). For this study, I take a more conservative ratio of less than 125 : 100 or 1.25 as a rough indicator of vowel neutralization. Vowel length was measured in ways suggested by Thomas (2011:133–44), especially in taking the offset as the point where the formants become indistinct. Table 2 gives the total number of neutralizations evident per group for each /iː/ : /ɪ/ pair, with column 3 showing the numbers for the actual tokens of fleece versus kit in the word lists, column 4 for beat versus bit, and column 5 for beat versus kit. Thus a proportion like 7/9 in column 3 indicates that of the pairs of fleece and kit uttered by nine speakers, seven showed neutralization.

<table>
<thead>
<tr>
<th>GROUP</th>
<th># OF SPEAKERS</th>
<th>FLEECE : KIT</th>
<th>BEAT : BIT</th>
<th>BEAT : KIT</th>
<th>TOTAL</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nm</td>
<td>9</td>
<td>7/9</td>
<td>4/4</td>
<td>6/8</td>
<td>17/21</td>
<td>81%</td>
</tr>
<tr>
<td>Nf</td>
<td>8</td>
<td>5/8</td>
<td>2/5</td>
<td>1/6</td>
<td>8/19</td>
<td>42%</td>
</tr>
<tr>
<td>Cm</td>
<td>11</td>
<td>5/11</td>
<td>1/5</td>
<td>0/11</td>
<td>6/27</td>
<td>22%</td>
</tr>
<tr>
<td>Cf</td>
<td>12</td>
<td>5/12</td>
<td>0/8</td>
<td>2/10</td>
<td>7/30</td>
<td>23%</td>
</tr>
</tbody>
</table>

Table 2. Proportions of ratios of /iː/ to /ɪ/ showing neutralization (< 1.25) of four tokens in word list style in four core groups of young Black speakers by school and gender.

Table 2 thus shows the hierarchy Cf = Cm > Nf > Nm, which accords with the hypothesized order. Overall, there are class differences involved, and gender differences within the non-model-C group. What was slightly unexpected was that Cf and Cm speakers should show any neutralizations at all, since this was not evident in their interview speech. I attribute this to the use of a ‘clipped’ style adopted for the word list by some speakers, rather than reflecting a tendency toward neutralization in their casual style speech. This admittedly nontechnical term refers to a certain playfully disdainful, truncated way in which some speakers treated the simple words in a simple list. On the whole, vowel-length considerations do support the findings for schwa.

6. The perception test for gender and ethnicity: more evidence of females in the lead. This section summarizes a perception study designed by the author to test
the hypothesis that middle-class Black females were crossing over into what used to be exclusively WSAE to a greater extent than males (Mesthrie, Chevalier, & McLachlan 2015). The focus was on the reactions of young South Africans from a range of ethnic backgrounds to short clips from a selection of the interviews. The test comprised twenty clips, thirteen of middle-class Black speakers excerpted from the interviews, plus another seven (White or non-middle-class Black interviewees) as a kind of control. Each clip was about twelve seconds long, chosen so as to be neutral with regard to content clues about the speaker’s ethnicity (e.g. avoiding names of places and people). It was therefore hoped that respondents would be making a general holistic judgment based on suprasegmental features like intonation, stress, rhythm, and articulatory setting in addition to vowel and consonant segments. The thirteen Black speakers were chosen by me and my two co-authors as the ones who overall sounded like candidates for crossover accents (and social dialect): only five of the males qualified (from group \textit{Cm}), and of the females, eight of a possible thirteen who qualified (from group \textit{Cf}) were selected.\footnote{All three of us authors, who are South Africans by birth or upbringing and familiar with SAE phonetics, were unanimous in our informal judgments about which speakers in the survey qualified as possible crossover speakers.} Respondents were asked to judge from two playings of each clip whether they thought the speaker was Black or White, or to indicate whether they were unsure. In order to soften the emphasis on race (still a sensitive topic in the country, though less so for the young respondents), two additional ‘distractor’ categories were added: respondents had to judge the sex of each speaker (male or female) and their age (under thirty or over thirty), which we do not report on here.

