Unconstraining Codeswitching Theories*

Jeff MacSwan

Codeswitching (CS) is the alternate use of two or more languages among bilingual interlocutors. Grammatical studies of CS focus on cases of intersentential CS, that is, language mixing below sentential boundaries, as illustrated in (1).

(1) Mi hermano bought some ice cream.

My brother bought some ice cream

The central problem for scholars interested in the linguistic properties of CS is the explanation of the contrast in grammaticality between cases such as (1) and (2).

(2) *El bought some ice cream.

Why would a bilingual's grammar permit switching between a lexical DP and verb but not between a pronoun and verb? What principles of grammar might account for the contrast, and what might these analyses tell us about the nature of bilingualism itself? The answers must account for CS between the particular languages in the data on hand as well as for switching between any other pair of languages.

This paper evaluates two approaches to CS, a constraint-based research program and a constraint-free research program, arguing in favor of the latter. The Minimalist Approach to CS is outlined with some illustrative analyses as a way of approaching a constraint-free program. We conclude with some comments on the paper by Schindler, Legendre and Mbaye (2009, this volume), which offers a critique of the PF Interface Condition.

MacSwan, J. (2010). Plenary address: Unconstraining codeswitching theories. *Proceedings from the Annual Meeting of the Chicago Linguistic Society 44*. Chicago: University of Chicago Press.

Two Research Programs for Codeswitching

In 1962 at the Thirteenth Annual Round Table Meeting on Linguistics and Languages at Georgetown University, Einar Haugen staked claim to the original use of the term codeswitching, although the word had first appeared in print in H. Vogt's (1954) review of Uriel Weinreich's (1953) Languages in Contact and two years before that in Haugen's Bilingualism in the Americas, a bibliography published by the American Dialect Association in 1956. Although much earlier work by Aurelio M. Espinosa (1911) had noted the phenomenon, an actual CS research literature did not emerge until the late 1960s and early 1970s, when work focusing on both social and grammatical aspects of language mixing began steadily appearing with scholarly engagement of previously published research. (See Benson, 2001, for further discussion of the early history of the field.)

Among the earliest to observe that there are grammatical restrictions on language mixing were Gumperz and his colleagues (Gumperz, 1967, 1970; Gumperz & Hernández-Chávez, 1970), Hasselmo (1972), Timm (1975), and Wentz (1977). For instance, Timm's list of restrictions noted that Spanish-English switching between a subject pronoun and a main verb is ill-formed but not so when the subject pronoun is replaced with a lexical subject, as the contrast between (1) and (2) shows.

While construction-specific constraints were typical of this early work, a literature soon emerged in which the grammatical mechanisms underlying these descriptive facts were explored. It is well known and uncontroversial that CS is constrained in the descriptive sense, meaning, simply, that CS behavior is itself rule-governed. Consider, for example, the contrast in (2) (Belazi, Rubin & Toribio, 1994).

(2a) The students *habían visto la película italiana*
The students had seen the Italian movie

(2b) *The student had *visto la película italiana*
The student had seen the Italian movie

Although the basic word order requirements are the same here for both English and Spanish, (2b) is judged to be ill-formed. Regardless of what account we might construct for the contrast, (2) is sufficient to show that CS behavior, like other linguistic behavior, is constrained or rule-governed.

However, research on CS conducted in the Aspects area (Chomsky, 1965) of Generative Grammar soon turned to the notion that CS, a broad, horizontal linguistic phenomenon, could be explained by positing the kind of theoretical constraints developed in the contemporaneous syntactic literature to impose vertical limits on transformations and phrase structure. As early as 1955, Chomsky had noted that the transformational component in a hybrid generative-transformational system had the disadvantage of vastly increasing the expressive power of the grammar, permitting the formulation of grammatical processes which did not seem to occur in any language. In response to the problem, Chomsky (1964, 1965) and other researchers such as John Ross (1967) posited constraints on transformational rules. Ross noticed, for instance, that an NP could not be extracted out of a conjoined phrase, as in (3a), accounting for the ill-formedness in (3b) but not in the semantically equivalent (but syntactically divergent) example in (3c).

(3a) John was having milk and cookies.

(3b) *What$_i$ was John having milk and t_i?

(3c) What$_i$ was John having milk with t_i?

Here, constraints were viewed as psychologically real restrictions on the application of transformations to phrase markers, and were therefore understood to be part of the grammar itself. In this context constraint refers not just to a description of intuitions about a set of constructions, but to an actual mechanism of grammar in the mind/brain of a language user – that is, the theoretical meaning of the term. We therefore arrive at two very distinct meanings of constraint – the descriptive sense, which simply tells us a phenomenon is rule-governed, and a theoretical sense in which actual grammatical mechanisms, called constraints, are posited as part of the actual grammar itself. That CS is constrained in the descriptive sense is not controversial. However, whether there should be CS-specific constraints in the theoretical sense has been at the heart of recent debates.

