Comparing Codeswitching and Borrowing

Article *in* Journal of Multilingual and Multicultural Development · September 2010

DOI: 10.1080/01434632.1992.9994481

1 author:

Carol Myers-Scotton
Michigan State University
97 PUBLICATIONS **3,421** CITATIONS

SEE PROFILE

1992

COMPARING CODESWITCHING AND BORROWING

Carol Myers-Scotton

Linguistics Program, University of South Carolina, Columbia SC 29208, USA

Abstract An issue regarding codeswitching discourse is the extent to which material from a donor language (the Embedded Language or EL) appearing in a recipient language (the Matrix Language or ML) shows internal differentiation. Three questions are relevant: (1) Are all singly-occurring EL lexemes in such discourse borrowed forms, or are some codeswitched forms (CS forms)? (2) If some are CS forms, how are they differentiated from borrowings? (3) What is the relationship of either established borrowings or singly-occurring CS forms to multiword stretches of codeswitching? Working within a model of the structural constraints on codeswitching (the Matrix Language Frame Model), this paper argues that while some important differences do exist among the various forms of EL material appearing in codeswitching discourse, in general the forms arise from related processes. Therefore, any model of the structural aspects of codeswitching must provide a unified account for all EL material in codeswitching utterances.

Introduction

This paper[1] attempts to clarify the status of lexical material from an Embedded Language (EL) which appears in codeswitching discourse in a frame provided by the Matrix Language (ML). Such a discussion is an important preliminary in assessing competing models of structural constraints on codeswitching, since a first step is to identify those linguistic items which properly must be accounted for as part of codeswitching utterances. Codeswitching involves at least two languages used in the same conversation. Of these languages, one is the ML, the language which sets the morphosyntactic frame for codeswitching utterances. The ML can be identified on the basis of relative frequency of morphemes. In turn, psycho- and sociolinguistic factors influence the choice of the ML. One or more languages may serve as embedded languages, providing both singly-occurring lexemes in

0143–4632/92/01/0019–21$1.80/0

constituents otherwise in the ML, and also EL islands, constituents entirely in the EL.

The major interest here is to discuss the similarities and differences between those singly-occurring lexemes from the EL which are borrowed forms (hereafter B forms) vs. those which are codeswitching forms (hereafter CS forms). The paper begins by providing examples of the type of data to be considered. Examples (1) and (2) illustrate sentences from codeswitching discourse. They include EL-origin lexemes in constituents otherwise in the ML. The ML is Swahili in (1) and Shona in (2).

(1) *Setting*: A man from the Luo ethnic group and another from the Kisii ethnic group are talking about the relative merits of some driving schools. They speak mainly Swahili, but with some switches to English.
Kisii: Kitu kizuri ni kwa wewe ku-*visit* karibu kila *shule* ili uone *difference*.
thing good is for you to-visit nearly each school so-that you-see difference
(The best thing for you is to visit almost every school so that you can see (the) difference.)
(Myers-Scotton Nairobi conversation, No. 15)

Note that *shule* 'school' is a long established borrowing into Swahili (originally from German); however the verb stem *-visit* and the noun *difference* are considered codeswitching forms, largely because they show no frequency of occurrence; they occur only in this one conversation in a set of 40.

(2) *Setting*: A welder, who has schooling only through grade seven, is being interviewed in Chitungwiza, a suburb of Harare, Zimbabwe. The interview is in Shona.
Interviewer: Ko kana pa-*tii time* munenge muchitaura nezvei?
(When you have tea time, what do you talk about?)
Welder: Tinotaura nyaya dzekumba nedze *weekend*, *especially Monday* unotaura zvaunenge uchiita pe-*weekend*.
(We talk about our homes and about the weekends, especially Monday you talk about what you did on the weekend.)
(Bernsten & Myers-Scotton Zimbabwe interview No. 21)

Note that *weekend* is a codeswitched form in this data corpus (it appears in two or fewer interviews). *Monday* is a borrowing, appearing in three conversations, as is *especially*, appearing in 10 out of 129 interviews. Frequency as a criterion to establish borrowing will be discussed below.

The Argument of This Paper

Singly-occurring EL lexemes (both B forms and CS forms) resemble each other more than they differ, but they are not identical. Their major similarity is that they undergo largely the same morphosyntactic procedures from the ML during language production. Also, some B forms arise as CS forms.

However, the constraints on their occurrence are different. This is a reflection of the fact that B forms have become part of the ML mental lexicon; whereas CS forms remain as EL material which only occurs in ML morphosyntactic frames during codeswitching discourse. Accordingly, the constraints on the occurrence of CS forms are specially related to those governing multiword codeswitching material (i.e. EL islands).

Therefore, a continuum of relationships exists between borrowing and all forms of CS material so that codeswitching and borrowing are not distinct phenomena, as some have suggested (e.g. Poplack, 1980; Sankoff, Poplack & Vanniarajan 1990: 97). It follows that an adequate model of the morphosyntactic constraints on codeswitching will account for both types of singly-occurring lexemes from the EL, B or CS forms.

The Data Base

The major data sets[2] considered are two.

1. A set of naturally-occurring conversations (N = 40) based on about 20 hours of audio recording in Nairobi, Kenya provides some examples of both B forms and codeswitching utterances. In these conversations, Swahili is the ML and the English is the EL.
2. Quantitative data on B forms is considered from a set of 129 interviews, for a total of about 22 hours of speech, audio-recorded in two different locales in Zimbabwe, one urban and one rural. All interviewees and interviewers were native speakers of Shona, and Shona was the main medium of the interview, but interviewees used many English B forms and also engaged in some codeswitching.

In both Kenya and Zimbabwe, data were collected by local residents serving as research assistants. In Kenya, most speakers were not aware they were being recorded; this step was taken to ensure that the conversations were indeed natural. However, speakers' permission to use the recordings anonymously was sought after the recording. In Zimbabwe, interviewees were not aware that the real purpose of their being interviewed was to study their use of English loan words; the interview dealt with matters in their everyday lives.

