

Codeswitching, Borrowing and Mixing in a Corpus of Xhosa English

Vivian de Klerk

*Department of English Language and Linguistics, Rhodes University,
Grahamstown, South Africa*

The paper analyses selected aspects of the codeswitching behaviour in a spoken corpus of the English of 326 people, all of them mother-tongue speakers of Xhosa (a local African language in South Africa), and all of whom would see themselves as Xhosa/English bilinguals. The corpus comprises approximately 550,000 transcribed words of spontaneous, relaxed, oral discourse in English between pairs of Xhosa-speaking interlocutors, discussing a wide range of topics. While the usual pattern in bilingual speech is to use the L1 as matrix language and the L2 as embedded language, in this corpus the opposite is the case, as interlocutors were interviewed in English (the L2). The corpus therefore offers a 'mirror image', in a sense, of normal codeswitching behaviour. Using Wordsmith (a concordancer programme), all incidences of codeswitching into Xhosa during these conversations were identified and analysed in an effort to reveal underlying patterns. Examination of the amount and nature of codeswitching in the corpus promised to throw some light on the extent to which participants are genuinely bilingual, in terms of their ability to converse comfortably in English.

doi: 10.2167/beb382.0

Keywords: codeswitching, Xhosa, English, bilingual, identity

The Context: Xhosa and English in South Africa

Since 1996 South Africa has had 11 official languages: English, Afrikaans (these used to be the only official languages) and 9 indigenous African languages (Zulu, Xhosa, Sotho, Tswana, Pedi, Venda, Swati, Ndebele and Tsonga). Speakers of these languages are widely scattered across a huge geographical area, mostly in big cities, but many live fairly isolated, rural existences, and seldom encounter speakers from other languages. In contrast to the demographic dominance of English in many countries, in South Africa, English is the first language of only 8.2% of the people (see Table 1). However, while English is not the numerical majority language in any of South Africa's provinces, it has the widest and most general distribution of all languages countrywide (although its speakers are mainly distributed in the Western Cape, KwaZulu Natal and Gauteng). Each of the indigenous languages (if Afrikaans is excluded) is found mainly in a particular province, but English is found (and generally understood) throughout the country, and it is not associated with the negative connotations that Afrikaans often brings with it, owing to its unfortunate historical links.

Table 1 Linguistic distribution in South Africa (Census 2001) (total 44.8m)

<i>Zulu</i>	9.2m (23.8%)	<i>Sotho</i>	3.1m (7.9%)
<i>Xhosa</i>	7.2m (17.6%)	<i>Tsonga</i>	1.8m (4.4%)
<i>Afrikaans</i>	5.8m (13.3%)	<i>Swati</i>	1.0m (2.7%)
<i>Pedi</i>	3.7m (9.4%)	<i>Venda</i>	0.9m (2.3%)
<i>English</i>	3.5m (8.2%)	<i>Ndebele</i>	0.6m (1.6%)
<i>Tswana</i>	3.3m (8.2%)	<i>Other</i>	0.2m (0.5%)

In educational contexts all over the world, at any time, students (many of them linguistic minorities) are acquiring and performing in new languages (Miller, 2004: 290), and their identities are constructed in their discursive practices, and their cultural and social contexts. Like many globalised societies, multilingual contexts are a fact of life in South Africa, and the most common pattern is for people to be bilingual in an indigenous African language plus English (or, less frequently, Afrikaans). While levels of competence among bilinguals are notoriously difficult to measure accurately, and although the problem is further complicated in the South African context by the fact that the variety of English used by many is not the 'standard' international variety best known to the world, nonetheless it is fairly uncontroversial to claim that most of South Africa's people are bilingual, and that English leads Afrikaans as the most widely spoken second language in the country (see Table 2). In South Africa, English is undeniably an instrumental asset, and a means to economic and social advancement, it undoubtedly carries positive connotations of modernity and internationalisation, and those who learn it would be likely to be motivated in terms of its effect on their persona and social identity (Gal, 1979). Increasingly, those with power and privilege are choosing English, given its status as a symbol of education, international mobility and modernity (de Klerk, 2000a, 2000b). In this respect, elite closure (Myers-Scotton, 2002: 35) is undoubtedly occurring, with standard English acting as 'a tactic of boundary maintenance' (Myers-Scotton, 2002: 35), revealing who has greater social mobility and power.

Counteracting this power and appeal of English is the fact that, since 1995, several governmental bodies (including the Pan South African Language Board, 9 Provincial Language Committees and 11 National Language Bodies and Lexicographical Units) have gone to great lengths to promote awareness of languages and to preserve and strengthen the indigenous languages and their associated heritage. Thus, while English continues to be an important marker of status and education in the country, at the same time the retention of Xhosa to mark ethnic status is equally important, promoting bilingualism rather than ultimate monolingualism. While bilingualism in South Africa is therefore regarded as totally normal, levels of bilingual competence depend heavily on individuals' acquisitional history and the functional uses of each language.