The respondents were 151 first-year students of linguistics, divided into nine tutorial groups of seventeen students on average. This summary gives the details of a subsample of only the Black and White respondents ($n = 127$). Implicational or Guttman scales (see Abdi 2010) showed that none of these 127 respondents gave judgments that placed them as outliers on an implicational scale. Table 3 gives the breakdown of the judgments for each of the thirteen speakers by gender, while Figure 10 presents this information in a more accessible graphic.

\begin{table}[h]
\begin{tabular}{|c|c|c|c|c|c|c|c|}
\hline
\textbf{FEMALE SPEAKERS} & \textbf{'BLACK'} & \textbf{'WHITE'} & \textbf{'NOT SURE'} & \textbf{MALE SPEAKERS} & \textbf{'BLACK'} & \textbf{'WHITE'} & \textbf{'NOT SURE'} \\
\hline
F1 & 126 & 0 & 1 & M1 & 125 & 0 & 2 \\
F2 & 120 & 2 & 5 & M2 & 118 & 2 & 7 \\
F3 & 72 & 30 & 25 & M3 & 117 & 4 & 6 \\
F4 & 54 & 53 & 20 & M4 & 115 & 2 & 10 \\
F5 & 49 & 55 & 23 & M5 & 76 & 20 & 31 \\
F6 & 45 & 51 & 31 &  &  &  &  \\
F7 & 14 & 96 & 17 &  &  &  &  \\
F8 & 14 & 107 & 6 &  &  &  &  \\
\hline
\textbf{TOTAL} & 494 & 394 & 128 & \textbf{TOTAL} & 520 & 28 & 56 \\
\hline
\% & (48.6\%) & (38.8\%) & (12.6\%) & \% & (86.1\%) & (4.6\%) & (9.3\%) \\
\hline
\end{tabular}
\end{table}

Table 3. Judgments of 127 respondents regarding ethnicity of each of thirteen Black speakers, grouped by gender.

The overall difference in responses to male versus female subjects is as follows: 48.6\% of responses took the female subjects to be Black, in comparison to 86.1\% taking the male subjects to be Black. Or equally starkly, 38.8\% of responses took the female subjects to be White, versus 4.6\% taking the male subjects to be White. To
ascertain the statistical evidence for the validity of this initial summary, it is necessary to consider the judgments of each of the 127 respondents regarding each of the thirteen subjects, and then to ascertain whether there is a consistent difference between their individual responses to female versus male subjects. The most appropriate test for this purpose is the Wilcoxon signed rank test (see Butler 1985:102–9), which ranks and compares the magnitude as well as the direction of differences between responses to the male versus female groups. The test takes into account the rank of each respondent in terms of the contrasts of relative frequency of correct responses to each gender group. The tests (whose statistics can be found in Mesthrie et al. 2015:399–402) show overwhelming evidence that the proportions of males exceed the proportions of females identified correctly. In fact, only two (of 127) respondents identified a greater proportion of females than males correctly.

The broader claim is thus amply supported: overall, young women at the University of Cape Town are in the lead over young males in being perceived to cross over from a traditional BSAE accent space into the space once associated with WSAE alone. This conclusion relies on the assumption that the selection of the thirteen Black speakers was not gender-influenced or biased. The final section considers the extent to which these patterns occur more widely in society, outside the university student sample.