The Matrix Language Frame Model (MLF)

Before a discussion of the characteristics of CS vs. B forms, a model to account for all codeswitching utterances will be presented in brief. (See Myers-Scotton forthcoming for a full discussion[3]). The model is called the Matrix Language Frame Model (MLF model) because its centrepiece argument is that codeswitching takes place within a frame set by the Matrix Language (ML). It places codeswitching restrictions at a relatively abstract level which is 'pre-syntactic' within any overall model of language pro-

duction. Among other things, this means that the provisions of current syntactic models are not considered adequate to account for codeswitching structures.

While the premise of the model is that structural configurations in code-switching utterances can be accounted for by a limited set of structurally-based hypotheses, codeswitching is a social phenomenon. That is, if code-switching occurs at all in a community, when it occurs, and to what extent—all of these matters have a psycholinguistic/sociolinguistic basis (cf. Myers-Scotton, forthcoming).

The model proposes that there are two hierarchies which shape code-switching utterances and are especially salient when constituents consisting of morphemes from both languages (ML + EL constituents) are produced:

(1) The ML is more activated than the Embedded Language (EL).
(2) There is differential accessing of *content* vs. *system* morphemes. Content morphemes are similar to 'open-class items' and system morphemes are similar to 'closed-class items'; noun and verb stems are prototypical content morphemes, and inflections and articles are prototypical system morphemes.

The concept of the matrix language

Since the ML plays such a major role in codeswitching utterances, identifying the ML vs. the EL objectively is essential. For the sake of brevity, let it suffice to say that a frequency metric will identify the ML:

> The ML in any CS utterance is the language of more morphemes in the type of discourse where the conversation in question occurs, if cultural borrowings for new objects or concepts are excluded from the morpheme count.

That is, the ML can be identified empirically, provided a discourse sample (i.e. more than just the sentence in question is considered).[4] Typically, the ML is the more unmarked choice for the interaction, in the sense of 'markedness' as discussed in Scotton (1988a) in connection with social motivations for CS. But, of course, there are interactions where codeswitching itself is the unmarked choice; this is why the frequency metric is proposed.

Types of constituents in codeswitching

The MLF model identifies three types of codeswitching constituents, governed by related constraints (ML + EL constituents, ML islands, and EL islands).

ML + EL constituents

These consist of any number of ML morphemes and (generally) single-lexeme EL forms. It is the comparison of these EL forms with B forms (also from the EL) which is the main subject of this paper, of course. In examples (3) and (4) from Swahili/English codeswitching, *u-na-m-time* and *plates tatu* illustrate these constituents. While the claim here is that these EL forms are *codeswitching material*, note that their status is controversial. For example, Poplack and her associates refer to (most of) them as 'nonce borrowings' (e.g. Sankoff, Poplack & Vanniarajan, 1990); but others, such as Nortier (1990: 209), see no motivation for categorising such forms as 'nonce borrowings', at least in their data.

(3) *Setting*: Several young men in Nairobi are discussing their attendance at a dance.
Luyia: . . .wewe ulikuwa umejikunja kwa *corner*
u-na-m-*time* tu.
2nds-PROG-her-time just
(. . .you had folded yourself in a corner (and) you were just 'timing' her.)
(Nairobi conversation No. 16)

(4) *Setting*: Later in the same conversation:
Nandi: Wewe, Ben, siku ile ulisosika[5] *plates* tatu safari moja!
plate(s) three journey one
(You, Ben, you really ate a lot that day—three plates at a go!)
(Nairobi conversation No. 16)

ML islands

ML islands in codeswitching utterances consist only of ML morphemes. They are well-formed according to the ML grammar; they must show internal structural dependency relations (i.e. show more than linear juxtaposition).

In (5), *inaanza usiku* 'it begins in the night' exemplifies an ML Island.

(5) *Setting*: Two Luyia men discuss the recent rains.
One man: Na nyumbani imezidi hapa. Inaanza usiku na kuendelea *throughout the day*.
(And at home it exceeds (the rains of) here. It begins at night and continues throughout the day.)
(Nairobi conversation No. 34)

EL islands

EL islands parallel ML islands. They are composed only of EL morphemes, following EL grammatical constraints, and show internal structural dependency relations. In (5), *throughout the day* is an example of an EL island.

The central hypotheses of the MLF model

The MLF Model is structured by three central hyypotheses, with one supplemental hypothesis.

The ML hypothesis

The ML hypothesis: The ML frames the morphosyntax of ML + EL constituents. Two empirically verifiable principles realise this hypothesis: *The Morpheme Order Principle*—states that morpheme order in these constituents must follow that of the ML. Swahili can be identified as the ML in the examples cited. Thus, in (4) *plates tatu* 'plates three' follows the head-first order of Swahili. Or, in (3) *u-na-m-time* (you-PROG-her-time) 'You were timing her', Swahili order prevails.
The System Morpheme Principle—states that 'active' system morphemes come only from the ML in these constituents. For example, in the VP in (3), *u-na-m-time*, the inflections are all from Swahili. ('Active' morphemes are those which participate in relationships within the sentence which are external to the head of the morpheme itself.)

System vs. content morphemes

What is a system morpheme? Space does not permit a complete discussion of the system vs. content morpheme distinction.[6] In brief, three features can distinguish the two types of morphemes.[7] (1) Categories of morphemes with a plus setting for the feature [Quantification] are system morphemes. Such categories pick out particular individuals or events. For example, this means that [+ Quantification] is a property of quantifiers, determiners, and possessive adjectives (since they pick out individuals). It also is a property of such morphemes as tense and aspect, as well as certain adverbs (since they pick out events). (2) Positive settings for two other features indicate content morphemes. These features are [Thematic-role Assigner] and [Thematic-role Receiver]. Most verbs and prepositions are thematic-role assigners. Nouns, free-form pronouns, and descriptive adjectives are typically thematic-role receivers. However, those morphemes with a minus setting for these features are system morphemes (e.g. the copula, which does not assign thematic role).

