Bilingualism, Language and Cognition 1 (1998) 39–50 © 1998 Cambridge University Press

A formal production-based explanation of the facts of code-switching*

DAVID SANKOFF
Centre de recherches mathématiques, Université de Montréal

We construct a production model of code-switched discourse, in order to explain empirical observations about switch-point distribution, well-formedness of monolingual fragments, conservation of constituent structure and lack of constraint between successive switch points. The two monolingual grammars are modeled by congruent context-free phrase structure grammars. Each code-switched sentence makes alternate reference to two virtual monolingual sentences, one in each language, and, in the immediate vicinity of switch points, to their constituent structures. Imposing conservative conditions on when a constituent should be labeled as to language, and invoking a constraint against real-time "look-ahead" from one code-switch to the next, we prove that the existence of a consistent tree labeling, easily monitored by the speaker, implies a constraint on local equivalence of constituent order in the two languages around a switch point. This constitutes an explanation, without invoking a "code-switching grammar", of the observed tendency for equivalence-point code-switching.

Introduction

The modern motivation for studying code-switching was initially to explain the observation that in bilingual communities, speakers tend to switch from one language to another intrasententially at certain syntactic boundaries and not at others (Gumperz & Hernandez, 1969). The first general explanation to account for this distribution was Poplack's argument that switching should be favored at the kinds of syntactic boundaries which occur in both languages, thus avoiding word order that might seem unnatural according to one or both grammars: the *equivalence constraint* (Poplack, 1978, 1980; see also Lipski, 1977; Pfaff, 1979).

Critics of this reasoning complain that in its focus on adjacency in surface structure – or worse, on only the linear order of lexical items, it does not derive from paradigms of linguistic explanation whereby surface adjacency is a secondary consequence of deep hierarchical structures or, more recently, of relations among categories and lexical items which must be satisfied through a scheme of hierarchical constraints (Rivas, 1981; Woolford, 1983; Di Sciullo, Muysken & Singh, 1986; Pandit, 1990; Belazi, Rubin & Toribio, 1994). Moreover, there is a tradition of trying to analyze code-switched sentences by permitting the substitution of a constituent from one language at a corresponding node in a phrase structure in the other language (Sridhar & Sridhar, 1980; Rivas, 1981; Woolford, 1983; Joshi, 1985; Mahootian & Santorini, 1994). In fact, there is no contradiction between the notion of hierarchy and the equivalence constraint; both can be formally incorporated into the same model (Sankoff & Mainville, 1986).

There are some other facts about code-switched sentences, however, that are perhaps more fundamental than equivalence. One is the *well-formedness of monolingual fragments* within such sentences. This is found whether a fragment constitutes a complete constituent or stretches across two or more possibly incomplete constituents. Another observation, formulated more precisely below, is the *conservation of constituent structure*. Insofar as a constituent may have the same structure in the two languages involved, in terms of the subconstituents it contains, even if these are ordered differently, a constituent containing a code-switch will also have the same structure.

The remaining fact about code-switching is that it is *unpredictable*. That is, even if we can determine where a code-switch can occur and where it cannot, there is no way of knowing in advance for any site whether a switch will occur there or not. In particular, if a switch occurs at some point in a sentence, this does not constrain any potential site(s) later in the sentence either to contain another switch or not to – there are *no forced switches*.

* Research supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Social Science and Humanities Research Council of Canada. The author is a Fellow of the Canadian Institute for Advanced Research.

Address for correspondence
Centre de recherches mathématiques, Université de Montréal, CP 6128 Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada
E-mail: sankoff@ere.umontreal.ca

Code-switching is essentially an oral discourse mode, and considerations of oral production must be given weight at least equal to those of syntactic well-formedness. In the model proposed here, processes and criteria pertaining to production are paramount. We do not assume that the mechanisms of switching from one language to another can be deduced entirely from the general principles of monolingual grammars. Thus we do not analyze the distribution of intrasentential switch points in terms of a grammar of any of the types ordinarily used for accounting for single languages, but by means of a left-to-right process that refers to two well-formed monolingual sentences (i.e. each satisfying the constraints of an "ordinary" monolingual grammar for one of the two languages) in producing monolingual sentence fragments and in evaluating potential switch points between these fragments.

The research reported in this paper is entirely deductive. Though we make assertions and hypotheses about bilingual discourse, and though we discuss their validity to some extent, we refer to the appropriate literature for their detailed justification and devote the body of our discussion to the mathematical derivation of our conclusions. Furthermore, our examples (of English/French mixing) are fabricated, and their well-formedness (or not) asserted, solely to illustrate our arguments; they do not constitute empirical data. To the extent that our more general "facts" about bilingual discourse discussed above are true, and to the extent that our simple model captures some essential aspects of the complex processes of natural language processing, it is an important objective to see whether the model explains the observations. This would set the stage for a wider range of empirical observations and more ramified mathematical linguistic models.

In the following sections, we first state explicitly the goals of our research in terms of the facts to be explained, the kind of explanation we seek, and our strategy for constructing it. We then set out some of the empirical motivations for this work and justify our use of monolingual context-free phrase-structure grammars in terms of their tractability and their ability to model the connection between hierarchy and linearity. We then elaborate a production-based model of code-switching which has the desirable properties of conserving well-formedness of monolingual fragments and coherent constituent structure. But it also permits unrealistically free juxtaposition of these fragments and has the implausible property of producing sentences in which certain code-switches are *forced*, or entirely predictable in advance. To solve these problems we first develop the notion of minimal language labeling of the non-terminal category nodes of a bilingual phrase structure. For binary constituent-subconstituent structure, this allows us to prove directly that a code-switched sentence satisfies a formal definition of the equivalence constraint. For more general constituent structures, we explicitly invoke a *no forced switches constraint* at the production level. This also leads logically to the equivalence constraint within our model.