7. Supportive media-based evidence. A third body of evidence regarding gender differences in the crossover variety is more circumstantial. While each piece of evidence may not stand the test of statistical rigor, it is cumulatively suggestive of grand changes in Black urban culture and lifestyle nationally, beyond the student sample of this article. Women feature more prominently in the reportage of this new class/status formation. In the wider society, crossover speakers are numerous and visible among young females; this is less true of young males. While we had no difficulties filling cells for both genders from our previous and current corpus for Cf, Nf, and Nm, it was harder to locate subjects for Cm. Moreover, within the thirteen Cm speakers, it was much harder to find males who sounded sufficiently crossover to use in the perception test. In addition to the earlier media evidence on the salience of gender changes, I cite a few further examples showing their broader significance. The novel Coconut, mentioned in §2, offers enormous insight into the sociopolitical and class/status changes experienced by young people in the post-apartheid era. The theme of language and cul-
tural loss is best exemplified by the female protagonist, Ofilwe, who has become totally immersed in a peer group in which White girls predominate (and within which she is rechristened Fifi). Toward the end of the main section of the novel, she makes a desperate and poignant attempt to compile lists of Pedi words so as to build up the competence that she never had (on account of parental choices over language use and schooling). The novel suggests a strong gender difference in these matters. Ofilwe’s brother, Thabo, an outspoken critic of his sister’s ‘whiteness’, solves the dilemma of cultural and linguistic loss much earlier—and more easily—by finding a part-time job at a fast food outlet. There he learns to speak to working-class Black people in Tswana and concretely reconnect with his roots. The theme of reconnecting away from the formerly White but now multiracial suburbs back to the Black townships is a motif one occasionally encounters in the press, where the exemplars are, this time, male.

Newspaper reports commenting on Black male crossover accents or using males as exemplars are rare. The only example in our database comes from 2013, in a sports gossip column (Sunday Times, 17 February 2013), alluding to the crossover accent of a soccer player and that of the ex-national captain: ‘One beer-swirling patron quipped “Geez, with that accent, the boy can replace Bongani Khumalo as Bafana Bafana captain”. In 2013 it was still newspaper copy to point to a male crossover accent, whereas this had long ceased to be news regarding women (by at least fifteen years).

Dramatic gendered differences can also be seen in the use of English on radio and television. Black females are prominent in news reading, as anchor persons, interviewers, sportscasters, and so forth. Black males in these roles generally have a noticeably non-crossover accent. In the early days of post-apartheid democracy, the few males with such accents were either not from South Africa or in one case had been adopted early in life by a White family.41 This observation of mine is worthy of closer empirical study, though it is also noticeable that at the time of the writing of the first draft of this article (2013), the number of male crossover accents on television is starting to grow—again reflecting a lag of about fifteen years.42 In one important domain, an empirical report is possible, from my regular monitoring of the country’s showpiece morning-radio news program. Originally begun in the 1970s as ‘Radio Today’, this was a program in which British-oriented or Respectable SAE accents predominated, with a ‘Whites only’ policy until the 1990s. In the post-apartheid era, the program was renamed ‘AM Live’, the race policy scrapped, and two anchor persons appointed, one male and one female. Since 1994 this prestigious post of anchor person has always had Black females, whose accents were impeccably crossover. The first of these in the early 2000s, Nikiwe Bhikitsha, was often taken to be White by new listeners of my acquaintance, especially since she was introduced with a flourish in ambiguous phonetics as (possibly) Niki Webekitsa (with an East European-sounding surname, with schwa at the end and not BSAE [a]). Almost all of her female successors had crossover accents and, like her, went on to even more prestigious careers in television, and in one case an international posting with CNN. In contrast, Black male presenters only appeared in 2007, and none had a crossover accent comparable to those of their female counterparts. It is striking that at the time of writing the ‘one male—one female’ desideratum has been replaced by two females, one of whom is Black with a crossover accent. It would appear that either there are not many Black males with crossover accents who qualify for the position, or that

41 This practice was common enough to warrant a two-page report in the Sunday Times, 10 April 2011.
42 Whether hiring policies of Black males versus females by accent is relevant is a question worthy of investigation by media and political specialists.
those who do are not deemed to project the right political nuance for the national morning news. I conjecture that a female crossover accent (with substrate erasure) is deemed sociopolitically more acceptable than a male equivalent to the Black majority of listeners. Keeping in mind these differences, I conclude by referring to some broader international debates.