Data to falsify the ML hypothesis

In general, all available data offer strong empirical support for the ML Hypothesis. That is, neither the data sets studied for this paper nor examples in any available published studies on CS show more than one or two counter-examples to the ML hypothesis. Evidence falsifying the ML Hypothesis would show either or both of two features:

(1) In ML + EL constituents, the morpheme order would violate that specified by the ML. However, only one such example occurs in the Nairobi corpus, while there are more than 100 examples showing ML morpheme order in mixed constituents, in support of the Morpheme Order Principle.
(2) In ML + EL constituents, 'active' system morphemes (i.e. those with governing relations external to the lexical head) from the EL would be present. Again, only one such example occurs in the Nairobi corpus in contrast with over 150 examples showing only ML system morphemes.

The blocking hypothesis

A Blocking Filter prevents surface realisations from ML lemmas (the building blocks of ML + EL constituents) by non-congruent EL content morphemes. Levelt (1989: 6) uses the term *lemma information* (*lemma* for short) to characterise all of the non-phonological part of an item's lexical information. Lemmas are in the *mental lexicon*, which Levelt refers to as 'the store of information about the words in one's language'.

Congruency between ML lemmas and EL content morphemes depends on matchings in regard to system vs. content morpheme status and also related subcategorisation restrictions.[8] That is, even if the EL realises a given grammatical category as a content morpheme, if it is realised as a system morpheme in the ML, the ML blocks the occurrence of the EL content morpheme in ML + EL constituents. The filter also blocks an EL content morpheme if it is not realised in the ML with a congruent content morpheme; non-congruence results when there is not a match for an EL morpheme in the ML regarding subcategorisation on its head in the maximal projection of which it is a complement.

Content morpheme insertion

This hypothesis takes on importance at the stage of content morpheme insertion, or after the frame is in place (consisting of ML morpheme order and ML system morphemes, in accordance with the ML Hypothesis). While any ML content morphemes may be chosen, the Blocking Hypothesis limits the set of EL content morphemes which are permitted. There is congruence between most nouns in most codeswitching examples, a reason why so many singly-occurring CS forms, and ultimately B forms, are nouns.

Examples of congruency: noncongruency

Noncongruencies in data studied to date are most apparent with prepositions and pronouns. For example, some English prepositions which are content morphemes (i.e. they are thematic role assigners) are not congruent with any Swahili counterparts (i.e. the Swahili verb ordinarily assigns thematic roles). Therefore, these English prepositions cannot occur in ML +

EL constituents in codeswitching. For example, *at* in *I wanted to find him at the stadium*, has no congruent counterpart in Swahili. Therefore, *at* cannot occur in the codeswitching equivalent of this sentence (6); its complement, *stadium*, occurs as a 'bare form'.

(6) *Setting*: Two Luyias and a Kikuyu are talking about a fourth person and his behaviour.
Kikuyu: Hee, yule *cousin* wangu, hata ni mjinga. Siku moja nilikuwa nataka ku-m-pata *stadium*
to-him-find
(Yes, that's my cousin, he's even a fool. One day I wanted to find him (at the) stadium.)
(Nairoi conversation No. 28)

But in the case of other prepositions, where there is congruence between Swahili and English (i.e. there are equivalent prepositions in both languages which are content morphemes), the Blocking Filter allows the English prepositions to appear as single CS forms. In (7) English *between* is accessed, because it is congruent with Swahili *kati ya*.

(7) *Setting*: Three women are talking about an accident.
Second person: Ilikuwa *between* saa mbili na saa tatu asubuhi.
it was between hour two (= 8am)
(It was between 8 and 9 in the morning.)
(Nairobi conversation No. 40)

Further, while pronouns are content morphemes in English, their usual counterparts in Swahili (agreement critics) are system morphemes. Thus, the prediction is that in Swahili/English codeswitching (with Swahili as the ML) no pronouns from English will appear in ML + EL constituents. Only one exception (i.e. only one English pronoun as a singly-occurring CS form) occurs in the 40 conversations in the Nairobi corpus. Reporting on a Tamil/English corpus, Sankoff, Poplack & Vanniarajan (1990: 80) state that there are no singly-occurring English pronouns appearing as direct objects in what the MLF model would call ML + EL constituents; yet there are many Tamil pronouns as direct objects (40% of the Tamil direct objects in the corpus as a whole are pronouns). This finding also is predicted by the Blocking Hypothesis.

The EL trigger hypothesis

The EL Trigger Hypothesis is this:
Accessing any EL morpheme not licensed under the ML or Blocking Hypotheses triggers the 'grammatical formulator' (i.e. 'control central' in post-conceptual stages) to inhibit all ML morphosyntactic procedures and complete the current constituent as an EL Island (i.e. by activating EL morphosyntactic procedures).

Relating CS forms and EL islands

Note that the EL Trigger Hypothesis is not an isolated (i.e. *ad hoc*) prediction, but is a corollary to both the Morpheme Order and the System Morpheme Principles. That is, if a CS form appears in an ML + EL constituent, it must conform to these principles. The Trigger Hypothesis states that CS forms which do not meet the principles must be accessed only as parts of EL islands. Thus, single CS forms and EL islands are governed by related constraints, a reason to argue they are related phenomena, as this paper does.

Exemplifying the EL trigger hypothesis

The Trigger Hypothesis predicts for the Nairobi corpus that if an English adjective is produced prior to its head, the result must be an EL Island (since the ML Hypothesis specifies ML morpheme order, which is head-first in NPs). Also, for example, if a quantifier from English (e.g. *all*) is accessed, an EL Island must follow (e.g. *all the clothing*), since quantifiers are system morphemes and only ML system morphemes are possible in ML + EL constituents. These predictions are supported.

Thus, the hypothesis predicts that an example such as (8a) is possible, but not (8b):

(8a) Nitafanya kazi hii *next weekend*.
(I will do this work next weekend.)
(8b) *Nitafanya kazi hii *next* mwisho wa wiki.
(I will do this work next weekend.)