Table 2 Distribution of first language by province (in %)

	Af	Eng	Nd	P	So	Sw	Tso	Tsw	V	X	Z
W Cape	59.2	20.3	0.1	0.0	0.4	0.0	0.1	0.0	0.0	19.1	0.1
E Cape	9.6	3.7	0.0	0.0	2.2	0.0	0.0	0.0	0.0	83.8	0.04
N Cape	69.3	2.4	0.0	0.0	0.9	0.0	0.0	19.9	0.0	6.3	0.3
Free State	14.5	1.3	0.2	0.2	62.1	0.1	0.5	6.5	0.1	9.4	4.8
ZK Natal	1.6	15.8	0.0	0.0	0.5	0.1	0.0	0.0	0.0	1.6	80
Gauteng	16.7	13.0	1.6	9.5	13.1	1.3	5.3	7.9	1.4	7.5	2.2
Mpumalanga	8.3	2.0	12.5	10.5	3.2	30.0	3.5	2.7	0.1	1.3	25
N Province	2.2	0.4	1.5	52.7	1.1	1.2	22.6	1.4	15.5	0.2	0.7
N West	7.5	1.0	7.5	4.0	5.1	0.5	4.7	67.2	0.4	5.4	2.5

Af, Afrikaans; Eng, English; Nd, Ndebele; P, Pedi; So, Sotho; Tso, Tsonga; Tsw, Tswana; V, Venda; X, Xhosa; Z, Zulu
 Source: Population census (1996: 11)

The Xhosa English Corpus

Among the frameworks that have been proposed to deal with the emergence of relationships among New Englishes is the distinction between English as native language (ENL), English as second language (ESL) and English as foreign language (EFL) (Schneider, 2003: 237). This corresponds loosely to Kachru's (1988) subsequent well known 'three circles of English' model, which differentiates between countries in the inner circle, the outer circle and the expanding circle. Norms and standards in these expanding circle varieties are less and less dependent on those of the inner (ENL) circle. These attitudes correlate fairly closely with the three phases in the emergence of standards in English identified by Gill (1999) in work on English in Malaysia. The phases range from an exonormative phase (pre-independence), with heavy emphasis on external norms, through a liberation and expansion phase, exhibiting a shift to internal norms, and finally end in an endonormative (independent) phase, in which robust internal norms have replaced external norms. More recently, Schneider's (2003) five-part cyclical, dynamic model seeks to explain the predictable patterns of evolution through which English develops all over the world, and to reveal the shared underlying process that drives the formation of New Englishes whenever the language is transplanted. This paper concerns itself with an ESL variety of English used in the expanding circle, namely Xhosa English (XE), the type of Black South African English (BSAE) that is used by the Xhosa-speaking people of South Africa (de Klerk, 2002, 2003, 2006).

The decision to collect a corpus based exclusively on the English of Xhosa speakers, as opposed to that of speakers of the other indigenous African languages, was based on the fact that although the nine languages form four natural linguistic groups (e.g. Zulu, Xhosa, Swati and Ndebele form the Nguni group) with some shared characteristics, the language groups themselves differ quite significantly, and are likely to have influenced English in different ways over time. Given the low levels of English competence and tuition during 60 years of neglect in the 1900s, and in light of the localised and isolated nature of the many separate linguistic communities that have evolved over the years, many without access to radio or modern electronic media such as TV or computers, it is argued that very different varieties of English are likely to have evolved among different linguistic communities, who have shared different mother tongues, values and traditions. Xhosa speakers of the Eastern Cape were therefore targeted for the XE corpus.

All contributors had to be at least 15 years old or in Grade 10. They either had to have been exposed to formal English tuition at school for at least 8 years or have had a more limited education but at least 20 years' exposure to normal use of English in their daily lives. Contributors also had to have been life-long residents of the Eastern Cape Province and currently reside there. The criteria regarding education level and years of exposure to English aimed to exclude those speakers whose English would better be defined as a learner inter-language that was still undergoing development. All conversations recorded in the corpus are spontaneous, unrehearsed, face-to-face spoken English, since, overall, it can be argued that natural, unmonitored, unprepared speech would

best reflect the real and authentic character of XE, especially as it is more than a little controversial to claim that any true variety of BSAE resides in the written mode at all (de Klerk, 2002).

In the case of the corpus data, although speech was between mother-tongue interactants who were known to each other, with no outsider (mother-tongue English) observation or participation, this use of English could be viewed as somewhat unnatural, under the circumstances, as two fellow Xhosa speakers would be likely to use Xhosa with each other. However, this formal bias in the corpus is not viewed as a problem, as these Xhosa speakers would typically only use English in slightly formal circumstances, when engaging with English speakers (usually strangers). In this sense, then, the corpus arguably does represent what is most typical for these bilinguals, when using English (de Klerk, 2006).

Field workers in the project were also mother-tongue speakers of Xhosa from the Eastern Cape Province. They were encouraged to arrange one-on-one informal 'chats' with members of their own linguistic community (friends or family members) in a quiet and private place, with a maximum of three participants.¹ The participants were reminded that they should simply speak in their normal English (i.e. not to try to speak 'well'), and could talk about any topic they liked, which they would normally talk about (e.g. friends, work, family, holidays, sport, school etc.). They were urged not to worry if they used the odd Xhosa word, and not to be concerned about laughing, colloquialisms or swear words.²

Language and Identity

The recent sociopolitical and socioeconomic changes in South Africa, after the declaration of its new 11-language policy, and all the efforts to uplift the formerly disadvantaged indigenous languages have resulted in a new range of identities being available to speakers. The complexity of identities in postmodern societies such as South Africa, where language is not only a marker of identity 'but also [a] site[s] of resistance, empowerment, solidarity and discrimination' (Pavlenko & Blackledge, 2004: 4), offers interesting insights into the close links between language and identity, and into those settings where languages function as markers of ethnic identity, those in which it functions as a marker of symbolic capital or a means of social control, and those where these multiple roles may be interlinked.