Objectives, background and strategy

Monolingually-based theory and bilingual discourse. While formal theories of grammar may well account for monolingual language in terms of general linguistic principles, there is no reason to believe that processes which juxtapose two languages can be explained in exactly the same way. The reasons implicit or explicit in attempts to do so have to do with explanatory economy, either of individual linguistic competence or of linguistic theories. This seems specious, since both are based on a notion of a "hard-wired" human linguistic faculty evolving in prehistoric monolingualism. Experience in widely diverse speech communities suggests instead that code-mixing strategies, including code-switching, evolve in the lifetime of particular communities, are only partly dependent on linguistic typology of the two languages and exhibit widely different patterns of adapting monolingual resources for incorporating linguistic innovation: borrowing, inflection-carrying "dummy" verbs, flags, tags and other discourse-level devices, but very little simultaneous access to the grammatical apparatus of the two languages. Nonetheless, intrasentential code-switching involving lengthy monolingual fragments in each of the languages is frequent in some communities, and understanding the mechanism for resolving inter-language word-order differences within such sentences is a fundamental problem.

Precursors to a production model. That monolingual fragments are not co-extensive with entire constituents is problematic for any model relying on hierarchical relations for deciding well-formedness, since such models are designed primarily to ensure well-formedness of entire constituents, monolingual or bilingual. They cannot ensure that adjacent same-language parts of neighboring constituents are compatible (i.e. yield a well-formed fragment when juxtaposed), since these parts may not even be in the same language as the rest of the constituents that contain them (Muysken, 1995). For example, an earlier model (Sankoff & Mainville, 1986), using the context-free grammars of two languages to account for code-switched sentences satisfying the equivalence

constraint, could not ensure the well-formedness of monolingual fragments, for the very reason Muysken has pointed out.

This problem is at the core of the conflict between hierarchical and linear modes of explanation. We will resolve it by ascribing ultimate responsibility for the well-formedness of monolingual fragments to production-level processes. These fragments, arbitrary substrings of pre-constructed well-formed monolingual sentences, are pieced together during linear production in a way which corresponds to the other general observations about code-switching and which is essentially neutral with respect to theories of monolingual grammar.

Projection in monolingual versus bilingual production. From the initial stages in the production of a monolingual sentence, speakers must project some aspects of the potential structure of the entire clause or sentence, in order to be able to convey messages of some semantic complexity. This probably includes a general idea of constituent structure, some of the lexical items to be used in particular constituents, and possibly more details such as final components of discontinuous elements, sentence-internal co-reference, remote grammatical agreement, etc. This projection is computationally costly, but it would not otherwise be possible to reconcile left-to-right processes of sentence production with the system of constraints constituting a monolingual grammar, in which left-to-right relationships are not predominant. Of course, it is the part of the sentence being constructed which is monitored in detail as to its grammatical compatibility with what has already been said and with the not yet concretized projections of the rest of the sentence.

What of code-switched sentences? Clearly the monolingual fragments within them must manifest the same details of internal planning as other monolingual discourse. But the situation is not as clear when it comes to the grammatical relationships *between* monolingual fragments, especially if they are in different languages. In fact, there is a distinct lack of evidence for any detailed grammatical projection or planning which constrains elements in the interior of two or more such fragments. Moreover, there is no planning of the code-switching itself; a speaker does not project a specific series of sites at which to change from one language to another and then back again.

In this paper, consequently, we will treat the connection between different language fragments more conservatively than may be necessary or desirable for intra-fragment structure. At the sentence level, projection or planning will be restricted to (largely unordered) constituent structure and to lexical content, without specifying language. Monolingual fragments will be produced in accordance with the full grammatical apparatus of the appropriate language. (This will remain implicit, since our model is independent of the specific monolingual grammatical theory.) For the assembly of these fragments into a code-switched sentence, the emphasis will be on monitoring local structure around the switch point.

Code-switching and borrowing. Though the equivalence constraint (formalized in the sixth section below) has been verified as a general tendency in several communities – Puerto Rican Spanish and English in New York (Poplack, 1980), Finnish and English (Poplack, Wheeler & Westwood, 1987), Tamil and English (Sankoff, Poplack & Vanniarajan, 1990), Wolof and French, and Fongbe and French (Meechan & Poplack 1995) and others – there are actually relatively few data on which it has been tested, since most of the voluminous literature on code-switching, especially that on "insertional" switching, is based on data which represent, properly speaking, lexical borrowing (Eliasson, 1991; Mahootian & Santorini 1994; Backus, 1996). Only the word-order of the recipient language is pertinent to borrowing, so attempts to understand code-switching based on a mixture of borrowing and true switching data (e.g. Myers-Scotton, 1993) are doomed to be unwieldy at best and descriptively inadequate.

In the model to be constructed here, the borrowing process is not relevant. Loanwords, including those *ad hoc*, "nonce", or momentary, uses, are not excluded, but simply considered to behave as native lexical items with respect to word order. This is in conformity with a large number of studies of these forms (Poplack, Sankoff & Miller, 1987; Sankoff, Poplack & Vanniarajan, 1990; Poplack & Meechan, 1995).

An outline of the research strategy. Our model is based on the assumption that bilinguals are fully competent in their two languages, that there is no convergence of the two monolingual codes even during bilingual discourse, and that a code-switched sentence consists of fragments of two monolingual sentences (each one a translation of the other) pieced together. This is done first through an otherwise unconstrained production model that simply copies part of one monolingual process followed by part of the other in such a way that constituent structure is conserved. The resulting process satisfies neither equivalence nor unpredictability. By adding two very simple rules for labeling some constituents according to their language, the existence of a consistent labeling – which can be easily monitored during real-time linear production – turns out to guarantee both equivalence and, for most situations, unpredictability.

To discover the logical consequences of the constraints and assumptions in a model, the best strategy is to see how they work out in the simplest possible model consistent with them, without the confounding effects of other restrictions and devices. We are interested in seeing how two hierarchically structured languages resolve their word-order differences during intrasentential code-switching, without the confounding effects of other linguistic phenomena. Thus we will construct a model where the only differences between two languages have to do with word order and the (phonological) form of lexical items, and work out the logical consequences for code-switching of various assumptions and constraints on the production of bilingual sentences. Linearity and hierarchy are the key structural aspects here, and the simplest class of recursive grammars accommodating these properties is that of context-free grammars. Thus in focusing on the relationship between word order and hierarchy, we are choosing a model not adapted to the treatment of tags and moveable elements such as many adverbials (whose "switchability" is uncontroversial), nor of remote relationships such as discontinuous constituents, internal coreference, etc., whose effects on code-switching, if any, have never been systematically documented.