The EL hierarchy hypothesis

While the EL Trigger Hypothesis predicts obligatory EL Islands, the EL Hierarchy Hypothesis predicts optional ones. It is best stated as two sub-hypotheses:

(1) The more peripheral a constituent is to the theta grid of the sentence (to its main arguments), the freer it is to appear as an EL Island.
(2) The more formulaic in structure a constituent is, the more likely it is to appear as an EL Island.

Examples (9a) and (9b) illustrate optional EL Islands (*every morning* and *the best*):

(9a) *Setting*: a schoolboy is talking about his daily schedule.
Schoolboy: *Mother* hunipeleka *every morning*.
(Mother takes me every morning.)
(Nairobi corpus No. 20)

(9b) *Setting*: Six young men are talking about local football players.

Kikuyu 1: . . .Na alikuwa *the best.*
(. . .And he was the best.)
(Nairobi corpus No. 19)

ML islands

The MLF Model has little to say about ML islands. They are the least considered aspect of codeswitching as a phenomenon, requiring more study. There appear to be no structural criteria requiring ML islands to be formed; therefore, ML islands differ from EL islands in this way. They also differ from ML + EL constituents in their composition. While the ML is active as the language setting the frame for both ML islands and ML + EL constituents, the two types of constituents differ at the content morpheme insertion stage. Only ML morphemes appear in ML islands while either ML or EL morphemes are present in ML + EL constituents.

Summary of the MLF model

The MLF model is based on a differential between the ML and the EL, with the ML taking the more prominent role in codeswitching utterances. The frame of ML + EL constituents comes from the ML: its morpheme order must not be violated and also 'active' system morphemes in these constituents must come from the ML. In addition, any ML content morphemes may be inserted into the frame.

The role of the EL is restricted: in ML + EL constituents only EL content morphemes appear, and even there, the ML Blocking Hypothesis prohibits some EL content morphemes. Further, the only way an 'active' EL system morpheme may occur in a codeswitching utterance is in an EL island.

Lexical Borrowing: Two Different Processes

Few of the codeswitching studies of the 1970s and 1980s consider relationships between borrowing and codeswitching; they simply treat all single EL lexemes as B forms (e.g. Reyes, 1976). Pfaff (1979) and Poplack (1980) at least raise the problem of distinguishing the two phenomena. But almost all of these early studies analyse only full constituents (or clauses or sentences) as 'true' codeswitching. However, increasingly, researchers are considering some (or all) singly-occurring EL-origin lexemes as material which must be accounted for in any codeswitching model (e.g. Eliasson 1989; 1991).

The problem is that those same codeswitching researchers who assume a sharp line between borrowing and codeswitching seem to presuppose that all lexical B forms are cultural B forms. Cultural loans, of course, are the usual textbook example of B forms. They stand for objects or concepts new to the ML culture. Evidence that many early researchers were thinking only

of cultural B forms is that they refer to B forms as 'filling lexical gaps' (cf. Sridhar & Sridhar, 1980: 409; Bentahila & Davies, 1983: 302).

Cultural B forms are undoubtedly most common in most cases of B, a claim supported by the distribution of cultural vs. core B forms in the Zimbabwe corpus.[9] Examples include *lunchhour* and *examination*. But the point is that such B forms are not the only singly-occurring lexemes from the EL and not even the only B forms; there are also CS forms and core B forms.

Cultural borrowings and the ML

In terms of their relationship to the ML, cultural B forms are best characterised as at the categorical end of any continuum of B. That is, once the lexeme encoding a new object/concept is used in the ML; it is predicted those speakers who first used it will use it again when the need to signify the same referent comes up. Therefore, the claim is that cultural B forms enter their ML lexicon *abruptly*.[10] In the Zimbabwe corpus, those interviewees referring to *school fees* categorically use the English compound to do so; the only variation is in its degree of phonological integration (as [skurufiz] or [skulfiz]).

Core borrowings and the ML

Because cultural loans do enter the ML abruptly, they are rather 'instant' B forms and unrelated to codeswitching as a phenomenon. The relation of core B form to codeswitching, however, is another matter. Core B forms are borrowed because certain types of contact situations promote desire to identify with the EL culture or at least aspects of it. As Haugen writes (1953: 373), 'Borrowing always goes beyond the actual "needs" of language'. Examples of core B forms include *problem* and *weekend* from the Nairobi corpus and *twenty* and *because* from the Zimbabwe corpus.

Of course *some* core B forms may enter the ML in the same way as some cultural B forms: a prestigious person uses them and others then follow suit. But one of the major arguments of this paper is that the unmarked mechanisms for introducing core B forms are the same production constraints on EL material when codeswitching takes place (contained in the MLF model).

Core borrowings and ML + EL constituents

The stage in CS production relevant to this discussion is the content morpheme insertion stage in ML + EL constituents. Recall that once the frame is set, content morphemes may be accessed from either the ML *or* the EL, as long as the EL morphemes pass the Blocking Filter. Thus, the lines are open at this stage to position EL morphemes in slots in basically

ML constituents. Thus, the fact that most B forms are content morphemes is entirely predictable under the MLF model, if they start out as CS forms in ML + EL constituents. That is, core B forms are first CS forms: some CS forms are ephemeral, but others are repeated until gradually they become B forms. This is the basis for speaking of B and CS forms as falling along a continuum.

A different explanation holds for cultural B forms. While they may well occur alongside core B forms in ML + EL constituents, they would enter the ML with or without CS,[11] since they indeed do fill lexical gaps. Many of them are nouns because they stand for new objects/concepts.

Other views of borrowing vs. codeswitching

There may be many language contact researchers who have always assumed such a connection between borrowing as a process and CS, even if they did not differentiate cultural and core B forms (e.g. Gardner-Chloros, 1987: 102). But within some codeswitching research circles, the idea of such a continuum is not so acceptable, nor is the relating of borrowing to longer stretches of CS material. While Poplack and her associates now do not distinguish between singly-occurring CS forms (largely isomorphic with their category, 'nonce borrowings') and B forms (what they call established loans), they are silent on the derivational history of loans. However, they do explicitly state (Sankoff, Poplack & Vanniarajan, 1990: 98), that 'nonce borrowing' is not 'an intermediate process situated somewhere between code-switching and borrowing in the more traditional sense.'

