Uniform structure: Looking beyond the surface in explaining codeswitching

Carol Myers-Scotton

This contribution looks beyond surface evidence of constraints on the grammatical structure in codeswitching data sets in discussing the key sources of structural uniformity in such data. In so doing, it emphasizes the relationship of the Matrix Language construct to maintaining such uniformity in the bilingual clause. The basic outlines of the Matrix Language Frame (MLF) model are reviewed, but emphasis is on explicating the newer 4-M model of morpheme classification and discussing ways in which outsider system morphemes differ from other morpheme types in various ways, including how they are accessed in language production. Their role in identifying the Matrix Language and indicating structural relationships within the bilingual clause provides insights into their nature of what is critical in language in general. Finally, the contribution challenges the notion that analysts can make correct predictions about well-formedness in codeswitching corpora without invoking the Matrix Language construct.

1. Introduction

The Matrix Language Frame (MLF) model is a comprehensive model that has been largely successful in predicting the permissible structures that occur within a clause showing codeswitching (Myers-Scotton 1997; 2002). The model relies on the distinction it makes between the roles of the participating languages. Specifically, the heart of the MLF model is that it restricts the participation of one of the languages in building the grammatical frame of the bilingual clause. Thus, the basic generalization that the model offers is that codeswitching (hereafter CS) is characterized by a basic asymmetry between the participating languages so that only one language accounts for the uniform structure that prevails in the bilingual clause. In addition, the model also shows how there are differences in the distribution of types of morpheme that are related to the asymmetry between languages. At the same time, any exposition of the model assumes that the main reason CS occurs at all is that a switch to another language better conveys the speakers' semantic and pragmatic intentions at some point in their conversation than speaking monolingually. That is, even though the MLF model is a model of the grammatical structure of codeswitching, in no way does the model

Rivista di Linguistica 17.1 (2005), pp. 15-34 *(ricevuto nel settembre 2005)*

disparage the notion that CS largely is driven by psycholinguistic and sociolinguistic motivations.

In earlier expositions of the MLF model, the surface consequences of the predicted asymmetry between the participating languages have been emphasized (cf. Myers-Scotton 1997; 2002). Two principles (the Morpheme Order Principle and the System Morpheme Principle) predict that only one language supplies the morpheme order to bilingual clauses as well as the morphemes that indicate grammatical relationships across maximal projections (e.g. subject-verb agreement). The principles are testable hypotheses because they state that only one language is the source of critical aspects of the morpho-syntactic frame of the clause. In corpora where these hypotheses are supported, the MLF model calls this language the Matrix Language (ML) and the other language is called the Embedded Language (EL).

Although a minor role for the EL is predicted, it participates in CS in two ways. First, it may supply content morphemes (e.g. nouns and verbs as well as some other elements) to those constituents within the bilingual clause that contain morphemes from both languages. Second, the EL may supply what are called EL islands. These islands are monolingual EL phrases that are grammatically well-formed in the EL (i.e. they include inflections). Similarly, phrases that are entirely composed of ML elements (ML islands) may also occur within the larger bilingual clause. Recent examples of corpora that largely conform to these predictions of the MLF model include Hlavac (2003) on Croatian-English CS in Australia; Deuchar (forthcoming 2007) on Welsh-English CS in Wales; and Myers-Scotton (2005b) on a Xhosa-English corpus in South Africa. Example (1) illustrates a typical example supporting the MLF model. By testing the hypotheses about morpheme order and source of critical frame-building elements, one can identify Xhosa as the Matrix Language. The example includes an English verb (*treat*) with Xhosa affixes (including the subject-verb agreement prefix *ba-*) as well as the Embedded Language island from English (*like human beings)*. All agreement prefixes come from noun class 2, the Xhosa plural class for persons. EL material is in italics.

(Myers-Scotton 2005b)

(1) Aba ba-ntwana ku-funek-aba-*treat*-w-e *like human beings*.
CL2/DEM CL2-child INF-need-INDICATIVE CL2-treat-PASSIVE-SUBJUNC
'These children need to be treated like human beings.'

In Hlavac's Croatian-English corpus, Croatian can be consistently identified as the Matrix Language. Still, the majority of English-origin items do not receive any Croatian inflection (nor are they phonologically integrated into Croatian), but neither do they receive English functional or inflectional elements. Example (2) is one instance in which an English-origin noun (tičera) does receive the expected Croatian suffix showing number and genitive case. Note that the MLF model does not specify that Embedded Language elements must receive Matrix Language inflections; the System Morprheme Principle only states that if there are any inflections of one type (now called outsider system morphemes), they must come from the Matrix Language.

(Hlavac 2003:73)

(2) ... i puno se sjećam tičer-a ['titʃera] i ...
'...and I remember the teachers a lot and...'

2. Goals of this contribution

Given that the predictions of the MLF model about well-formed surface configurations are largely supported, the goals of this contribution are to consider more fully the abstract levels behind these surface structures. The discussion assumes what I have stated elsewhere is the scope of the MLF model, classic codeswitching. In such CS, only one language, called the ML, is always the source of the elements that build the morpho-syntactic frame of the bilingual clause (with the exception of the internal structure of EL islands). Another type of CS not referred to here is called composite CS. In this type, most of the abstract structure underlying the morpho-syntactic frame of the bilingual clause comes from the putative ML, but some comes from the putative EL as well. Composite CS may be more common than classic CS. The two types of CS clearly share structural features and some of the discussion here about abstract levels underlying surface structure should apply to composite CS, too.