The syntactic mode

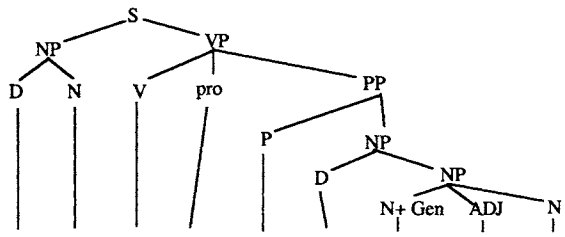
Consider a context-free grammar consisting of a set C of "categories" or non-terminal symbols, including one distinguished symbol S (for "sentence"); a set of terminal symbols T ("lexical slots"), none of which are also in C, a lexicon L which is basically a set of words and an indication of the kind of lexical slot each word can fill, and a set of rewrite rules R of form $c \rightarrow v_1 \ldots v_n$ where c is a symbol in C, and the string on the right hand side $v_1 \ldots v_n$ consists of one or more symbols in T or C. Recall that a sentence derivation is formed by writing down S, then rewriting S by the string $u_1 \ldots u_m$ on the right-hand side of any rule in R of the form $S \rightarrow u_1 \ldots u_m$, then rewriting any u_i which is non-terminal by some rule of the form $u_i \rightarrow w_1 \ldots w_p$ and so on. Rewriting a symbol does not affect the other terms in the string, so that it does not matter which non-terminal symbol of a string is rewritten first. Whenever a lexical slot appears in the string it can be filled with words of the appropriate category from L. When there are no more non-terminal symbols in the string (and R must be such that this is always possible), the derivation stops, and the resulting word string is just a sentence generated by the grammar.

We will make use of the familiar phrase structure tree representation for the derivation of a sentence. Each symbol appearing in the derivation is represented by a node of the tree. This symbol is said to dominate the constituent or subtree of which it is the highest node, and may be used to represent that constituent. In this definition each terminal node is itself a constituent. When a rule is used to rewrite a non-terminal symbol, the element on the right-hand side of the rule are disposed immediately below it, in the same left-to-right order as appears in the rule, and are connected to it by branches of the tree. Note that the words in each constituent form a (contiguous) substring of the sentence.

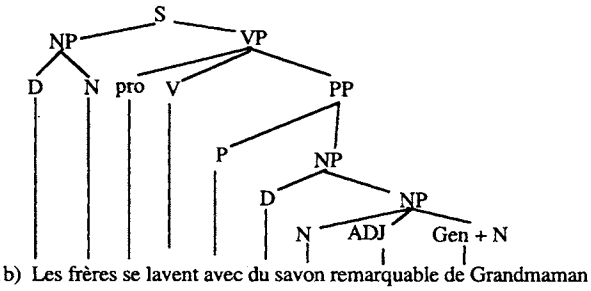
In order to speak of code-switching between two different grammars, it is necessary to have some connection between the categories of one and the categories of the other. In natural languages, such correspondences will be incomplete and will admit many exceptions (Muysken, 1995), but since this is peripheral to our interest in word order, we will simply make the strong assumptions of *lexical translatability* and *categorial congruence*, meaning that there is a one-to-one correspondence between the lexicon L_A of language A and the lexicon L_B of language B, though the words are all recognizable as coming from one language or the other, and a one-to-one correspondence between the categories in C of the two languages and between the lexical slots in T of both languages; for the elements of C and T we use the same symbols for both languages except in certain cases when they carry language labels (see the fifth section below). Furthermore we assume *grammatical congruence*: there is a one-to-one connection between the rules R_A of language A and R_B of language B – if R_A contains a rule $c \rightarrow v_1 \ldots v_n$, then R_B must contain a rule $c \rightarrow u_1 \ldots u_n$, where each symbol in $v_1 \ldots v_n$ has its counterpart in $u_1 \ldots u_n$, and vice-versa, though the order of the terms in one string will not in general correspond to the order of the terms in the other. Finally, for convenience, we will assume *fixed word order*, that is if $c \rightarrow v_1 \ldots v_n$ in a given grammar, then there may be other rules rewriting c in that grammar, but none where the right-hand side contains exactly the same set of symbols $\{v_1, \ldots, v_n\}$.

To simplify our presentation in this article, we do not allow ambiguity in our grammars, though this is much more restrictive than necessary for the model. Not only must each monolingual sentence be derivable in exactly one way, but each rule may contain any one symbol only once on its right-hand side.

In comparing the structure of two sentences, we say that they have the same constituent structure if there is a one-to-one correspondence between their constituents such that if x in one sentence corresponds to y in the other, then the (unordered) set of



(1a) The brothers wash themselves with some of Grandma's remarkable soap



(1b) Les frères se lavent avec du savon remarquable de Grandmaman

subconstituents of x corresponds to the set of subconstituents of y.

The consequences of our assumptions are summarized in:

Theorem 1. *Every sentence in language A has a unique counterpart in language B which has the same constituent structure and whose lexical items are translations of those in the language A sentence.*

Proof. There being no ambiguity, the sentence from language A can be parsed uniquely. By categorical and grammatical congruence, the derivation thus obtained has a unique counterpart using rules from the grammar of language B, and the translatability assumption provides the unique words for each lexical slot in the resulting sentence.

Examples (1a,b) are two (fictitious) sentences in English and French which we may imagine to be counterparts of each other according to some grammatical analysis of the two languages, in the sense of Theorem 1.

Despite differences in word order, the constituent structure is identical and the lexical items are word-for-word translations (without quibbling about the questionable lexical status of the reflexive clitic and the genitive particle and the somewhat different internal structure of determiner in the PP).

We will have occasion to assess monolingual "fragments" (i.e. strings of words which are not complete sentences) for well-formedness. We define any contiguous substring of a well-formed sentence string to be a well-formed fragment.

The production model

In the model, the production of a code-switched sentence presupposes the existence of two virtual sentences, one in language A and one from language B, counterparts of each other as in Theorem 1. For each of (2a,b,c,d) the pair of virtual sentences is the one illustrated in (1a,b).

(2a) The brothers wash themselves | avec du savon remarquable de Grandmaman
(2b) Les frères se lavent | with some of Grandma's remarkable soap
(2c) The brothers | se lavent avec | some of Grandma's remarkable soap
(2d) The | frères | wash themselves | avec | some of Grandma's remarkable soap

That we require two complete sentences is an analytical convenience such as will be discussed later; we could do with much less, but this would greatly complicate the presentation.