Comparing Borrowing and Codeswitching

If single B and CS forms fall along a continuum, is it possible to differentiate the two? Arguments typically advanced will be reviewed, although many of them apply implicitly to distinguishing single B forms from longer stretches of codeswitching. I will suggest that absolute frequency/relative frequency is the single criterion best linking B forms more closely with the ML mental lexicon than single CS forms. This difference is related to how the forms are handled by the 'formulator' (i.e. directing morphosyntactic procedures during language production) and how they are entered in the mental lexicon, I argue.

But I will also conclude that there is little reason to differentiate the two forms *as processes*, that is, either in terms of their derivations *or* the ML morphosyntactic processes they undergo for surface realisations. Certainly, there are more similarities than differences. That is, I agree with Treffers-Daller (1991) that 'both code-switching and borrowing may be considered in the first place as the interaction of lexicons.'[12] I would put it a little differently in terms of the MLF model: while codeswitching and borrowing

are always possible between any two languages, the specific outcomes depend on the existence of 'lexicon-driven congruencies'.

Comparing phonological integration

An early position (e.g. Poplack, 1980; Bentahila & Davies, 1983) was that B and CS forms could be distinguished on the basis that B forms typically undergo phonological adaptation to the ML. But it soon became apparent that while *most* established B forms may well be phonologically integrated into the ML, by no means do *all* B forms show such integration. The same psycho-sociolinguistic factors favouring the borrowing of core lexemes also often favour the non-integration of any type of B form from the donor language. Such a situation arises, I suggest, when the EL is the language of a group with more socio-economic prestige than the ML. Also, I have argued elsewhere (Myers-Scotton, 1990) that educated bilinguals may practice 'elite closure' by making their speech and use patterns different from that of the masses. Surely one form of elite closure is for bilingual elites to pronounce loans as close to the originals as possible.

This is not to deny that the majority of B forms in most data sets show a good deal of phonological integration into the recipient language, or that *multiple-word* codeswitching utterances rarely show much phonological integration into the ML. But the point remains that one must conclude that far from *all* B forms can be distinguished from *single* CS forms on the basis of their phonological integration into the ML.

Comparing morphosyntactic integration

If phonological criteria do not clearly distinguish borrowing and codeswitching, what about morphosyntactic criteria? If the issue is distinguishing single B forms and multiple-word codeswitching, it is a non-issue. Codeswitching material of full-constituent length or longer (i.e. EL islands) regularly shows *no* morphosyntactic integration into the ML, taking only the morphosyntax of the EL. But the problem is that *much* EL material in what many researchers call codeswitching consists of single lexemes or bound stems (in ML + EL constituents). These single CS forms are always integrated into the syntax and often into the morphology of the ML. (Recall, for example, *u-na-m-time tu* (SUBJ2ndS-PROG-OBJ3rds-time just) 'you were just timing her' in example (3) above. It is hard to see how *time* could be more morphosyntactically integrated than it is.)

While earlier researchers often excluded single EL lexemes as instances of codeswitching, more recent studies now include them. For example, Nortier (1990: 140) states, 'By far the largest group of switches concerns insertion of single words in one language (usually Dutch) in sentences of the other language [Moroccan Arabic]. 402 such insertions were found.' Among others, Berk-Şeligson (1986) also includes single lexemes.

Different procedures for EL islands

Why do these singly-occurring CS forms resemble ML material in their morphosyntax and not EL islands? The answer is that they undergo different morphosyntactic procedures (those of the ML) because they are parts of different types of codeswitching constituents. Recall that the MLF model assigns different morphosyntactic requirements to these different types of constituents (those of ML + EL constituents vs. EL Islands). ML morpheme order and ML system morphemes are the rule for all material in ML + EL constituents. Thus, it is no surprise that Sankoff *et al.* (1990) find that English 'nonce borrowings' show the same morphosyntactic features (e.g. Tamil inflections) as indigenous Tamil word in their Tamil/English corpus.

Nonce borrowings

The category 'nonce borrowing' originates with Poplack and her associates. These are singly-occurring EL forms which occur infrequently or even only once in a data corpus, but show morphosyntactic integration into the ML. According to Poplack, Sankoff & Miller (1988: 50), they are not part of the phenomena of codeswitching. But these authors also distinguish between forms 'that occur only once in our corpus ("nonce" borrowings) and those used by many speakers (widespread loans).' Two examples of 'nonce borrowings' appear in (10), *car* and *drive*:

(10) anta *car*-ei *drive* paNNanum
that ACC do must
(we must drive that car.)
(Tamil/English Sankoff *et al.*, 1990: 80)

The problems with creating such a category as 'nonce borrowings' for such forms are:

(1) No explanatory value is gained in exchange for adding another category of description which is unrelated to other categories and constraints within the codeswitching model.
(2) Creating a category of quasi-borrowings masks recognised similarities between either ML-origin or EL-origin material in ML + EL constituents *in the production processes they undergo.*
(3) It also blurs distinctions between B forms and CS forms (assuming CS forms include 'nonce borrowings') as *end products*. They differ in that borrowing is a phenomenon open to monolinguals while codeswitching is not. More important, they differ in their freedom of occurrence.

Why singly-occurring lexemes may be CS forms

Thus, the argument of this paper is that there is no reason to remove from the CS arena the single-lexeme EL material which cannot be identified as established B forms. True, these single CS forms are related to B forms. That core B forms arise through codeswitching has already been suggested; in the next section, the similarity of CS and B forms in the morphosyntactic procedures they undergo will be highlighted.

But the point remains that single CS forms also are related to multiword CS utterances (EL Islands) in regard to their frequency and freedom of occurrence. Ultimately, this relation arises because CS material (i.e. single CS forms and EL islands) are not part of the ML mental lexicon, as are B forms.

Incomplete morphosyntactic integration

An examination of any codeswitching data corpus confirms that both established B and single CS forms do show morphosyntactic integration into the ML. Recall the example sentences cited above, with both single B and CS forms taking ML system morphemes and following ML morpheme order.