The Constraint-based Research Program

The idea of a constraint in the syntactic/grammatical sense appealed to a number of researchers in CS, and was used to articulate the grammatical restrictions observed in CS data. However, there are good reasons for avoiding such mechanisms altogether. For concreteness, consider Joshi's (1985) Constraint on Closed-Class Items.

(4) Constraint on Closed-Class Items

Closed-class items (e.g., determiners, quantifiers, prepositions, possessives, Aux, Tense, helping verbs) cannot be switched.

Joshi's constraint, like many similar mechanisms, makes explicit reference to (code)switching; codeswitching, of course, denotes a change from one language to another, say, from Hindi to English, or Spanish to Nahuatl. However, all such entities – languages – are each a class of expressions defined by the grammar. So a grammar G

defines a class of expressions L. We cannot insert L as part of any function of G, as L is itself defined by G. Hence, explicit constraints on CS are not theoretically well defined because they reference language switching, and grammars are formally blind to the languages they generate.

Furthermore, constraints so formulated may serve to provide good linguistic description (to the extent it is empirically correct), but they do not serve to explain or enlighten. Constraints on CS, in the theoretical sense, restate the descriptive facts by telling us which grammatical constructions or properties are evident in CS or not. While linguistic description is an important first step, it does not constitute a linguistic theory. Hence, the more serious problem is with CS-specific mechanisms is that they threaten to trivialize the enterprise: Rather than explaining descriptive restrictions observed in CS data, CS-specific mechanisms simply note these restrictions within the grammar itself so that no explanation is needed, and one is left still wondering what general principles of grammar might underlie the observations and descriptions. We might define a CS-specific constraint, then, as a proposed grammatical mechanism which makes explicit reference to (code)switching or language(s), and which is understood to be part of the actual linguistic competence of a bilingual.

As a further illustration of a current approach to CS within the family of constraint-based systems, let us example a class of proposals built around Levelt's (1986) Speaking model. Although similar proposals surfaced independently (de Bot, 1992; Azuma, 1991, 1993), Myers-Scotton's (1993) Matrix Language Frame (MLF) Model has had a very strong influence on the field of CS. The MLF Model differentiates the languages involved in CS, as other models (e.g., Joshi, 1985) have also done; one

language is known as the matrix language (ML), the other as the embedded language (EL). According to this approach, the matrix language defines the surface structure positions for content words and functional elements.

The MLF Model includes two basic components – the Morpheme Order Principle, which requires that morphemes within a bilingual constituent follow the order prescribed by the ML, and the System Morpheme Principle, which states that all "system morphemes" – defined as morphemes which have grammatical relations with other constituents outside their maximal projections – come from the ML in any CS utterance. From a theoretical point of view, we see that we immediately encounter the same difficulties as found in constraint-based approaches generally: The grammatical principles responsible for defining the distribution of CS explicitly refer to the separate languages involves in CS without specifying a formal definition of the languages in interaction. Moreover, the general mechanisms are defined in very vague terms, making it difficult to identify the specific empirical predictions of the MLF Model.

Concern over the vagueness of the definition of the ML, in particular, has been voiced in Muysken and de Rooij (1995), Bentahila (1995), MacSwan (1999, 2000, 2005), and Muysken (2000). Jake, Myers-Scotton and Gross (2002) resolved some of the ambiguities associated with Myers-Scotton's (1993) original definition of the ML as the language contributing the majority of the morphemes in an utterance, which "may change across time, and even within a conversation" (p. 69). In contrast, the Uniform Structure Principle of Jake, Myers-Scotton and Gross offered a more structurally-oriented definition of the ML: "The ML may change within successive CPs, even within a multi-clausal sentence, but we stress that the ML does not change within a single bilingual CP"

(p. 73). In other words, within single CP, the MLF Model predicts that all grammatical morphemes will be from one language only (System Morpheme Principle), and that the language contributing the grammatical morphemes will also define the surface order of the utterance (Morpheme Order Principle). Jake, Myers-Scotton and Gross (2002) furthermore stress that the ML/EL distinction is universal, existing "in monolingual language as well as bilingual language" (p. 88), though the relevance associated mechanisms can only be detected in bilingual data.