First, I address this question, what is there about the nature of language in general that only one of the participating languages is selected to frame the bilingual clause? In support of the ML construct, I will be referring to a principle of uniformity that seems to apply across grammatical structure universally. In explicating morpheme types according to the 4-M model, I will emphasize the role of the structure-building late system morphemes in maintaining unifor-

mity in the bilingual clause. As a second goal, I will support the need for a construct, such as the ML, against claims that models intended to explain monolingual structure, such as the Minimalist Program, can correctly predict CS structures.

3. The Uniform Structure Principle

In language in general, well-formedness conditions apply both within and between maximal projections (i.e. phrases and clauses). For example, in both monolingual and bilingual speech, noun phrases have a certain structure in a given language and phrases that are maximal phrases can only be combined in a specified way in that language. The Uniform Structure Principle (USP) formalizes this notion and specifically applies well-formedness conditions to classic CS. The USP follows

> A given constituent type in any language has a uniform abstract structure and the requirements of well-formedness for this constituent type must be observed whenever the constituent appears. In bilingual speech, the structure of the Matrix Language prevails. Within the bilingual clause framed by the Matrix Language, Embedded Language structures appear only in Embedded Language Islands.
> (wording based on Myers-Scotton 2002: 8-9)

Of course analysts of monolingual data recognize the notion of uniform structure for constituents as a long-standing principle of linguistic analysis and may see no need to state it anew. This principle underlies most theoretical analyses of individual languages and references to universal grammar (Chomsky 2001; Bresnan 2001). That this principle applies to bilingual data is not so obvious, but the same principle applies everywhere. It needs to be stated in a discussion of bilingual data because, theoretically, the source of grammatical structure could be shared in any number of ways by the participating languages in CS. But this does not happen. Empirical testing of the Morpheme Order and System Morpheme Principles shows that only one of the languages prevails in regard to these basic structural constraints. One can go further and argue that the provisions of the System Morpheme Principle are the primary way that uniform structure is maintained in bilingual clauses in language contact phenomena in general (cf. Myers-Scotton 2002:59; 87-91). This principle

restricts the source of what are now called outsider (late) system morphemes under the 4-M model to only one of the participating languages, and empirical evidence supports the principle.

Further study shows, in fact, that this language, the ML, provides all the structure underlying the morpho-syntactic frame of classic CS, not just morpheme order and outsider morphemes. For example, in Swahili-English CS, with Swahili as the ML, a color adjective + noun construction follows an associative pattern that is identical to what one finds in monolingual Swahili.

This is exemplified in example (3a) and (3b). Constructions such as **black kalamu* or **red Datsun* that would be possible in the EL (English) do not occur.

(Myers-Scotton 1988)

(3a) *Daddy*-hi-lo ø-shati l-ako
Daddy CL5-DEM CL5-shirt CL5-your
li-na ø-kalamu y-a *black* ama *red.*
CL5-with CL5 pen CL9-ASSOC black or red
'Daddy-this shirt [of] yours has [a] pen of black or red.'

(3b) *Father* y-ake a-na Datsun y-a *red*
father CL9-his 3S-with Datsun CL9-ASSOC red
'His father has a red Datsun.'

4. Four types of morpheme

In order to discuss CS data further, the 4-M model is introduced briefly (cf. Myers-Scotton & Jake 2000 & Myers-Scotton 2002 for more details). The first thing to note is that this is a model of morpheme classification, not a model that makes any predictions about CS. However, the various implications of the 4-M model do indicate why the prediction that the System Morpheme Principle captures turns out to be the cornerstone of the MLF model.

The MLF model already classified morphemes by distinguishing content morphemes from system morphemes (cf. Myers-Scotton 1997). From the standpoint of semantic-syntactic structure, content morphemes are the only ones that assign or receive thematic roles. From a language production point of view, they are the only ones directly activated by semantic and pragmatic features that match the speaker's pre-linguistic intentions. Verbs are prototypical assigners of thematic roles and nouns most typically receive these roles.

Content morphemes are activated early. As a first step in actual linguistic production, the semantic-pragmatic feature bundles that match the speaker's intended meanings point to the lemmas (abstract lexical elements) in the mental lexicon that underlie content morphemes. In contrast, system morphemes are not directly activated and none assigns or receives thematic roles. Note that the 4-M model uses "morpheme" in two ways, as the abstract entries underlying morphemes and as the actual surface level morphemes themselves.

The 4-M model divides system morphemes into three types; any post-2000 discussion of the MLF model refers to them by these types, as either early or bridge or outsider morphemes. Like content morphemes, early system morphemes are conceptually activated. This doesn't mean that content and early morphemes are the only morphemes that convey ideas or concepts; instead, it means that they are the ones that are activated in the interface between pre-linguistic ideas or concepts and language-specific lemmas of the mental lexicon. Thus, in this sense, content morphemes can be referred to as directly elected (Levelt 1989). Early morphemes are indirectly elected by their content morpheme heads to add additional meanings. In English, *up* changes the meaning of *look* on its own (compare *I look at the dog* with *I looked up the word*).. Also, determiners are early morphemes because they add a reading of definiteness to nouns. Derivational affixes are also early morphemes because they alter the meaning of the content morpheme head.