But both *also* show *incomplete* integration in some cases. However, there may well be a difference between B and CS forms in this regard, with B forms showing generally more integration. Still, the difference seems to be one of degree, not kind.

This subject requires further study; here I only offer some evidence and two tentative hypotheses, one concerning 'bare forms' and the other the relativity of incomplete morphological integration.

The 'bare form' hypothesis

First, it seems that non-congruence between the ML and the EL in regard to subcategorisation in general (including that discussed above under the Blocking Hypothesis) sometimes results in single EL forms which are 'bare forms'. That is, especially nouns (but also some verb stems) appear without the system morphemes they would take in an EL utterance. In example (9) above, *stadium* is such a 'bare form'; it lacks a locative suffix (which it would have in Swahili). For another example, see *escort* in (11); it lacks any modifier. An article would be expected in English, but not Swahili; but some modifier would normally accompany it in Swahili, possibly *mmoja* resulting in *escort mmoja* (escort-one) 'an escort'.

(11) *Setting*: Three young men from different ethnic groups are discussing problems in getting home at night.
Luo man: Tu-ta-ku-pat-i-a *escort* mpaka nyumbani.
(We will get for you (an) escort up to home.)

(Nairobi conversation No. 26)

Such 'bare forms' have been reported in many codeswitching corpora (e.g. Berk-Seligson, 1986; Backus, 1990; they are discussed at length in Nortier, 1990 and also in Myers-Scotton, forthcoming, b).

My speculation is that 'bare forms' are a type of 'speech error', resulting when an EL content morpheme is accessed under the command of an ML lemma. I suggest that the error (in the application of the ML morphosyntactic procedures) results because the congruence between the EL form and the ML lemma is less than perfect.

This speculation leads to a 'Bare Form' Hypothesis: such 'bare forms' are much more frequent among single CS forms than among B forms. The motivation for this hypothesis is the claim that B forms have their own ML lemmas in the ML mental lexicon, an entry which normally includes access to all the relevant ML morphological procedures. In contrast, CS forms are only accessed through ML lemmas (at the content morpheme insertion stage). Whether a singly-occurring EL-origin lexeme is a CS or a B form would be determined by its frequency, an issue discussed below.

The morphological integration hypothesis

Second, when morphological integration is incomplete, a Morphological Integration Hypothesis suggests this: Morphological Integration will be more incomplete on CS forms than on B forms, even though they have the same EL origin. Further, following Hopper & Thompson (1984), it is predicted that the more functional peripheral inflections are more likely to be lacking in general; however, if present, they will only occur on CS forms if they also occur on B forms.

Two cases support this claim.

1. In Bantu languages, verbs carry a number of affixes. The two most central affixes are the subject prefix indicating agreement with the subject of the verb and the tense/aspect prefix. These are categorically present in all CS and B forms in the Nairobi and Zimbabwe corpora. However, the final vowel in the verb, an affix carrying little functional load in most cases, is often absent *in CS forms, although it is almost always present in B forms.*[13] For example, (12) from the Nairobi corpus and (13) from the Zimbabwe corpus illustrate CS forms:

(12) matatu i- -li- *-overturn* na kugonga basi.
small bus it-PAST-overturn and to-hit bus
(The matatu overturned and hit a bus.)
(Nairobi conversation No. 40)

(13) . . .Ukaona achifamba apa u- -nga- *-guess*. . .
if you see she-walking 2ndS-able-guess
(. . .if you see a woman walking, you can guess. . .)
(Shona/English Crawhall unpubl. data)

2. Additional evidence that B forms tend to show more morphological integration than single CS forms is supplied from an analysis of plurality marking on Shona nouns in the Zimbabwe corpus. Placing foreign nouns in Bantu class six (the prefix *ma-* marks both the class and plurality) is a favourite strategy in southern Bantu languages in general. Bernsten (1990: 82) compared the number of B forms (defined as occurring in at least three different interviews) with CS forms (defined as occurring in two or fewer interviews) which received *ma-* and also showed the English plural suffix (e.g. *ma-day-s*).[14] She found that of the 136 tokens taking *ma-* which are B forms, only 17% occur with English *-s* as well. But 69% of the CS forms taking *ma-* also show English *-s* (86 out of 124). Both CS and B forms undergo the same ML morphosyntactic process, and both equally govern class six agreements on other elements in the sentence (i.e. agreements congruent with *ma-*). But one can say that the CS forms are less morphologically integrated because they are still open to EL morphological processes as well.

Differences in Frequency

On another front, there are several motivations to hypothesise that relative frequency should separate single B and CS forms.

First, recall that the ML and Blocking Hypotheses predict that an important distinction exists between the two types of EL forms: the formulator's producing surface sentences puts no restrictions on B forms in ML + EL constituents, whether they are content or system morphemes, but it permits only content morphemes as CS forms and only those meeting the requirements of the Blocking Filter.

Second, if one accepts the speculation that B forms are accessed via their own ML lemmas in the ML lexicon because they have become part of the ML (even while their originals, of course, may maintain membership in the EL mental lexicon), then B forms, as 'naturalised' ML forms, should have a similar frequency to indigenous ML forms. In contrast, if there is EL material which is accessed only through ML lemmas if there is a congruence between the ML lemma and the indigenous EL lemma, then these forms should have a more limited frequency (cf. Scotton, 1988b).

The absolute frequency hypothesis

These observations/speculations motivate this prediction: not all EL material has the same frequency in a given corpus. This prediction is supported in the Nairobi and Zimbabwe corpora. Taking an admittedly arbitrary cut-off point, Bernsten (for the Zimbabwe corpus) and I (for both corpora) were able to distinguish EL-origin forms, based on whether they occurred in three or more conversations/interviews. Those forms occurring in three or more samples are considered B forms; others are considered CS

forms. That is, CS and B forms are defined by frequency. (In addition, all forms for objects/concepts new to the culture were considered cultural B forms; establishing their frequency is not at issue, as discussed above.)