Given the two virtual sentences in languages A and B, the code-switched sentence is produced by taking part of one of them, followed by part of the other, and so on, without using any word (or its translation) more than once, until every lexical element (or its translation) has been used up. This description applies not only to plausible code-switched sentences such as those in (2), but also to any combination of elements in any order as in (3).

(3) de Grandmaman | the with | remarquable frères | soap themselves wash some of

To arrive at an empirically and conceptually satisfactory model, we must add constraints. The first constraint is motivated by the empirical observation that each monolingual fragment in bilingual discourse tends to be well-formed in its lexifier language. We will assume that the production of the sentence starts with a word in (either) one of the virtual sentences, and copies successive words from left to right in that sentence without skipping any until there is a code-switch to some word in the other virtual sentence. From this point in the other virtual sentence, production continues from left to right, and so on. When the left-to-right production arrives at a word (such as *se* after *frères* in example 4) which has already been used in the current or the other language (also *some of* after *with*) or at the end of one of the virtual sentences, there must be a switch to the other virtual sentence or, if all the words have been used (after *remarkable*), the production must stop.

(4) du savon | wash themselves with | les frères | Grandma's remarkable

Note that although repeat-translations, or *palindromic* constructions (**se lavent themselves*), do occasionally crop up in some corpora (Nishimura, 1986; Poplack, Wheeler & Westwood, 1987), these tend to

be rare and evidence production-level difficulties (hesitations, autocorrection, etc.).

Theorem 2. *The monolingual fragments in a code-switched sentence produced by left-to-right copying are well-formed.*

Proof. Because no "skipping" is allowed in the left-to-right copying process, every fragment is a contiguous substring of one of the (well-formed) virtual sentences. Thus these fragments are by definition themselves all well-formed.

The *left-to-right* assumption ensures that monolingual fragments are well-formed, as illustrated in (4), as well as (2). By itself, however, it does not constrain how the alternating fragments in one language and the other are related; indeed it allows them to be juxtaposed in any order, as long as the fragment languages alternate; sister elements in the same constituent in a virtual sentence may find themselves remote from each other in the code-switched sentence, as with *remarkable* and *savon* in example (4). As mentioned in Section 1, however, empirical research confirms that constituent structure, insofar as content and embedding or nesting relations are concerned, is conserved even if the constituent contains a code-switch.

To conform to this observation, we make a second assumption, that once the production process enters or switches into a constituent, it must exhaust all the lexical slots in the constituent, in one language or both, before returning into a higher-level constituent or entering a sister constituent. In other words, each time it enters a deeper, or more nested, subconstituent, it cannot "get out" until that subconstituent is exhausted. Note that this assumption is independent of whether or not the production follows the left-to-right process described above, so that the sentence (5) satisfies *most nested first*, but not *left-to-right*.

(5) (wash((some of (Grandma's soap remarkable)) with) themselves) | (frères les)

Theorem 3. *Lexicalizing constituents in a "most nested first" way is a sufficient condition for conserving the same constituent structure in the code-switched sentence as in the two virtual sentences.*

Proof. Constituent structure determines a (sub)-string of words defined by each constituent, and it is the content of this string (not the order) which identifies the constituent. These substrings are non-overlapping unless one is completely included in the other, representing a subconstituent or a subconstituent of a subconstituent, etc. Note that by changing the order of its subconstituents, the order of the words in the string corresponding to a constituent will change, but not the content of the string. Thus, by Theorem 1, the two virtual sentences have the same constituent structure.

We use a kind of induction argument on the tree structure represented by the phrase structures. We first show how low-level constituents in the virtual sentences can be identified in the structure of the code-switched sentence, and then build on this to show the same thing for successively higher order constituents.

For any constituent in the two virtual sentences whose only subconstituents are lexical slots (and there must always be at least one such, for a finite tree structure), once one of these slots is used in the production of the code-switched sentence, the assumption ensures that the rest must follow immediately, in some order. Thus a constituent can be identified as including exactly the resulting string. All such lowest-level constituents in the virtual sentences are thus reflected in constituents of the code-switched structure, and it is clear that they do not overlap.

For a higher-order constituent in the virtual sentences, suppose that all its subconstituents have been identified in the code-switched sentence (and there must always be at least one such, for a finite tree structure). They do not overlap, by previous construction, but must all be contiguous, otherwise some word in one of the virtual sentences, out of the constituent, would have been used in the code-switched sentence before exhausting that constituent. Thus a constituent can be identified as including exactly the union of the strings determined by the subconstituents. By the contiguity of the subconstituents, we can also conclude that this constituent does not overlap with any previously constructed constituent not nested in it.

The tree structure ensures that eventually each such constituent (including *S*) in the virtual sentences will be reflected in a contiguous substring of the code-switched sentence, and that none will overlap with any other unless one is nested in the other, in the same way as in the virtual sentences.

The *most nested first* condition and the *left-to-right* assumption are independent, in the sense that neither implies the other, as is clear from (4) and (5). Neither excludes the other, and together, as in (6) and (2), they produce code-switched sentences with well-formed monolingual fragments and the same constituent structure as the virtual sentences.

(6a) (se lavent | (with (some of (Grandma's remarkable soap)))) | (les frères)

(6b) (frères | the) | ((avec | (some of (Grandma's | savon remarquable))) | wash themselves)

The language of constituents and subconstituents

The model in the previous section, though it produces sentences like those in (2) and (6) with some desirable properties, is not complete. With only the two conditions, certain configurations may occur, such as those in (6), that are clearly unrealistic. For example, if a monolingual fragment being copied includes a word which must be final position in a constituent, like *frères* in (6b), and if the words of the constituent in one virtual sentence or the other have not yet been used up, then an immediate code-switch, to *the* in this instance, is obligatory to satisfy *most nested first* – the monolingual fragment cannot continue, even though it may have a natural continuation into another constituent. This, and other instances of forced code-switching, are clearly not phenomena observed in real bilingual discourse.

A type of construction not found in natural bilingual corpora but permitted in the simple production model might include a code-switched sentence which begins with a fragment of the virtual sentence in language A which would never occur in sentence-initial position in monolingual discourse in language A, such as *se lavent* in (6a). More generally, if there are several sister subconstituents in a constituent, they may be permuted in any order, as long as there is a switch between each adjacent pair. Another possible output from this model, even if the two monolingual grammars contrast completely, is a code-switch between every adjacent pair of words in the sentence.