Studies in the South African context suggest just such a complex situation, as evidenced by Ramsay Brijball's (2004) analysis of the codeswitching behaviour of Zulu first language students on the multilingual Durban Westville campus. She shows that these speakers use Zulu English code-switching, which she describes as a mixed and stigmatised variety in the eyes of the Zulu speakers, in a way that exposes the hybrid nature of their linguistic, cultural and social identity. This codeswitching, she argues, expresses a dual identity, and the greater use of a mixed variety in comparison to the use of either a monolingual English or Zulu variety shows that the speakers seek a negotiable identity for themselves.

Speakers in the XE corpus, however, are different from these Zulu speakers in Natal, because they have experienced constrained circumstances. Living in the Eastern Cape, which has high levels of unemployment and concomitant socioeconomic problems, their educational opportunities and exposure to English are very likely to have been fairly limited, compared with more urban South African bilinguals, who would arguably be more vulnerable to language shift to English. Also, because of their own constrained personal circumstances, they are not involved in the international linguistic marketplace to any significant degree (Bordieu, 1991). Despite the multilingual nature of the country as a whole, in their day to day life they interact mainly with other Xhosa speakers, and they are less likely to see language as political capital in the same way as people living in multilingual environments would. The dominant language of the speakers in this corpus is Xhosa, and it is very likely to remain Xhosa. Nonetheless, English has a strong presence in their lives as well (the database reflects high reported levels of English usage) and codeswitching occurs frequently.

Codeswitching and Bilingualism

Codeswitching is the use of more than one variety or language in the same conversation. Pavlenko and Blackledge (2004) provide an excellent overview of the theoretical progression in the linguistic description of codeswitching behaviour over the past 20 years, moving from interactional sociolinguistic approaches, such as those of Le Page and Tabouret-Kelly (1985), through to poststructuralist approaches, to negotiation of identities. Gumperz (1982) stresses the importance of discourse in codeswitching, giving clues about conversational contributions such as certainty and change of topic. Using data mostly from conversations studied in Kenya, Myers-Scotton (1995) focuses on the social motivations for codeswitching, in which speakers use two or more linguistic varieties in the same conversation, and shows how speakers engaging in codeswitching exploit the sociopsychological values which have come to be associated with different linguistic varieties in a specific speech community, relying on the notion of markedness (Myers-Scotton 1998; 1999) to negotiate a change in social distance between themselves and other participants in the conversation.

While many people who know two languages are sufficiently bilingual to switch codes strategically and skilfully, following underlying social or stylistic motivations such as commenting on perceptions of self, topics or context, expressing nuances such as these is not the only important factor in codeswitching. Other scholars (e.g. Auer, 1998; Jacobson, 1998) have pointed out that in many cases, shift between two languages is explained not so much by negotiation of social identity, but rather by examining the relative linguistic abilities of the interlocutors. Zentella (1997: 98) refers to this as the 'crutch' syndrome, where 'a bilingual who is stumped in one language can keep on speaking by depending on a translated synonym as a stand-by.' In similar vein, Giampapa (2004: 193) also makes the point that it is not only multiple and shifting identities that are the root cause of codeswitching behaviour, and suggests that incomplete competence in certain linguistic domains may also

play a fairly important role. For this reason, the analysis of macro-linguistic aspects of codeswitching, while valuable, is not sufficient in and of itself; close attention to micro-linguistic aspects is also important (Pavlenko & Blackledge, 2004: 10). Using such a microlinguistic approach, this paper attempts to offer some different explanations for the codeswitching behaviour in the corpus of Xhosa English, keeping in mind that the speakers in the corpus are possibly not fully bilingual.

Bilingualism seldom involves the 'the habitual, fluent, correct and accent-free use of two languages' (Paradis, 1986), as using this definition few individuals would qualify as complete bilinguals. More often, bilinguals have different abilities in their languages, or use them in different domains, and Grosjean's (1982) definition of a bilingual as someone who uses two or more languages in everyday life (i.e. a definition in terms of language use rather than proficiency) is probably more realistic. While generalisations can only be made with caution, it is probably safe to assume that the speakers in this corpus are not 'equal' bilinguals, as their competence in Xhosa (the L1) is significantly better than in English (the L2). Nonetheless, they are undoubtedly bilinguals in terms of Grosjean's definition, and they use Xhosa (L1) and English (L2) with some regularity. Kachru (1986) makes the point that codeswitching reveals one's linguistic repertoire and one's cline of bilinguality and degree of proficiency.³ Codeswitching behaviour in the corpus will be unlikely to be a genuine matter of choice (of the marked versus unmarked code) to make a social point; it is likely that speakers will also be constrained by their actual ability and level of fluency in the two languages.

One also needs to take account of the other unique circumstances that prevailed in the collection of the corpus itself: firstly, in terms of the classical codeswitching model, in this corpus English is the matrix language (or structurally dominant language) and Xhosa is the embedded language. Thus English is the source of the abstract morphosyntactic frame for the bilingual clause, providing the morpheme order. This is the opposite of the normal situation prevailing for these speakers, and what is interesting in this corpus is to see how much codeswitching occurs when these bilinguals are *not* using their usual matrix language – Xhosa.