8. CONCLUSION: GENDER, STATUS, ETHNICITY, AND SOCIOSYMBOLIC MEANING. Drawing mainly on Labov’s early New York City studies, Bourdieu (1977) wrote insightfully about the symbols and distinctiveness of the dominant classes in Western society. In their speech is evident the confidence of those who are taken as the embodiment of society’s overt notions of prestige. Bourdieu stresses the ‘relaxation in tension’ characteristic of the speech of this class. This relaxed production of prestige forms of speech can only be acquired by protracted and ‘precocious’ familiarity with the dominant educational system in its more formal manifestations, as well as in the informal peer group interaction outside the classroom. Bourdieu (1991:85) uses the term ‘control’ for the upper-middle-class attention to the norms of the prestige code, and to practices like the avoidance of prescriptive ‘error’ and exaggeration, and keeping a distance from one’s utterances, rather than ‘surrendering without restraint or censorship to their expressive impulse’.

There appear to be several kinds of status operating in this study. The Cf females are largely part of an upper-middle-class-to-be, who do seem to invest more readily in the present status hierarchy of South African society, previously dominated by White, Western practices. Such practices promote certain ideals regarding womanhood in the media, advertising, and to some extent in everyday interaction too. The status of Cf females reflects a real foothold in the new class hierarchy afforded by early education and social interaction at the best schools, which is in turn afforded by parents’ ability to pay high fees. Given the limited options available to Black children in the past, their easy integration into the social and educational worlds of school, university, and the world beyond is already a success story (see n. 46 for a cautionary qualification). Some female students in our interviews spoke confidently of aspirations for the future in ways inconceivable to their parents’ generation: positions at the United Nations, national political leadership, being an anchor person on national television, and so forth. They display the attributes associated with such status positions already: relaxation in tension associated with not just accent, but also dress, social attitudes, knowledge of the wider world, and friendship across ethnic groups. For the Cf females particularly, status is thus a reflection of real achievement, not a symbolic compensation by women for economic, occupational, and other restrictions placed on them in male-dominated societies, as theorized by Trudgill (1972) in his study of gender and variation in the United Kingdom. A crossover accent is a kind of cultural capital in itself, which will enable entry into important prestigious jobs associated with the talk market.43 At the same time it is a symbol of status and distinctiveness, of belonging or aspiration to the highest rungs of a new social ladder.

The second kind of status, which is more aspirational, relates to Trudgill’s (1972) conjecture cited above (of which see criticisms by, inter alia, Cameron 2005). This kind of status might well apply to the Nf females, who have not been to model-C schools but who line up with Cf norms far more than Nm males line up with Cm norms. (A note of caution must be sounded that this analysis would apply to English only, rather than a full multilingual repertoire.) The greater attraction of women to high-status forms of English may also be associated with a somewhat better role and status allocated to edu-

43 Bourdieu (1991) distinguishes different types of capital: economic (material wealth), social (based on connections and group membership), cultural (based on knowledge, skills, and education), and symbolic (accumulated prestige, honor).
cated women within Western and English society groups than within traditional African language groups (see Herbert 1990, Dowling 1998).

Where does this leave the men, who in the interviews are versatile too, if in other ways? Though their range of prestige variants lags slightly behind the women, it is unlikely to be related to ability (as the section on vowel length suggests). The reasons must relate to indexicality too: men are expected to be strategically streetwise when necessary, even if they come from the new middle classes. This entails using not just an urban form of a local African language, but a streetwise instantiation known variously as Tsotsitaal, Iscamtho, and so forth, which all of our male interviewees claimed to command to varying degrees. An African solidarity for males discourages substrate erasure, at least in the early stages of post-apartheid society. For example, despite a Western education at a prestige school, several male interviewees said that they had to undergo a traditional initiation ceremony at a week-long rural retreat and be taught traditional values pertaining to African ideas of manhood. Overall, the unfolding class and gender pressures work toward a greater status orientation for young middle-class women, as against a combination of the overt prestige of English and the covert prestige of multilingualism for men. This loyalty to an African language is most strongly expressed by the Nm males in the interviews, two of whom openly criticized young females in their neighborhood for speaking more English ‘than they needed to’. In this high language loyalty the Nm males are reminiscent of their male elders. De Klerk (2000) provided statistical evidence from Grahamstown (Eastern Cape) that it is mothers rather than fathers who are the prime movers in sending their children to model-C schools. De Klerk points to the conservatism of the males, evident in comments like the following: ‘my husband was hesitant; he didn’t want them to lose their language’ and ‘my husband is very indigenous, he doesn’t have very many ambitions for his children’ (2000:104).