Thus the output of the production process as is hitherto formulated seems unduly constrained from the production point of view (by forcing switches) and not constrained enough from the perspective of the output structure. In modeling bilingual discourse, it is not realistic to require speakers to switch at a certain point or before a certain point in the sentence construction as a consequence of previous switches in the sentence. Nor is switching a license to abandon syntagmatic ordering completely. We thus come to the main point of this research: short of the equivalence constraint itself, can we motivate some constraint to account for the observation that switching occurs almost exclusively at equivalence points (notion to be formalized in the next section) and virtually all equivalence points seem to be eligible switch points? And can this be done in such a way that during production, the model speaker avoids any switching which will obligatorily require compensatory switches later on in the sentence construction (switch planning or forcing)? Furthermore, can we do this without referring to facts of particular languages, properties of particular grammatical categories, or even the mechanisms of particular theories of monolingual grammar?

Our approach to this problem is to postulate a limited degree of structural monitoring. Monitoring the monolingual fragments for well-formedness is uncontroversial, and we need not enter into the details. What we propose for monitoring at the switch points is the "language label" of the constituent in which the switch occurs and of its immediate subconstituents.

We first ask: what parts of the constituent structure of the code-switched sentence may be with certainty ascribed to one language or the other only, and hence should be labeled accordingly? Certainly (i) all terminal symbols – lexical slots – are labeled according to whether the word filling the slot comes from L_A or L_B, since all words are identifiable as to their language, by definition in our model of congruent grammars; and (ii) at the constituent level, any constituent, all of whose immediate subconstituents have the same label, should itself have this label. Anything else would be inconsistent. We will propose a third criterion for labeling constituents, but we first prove the following:

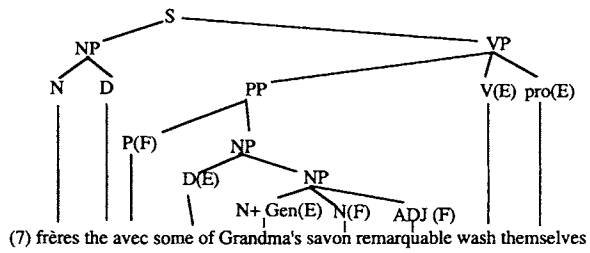
Theorem 4. *Any non-terminal node carrying a label for whatever reason must have at least one immediate descendant node which has this same label or is unlabeled, and one descendant (possibly a lexical slot) which has the same label.*

Proof. By (ii), not all the immediate subconstituents of a labeled constituent can have the contrasting label, so at least one must have the same label or be unlabeled. By the same token, not all the lexical slots in the constituent can have the contrasting label.

Different instantiations of this model may actually specify that certain subconstituents "inherit" the label – in one theory the determiner may inherit the label of the noun phrase, in another theory it may be the noun itself.

Requirement (ii) depends on constituent content, but not on constituent order, so that it applies meaningfully to any sentence satisfying most nested first such as (5) or (6b). The labeling of (6b) is illustrated in (7).

To constrain the order, we are motivated to try to exclude situations such as a declarative sentence which begins with a well-formed verb phrase entirely in English followed by a subject noun phrase in another language. More generally, (iii) any subconstituent which is out of rank order position among its sister subconstituents according to one of the languages, must receive the label of the other language. For example, if the languages A and B are SVO and VSO, respectively, and the code-switched sentence has order VSO, then the labeling should be V^BS^BO; if



(7) frères the avec some of Grandma's savon remarquable wash themselves

the code-switched sentence were SOV, then condition (iii) could not be satisfied, since O is out of order according to *both* languages, as is the V.

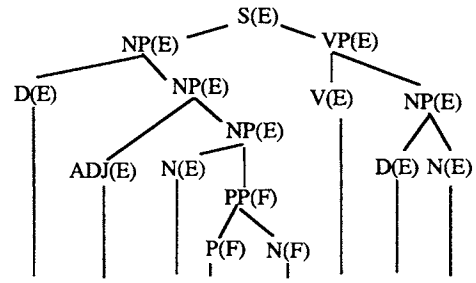
If conditions (i)–(iii), all of which are well-motivated, cannot be simultaneously satisfied, the code-switched sentence cannot be considered well-formed. Thus in our hypothetical example with a sentence-initial English verb phrase, requirements (ii) and (iii) conflict, so the sentence is not well-formed. And in example (7) condition (iii) cannot be satisfied with respect to any of the categories lexicalized by *frères*, *the*, *savon*, *remarquable*, *wash* and *themselves*, nor the PP node – according to the extremely constrained grammars responsible for examples (1a,b). On the other hand, each of examples (2a,b,c,d) satisfy all of the conditions (i)–(iii).

For a model speaker, all that has to be monitored is the language label of the constituent in which a potential switch occurs and that of its subconstituents. One detail, which we will resolve in the seventh section, is that it would be preferable not to have to monitor the label of all subconstituents *after* the potential switch point, except of course the one immediately following this point, i.e., we would prefer to avoid "planning" switches. This is automatically (and trivially) achieved for binary constituents, but lacking further constraints, does not hold true for constituents with more than two subconstituents.

No additional labeling is warranted within this framework, though each theory of code-switching may have its own particular way of assigning a label to each and every constituent (Rivas, 1981; Woolford, 1983; Joshi, 1985; Di Sciullo, Muysken & Singh, 1986; Myers-Scotton, 1993). In particular, there is no conceptual justification or need for postulating an underlying (or "matrix") language for the entire sentence itself when this is not motivated by criteria (ii) and (iii).

The equivalence constraint

In the string of words which constitute a code-switched sentence in our model, there is no problem in identifying where a fragment in language *A* stops and one in language *B* starts. On the constituent level, however, it is not as obvious where this switch



(8) The original Duke | de Lorraine | had 10 000 men

should be located and sometimes even whether or not there is an inter-constituent switch, as in (8).