A second difference comes from the unique context in which the data was gathered: interlocutors were not holding conversations for any obvious or particular social purpose or function, such as to exchange goods or services, or to give or request information (using Halliday's (1994) model of the functional purposes of language). The usual underlying motivation for informal conversations is a genuine social need to reaffirm social links and to catch up with gossip or news, alongside more subtle attempts to influence opinions. In contrast, these speakers were engaging in an informal interview of sorts, with the interviewer (an L1 Xhosa speaker) aiming to obtain as much 'natural' conversation from the interviewee (also an L1 Xhosa speaker) as possible, which would inevitably have added a touch of *irrealis* and formality to the encounter, given that both participants were fully aware of being recorded. In addition, while they knew that it was fine to use a little bit of Xhosa here and there, they were following an explicit instruction to speak English.

Thus the sociopsychological aspects of using different languages in order to achieve such aims as to negotiate a change in social distance between themselves and the other participant are far less likely to prevail. And the codeswitching in the XE corpus needs to be viewed through different lenses, as it were, with the range of functions that it serves likely to be reduced. Prime among these would be its use as a marker of identity, and its use as a strategic coping device because of linguistic limitations. In other words, while codeswitching can convey procedural meaning (in that every time we make a linguistic choice, our previous experiences remind us of what the 'safe' or unmarked option would be), marked choices are only a skilled negotiation of identity if one does have an option in the first place.

Methodology

All instances of a switch to Xhosa were annotated as such during transcriptions of the corpus, making it possible to identify them during analysis. From all of this data, three categories of codeswitching were selected for closer analysis, all of which aimed to throw some light on the actual degree of bilingualism manifested by the speakers in the corpus. The first category was instances of codeswitching on selected content morphemes (further details follow below). The second category included instances of codeswitching that might be obviously related to lack of competence in English, linked to overt signals of hesitation or uncertainty in the speech (e.g. long pauses, the use of 'um', or 'uh', and other markers of vagueness). The third category comprised instances of codeswitching which appeared to serve a genuinely communicative purpose, such as affirming identity, and which did not suggest heavy reliance on Xhosa. Such usages would indicate confidence in both codes. All three categories would, it was hoped, ultimately throw some light on general levels of bilingualism in the speakers.

The first of these categories is based loosely on Myers-Scotton's (2002) 4-M model (a model of morpheme classification), in which content morphemes would be nouns and verbs and prepositions (those words which assign or receive thematic roles such as agent, patient etc.). (Other morpheme types are early system morphemes, bridge late system morphemes and outsider late system morphemes.⁴)

The lower down in the order (and the further distant from the mental lexicon), the harder these morphemes are to learn, and the more likely, in terms of the model, that they will come from the matrix language. These different types of morpheme are accessed differentially during speech production, with the first two types being accessed at the level of the mental lexicon, and the latter two types later on, at the level of the 'mental formulator' (Myers-Scotton, 2005: 3). Because the basic split between content morphemes (the lexicon) and the grammatical elements (system morphemes) is usually revealed in codeswitching data in contact across different languages, these subsystems of the abstract lexical structure are likely to be distributed differently.

In terms of Myers-Scotton's model, even in these reverse circumstances, the matrix language principle will be upheld, and if there is sufficient congruence,

the embedded language content morphemes and conceptually activated system morphemes can be integrated into the ML framework. But when there is not enough congruence between the languages in contact (as is the case with Xhosa and English), then compromise strategies are likely, such as the use of bare lexical forms from the embedded language in the matrix language, or the occurrence of embedded language 'islands', framed by the ML. As Xhosa and English are not congruent languages at all, one expects to find such strategies in the corpus.

High levels of usage of Xhosa nouns, verbs and prepositions (content morphemes) would be evidence of a *lack* of bilingual ability in the speakers. If, on the other hand, there is minimal evidence of this, then one might argue that the speakers are sufficiently comfortable in English not to rely on Xhosa.

According to Myers-Scotton (2002: 206), the lexicon is viewed as most susceptible to loss, and 'Content morphemes are not only first in language acquisition and in contact situations promoting borrowing, but they are also first out in language attrition', with speakers of a waning L1 showing evidence of forgetting content morphemes. She goes on to say 'there is little or no quantitative evidence to support this assumption; however the attritional literature is full of anecdotal evidence that content words dominate as the elements that are "forgotten" in attrition' (Myers-Scotton, 2002: 206). This means that if XE speakers were in the process of shifting to English, they would not be using many Xhosa nouns in their English, as these would be in the process of being substituted by English equivalents. High frequencies of this sort of codeswitching, on the other hand, would imply a fair degree of reliance on Xhosa while speaking English. Also, if they do not use many Xhosa nouns (other than traditional words for which there is no English equivalent), this could be interpreted as showing a high level of bilingualism generally.

Results

Altogether there were 1117 instances of the use of Xhosa in the corpus. Given a total of 548,940 words, this is a very low rate of codeswitching (approximately 2 switches every 1000 words). This already suggests a fair level of confidence and competence in using English.