Example (4) from Palestinian Arabic-English CS shows an early morpheme from Arabic, a determiner, with an English noun (*el students*).

(Okasha, cited in Jake & Myers-Scotton 1997: 31)

(4) šaɣilti inni *advise* el *students* taba ʕuuhum
work.1S that advise DET students own/3PL/M
'My job is to advise their students.'

The two other types of system morphemes are call 'late' because they only become activated when larger structures are assembled at the level of the formulator in our production model. The formulator receives information from content morpheme lemmas and their accompanying early morphemes in the mental lexicon. Among other things, this information refers to well-formedness requirements within and between phrases in the clause under construction. Thus, the formulator puts together the strings of morphemes that receive

phonological representations at the surface level. (This discussion relies on basic views about language production in Levelt 1989).

One type of late system morpheme is called a bridge. Most typically, it meets specific well-formedness requirements between phrases that are combined to form a larger constituent or requirements in the larger clause itself. For example, to create an English constituent encoding possession via noun phrases, either *of* is needed to join two noun phrases or *'s* is suffixed to the possessor noun (e.g. *book of the student* or *the student's book)*. In 'weather clauses' in a variety of languages, the language's equivalent of existential *it* functions as a bridge; it serves as a dummy subject (e.g. French *il* in *il pleut chaque jour*). Note that *il* in this construction is a different type of morpheme from *il* as a marker of subject-verb agreement where it is an outsider morpheme, as discussed below.

Example (5) illustrates a bridge in CS. In noun phrases in a corpus of Xhosa-English CS, what could be a compound noun in English (*crime rate*) can only appear in an associative construction of NP-associative element-NP when Xhosa is controlling the grammatical frame as the ML. Note that it happens that both of the nouns themselves come from English, the EL. This structure is exemplified in (5) in which English elements are in italics and numbers refer to Xhosa noun class prefixes. This corpus of bilingual speech from 50 Xhosa-English bilinguals shows that the only time a compound noun appears is when English frames a clause (e.g. *business person* in the clause, *let's say I am a business person*). In this example, the prefix on *i-rate* is an early morpheme (it specifies the noun as a class nine noun), but the prefix on the verb (*i-nyukile*) is an outsider morpheme, indicating subject-verb agreement. The bridge morpheme is *-e* and it receives a prefix from noun class prefix nine that is also an outsider morpheme.

(Myers-Scotton 2005b)

(5) .... i-*rate* y-e *crime* i-nyuk-ile
CL9-rate CL9-ASSOC crime CL9-go up-PERF
'.... [the] rate of crime is high.'

The second type of late system morpheme is called a late outsider or just an outsider. Outsiders do not join constituents; rather, they co-index relations between elements, but across phrasal boundaries. They are called outsider morphemes because their form depends on information outside the phrase in which the outsider morphemes themselves occur. For example, if subject-verb agreement

is signaled on a verb, its form typically depends on the noun with which the affix or clitic is co-indexed. In example (1) *ba-* on the verb ba-*treat*-w-e 'they should be treated' is an example of an outsider morpheme indicating subject verb agreement, as is the prefix on the verb in example (5) as mentioned above. In example (2) above, the suffix for Croatian genitive case (and plural) appears on the English-origin noun *tičer-a* 'teacher' but the selection of genitive case depends on information outside the noun phrase in which it occurs (it depends on the verb).

Example (6) illustrates an example of an outsider morpheme that marks case in CS and comes from the ML. In the Australian Finnish-English CS corpus studied, the English noun *country* is marked with Finnish genitive case. The postposition *puolessa* 'side' governs the Finnish genitive case.

(Kovács 2001: 160)

(6) kun se oli vähän niinku tuolaa *country*-n puole-ssa
Cause it be.3S bit like there country-GEN side-INNESS
'Cause it was a bit like on the country-side.'

5. The Differential Access Hypothesis

The nature of the role of late system morphemes, as well as empirical evidence about their distribution in various types of data, implies that these morphemes are not accessed in language production as the same level as the directly elected content morphemes and their indirectly elected modifiers, early morphemes. The notion that this difference exists is formalized in the Differential Access Hypothesis:

> The different types of morpheme under the 4-M model are differentially accessed in the abstract levels of the production process. Specifically, content morphemes and early system morphemes are accessed at the level of the mental lexicon, but late system morphemes do not become salient until the level of the formulator (Myers-Scotton, 2002, p. 78; Myers-Scotton, 2005a).