Relative frequency as a criterion

The frequency of a core EL form relative to its indigenous counterpart can be studied as another metric to differentiate single EL forms. Two cases from Shona, with forms from English which theoretically could be either B or CS forms, are discussed briefly. (For details, see Myers-Scotton, forthcoming, b.)

Numbers in Shona: English-origin B forms

In the Zimbabwe corpus, analysis showed that of all the numbers used by interviewees (N = 1, 257), 86% were in English. The use of numbers was studied by a number of variables; but in both urban and rural samples, for both males and females, for a variety of semantic functions, English numbers far outnumbered those in Shona. Based on this study of relative frequency, the conclusion that English numbers are B forms in Shona seems non-controvertible.

Discourse markers: More English-origin B forms

In addition, the frequency of the discourse markers *because* and *but* was studied relative to their Shona counterparts. (Because of their discourse 'prominence' such markers often seem to be borrowed in many language contact situations (cf. Scotton & Okeju, 1973 on Swahili borrowings into Ateso).) The Shona/English discourse marker findings are much less dramatic than those for numbers, but still convincing: 8% (26 out of 367) of the encodings of 'because' were with the English morphme *because*; 9% (26 out of 315) for encoding 'but' are with the English morpheme *but*.

Is frequency too arbitrary as a criterion?

There are clearly problems with using frequency to differentiate single B forms and CS forms, the main one being the issue of arbitrariness. But many thorny issues disappear by approaching the issue from the standpoint of CS forms. It is not that a B form *must* recur, it is that a CS form must *not* recur in order to be a CS form. This relates to my final point.

ML Forms as on 'The Preferential Path'

That morphemes from the two or more languages active in codeswitching are not equally accessed, that there is a 'preferential path', is the claim of the

MLF model. This argument is supported by the numerical and grammatical dominance of one language. One language is identified as the ML by the greater frequency of its morphemes; it turns out this same language sets the morphosyntactic frame of ML + EL constituents and also is accessed in the numerous ML Islands. Thus, there is ample quantitative evidence that both the morphemes and morphosyntactic procedures of one language are preferred. This language is called the ML. And, since B forms are, in fact, ML forms, it follows they are also preferred, with preference expressed by frequency. The outcome is that all EL material, whether single CS forms or EL Islands, is less frequent in CS utterances than ML material. This empirical finding indicates there is a clear link between all types of EL material as 'off the preferential path'.

Conclusion

This paper has argued that there are both similarities and differences between singly-occurring forms originating in an embedded language (EL) which occur in a recipient or matrix language (ML), whether they are B forms or CS forms. A crucial similarity is that both undergo similar, if not identical, ML morphosyntactic procedures; therefore, any CS model adequately accounting for single CS forms also necessarily accounts for B forms. The similarities in morphosyntactic treatment largely motivate the argument that core B forms arise originally as CS forms; that is, at any point in time, B forms and CS forms fall along a process continuum.

Just as important, the paper claims that, for a variety of reasons, multiple-word codeswitching utterances also share a relationship with single CS forms. Both types of codeswitched material remain with entries only in the EL mental lexicon; their accessing in codeswitched utterances is possible only because the more dominant ML procedures are inhibited under special provisions outlined in the MLF model. In contrast, B forms have entries as ML lemmas and, accordingly, more accessibility to recurrence.

Thus, what may be the most telling feature of all codeswitching material is its relative lack of predictability; specific codeswitching forms may appear in codeswitched discourse, but then surface again only in monolingual EL discourse. This absence of predictability supports the unity of all CS morphemes as 'aliens'. They are EL forms appearing in ML frames, distinct in this sense from B forms which have become part of the ML.

Notes

1. Fieldwork for this paper was conducted under a Fulbright Research Grant to Kenya and Zimbabwe (1983) and a Social Science Research Grant to Zimbabwe (1988) as well as a University of South Carolina Research and Productive Scholarship Grant (1988); support is gratefully acknowledged. Claims here supersede a somewhat different argument in Scotton, 1988c.
2. My research associates in Zimbabwe were Janice Bernsten, who supervised much of the

interviewing for the Zimbabwe corpus, and also Nigel Crawhall, who collected some naturally-occurring conversations showing Shona/English codeswitching. Shem Lusimba Mbira has been my primary research assistant in Kenya.

3. The forerunners of the MLF Model appear in Myers-Scotton & Azuma, 1992; and in Myers-Scotton, 1991.
4. The ML is not fixed in any sense; that is, which language is the ML may even change within the course of a conversation (depending on a change in situational factors, such as topic or participants, in most cases). It also may change over historical time.
5. This verb form includes the stative extended form of a slang word for eating *-sosa/-soza*, probably derived from *sauce*; it is not part of Standard Swahili.
6. Joshi, 1985 was one of the first codeswitching researchers to use the ideas of Garrett (e.g. Garrett, 1990) about differences in the ways 'closed class' vs. 'open class' items are accessed in monolingual speech. Basing his argument on evidence from speech errors, Garrett hypothesises that these two different types of morphemes may be accessed at different points in the language production process. I follow Joshi in using the terms 'Matrix Language' and 'Embedded language'.
7. For a fuller discussion, see Myers-Scotton (forthcoming).
8. Although their approach is different, Bentahila & Davies (1983) discuss subcategorisation restrictions as possibly limiting which forms could occur in codeswitching utterances. Azuma (1991) also discusses constraints on codeswitching material in terms of subcategorisation restrictions; his model is similar to the MLF model, but not identical.
9. I analysed Bernsten's (1990) listing of B forms in the Zimbabwe corpus. Out of 391 B forms, 68% are clearly cultural loans. Some of the remainder are borderline cases (e.g. [beg]/[begi] 'bag'; [oda] 'order').
10. Claiming that cultural B forms enter the ML abruptly does not mean that *all* speakers immediately adopt such forms. Some B forms are always only part of the repertoire of only certain sub-groups within a speech community.
11. Their appearance in codeswitching utterances may facilitate the borrowing of cultural B forms; however, no evidence is available.
12. Treffers-Daller refers to Muysken (1990) for a more detailed discussion in regard to codeswitching of how syntactic properties may derive from the lexicon.
13. Of course those English-origin verb stems showing no final vowel also violate Bantu phonotactics which call for open syllables only.
14. In many codeswitching data sets, some single CS forms show 'double morphology'. Often, this means there are plural affixes present from both the ML and the EL (e.g. *ma-watchmen* in Swahili/English codeswitching). An argument as to why the EL affix is not 'active', but simply accessed as part of the EL stem, appears in Myers-Scotton forthcoming.