Codeswitching on content morphemes (nouns, verbs and prepositions)

Nouns and verbs for cultural practices and concepts

The corpus revealed a fair number of uses of isolated Xhosa nouns that represent cultural or traditional practices, for which there is no exact equivalent in English. The word *ubuntu*, for example, refers to a quality of humanity and goodness, and has been borrowed into South African English because it has no English equivalent. It was used 10 times in the corpus. Similarly, words like *umqombothi* (traditional Xhosa beer) was used 13 times, and has no translatable equivalent. Use of Xhosa in such instances is unexceptional, and should not, strictly speaking, be regarded as codeswitching at all. Other examples of such insertions include *um-fabakhwetha* (× 11); *icheya* (× 6); *ukujiswa* (× 7); *umbhaqanga* (a type of music) (× 11); *umgidi* (× 3); *umguyo*

(× 5); *imphempe* (carrier messenger); *umsobomvu*; *umteyo* (× 4); *umsinqowe* (a traditional ritual); *ubuhlanti* (kraal); *ubuxhwele*; *ubuGqirha* (× ?); *imbola*; *imitshotsho*; *isikaka* (a bed- beaded dress/skirt); *izifombo*; *ukulwakwentonga* (stick fighting); *uqamatha*. Verbs of this type included *ukuqobhoza* and *ukuvumisa*. A few examples of how such words are used in the corpus are given below, and they show how English determiners often preceded the Xhosa word:

- the *ingcibi* supposed to be trained on that thing, the heritage ...
- okay so even there is this this this *nqalathi*
- Secondly the *amakhankatha* must remain ...
- most of the *abakhwetha* they go to the hospital
- ja he is a *madala* (old man) (× 11)

In other cases, the Xhosa nouns that are used are not so much 'traditional cultural' terms, as such, but they do refer, nonetheless, to concepts which have developed in urban black culture. Such words include *tsotsi* (street thug) and *amakwerekweres* (black (illegal) immigrants, usually from Zimbabwe or Mozambique), e.g.

- you know we are not working, so the the people of kazi looking you and say hey the *tsotsi unqivu* because of we are not working.
- when they come here they are called the the *amakwerekweres*

As far as verbs are concerned, there was a fairly high level of usage of the verbs *ukuthi* ('to say') (used as first word 19 times) and *uyabo* ('you know'), most often used singly and somewhat formulaically, rather than as a codeswitch *per se*. This is very similar to the way 'you know' is used by L1 English speakers, e.g.

- but if you say, *ukuthi ja* ...
- no, he feels, *ukuthi* I would ...
- then you realise, *ukuthi* no I have ...
- but I don't know, *uyabo*, but the thing is, I am not so interested.

There were only 10 other verbs used as the initial word of the codeswitch, including the following:

- you see but uh I I must tell you that *ukutyabekha emveni kokhutyabekha*
- and then *ukulapha ezinye izinto*
- ... maybe *ukuhamba ne* girlfriend
- I say I'm HIV negative cause *ndiyil'ukuyozi testa mfana'kithi* as ...

On rare occasions, a Xhosa prefix was attached to an English verb stem to show subject concord (e.g. 'and then *ubuviolet* some of us'; '*abasocialize*'). These would represent late system outsider morphemes in terms of Myers-Scotton's (2002) model, and they are the most likely morphemes to be 'allowed' into the English matrix. Verbs with Xhosa suffixes also occurred a few times (e.g. *developa*; *recondisha*).

Another category of content morphemes is prepositions. English has a great number of phrasal verbs, containing prepositions that act as early system

morphemes, e.g. *give in, look up, come through, come across*, and it is widely agreed that the lexical range of prepositions available in English is far greater than in Xhosa, which relies heavily on locatives such as *e-* and *kwa-*. Uses of these locatives in the corpus was limited, with only eight uses of *e-* and one with *kwa-*:

- there is a work like a farmer like at *k-* at King Williamstown, *eBhayi*
- in my language eh *ecaweni khubalulekhile* because we have ...
- those who are staying there *ecementeni* location
- here here here *ekazi* are not correctly right you know
- even our curtains *ezindlini, uyabo*, you find out
- like when you have to go to *emlanjeni* what you are you have to carry ...
- do you think that is okay, to have ten ministers in *kwa* education?

The last two examples in this list are particularly interesting, in that the locatives are duplicated in both languages ('to *emlanjeni*'; 'in *kwa* education'). What emerges is that, instead of Xhosa prepositions occurring in the embedded language, there is a regular pattern (over 20 instances) of Xhosa nouns immediately following English prepositions.⁵

- what about the things, about *i-future yamaGqirha* eh?
- you get them through *ilanto i-theatres*
- the only thing after *yona* after *i-alcohol* ...
- just because most of *i-youth yethu ihleli apha elokishini uyabo*
- ja I mean like to the question of *lento i-music*
- even the the the rate of *i-crime inga decreasa uyabo* so that ...
- ja ja I know compared to *pha* ne? compared to *le chicki yakho?*
- they must rather concentrate to *i-education* or doing *i-sport*
- we're playing with *utiti uleza no marcus* mm

Words as direct translations

Sometimes in the corpus, Xhosa nouns and verbs are offered as a direct translation straight after or before the equivalent English noun, in an apparent effort to clarify what the speaker is saying – and in most cases the discussion revolved around cultural practices. This type of codeswitching is clearly not caused by linguistic incompetence, but rather it seems to be used in order to remind interlocutors of common or shared understandings:

- I had to go to the sort of *ubuhlanti*, kraal
- ... erosion like *ukhukutise*
- ... to love, *ukhuthandana*
- still eat African salads, *umphokoqo*, you know
- traditional beer, which is *umqomboti* as you know
- is just like *uGqira*, the doctor

In similar vein, in the following two exchanges, speakers echo previous speakers' words in Xhosa, as if to confirm them (in the first example) or to clarify matters (in the second):

IT: they drink there
 NL: *abasocialise*
 AN: everything like what?
 SV: eh eh I like uh
 AN: *intoni?*
 SV: like uh *i-video games*