This is an example of string-level code-switching with and without corresponding constituent-level switches. The rule NP → ADJ+NP in English and NP → NP+ADJ in French results in the lowest NP being labeled E, by requirement (iii). The higher NPs, the VP, the object NP and the S are labeled E, and the PP labeled F because all of their subconstituents are (requirement (ii)). The code-switch between *Duke* and *de* is also a constituent-level switch between *Duke* and the PP, but the switch between *Lorraine* and *had* is not reflected by a switch between the highest NP and its sister VP since they are both labeled E.

In general, what constitutes a code-switch between two adjacent sister constituents? The only reasonable answer is that one constituent is labeled *A* and the other *B*. What happens if two differently labeled sister constituents are separated by one or more unlabeled constituents? Once again it is clear that there has been a code-switch at the constituent level, but the site cannot be pinned down, other than by saying that it occurred in the interval between the two labeled constituents. Note that there also must be switches at some lower levels *within* each of the intervening unlabeled constituents; otherwise they could not be unlabeled, by criterion (ii). Thus we could imagine that the switch occurred gradually over the interval between the two labeled constituents. However, we will instead say that this labeling indicates a switch between one of the adjacent pairs of constituents in the interval, which includes the constituent labeled *A* and the one labeled *B*, but which pair in particular cannot be determined.

We can now state the equivalence constraint. Consider the corresponding rules for ordering the n subconstituents of a given constituent in language *A* and language *B*. If the sets of the first i symbols on the right-hand side of the rules are different in the two grammars (and hence the set of the last $n - i$ symbols are also different, since the two rules are congruent), then the equivalence constraint prohibits a code-switch between the *i-th* and *i+1-st* subconsti-

tuents. Otherwise the boundary between the two subconstituents is an *equivalence point* and a code-switch is permitted. In the case $n = 2$, this reduces to the prohibition of a code-switch between the two subconstituents unless the two languages order them in the same way.

The main theorems

The production model in the fourth section and the labeling rules in the fifth section ensure that monolingual fragments are well-formed, constituent structures are correct and that no constituent labeled X appears within a higher constituent in a rank order position not permitted in language X. This does not mean that the equivalence constraint holds. Consider for example the rules $c \rightarrow$ x y z and $c \rightarrow$ z y x in languages A and B respectively. Then the model as it is now constituted would permit the constituent order $x^A\ y^B\ z^A$, with two constituent-level code-switches, both of which violate the equivalence constraint. (cf. *Grandma's* | *remarquable* | *soap* or *savon* | *remarkable* | *de Grandmaman.*)

In one important case, however, equivalence always holds. Monolingual grammars are said to have binary constituent-subconstituent structure if there are at most two symbols on the right-hand side of every rule. Then the following holds:

Theorem 5. *Given a well-formed code-switched sentence where the monolingual grammars have binary constituent-subconstituent structure, if two sister subconstituents are labeled A and B, respectively, the code-switch between them satisfies the equivalence constraint.*

Proof. The constituent must have the same subconstituent order according to the rules of both languages, otherwise labeling convention (iii) would be in conflict with the labels hypothesized. Thus the constituent to the left of the switch point is the same in both languages and the constituent to the right of the switch point is the same in both languages, verifying the equivalence constraint.

As we have seen, however, the theorem may not hold if rules may have more than two terms on their right-hand sides. The counter-example shown above, for example, represents the insertion of a language B subconstituent into an otherwise language A constituent. This requires two code-switches, one before and one after the inserted subconstituent. If a code-mixing strategy were to be based on the insertion of constituents in this way, every code-switch before an insertion would require the speaker to plan for an appropriate second code-switch later on in the sentence.

We have argued above that while such planning ahead must be incorporated into monolingual production models, the distribution of code-switches in bilingual corpora is more consistent with a hypothesis of the independence of successive code-switches. Given the switch from one language to the other at a particular site, there seems no way of predicting where the next switch will occur. There are some exceptions: in rare bilingual communities, the code-mixing mode of discourse may include the possibility of inserting one specific type of constituent into positions where it would not occur monolingually. This has been characterized quantitatively in Naït M'Barek and Sankoff (1988) and Sankoff et al. (1990). Nevertheless, in a production model of code-switching, it would seem more generally valid to assume a *no forced switches* (or *no planning*) regime, as discussed in the fifth section. Thus, we will assume that where a switch takes place between two constituents, that the well-formedness of the code-switched sentence does not require further switches later on in the left-to-right order.

Theorem 6. *Given a well-formed code-switched sentence, if two sister subconstituents are labeled A and B, respectively, and there is no labeled subconstituent between them, then under the no forced switches assumption, this indicates that there is a code-switch satisfying the equivalence constraint between some two adjacent subconstituents in the interval between the labeled subconstituents.*

Proof. If there are several such pairs, we first examine the leftmost. Suppose the language A subconstituent is the i-th and the language B one is the j-th, where $i < j$, out of a total of n sister subconstituents. Note that $n > 2$. Consider the rules in language A and language B that determine the order of these n subconstituents. Between the i-th and j-th subconstituent, there are no labeled sister subconstituents, so both rules coincide at these positions, and a switch can be considered to occur between any consecutive two of the i-th, $i+1$-st, $i+2$-nd, . . . , $j-1$-st, j-th subconstituents.

Now we focus on some one of these adjacent pairs, say the h-th and the $(h+1)$-st subconstituents. Suppose the equivalence constraint did not hold here. Then there would be $r > 1$ terms among the first h terms on the right-hand side of the rule in language A not among the first h terms on the right-hand side of the rule in language B and vice versa, by congruence and fixed word order assumptions.

By the supposition above that i and j are the leftmost subconstituents with contrasting labels, the language A version of the r subconstituents would have to appear in the structure of the code-switched sentence and, by labeling convention (iii), would be labeled as such. Conversely, the r language B subcon-

stituents with no language *A* versions among the first *h* subconstituents could not be used in the structure. Versions of these *r* subconstituents would have to appear, labeled or not, somewhere among the last $n - h$ constituents, and wherever they appear they must be labeled *A*, by (iii), since the corresponding *B* versions cannot be in the same rank order positions; indeed the latter all occur in the first *h* positions.

In other words, a switch from *A* to *B* between the *h*-th and the *(h+1)*-st subconstituents requires a switch back to *A* later in the constituent, a requirement which contradicts the no forced switches condition. Thus the equivalence constraint must hold if the switch occurs between the *h*-th and the *(h+1)*-st subconstituents, and similarly for all the other boundaries between the *i*-th and *j*-th subconstituents.