One rationale in support of this hypothesis is that larger constituents can hardly be assembled until the level of the formulator because their parts first have to be made salient at the level of the mental lexicon. Also, differences in the patterns of distribution for outsider morphemes in comparison with other morpheme types, in

CS as well as in other types of data, implies the Differential Access Hypothesis. For details about outsider morphemes in data outside of CS, see Myers-Scotton & Jake (2000) on data from Broca's aphasics and second language acquisition; see Myers-Scotton (2002) on outsider morphemes in speech errors and first language attrition. Also, outsider morphemes are very resistant to transfer from one language to another; Myers-Scotton (2003) argues that only a few languages qualify as split (mixed) languages under the criterion that at least some of such a language's outsider morphemes must come from two different sources.

Recall that under the MLF model at least, outsiders are the only system morphemes which must come from the ML. True, most or all other system morphemes generally also come from the ML, a fact that supports the Uniform Structure Principle. But they can pattern differently from outsiders.

For example, only early morphemes seem to be able to double. The Early System Morpheme Hypothesis (Myers-Scotton 2002: 91-3) states that, among system morphemes, only early morphemes may double (one form is accessed with its head from the EL and then the ML supplies its form as well). Note that early morphemes do not always double and this motivates Myers-Scotton's argument that they are accessed by chance along with their content morpheme heads through a 'mistiming' error.

Also, both early and and bridge morphemes seem susceptible to what can only be called a type of non-lexical borrowing, while outsider morphemes are not. For example, at least in the United States, Spanish determiners (early morphemes) sometimes are used in otherwise monolingual English clauses in a type of word play (e.g. *I lost el textbook that I need for my Spanish class*.) Also, in some cases, early morphemes have been borrowed along with their content morpheme heads in various European languages (e.g. established loan words from Arabic such as *alcohol* or *algebra*, show such an early morpheme). Also, in Tanzania, at least in the 1960's, the English word *mudguard* (on cars), which was phonologically integrated into Swahili as *madigadi*, was treated as if the first syllable was the early morpheme prefix for Swahili noun class six. This is a class of plural nouns and so when speakers wished to refer to one mudguard they called it a *digadi*. (This route was followed because *digadi* could be considered a member of noun class five, a class of singular nouns pairing with plural class six, in which one of the allomorphs of the noun class prefix is a zero). The noun class prefix is an early morpheme. Another English word beginning with *ma-* received the same

treatment; that is, *ma-ture* (pronounced with the Swahili CVCV syllabification pattern) was used to refer to 'mature entrants' to the university. These are persons who entered via taking special examinations or by meeting other requirements rather than through the usual route of completing all the stages of secondary school. Not surprisingly, one such 'entrant' was called a *ture* (again, it was put into noun class five with no class prefix marker).

Bridge morphemes seem less likely to be borrowed, but in CS corpora with Arabic as the ML (e.g. Ziamari 2004), the Arabic bridge *dyal*, which translates more or less as 'of', appears in monolingual EL (e.g. French) islands or monolingual clauses and looks very much like a borrowing.

Note that even though the Differential Access Hypothesis groups bridge and outsider system morphemes together, the differences in the roles they play and their differences in distribution in various types of bilingual data indicate they differ in as yet poorly understood ways. One structural difference seems fairly clear: Bridges typically have only one form of the morpheme in question (i.e. they consist of one allomorph each) while outsiders seem always to form a paradigm or conjugation, with the individual members structurally-assigned. For example, specific cases are assigned by specific prepositions or verbs in such languages as German and Russian. In the Bantu languages, outsider morphemes are assigned to agreement roles (e.g. subject-verb agreement) by nouns in specific classes.

Also, some studies of creoles show another way how bridge morphemes differ from outsider morphemes. In Gullah, a creole still spoken on the South Carolina coast, bridge morphemes from the lexifier (English) seem to be able to appear in a morpho-syntactic frame which is largely supplied by the substrate languages, but outsider morphemes from the lexifier do not. In Gullah, English *of* as a bridge in the possessive NP + NP construction appears very frequently in a Gullah data set studied by Myers-Scotton and Jake (2002). In the set of 27 Gullah narratives from Turner (2002), *of* occurs 51 times in this construction, as in *baskɪt o bin* 'basket of bean') One speculation is that because of its invariant nature, *of* is more accessible than any outsider morphemes. In this same data set, regular past tense *-ed* (which is a member of a past tense conjugation including *did* in clause-initial position) and third person singular *-s* (which is a member of a conjugation paradigm including a zero suffix on other persons and numbers) never appeared. Instead, if any outsider morphemes appear in creoles, such as those marking tense or aspect, they are reanalyzed content morphemes from the lexifier, not recyled

outsider morphemes from the lexifier (cf. Myers-Scotton 2001). Note that whatever the differences or similarities between bridge and outsider morphemes, the Differential Access Hypothesis calls into question the notion implied in some treatments that all constructions with regular morphological elements (i.e. early morphemes as well as the late morphemes) undergo language production in the same way.

6. What does the special status of outside morphemes in codeswitching imply?

As noted in section 3, the Uniform Structure Principle requires uniformity in frame-building, with special reference to the ML and its role in achieving uniform structure in bilingual speech.

Although one can argue that all three types of system morpheme build phrases with uniform results, the basic relationships *within a clause* are built by outsiders. True, some languages rely on word order to varying degrees for this function; for example, English gets by seemingly with only two outsiders (the markers of present tense subject-verb agreement and the regular past tense markers). Note that word order, or rather juxtaposition, is the main means by which early and bridge morphemes build structure. Recall that early morphemes depend on their content heads for their presence and form and they occur with them; bridge morphemes meet language-specific requirements to form a larger constituent out of smaller ones, again occurring with these smaller units.