Nouns with Xhosa prefixes

In addition to the use of full (stand-alone) Xhosa words, there were many cases of the use of Xhosa prefixes on English nouns. In this regard, the *i-* prefix was consistently used (239 times) with non-human entities, and the *u-* and *ama-* prefixes typically for humans. Examples of nouns with the *u-* prefix include personal names and references to people, as well as cases such as *u-fourteen*, *u-poverty*, *u-days*, *u-generations*, *u-government* (× 6) and *u-rush hour*. On 23 occasions, the noun started with *aba-*. Examples of stand-alone usages of the *i-* prefix include *i-crime*, *i-capacity*, *i-computer science*, *i-faces*, *i-point*, *i-political*, *i-churches*, *i-lack*, *i-education*, *i-tutor*, *i-drugs*, *i-pool table*, *i-problems*, *i-rubbers* (condoms), *i-rate*, *i-standard*, *i-video*, *i-sport*, *i-drugs*, *i-sex education*, *i-aids*.

Occasionally there is some doubt as to which language is acting as matrix language, as one can see in the following extract: 'ja Simon let's let's, I mean like, let's talk about *i-i-i-question* of *i-i-lento i-employment ...*'.

When one examines these nouns, it becomes evident that they tend to refer to fairly 'modern' concepts, often with strong sociopolitical links. One interpretation of this would be that the speakers add these prefixes if they judge the words to be borrowings into Xhosa, and that one should therefore interpret such words as Xhosa content morphemes in an English matrix. Another would be that the speakers see themselves as using Xhosa words that have been borrowed into English, in which case they would not be codeswitching at all. The occasional use, in the corpus, of the English determiner in phrases like 'this *i-facilitator*' in a sense duplicating the function of *i-*, suggests that speakers see these Xhosa-ised nouns as English.

Switching at a time of vagueness or uncertainty

When speakers switched to Xhosa after overt pausing or signals of hesitation, it would be reasonable to assume that they were experiencing a word-finding difficulty in English, and resorted to Xhosa to solve the temporary problem. There were 19 clear cases of such switches after the use of the discourse marker 'like', overt signals of hesitation such as 'um/uh', and extended pauses. There were also other cases in which repetitions and other signs indicated word-finding difficulties. The examples below illustrate the nature of some of these codeswitches:

- rather than go to shebeens drinking and and and, like *wabona abafumane* these sexually transmitted diseases ...
- he was in that project of gateway he has done he he like *andiyazi izinto zamagcingo ...*

- you wake up during the morning you kind a like have to pop up maybe hundred rands, to me that is like *i- i- lento i-prostitution* you see.

The word *nantsika*, which means 'whoever'/'whatever', and is used when one cannot remember a word or name, is also used three times, and shows the same speaker hesitation:

- we should be marketing this in a business-minded form you see, but it will remain in eh in a in a *nantsika* in a sort of a non-profit basis you see.
- it became eh a process that that that failed because the community the *nantsika* the funds also for the community media were shrinking.

Similarly, repetition of words also suggested word-finding problems, and tended to herald a codeswitch:

- in the bible that there will be these kind of things you know we are approaching the the the *landuka* the end of the world man
- and she is end-upping ye end-upping to end-upping to *uyabo end up uba bezukwalomtu mani uquba letrucka* since eh ja end-upping ...

Other clear cases of hesitation or repetition that preceded a switch to Xhosa were overt pauses (on 10 occasions) and the use of fillers such as 'uh', 'uh' or 'eh' (9 times):

- with my energy for twenty four hours < 2 sec pause > *uyabo* doing this thing its mm < 2 sec pause > *ngu* hope ...
- i don't know how how can you prevent this < 3 sec pause > *hayi* i can't say when the time comes < 3 sec pause > *ewe kalokhu*
- you have to like study like all all all of the time and stuff and but eh *kumnandi*.

Codeswitching that serves a communicative purpose

The discourse of the interaction also had a significant effect on codeswitching behaviour. 183 of all instances of a switch occurred at the beginning of a new speaker's turn. Often the first word would be a signal of disagreement (*hayi* = no) (43 times), agreement (*ewe* = yes) (24 times) or astonishment (*jo*) (6 times), with the word serving an emotive function before reverting to English, e.g.

TM: to their children?

MJ: *ewe* they don't have a problem with their minds

MJ: ... some of the people who are doing this are very known people

DK: *jo*

MJ: ja ja that's right

KN: people re- relates to stories not so much on characters

MG: *hayi* then what's your point?

Occasionally such uses of Xhosa seem to underline the speaker's personal identity:

MG: at the library?

KN: mhhm

MG: *mna* I use the upstairs section

Another technique for possibly affirming shared identity was the occasional use of Xhosa as an echo of a previous speaker's codeswitch into Xhosa:

TG: I just assumed *ukuba ne*

LP: *ukuba* I knew so you told me

TG: ja ja exactly ja

As mentioned earlier, interlocutors also often inserted a single discourse marker, such as *uyabo(na)* or *ukuthi*, not strictly as codeswitching, but more as familiar signals of engagement, reaffirming solidarity and shared identity and encouraging the conversational partner. While *uyabo(na)* occurred 150 times in the corpus, its use as a discourse marker in this way was particularly noticeable, occurring as the initial word in a turn 11 times and as final word 16 times.

- I've changed my life *uyabona* ja even here the community ...
- and I was like *nam uyabona* like that *uyabona* I was prince charming
- they don't want to listen to guys *uyabo* they just take this democracy ...