The sociolinguistic literature in the United States has been grappling with notions of race, class, and gender identities, which offer pointers to understanding these South African dynamics. ‘Whiteness’ in the United States is associated with middle-class affiliation (Barrett 1999), education, and/or intellectual orientation (Trechter & Bucholtz 2001). Fought (2006:116) notes further that ‘one ideological pattern that seems a bit more complex, and does not follow directly from a social position of dominance, is that whiteness does not seem to be associated with masculinity’. In the United States masculinity appears to be associated more with ‘blackness’ (though the studies appear to be ignoring the masculinities of the White and Hispanic working classes). These characterizations are resonant in multilingual South Africa, though there must surely be many shades of gray between the polarities. The Cf females of this study do not appear to be conflicted by considerations of whiteness versus blackness. They appear to value their current status while also valuing the cultural norms of an older—usually grandparental—generation. They also do not see their linguistic behavior as ideologically ‘White’, stressing instead that their English is ‘normal’ and of a standard South African and international/global nature, rather than tied to an older generation’s L2. Gender and realignments of class appear to trump ethnicity here.

The effects noted in this article are thus not purely ‘mechanical’, in the sense of women using more prestige forms of English simply on account of a greater degree of contact with White students. The Nf females confirm this since they differ from their

---

44 Indigenous here means ‘(over)attentive to traditional culture’.
45 This might be true, though, of a small proportion of Cf speakers and their networks in which Whites feature prominently.
male counterparts without having had any model-C schooling. Labov (2001:291) proposed that ‘in stable situations, women perceive and react to prestige or stigma more strongly than men do, and when change begins, women are quicker and more forceful in employing the new social symbolism, whatever it might be’ (my emphasis). This African case study shows that Labov’s proposition can be extended to the acquisition of a new elite crossover variety that accompanied and entrenched middle-class formation within a short period of a post-apartheid order.46

REFERENCES


---

46 Since completion of the final draft of this article there has been something of a backlash against the ideological certainties and aspirations of the middle classes, culminating in the #RhodesMustFall postcolonial campaign, ironically at the very university of this study. The force of this campaign cautions us not to assume the perpetuation of these middle-class ideals; perhaps the Cm group were prescient in their sociolinguistic ambivalence. Moreover, older BSAE—a clearly ‘African’ variety in its substrate fidelity—remains a significant and potent variety in politics, Parliament, radio, television, and public events.


thoughtleader.co.za/silwane/2008/03/19/im-a-coconut-and-im-proud-of-it-say-it-with-me/.


Rosenfelder, Ingrid; Josef Fruehwald; Keelan Evanini; and Jiahong Yuan. 2011. FAVE (Forced Alignment and Vowel Extraction). Online: http://fave.ling.upenn.edu.


Rosenfelder, Ingrid; Josef Fruehwald; Keelan Evanini; and Jiahong Yuan. 2011. FAVE (Forced Alignment and Vowel Extraction). Online: http://fave.ling.upenn.edu.


Trudgill, Peter. 1972. Sex, covert prestige and linguistic change in the urban British English of Norwich. Language in Society 1.179–95. DOI: 10.1017/S0047404500000488.

UCT Unilever Institute of Strategic Marketing. 2012. 4 million & rising: Black middle class expanding. (University of Cape Town strategic marketing report.) Online: http://www.uctunileverinstitute.co.za/research/4-million-rising/.


[rajend.mesthrie@uct.ac.za] [Received 23 November 2013; revision invited 14 September 2014; revision received 29 September 2015; revision invited 9 January 2016; revision received 21 March 2016; accepted 16 June 2016]