In contrast, outsider morphemes function at a more abstract level because of these two characteristics. First, they build structure through various means of co-indexing that operates across phrases so that the source of an indexing and the outsider morpheme itself do not occur together. Even so, one can argue co-indexing provides a more precise indication of the relations that hold beyond word order. Second, more than just meeting well-formedness requirements in regard to syntax, outsiders knit together elements at another level. Consider, for example, what indicating subject-verb agreement or object agreement accomplishes in regard to semantic coherence within the clause or even the larger discourse.

These characteristics are the basis for an argument that outsider morphemes are the main bastion for maintaining uniform structure. One can argue that the other system morphemes convey 'local' relationships within the clause. Also, content morphemes and earlies are 'free' in the sense that which ones will appear depend on

the specific semantic-pragmatic feature bundle that the speaker's intentions activate. Not only do outsider morphemes operate on a different level within the clause, their presence is not 'free'. That is, activating language X means that the speaker must produce specified outsider morphemes to satisfy that language's conditions of uniform structure. Their selection among a set of possibilities is structurally-assigned; choice is not 'free'.

Content morphemes, and to a lesser extent the non-outsider system morphemes, are free in another sense: They can take on new meanings or functions through language contact. Outsiders have more permanence; but the quality of permanence is what is required if uniformity is to prevail. Elsewhere (Myers-Scotton 2002) I have suggested that when Sprachbund areas show similarities across languages, they are almost always in regard to convergence in regard to content and early morphemes, although I have not studied this systematically. Also, when content morphemes become parts of calques, their meaning often change. (e.g. French *gratteciel* literally means 'scratch sky' and German *Wolkenkratzer*, literally means 'cloud scratcher', but they are meant to approximate the meaning of English *skyscraper*). In addition, more so than that for other morphemes, when languages are in contact, the lexical-conceptual structure of content morphemes may be split and combined with the lexical-conceptual structure of another content morpheme in a different language. For example, Türker shows how the Turkish ML in Turkish-Norwegian CS accepts the Turkish verb for 'drive' instead of the unmarked choice, which would be the Turkish verb for 'mount' bicycles. She argues that this acceptance is based on the fact that Turkish is in contact with Norwegian in Norway (via Turks who have become bilingual in Norwegian). In Norwegian, one would say the equivalent of 'drive (on) a bicycle' (kjøre sykkel). This can be seen as a case of convergence to Norwegian at the lexical-conceptual level, thus building a new meaning to convey 'riding bicycles' in Turkish. Note that the absence of the expected Turkish dative case marker on *bisiklet* indicates convergence at the level of morphological realization patterns, too.

(7) ... .git-ti-k bisiklet sür-dü-k
go-PAST-1PL bicycle drive-PAST-1PL
'... We went, we rode bicycles.' (Türker 2000: 172)

What does this discussion tell us about the nature of CS and how does it enlarge our understanding of the nature of language in

general? First, the notion that the Uniform Structure principle is salient in CS corpora, coupled with the role of ML outsider morphemes in maintaining uniform structure, offers an explanation for the distributions of CS types across CS corpora. In general, there are more mixed constituents than EL islands in any data set. That is, in a bilingual clause, there are more constituents consisting of one or more singly-occurring EL elements, as well as ML elements than there are constituents consisting of monolingual EL islands and ML islands. In these mixed constituents, the EL elements are morphologically integrated into ML frames via ML outsider morphemes. From the standpoint of uniformity in structure, the result is an obvious economy. That is, if there were more EL islands, the result would be less uniformity in structure. Second, CS, with its many mixed constituents, calls attention to the critical role of outsider morphemes in frame-building and maintaining uniformity in any language in a way that studying morpheme types only through analyzing monolingual data (i.e. studying one language at a time) does not. That is, looking at morpheme types through the lens of CS (from the perspective of the MLF and 4-M models) can give us new insights about outsider morphemes as key elements in the uniform structure that is so much a hallmark of human languages.

7. A Matrix Language as part of uniform structure

One major issue about uniform structure requires discussion; this is whether there is a need to posit the ML and EL division of labor in CS, as the MLF model does. Some linguists, especially some within the tradition of generative grammar, have argued that CS can be explained without invoking the construct of the ML. For example, MacSwan (2005) argues that a Minimalist analysis can account for the Spanish-English CS discussed in Jake, Myers-Scotton, and Gross (2002:19) "without reference to the ML construct". MacSwan's analysis of the Spanish-English corpus is based on considering *phi* features (such features include person, number, and gender). He points out how Spanish and English differ in regard to these features: Spanish determiners and nouns are marked for gender and number, "but in English only PERSON and NUMBER are marked while GENDER is absent" (p. 18).