In similar vein, *wena* (you) was used 30 times in the corpus as an isolated discourse marker:

- what are you thinking, what do you think *wena* must be must be done?
- because they do not have food and *wena* you are starving because ...

Following the same pattern, *yebo* (yes) was used 11 times in the corpus, and every time it was used singly:

- this day will come *yebo* so I will say I'm not I'm not going
- not as involved as they they have to be *yebo* because at school ...

Playing a similar role in the corpus are repeated instances of idiomatic expressions or formulaic phrases, which were frequently used almost as markers of shared identity, possibly used to maintain the flow of conversation while simultaneously endorsing shared solidarity. For example, *ukuza kuka nxele* means 'something that will never happen', and it was used 12 times in the corpus, usually as a formulaic island between English clauses.

While the corpus also contains fascinating examples of other types of codeswitching, such as switches after verbs that are typically followed by a complementiser (e.g. 'say', 'promise' and 'remember'), space does not permit further discussion.⁶ Switches after 'when' and 'why' also seem interesting (e.g. 'because the people they lose jobs when *ilantika ithegisiwe ilantika*'; 'okay that's why *mabudede ukukhanya ...*'), as does the use of relative pronouns in the corpus, in light of the fact that there are very few 'true' adjectives in Xhosa, so the likelihood of codeswitching and using a

relative clause instead (e.g. 'the boy who is innocent' versus 'the innocent boy') might well increase.

Final Comments

The most obvious aspect of bilingual speech is lexical borrowing, with items going from the more prestigious language to the less prestigious language (Myers-Scotton, 2002: 41). Those items that are very frequently used in the second language eventually become borrowings and arguably achieve lexical status, mentally, and are tagged as legitimate L2 lexical items for the speakers. The high number of English nouns with Xhosa prefixes in this study suggests that a similar process has occurred, and that speakers are probably regarding the Xhosaised nouns they use as fully legitimate English words. In other words, such usages are not actually true instances of codeswitching at all. This exploratory analysis of the corpus also confirms Myers-Scotton's prediction (2002: 41) that the two types of content morphemes that are more frequently borrowed into L2 are words for objects and concepts that are new to the L2 (cultural borrowings that are rapidly integrated into the target language, although often retaining phonological features), and 'core borrowings', which come more gradually into the target language via codeswitching. The data also eventually needs to be examined to see whether speakers are more predisposed to switch codes at the point at which content and early system morphemes occur.

Levels of bilingualism among the speakers in the corpus seem fairly high, in terms of the measures adopted in this analysis: given that many of the original 1117 instances of codeswitching in the corpus are arguably borrowings, rather than codeswitches, and that high numbers of the Xhosa content morphemes that were used tended to refer to cultural traditions or items for which no English equivalent exists, the overall rate of codeswitches is much lower than it might initially appear to be. In addition, many of the single-word codeswitches in the corpus (such as *ewe*, *haai*, *uyabo*, *wena* etc.) appear to serve to affirm solidarity and assert identity rather than to signal difficulty in the target language. The only evidence of codeswitching that might arguably signal less than native-speaker competence in the target language was those cases when speakers appeared to be undergoing some sort of word-finding difficulty, resorting to Xhosa after a clear signal of hesitation or uncertainty in the speech (e.g. long pauses, repetitions, the use of 'um', or 'uh', and other markers of vagueness). It is only these cases that could be used as evidence of low levels of bilingualism, and there are not sufficient cases to make such a claim. In more than half of the 2000-word files, there was no Xhosa used at all, and overall, then, the Xhosa speakers in the XE corpus show confidence in English, and exhibit reasonable levels of bilingualism. Obviously there were a few speakers who tended to use more Xhosa in their speech (e.g. 'and some of the houses *kule area zikhona ezi khuku lisekhayo* you see uh *zi khona ezi canekha kwi floods zimane zibakhona i- i- i- ezi nye izinhlu nazo*'), but these were in the minority, and most of the Xhosa that occurred related to discussions that revolved specifically around cultural and traditional practices.

According to Myers-Scotton (2005), the bilingual speech of L1 speakers (such as these Xhosa speakers) who are in contact with a community-dominant L2 (in this case English) is likely to proceed through quantifiably identifiable stages which form an implicational scale of language shift to the L2. In this case, I would argue that, despite their fairly high levels of bilingualism, the likelihood of imminent language shift to English for these speakers is low for various reasons, primary among them social and economic. Despite the fact that Xhosa could be viewed as a minority language in terms of its socioeconomic standing in the country, these speakers form a tight-knit and cohesive linguistic community, and such a context does not promote language shift (Fase *et al.*, 1992). They have dense and multiple daily contacts that are likely to sustain Xhosa; they are in a numerical majority where they live, and therefore they are not vulnerable to language loss. Although access to well paid jobs and resources are undoubtedly linked to English, increased exposure to the standard variety of the language and its speakers is problematic for them, and it is Xhosa that serves these bilinguals in most domains of their daily lives. In the mean time, concerted government efforts to strengthen local indigenous languages and improve translation and interpreting facilities in the country are increasing, which would reduce the urgent need for speakers of these languages to master English at all. They value their vernacular, and they need English: a situation that promotes bilingualism but not language shift – at least for now.

Correspondence

Any correspondence should be directed to Dr Vivian de Klerk, Department of English Language and Linguistics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa (v.deklerk@ru.ac.za).