When there is more than one switch between the subconstituents within a constituent, the above argument may be repeated, switch by switch, each time considering only those subconstituents occurring after the previous switch.

The psychological and linguistic significance of the model

Four aspects of the model must be evaluated, the production model based on alternate reference to two virtual monolingual sentences under the *left-to-right* and *most nested first* constraint, the use of perfectly congruent context-free grammars for languages *A* and *B*, the constituent labeling rules, and the *no forced switches* hypothesis.

The idea of using virtual sentences is an extension of concepts implicit in Poplack's original discovery (1978, 1980) of the importance of equivalence sites to code-switching and is what distinguishes it from attempts to account for code-switching using purely distributional data – examples of sentences thought to be either well-formed or not. The notion of equivalence implies the comparison of the sentence actually produced with what "could have been said" in either of the monolingual modes. Though the comparative data are of course not directly accessible, since only one sentence is uttered, controlled inference about unrealized possibilities is consistent with rigorous methodology (cf. the notion of the linguistic variable (Labov, 1969; Sankoff, 1988)).

Postulating two *complete* virtual sentences, however, is an analytical convenience. All that would really be needed in a more realistic analysis are the parts of each sentence that are actually used plus some additional details (subconstituent structure) about the constituent within which a switch occurs. Constructing a model on this basis would complicate the formulation of the *most nested first* procedure and the *no forced switches* hypothesis, but would not essentially alter them. The importance of this device, and its realism, lie largely in the way the monolingual fragments are produced "on the fly" by the monolingual grammars, however these grammars are conceived in theory. Indeed, from the analytical point of view, no reference need be made to particular linguistic theories for this aspect.

The study of code-switch sites is basically an attempt to understand how the speakers in a bilingual community contrive to reconcile the two different ways in which hierarchical syntactic structure is linearized by languages *A* and *B*. The simplest class of languages for modeling the recursive hierarchical sentence structures of natural language is the class of context-free grammars. Of course, these do not constitute a theory of natural language, only a model of one aspect of language, that of linearized hierarchy, which is exactly the aspect most important to the investigation of linearly and hierarchically constrained switch points.

Context-free grammars are not suited to the investigation of discontinuous sentence elements, intrasentential co-reference, and other remote syntactic phenomena, but no systematic investigation has ever shown the pertinence of these to switch sites. On the other hand, context-sensitive phenomena such as subcategorization (Bentahila & Davies, 1983), cliticization or certain deletion processes also escape the scope of context-free modeling, as do other null elements, agreement rules and other features which may, in particular communities, be important to understanding switch sites (Muysken, 1995). Both remote and local relationships are, however, incorporated into extensions of context-free grammars, notably Head-driven Phrase Structure Grammar (Pollard & Sag, 1994). The strong surface-orientation of these grammars and their context-free architecture would facilitate their integration into incremental production (processing) models such as ours.

This reasoning applies equally well for our assumptions of congruence. Lack of one-to-one lexical translatability, non-congruent grammatical categories, differing constituent structures and variable word order may all be pertinent to where code-switches occur. Some of these, such as lexical differences, are separate from word-order considerations and would impinge only sporadically on the susceptibility of syntactic boundaries to switching. The others, such as the bilingual confrontation of a "binary" and a "flat" style of constituent structure or the weakening of the fixed word order assumption, are pertinent to switch boundaries but seem amenable in principle to an extension of the present model. In any case, they would be unlikely to alter our

results except to increase somewhat the inventory of boundaries at which switching can occur.

Our assumptions of congruence may be too strong, but they are not inconsistent with the explanatory goals of linguistic theory. Sag and Wasow (1997, p. 184) argue, for Head-driven Phrase Structure Grammar, that the rules "... are sufficiently general that, aside from the linear ordering of the constituents, they are natural candidates for universality. It would not be hard to factor out the ordering, so that versions of these rules could be posited as part of universal grammar." To the extent that this could be achieved, our code-switching model could be applied directly to the different "factored-out" orders and otherwise common rules.

The constituent labeling criteria are of two types. Criteria (i) and (ii), lexical labeling and labeling of monolingual constituents, seem unavoidable for a meaningful notion of the language of a constituent. Theorem 4, about the compulsory transmission of a label to at least one subconstituent (not necessarily immediate), could be strengthened by imposing a more specific requirement depending on the theoretical orientation of the analyst or, perhaps, the nature of the two languages involved. Neither (i) nor (ii) nor the theorem pertain to word order or constituent order.

Criterion (iii), however – labeling of subconstituents in language-specific slots – is basic to our considerations of constituent order. Though it is a strong constraint, it seems empirically unimpeachable, especially in the binary case where it directly implies the equivalence constraint, or to avoid pathological constituent-initial examples as discussed in the fifth section. This criterion could be weakened, to take into account variable word order, for example, or to permit a restricted range of constituent insertion.

Some linguistic theories overgenerate the set of possible sentences that can have a given structure and then filter out all those that do not satisfy certain constraints. There is little motivation for classical mathematical modeling because the generation of structures is seen as but a small part of syntax, while the feature-matching or other verification constitutes the essence of the theory itself, and loses its power in any attempt to simplify or model it.

The *no forced switches* hypothesis in the seventh section is thus of particular interest, not so much for its impact, which is limited to non-binary constituents, but because it represents a plausible type of production-level consideration new to the modeling of intrasentential code-switching. Both criterion (iii) and the *no forced switches* hypothesis seem necessary to prove the equivalence constraint, but it would be worthwhile to try to strengthen the latter while weakening the former, to increase the transparency of the production model and to minimize the filter-like test of well-formedness inherent in the labeling criteria. In any case, the labeling criteria are very simple, and it is psycholinguistically plausible that they too could be monitored by the speaker in the ongoing production of a code-switched sentence.

Discussion

It is false to assume, as do many analysts, that explanation of bilingual syntax must derive from general principles inferred from the study of monolingual grammar and hypothesized to have evolved with the human (monolingual) language faculty in primitive, and presumably monolingual, communities. Why these principles should apply directly and strictly to bilingual discourse is a question that has been overlooked, but of key importance, since bilingual communities show an extreme diversity of patterns of code-mixing not predictable through language typology considerations.

In our model, then, we have tried to separate out considerations of monolingual well-formedness from the "real-time" production of a code-mixed sentence and to find the minimal locus of interaction of the two monolingual grammars.