In the corpus that Jake et al. discuss, there are no mixed NPs with English determiners and Spanish nouns, but there are 161 mixed NPs with Spanish determiners and English nouns. Such NPs

represent 70% of a corpus of 230 NPs that include any English elements. For example, *el career fair* occurs. Jake et al. (2002) account for this distribution by following the MLF model in arguing that only one of the participating languages is the ML. They identify Spanish as the ML because it is the source of the required outsider morphemes, such as subject-verb agreement, in any constituents showing both Spanish and English morphemes. Even though, the model does not require early morphemes, such as determiners, to come the ML, as noted above, the ML is generally their source.

To explain why English determiners do not occur with Spanish nouns, MacSwan makes an argument that the *phi* features of English nouns are a subset of the *phi* features of Spanish nouns. He presents four possible configurations of determiners and nouns and says that one of these, the configuration of English determiner + Spanish noun (his example (37d)) is "an ill-formed case." (p. 18). MacSwan considers the other three configurations (in his example 37) as well-formed; they are (1) Spanish determiner + Spanish noun, (2) English determiner + English noun, and (3) Spanish determiner + English noun. As indicated above, the Jake et al. (2002) data set does include NPs consisting of a Spanish determiner and an English noun. MacSwan's account of why such NPs can occur is that "we note that the configurations in (37) are well-formed if the *phi* set of N is included in the *phi*-set of N is included in the *phi*-set of D" (p. 18). But of course the Spanish-English corpus in Jake et al. (2002) is also explained by the MLF model under the analysis with Spanish as the ML.

A problem with MacSwan's feature-counting analysis in reference to how determiners and nouns are distributed in mixed NPs is that it implies that counting features also will predict distributions in other data sets beyond the Spanish-English corpus in Jake et al. (2002). This prediction is not supported. Using the criterion of *phi* features does not take account of a number of other factors.

First, it does not take account of the possibility of English-Spanish CS with evidence (based on morpheme order and outsider morphemes) that English is the ML. Such CS would be likely to contain some mixed NPs with English determiners and Spanish nouns; these are the NPs MacSwan classifies as ill-formed.. True, data sets with English as the ML and Spanish as the EL are not common. They are probably rare because most subjects of Spanish-English CS research seem to be Spanish L1 speakers. Based on the profiles of speakers in other CS corpora, one can predict that these Spanish speakers would select Spanish as the ML in any CS. However, Flores-

Ferrán (2005, personal email communication), who is an L1 speaker of Spanish and a linguist, offers two examples in which English is the ML. That is, Spanish nouns occur with English determiners in clauses that are otherwise in English.

See example (8) for one example.

(8) An *abeja* bit me.
'A bee bit me.'

Also, Callahan's (2004) study of CS in written (fiction) texts includes examples in which the writer has the speaker switch from Spanish to English as the ML; in these cases, there are examples in which the determiner is from English and the N is from Spanish.

Second, other corpora exist that cannot be accounted for by MacSwan's analysis. In one corpus, feature-counting clearly does not predict the results. This is a corpus of Italian-Swiss German from Preziosa-DiQuinzio (1992) in which Italian is sometimes the ML and Swiss German is sometimes the ML (the source of outsider morphemes identifies which one is the ML). Myers-Scotton and Jake (1995; 2001) report on this corpus. When Italian is the ML, mixed NPs with Italian determiners and Swiss German nouns occur freely in examples. In comparison with full German NPs (which are EL Islands) when Italian is the ML in this corpus, Swiss German nouns occur with Italian determiners twice as frequently (9/13 or 69% of the occurrences including German determiners).

If feature counting is the criterion that predicts well-formed NPs, then Swiss German determiners should prevail because Swiss German determiners have more features than Italian determiners. Swiss German determiners are marked for CASE as well as for the person, number, and gender that Italian determiners show. When Italian can be identified as the ML, it is no surprise that mixed NPs of Italian determiners and German nouns occur; as noted in section (6), for early morphemes (i.e. determiners) to come from the ML is expected. The fact that German determiners have more features is not relevant, but instead it is relevant that Italian is the ML.

For a fuller explanation of what occurs in CS, again, one must look to the abstract level for a viable explanation. As noted in section (4), determiners include early morphemes because they encode definiteness. These are indirectly elected by their content morpheme heads at the level of the mental lexicon (Bock & Levelt 1994). The *phi* features that they often carry (person, number, and gender) are base-generated at this level. But in some languages, definiteness interacts

with other grammatical features that do not become salient until the level of the formulator. Swiss German determiners are multi-morphemic; they are inflected for person, number, gender, but also for case. Because they include case, which is an outsider morpheme, these determiners cannot be spelled out until grammatical information about case is available (cf. Myers-Scotton 2002:305 for more on such multi-morphemic units). Case is assigned in German varieties by verbs and prepositions. Thus, until constituents larger than the NP are assembled at the level of the formulator, unlike Italian determiners, Swiss German determiners are not available.

MacSwan's argument that aspects of Minimalism alone can explain what occurs and what does not in CS does not explain the data as well as a model that recognizes the asymmetry between the participating languages; the ML construct formalizes this asymmetry.[1] Minimalism, after all, is a model designed to explain structures in monolingual data.