Notes

1. Conversations in bigger groups were avoided, and a maximum of 3 participants was advised, because of difficulties in ascertaining speaker identity during transcription.
2. All relevant details regarding the ethnicity, gender, age, occupation and educational background were recorded. The total corpus of spontaneous speech currently consists of 548,940 words and represents the transcribed speech of 326 speakers.
3. In this regard, Finlayson *et al.* (1998: 415) have argued that more proficient bilinguals tend to produce larger embedded language constituents from L2 in the matrix language, but in this corpus the L2 is the matrix language, so the situation should be different.
4. The early system morphemes, which depend on their content morpheme heads for their form, are typically plural affixes and determiners (*those dogs*) and the prepositions of phrasal verbs (*threw up*), which do not assign thematic roles. These first two types of morphemes are more likely to come from the embedded language. Bridge late system morphemes are those which depend on their immediate maximal projections for their form, such as possessive 's', or expletive "it", and outsider late system morphemes are those which depend for their form on information outside their immediate maximal projections, such as subject-verb concord and case ('she gives *him* a lot of problems').
5. Instances in which Xhosa names or terms for untranslatable cultural traditions were used have not been included.

6. There were 15 switches to Xhosa on *ukuba* which means 'if', some of them containing a full embedded Xhosa clause, and others switching immediately back to English. e.g. 'you see because ne if *ukuba ukuba mhlawumbi ngumama okanye utata*'; 'that's why I'm asking *ukuba* what are these hospitals for?'

References

- Auer, P. (1998) *Codeswitching in Conversation: Language, Interaction and Identity*. London: Routledge.
- Bourdieu, P. (1991) *Language and Symbolic Power*. Cambridge: Polity Press.
- de Klerk, V. (2000a) Language shift in Grahamstown: A case study of selected Xhosa speakers. *International Journal of the Sociology of Language* 146, 87–110.
- de Klerk, V. (2000b) To be Xhosa or not to be Xhosa: That is the question. *Journal of Multilingual and Multicultural Development* 21 (3), 198–215.
- de Klerk, V. (2002) Xhosa English as an institutionalised variety of English ... In search of evidence. *English World-wide* 24 (2), 221–243.
- de Klerk, V. (2003) Towards a norm in South African Englishes: The case for Xhosa English. *World Englishes* 22 (4), 463–481.
- de Klerk, V. (2006) *Corpus Linguistics and World Englishes: An Analysis of Xhosa English*. London: Continuum.
- Fase, W., Jaspaert, K. and Kroon, S. (1992) Maintenance and loss of minority languages: introductory remarks. In W. Fase, K. Jaspaert and S. Kroon (eds) *Maintenance and Loss of Minority Languages* (pp. 3–13). Amsterdam: John Benjamin.
- Finlayson, R., Calteaux K. and Myers Scotton, C. (1998) Orderly mixing: Codeswitching and accommodation in South Africa. *Journal of Sociolinguistics* 2 (3), 395–420.
- Gal, S. (1979) *Language Shift. Social Determinants of Language Change in Bilingual Austria*. New York: Academic Press.
- Giampapa, F. (2004) The politics of identity, representation and the discourse of self-identification: Negotiating the periphery and the center. In A. Pavlenko and A. Blackledge (eds) *Negotiation of Identities in Multilingual Contexts* (pp. 192–218). Clevedon: Multilingual Matters.
- Grosjean, F. (1982) *Life with Two Languages: An Introduction to Bilingualism*. Cambridge: Cambridge University Press.
- Gumperz, J. (1982) Conversational code-switching. In J. Gumperz (ed.) *Discourse Strategies* (pp. 55–99). Cambridge: Cambridge University Press.
- Halliday, M. (1994) *Functional Grammar*. Edward Arnold: London.
- Jacobson, R. (ed.) (1998) *Codeswitching Worldwide*. New York/Berlin: Mouton de Gruyter.
- Kachru, B.B. (1986) *The Alchemy of English*. Urbana: University of Illinois Press.
- Le Page, R. and Tabouret-Kelly, A. (1985) *Acts of Identity: Creole-based Approaches to Language and Ethnicity*. Cambridge: Cambridge University Press.
- Miller, J. (2004) Identity and language use; the politics of speaking ESL in schools. In A. Pavlenko and A. Blackledge (eds) *Negotiation of Identities in Multilingual Contexts* (pp. 290–315). Clevedon: Multilingual Matters.
- Myers-Scotton, C. (1995) *Social Motivations for Codeswitching: Evidence from Africa*. Oxford: Oxford University Press.
- Myers-Scotton, C. (2002) *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press.
- Myers-Scotton, C. (2005) The grammatical abruptness of language shift: Why creolists should care. Unpublished manuscript, presented to Society for Pidgin and Creole Linguists, Oakland, CA (USA), January 7, 2005.
- Myers-Scotton, C. (1998) *Codes and Consequences*. Oxford: Oxford University Press.
- Paradis, M. (1986) Foreword. In J. Vaid (ed.) *Language Processing in Bilinguals: Psycholinguistics and Neurolinguistic Perspectives*. Hillsdale, NJ: Erlbaum.
- Pavlenko, A. and Blackledge, A. (2004) *Negotiation of Identities in Multilingual Contexts*. Clevedon: Multilingual Matters.

- Ramsay Brijball, M. (2004) Exploring identity through codeswitching. A post structuralist approach. *Alternation: Journal of the Centre for the Study of Southern African Literature and Languages* 11 (2), 144–164.
- Zentella, A. (1997) *Growing up Bilingual*. Oxford: Blackwell.