References

Azuma, S. (1991) Two level processing hypothesis in speech production: evidence from intrasentential code-switching. Paper at 27th Chicago Linguistic Society meeting (5/91). To appear in *Proceedings*.

Backus, A. (1990) Turkish–Dutch codeswitching and the frame-process model. Paper at International workshop on ethnic minority languages in Europe. Tilburg, The Netherlands (12/90).

Bentahila, A. and Davies, E. E. (1983) The syntax of Arabic–French code-switching. *Lingua* 59, 301–30.

Berk-Seligson, S. (1986) Linguistic constraints on intrasentential code-switching. *Language in Society* 15, 313–48.

Bernsten, J. (1990) *The Integration of English Loans in Shona: Social Correlates and Linguistic Consequences*. Ph.D. dissertation. East Lansing MI: Michigan State University.

Eliasson, S. (1989) English–Maori language contact: code-switching and the free-morpheme constraint. *Reports from Uppsala University Department of Linguistics* 18, 1–28.

——(1991) Models and constraints in code-switching theory. *Papers for the Workshop on Constraints, Conditions and Models*. (pp. 17–50.) Strasbourg: European Science Foundation.

Gardner-Chloros, P. (1987) Code-switching in relation to language contact and convergence. In G. Ludi (ed.) *Devenir bilingue-parler bilingue*. (pp. 99–113.) Tubingen: Max Niemeyer Verlag.

Garrett, M. F. (1990) Sentence processing. In D. N. Osheron and H. Lasnik (eds) *An Invitation to Cognitive Science, Vol. 1*. (pp. 133–75.) Cambridge MA: MIT.

Haugen, E. (1953) *The Norwegian Language in America: A Study of Bilingual Behavior*. Philadelphia: University of Pennsylvania Press. (Reprinted 1969 Bloomington IN: Indiana University Press.)

Hopper, P. and Thompson, S. (1984) The discourse basis for lexical categories in universal grammar. *Language* 60, 703–52.

Joshi, A. K. (1985) Processing of sentences with intrasentential code-switching. In D. Dowty *et al.* (eds) *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*. (pp. 190–205.) Cambridge: Cambridge University Press.

Levelt, Willem J. (1989) *Speaking: From Intention to Articulation*. Cambridge MA: MIT Press.

Muysken, P. (1990) Concepts, methodology and data in language contact research: remarks from the perspective of grammatical theory. In *Papers for the Workshop on Concepts, Methodology, and Data*. (pp. 15–31.) Strasbourg: European Science Foundation.

Myers-Scotton, C. (1990) Elite closure as boundary maintenance: the case of Africa. In B. Weinstein (ed.) *Language Policy and Political Development*. (pp. 25–42.) Norwood NJ: Ablex.

—(1991) Intersections between social motivations and structural processing in code-switching. In *Papers for the Workshop on Constraints, Conditions, and Models*. (pp. 57–82.) Strasbourg: European Science Foundation.

——(forthcoming, a) *Social Motivations for Codeswitching: Evidence from Africa* (in press). Oxford: Oxford University Press.

——(forthcoming, b) *Accounting for Structures in Codeswitching: the Matrix Language Frame Model* (in press). Oxford: Oxford University Press.

Myers-Scotton, C. and Azuma, S. (1992) A frame-based process model of codeswitching. In K. Deaton *et al.* (eds) *Papers from the 26th Chicago Linguistic Society Meeting*. Chicago: University of Chicago Linguistics Department.

Nortier, J. (1990) *Dutch–Moroccan Arabic Code Switching*. Dordrecht: Foris.

Pfaff, C. (1979) Constraints on language mixing: intrasentential code-switching and borrowing in Spanish/English. *Language* 55, 291–318.

Poplack, S. (1980) 'Sometimes I'll start a sentence in Spanish y termino en espanol': Toward a typology of code-switching. *Linguistics* 18, 581–618.

Poplack, S., Sankoff, D. and Miller, C. (1988) The social correlates and linguistic process of lexical borrowing and assimilation. *Linguistics* 26, 47–104.

Reyes, R. (1976) Language mixing in Chicano bilingual speech. In J. D. Bowen and J. Ornstein (eds) *Studies in Southwest Spanish*. (pp. 182–88.) Rowley: Newbury House.

Sankoff, D., Poplack, S. and Vanniarajan, S. (1990) The case of the nonce loan in Tamil. *Language Variation and Change* 2, 71–101.

Scotton, C. M. (1988a) Codeswitching as indexical of social relationships. In M. Heller (ed.) *Codeswitching: Anthropological and Sociolinguistic Perspectives*. (pp. 151–86.) Berlin: Mouton de Gruyter.

——(1988b) Code-switching and types of multilingual communities. In P. Lowenberg (ed.) *Language Spread and Language Policy*. (GURT '87). (pp. 61–82.) Washington: Georgetown University Press.

——(1988c) Differentiating borrowing and codeswitching. In K. Ferrara *et al.* (eds) *Linguistic Change and Contact* (NWAV–XVI). (pp. 318–25.) Austin: University of Texas Linguistics Department.

Scotton, C. M. and Okeju, J. (1973) Neighbors and lexical borrowings. *Language* 49, 871–89.

Sridhar, S. N. and Sridhar, K. (1980) The syntax and psycholinguistics of bilingual code-mixing. *Canadian Journal of Psychology* 34, 407–16.

Treffers-Daller, J. (1991) Towards a uniform approach to code-switching and borrowing. In *Papers for the Workshop on Constraints, Conditions and Models*. (pp. 259–79.) Strasbourg: European Science Foundation.