Any explanation of CS must consider which participating language is setting the morpho-syntactic frame of the bilingual clause (the ML). When Italian is the ML and the clause is being assembled, Italian determiners can meet the requirements of the Italian frame, but German determiners are not available at the level when the Italian frame is assembling NPs. German determiners can and do occur in EL German Islands, but as noted above, the majority of German nouns occur in mixed NPs with Italian determiners. This scenario recognizes that morphemes differ at the abstract level in regard to when they are accessed in language production.

Finally, here is another piece of evidence that shows how it is unlikely that comparing feature counts across languages in CS will predict which bilingual phrases are well formed in a corpus. Even though CS data sets of languages in which the ML changes from one clause to the next are rare, I am currently analyzing such a data set. This is a corpus of informal interviews of Xhosa-English bilinguals who are living and working in industrialized Gauteng Province (South Africa), an area that is far from their Xhosa homeland. Xhosa is not an endangered language by any means; it has at least six million speakers. But Xhosa speakers in Gauteng are living in a multiethnic and multilingual setting where English is clearly seen as the language of socio-economic mobility. Further, many Xhosas there must use English in their everyday working lives. The result is that some of these Xhosa may be shifting to English as their main public language even while they maintain their fluency in Xhosa.

In this situation, it is no surprise that many of the Xhosa-

English bilinguals interviewed chose to speak a good deal of English, even though the interviewer biased the interview toward Xhosa by speaking mainly Xhosa her/himself. Example (9) is an extract from one speaker's turn in which she switches from Xhosa to English to frame alternate clauses. The speaker is a young woman who is a student at a technical college, but who runs a business on the side. The clauses are bracketed to make this clear divisions clear.

In such an extract, is there a basis for predicting those clauses that are well-formed in either Xhosa or English? The MLF model does this by examining morpheme order and any outsider morphemes. If they both come from the same language, then the model identifies that language as the ML and predicts that this language will provide the morpho-syntactic frame for the entire clause (except for the possibility of EL islands).

In this extract, it is easy to see that some clauses are entirely in English. There are also a few that are entirely in Xhosa. In these cases, monolingual structure is the rule, of course, and the MLF model does not apply. One English framed clause has an EL island from Xhosa (*e-khaya*). Some of the clauses that are framed by Xhosa include some English elements, such as *improve* in the infinitive U-KU-IMPROV-a 'to improve' and *new* in the word i-NEW. [2] Still, all the outsider morphemes and the morpheme order in these clauses come from Xhosa. These features identify the ML as Xhosa, and, as this contribution has argued, identifying the ML is a way to predict the of overall structure of any clause; it also is a way to predict that this structure will be uniform.[3]

(9) [... mna ndi-phum-a e-Komani eh]
TOPICALIZER 1S-come from-INDIC LOC-Queenstown
'As for me, I come from Queenstown'

LET'S SAY I AM A BUSINESS PERSON AND

[Ndi-fun-a u-ku-*improv*-a [WHERE I'M FROM]]
1S-want-INDIC CL15-CL15improve-INDIC
'I want to improve where I'm from'

[i-Komani a-yi-kho *developed*] LIKE
CL9-TOPICALIZER NEG-CL9-COP developed filler
'Queenstown, it is not developed -- like–
[ndi-z-e nd-a-bon' i-Pretoria i-njani]
1S-COME.PAST-SUBJUNC 1S-PAST-see CL9-Pretoria CL9.COP-how
'I came, I saw Pretoria how it is'

AND [I HAVE BEEN IN OTHER TOWNS]
[BUT EVERY TIME I GO BACK TO E-KHAYA]
LOC-home area
[IT'S STILL THE SAME THING]

[A-ku-kho nto i-*new*]
NEG-COP-LOC thing COP.CL9-new
'But every time I go back to [the] home area, it's still the same thing, there is no thing [that] is new.'

Thus, to the extent that other such data sets exist, they also provide evidence against the notion that counting any type of feature and then comparing the count across participating languages can predict the structure of a given clause. In contrast, identifying the ML is a way to predict structure. For example, if a corpus would include any associative constructions, identifying Xhosa as the ML of a clause would predict the structure of such constructions. An associative construction is illustrated in example (5).

8. Conclusion

The goal of this contribution has been to make an argument about the criterion of uniform grammatical structure as it applies to CS. Specifically, the discussion has emphasized the key role of outsider system morphemes in maintaining this structure in CS. Thus, in order to make clear the abstract aspects of these outsider morphemes, a good deal of space has been devoted to differentiating types of morpheme, paying attention to their abstract as well as surface-level features.

Part of the argument about uniform structure in CS has been devoted to arguing that identifying a Matrix Language (via the testable principles of the MLF model) is the clearest way to predict what constitutes a well-formed bilingual clause in any given CS corpus. This involves taking account of the hypothesis that late morphemes are accessed at a different level in language production from content and early system morphemes. It also involves recognizing that CS corpora do exist in which the ML can change from one clause to the next.

Address of the Author:

Carol M. Myers-Scotton, Linguistics Program, University of South Carolina, Columbia SC 29208 USA <carolms@gwm.sc.edu>

Notes

[1] Jake, Myers-Scotton and Gross (forthcoming 2005) provides a longer, more detailed reply to MacSwan (2005).
[2] Such words, with morphemes from both English and Xhosa, are counter-examples to MacSwan's PF Disjunction Theorem (cf. MacSwan, 2005).
[3] In this example, English is in italics when it occurs within a Xhosa-framed clause; in caps when the speaker uses full English clauses.