We have identified the aspects of structure and production which are responsible for equivalence point switching in contrast to insertional modes of switching, in line with the empirical observation that the latter is restricted to a small number of bilingual communities and then to only one or two types of constituent (Naït M'Barek & Sankoff, 1988; Sankoff et al., 1990). Unfortunately, much of the literature treats, implicitly or explicitly, the process of lexical borrowing as if it were insertional code-switching. This is descriptively inadequate, however, since insertional code-switches, where they exist, are well-formed constituents in the language of their lexical items, while borrowings are integrated morphologically and syntactically into the recipient language. The effect of this conceptual and methodological error is to exaggerate greatly the scope of insertional switching and to blur the real distinction between the very different processes of switching and borrowing.

Our model is built on an earlier formulation of the equivalence constraint (Sankoff& Mainville, 1986). It is the production aspect here, however, that allows us to achieve the all-important *well-formedness of monolingual fragments*, not strictly guaranteed in the earlier work, and to model the essential *unpredictability* of code-switching. The present version has a more economical protocol for constituent labeling and a more complete account of the coincidence (or

lack thereof) between word-level switching and constituent-level switching. Our strong results in Theorems 5 and 6 are of course due in part to our use of a mathematically rigorous and simple model of surface structure. Nonetheless we would claim that it captures the major phenomena of real code-switched discourse, and that those effects it does not explain have only to do with the relatively peripheral phenomena not in the scope of the model.

References

Backus, A. (1996). *Two in One. Bilingual Speech of Turkish Immigrants in the Netherlands*. Tilburg: Tilburg University Press.

Belazi, H. M., Rubin, E. J., & Toribio, A. J. (1994). Code switching and X-bar theory: The functional head constraint. *Linguistic Inquiry*, 25, 221–37.

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

Di Sciullo, A.-M., Muysken, P., & Singh, R. (1986). Government and code-mixing. *Journal of Linguistics*, 22, 1–24.

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

Gumperz, J., & Hernandez, E. (1969). *Cognitive Aspects of Bilingual Communication*. Working Paper Number 28, Language Behavior Research Laboratory. University of California, Berkeley.

Joshi, A. K. (1985). Processing of sentences with intrasentential code-switching. In D. R. Dowty, L. Karttunen & A. M. Zwicky (eds.), *Natural Language Parsing*, pp. 190–205. Cambridge: Cambridge University Press.

Labov, W. (1969). Contraction, deletion and inherent variability of the English copula. *Language*, 45, 715–762.

Lipski, J. (1977). Code-switching and the problem of bilingual competence. In M. Paradis (ed.), *Aspects of Bilingualism*, pp. 250–63. Columbia, SC: Hornbeam Press.

Mahootian, S., & Santorini, B. (1994). Adnominal adjectives, codeswitching and lexicalized TAG. In A. Abeille, S. Aslanides & O. Rambow (eds.), *3e colloque international sur les grammaires d'arbres adjoints* (TAG+3), Technical Report TALANA-RT-94–01, pp. 73–76.

Meechan, M., & Poplack, S. (1995). Orphan categories in bilingual discourse: Adjectivization strategies in Wolof-French and Fongbe-French. *Language Variation and Change*, 7, 169–94.

Muysken, P. (1995). Code-switching and grammatical theory. In L. Milroy & P. Muysken (eds.), *One Speaker, Two Languages*, pp. 177–98. Cambridge: Cambridge University Press.

Myers-Scotton, C. (1993). *Dueling Languages*. Oxford: Clarendon Press.

Naït M'Barek, M., & Sankoff, D. (1988). Le discours mixte arabe/français: des emprunts ou des alternances de langue? *Revue Canadienne de Linguistique*, 33 (2), 143–54.

Nishimura, M. (1986). Intrasentential code-switching. The case of language assignment. In J. Vaid (ed.), *Language Processing in Bilinguals: Psycholinguistic and Neuropsychological Perspectives*, pp. 123–43. Hillsdale, NJ: Lawrence Erlbaum.

Pandit, I. (1990). Grammaticality in code-switching. In R. Jacobson (ed.), *Codeswitching as a Worldwide Phenomenon*, pp. 33–69. New York: Peter Lang.

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

Pollard, C. A., & Sag, I. A. (1994). *Head-driven Phrase Structure Grammar*. University of Chicago Press.

Poplack, S. (1978). Syntactic structure and social function of code-switching. In R. Duran (ed.), *Latino Discourse and Communicative Behavior*, pp. 169–84. New Jersey: Ablex.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching. *Linguistics*, 18, 581–618.

Poplack, S., & Meechan, M. (1995). Patterns of language mixture: Nominal structure in Wolof-French and Fongbe-French bilingual discourse. In L. Milroy & P. Muysken (eds.), *One Speaker, Two Languages*, pp. 199–232. Cambridge: Cambridge University Press.

Poplack, S., Sankoff, D., & Miller, C. (1987). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, 26, 47–104.

Poplack, S., Wheeler, S., & Westwood, A. (1987). Distinguishing language contact phenomena: Evidence from Finnish-English bilingualism. In P. Lilius & M. Saari (eds.), *The Nordic Languages and Modern Linguistics 6*, pp. 33–56. Helsinki: University of Helsinki Press.

Rivas, A. (1981). On the application of transformations to bilingual sentences. MS. Amherst, MA: Dept. of Spanish and Portuguese, University of Massachusetts.

Sag, I. A., & Wasow, T. (1997). *Syntactic Theory: A Formal Introduction* (on-line draft). CSLI Publications, Stanford University.

Sankoff, D. (1988). Sociolinguistics and syntactic variation. In F. Newmeyer (ed.), *Linguistics: the Cambridge Survey. IV Language: the Socio-cultural Context*, pp. 140–61. Cambridge: Cambridge University Press.

Sankoff, D., & Mainville, M. (1986). Code-switching of context-free grammars. *Theoretical Linguistics*, 13, 75–90.

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

Sridhar, S. N., & Sridhar, K. (1980). The syntax and psycholinguistics of bilingual code-mixing. *Canadian Journal of Psychology*, 24, 407–16.

Woolford, E. (1983). Bilingual code-switching and syntactic theory. *Linguistic Inquiry*, 14, 520–36.