Bibliographical References

Bock Kathryn & William J. M. Levelt (1994), "Language production: Grammatical encoding", in Morton Ann Gernsbacher (ed.), *Handbook of psycholinguistics*, London, Academic Press: 945-984.

Bresnan Joan (2001), *Lexical-functional syntax*, Malden MA, Blackwell.

Callahan Laura (2004), *Spanish/English codeswitching in a written corpus*, Amsterdam, Benjamins.

Chomsky Noam (2001), "Derivation by phase", in Michael Kenstowicz (ed.), *Ken Hale: A life in language*, Cambridge MA, MIT Press: 1-51.

Deuchar Margaret (in press), "Welsh-English code-switching and the Matrix Language Frame model", *Lingua* 2007.

Hlavac Jim (2003), *Second-generation speech: Code-switching and morphosyntax of Croatian-English bilinguals*, Bern/New York: Peter Lang.

Jake Janice L. & Carol M. Myers-Scotton (1997), "Codeswitching and compromise strategies: Implications for lexical structure", *International Journal of Bilingualism* 1: 25-40.

Jake Janice L., Carol M. Myers-Scotton & Steven Gross (2002), "Making a minimalist approach to codeswitching work: Adding the Matrix Language", *Bilingualism, Language and Cognition* 5: 69-91.

Jake Janice L., Carol M. Myers-Scotton & Steven Gross (2005). "A response to MacSwan (2005): Keeping the Matrix Language", *Bilingualism, Language and Cognition* 8: 271-276.

Kovács Magdolna (2001), *Code-switching and language shift in Australian Finnish in comparison with Australian Hungarian*, Åbo, Åbo Akademi University Press.

Levelt William J. M. (1989), *Speaking: From intention to articulation*, Cambridge MA, MIT Press.

MacSwan Jeff (2005), "Codeswitching and generative grammar: A critique of the MLF model and some remarks on "modified minimalism"", *Bilingualism, Language and Cognition* 8 (1): 1-22.

MYERS-SCOTTON Carol M. (1988), *Nairobi Corpus*, Unpublished Swahili-English bilingual corpus.

MYERS-SCOTTON Carol M. & Janice L. Jake (1995), "Matching lemmas in a bilingual language production model: Evidence from intrasential codeswitching", *Linguistics* 33: 981-1024

MYERS-SCOTTON Carol M. (1997), *Duelling languages: Grammatical structure in codeswitching* (2nd edition), Oxford, Oxford University Press.

MYERS-SCOTTON Carol M. (2001), "Implications for abstract grammatical structure: Two targets in creole formation", *Journal of Pidgin and Creole Languages* 16: 217-273.

MYERS-SCOTTON Carol M. & Janice L. JAKE (2001), "Explaining aspects of codeswitching and their implications", in Janet Nicol (ed.), *One Mind, two Languages: Bilingual Language Processing*, Oxford, Blackwell: 84-116.

MYERS-SCOTTON Carol M. (2002), *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

MYERS-SCOTTON Carol M. (2003), "What lies beneath: Split(mixed) languages as contact phenomena", in Yaron MATRAS & Peter BAKKER (eds.), *The mixed language debate, theoretical and empirical advances*: 73-106, Berlin, Mouton de Gruyter.

MYERS-SCOTTON Carol M. (2005a), "Supporting a differential access hypothesis, code-switching and other contact data", in Judith Kroll & Annette DeGroot (eds.), *Handbook of bilingualism, psycholingusitic approaches*, New York, Oxford University Press: 326-348.

MYERS-SCOTTON Carol M. (2005b), *Gauteng Corpus*, Unpublished Xhosa-English bilingual corpus.

MYERS-SCOTTON Carol M. & Janice L. JAKE (2000), "Four types of morpheme: Evidence from aphasia, codeswitching, and second language acquisition", *Linguistics* 38: 1053-1100.

Myers-Scotton Carol M. & Janice L. JAKE (2002), *Sources of inflection: Testing the Creole System Morpheme Hypothesis*, Paper presented at the annual meeting of the Society for Pidgin and Creole Linguistics, San Francisco CA.

PREZIOSA-DI QUINZIO Ivana (1992), *'Teoreticamente la firma fa indietro': Frammistione di italiano e Schwyzertüsch nella conversazione di figli di emigrati*, Lavoro di licenza in linguistica italiana, Facoltà di Lettere e di Filosofia, Università di Zurigo.

TÜRKER Emel (2000), Turkish-Norwegian codeswitching: Evidence from intermediate and second generation Turkish immigrants in Norway, Unpublished Ph.D dissertation, Oslo, University of Oslo.

TURNER Lorenzo D. (2002), Africanisms in the Gullah dialect, Columbia SC, University of South Carolina Press [earlier editions, 1949, 1974; University of Chicago Press].

ZIAMARI Karima (2004), Le codeswitching dan les conversations de l'Ensam, Unpublished dissertation, Morocco, L'École Nationale Supérieure d'Arts et Métiers (ENSAM).