Myers-Scotton and colleagues further stipulate that an "EL island" may occur below the CP: "[A]s well-formed maximal constituents in the EL, [EL islands] are not inflected with ML system morphemes, although they occur in positions projected by the ML, following the Morpheme Order Principle" (Jake, Myers-Scotton & Gross, 2002, p. 77). That is, EL islands are essentially lawful violations of the System Morpheme Principle because they contain grammatical morphemes that are not in the ML, but an EL island must be a maximal projection and must remain true to the Morpheme Order Principle (that is, its position within the utterance must be dictated by the ML). Let us briefly some empirical evidence bearing on these predictions, focusing on the System Morpheme Principle. The French-Italian data in (5) are reported in Di Sciullo, Muysken and Singh (1986).

(5a) No, *parce que* hanno *donné des cours*
no, because have given of the lectures
'No, because they have given the lectures'

(5b) Oui, alors j'ai dit quie *si potev* aller comme ça
yes so I have said that REF could walk like that
'Yes, so I said that we could go like that.'

Note that in both cases we observe a switch between an auxiliary or modal and its complement. Because these forms have grammatical relations with other lexical heads within the structure, they meet the MLF Model's definition of a system morpheme. Yet, contrary to the requirements of the System Morpheme Principle, each utterance involves system morphemes from different languages below the CP.

To rescue the MLF Model, one might argue that [*donné des cours*] in (5a) is an EL island, projected as a VP complement of the auxiliary verb, and that (5b) similarly involves an EL island [*aller comme ça*], an IP complement of the modal. However, note that the examples in (5) contrast with Spanish-English data in (2), where a switch between an auxiliary and a participle is ill-formed. The construction in (2b) is eligible for the same structural analysis as (5a), in which an EL island is hypothesized, yet it is ill-formed, contrary to the predictions of the MLF Model.

As an additional example, consider the Spanish-Nahuatl examples in (6). Notice that Spanish negation (*no*) does not tolerate a Nahuatl complement, while Nahuatl negation (*amo*) permits a Spanish complement. Both the agreement morphology on the verbs and negation count as system morphemes since they enter into grammatical relations with other morphemes (in the less obvious case of negation, it c-commands a negative polarity item and may form a syntactic clitic with its verb). Hence, according to the System Morpheme Hypothesis, both (6a) and (6b) should be ill-formed because system morphemes are mixed below the CP, yet this is not so. Remarkably, the constructions contrast in acceptability, even though they appear to have identical underlying structures.

(6a) *No *nitekititoc*
no ni-tekiti-toc
not 1S-work-DUR
'I'm not working'

(6b) Amo *estoy trabajando*
amo estoy trabaja-ndo
not be/3Ss work-DUR
'I'm not working'

Myers-Scotton and colleagues might argue that NegP is an EL Island in (6a) but not in (6b), but with no independent evidence of the status of island entities these claims appear to be mere rationalizations. Myers-Scotton (1993) and Jake, Myers-Scotton and Gross (2002) furthermore allow "internal EL islands," defined as "a constituent in the EL made up of EL morphemes following EL morpheme order, but smaller than a maximal projection" (Jake, Myers-Scotton & Gross, 2002, p. 76). In other words, not only can maximal projections be "islands," but structural units smaller than EL islands can too – sanctioning essentially any and all CS examples. One must recall that the constraint mechanism proposed as part of a model of CS cannot be selectively applied; that is, we cannot reasonably claim that negation is an island, immune from the System Morpheme Principle in (6b) but not in (6a) where the result is ill-formed. Rather, once created, these mechanisms must operate in all cases, and as such they create a universe of expectations where essentially all CS is well-formed.

Of course, all CS is not well-formed, as is abundantly clear. But there is a historical tendency for CS researchers to rely primarily on naturalistic data alone – evidence of what does occur in CS – making it impossible to discover a model that over-identifies well-formed constructions. In the absence of negative evidence, generally available through grammaticality judgment tasks, one will not be in a position to

construct a generative theory of CS with the capability of generating all and only the well-formed cases. (For further discussion in the specific context of the MLF Model, see MacSwan, 2005.)

The Constraint-free Research Program

The history of CS reveals a common intuition among researchers that theories about CS should be free of grammatical mechanisms and constraints specific to it. Poplack and Sankoff (1981) noted complications associated with the free (unconstrained) union of two phrase structure grammars, then reluctantly introduced a CS-specific tagging mechanism in response to what they perceived to be the empirical demands of the data. Others did not address the issue and formal problems directly, while some proposed CS-specific constraints but argued that the mechanisms applied to monolinguals as well as bilinguals.

However, a true constraint-free approach to CS will permit no such mechanism, and will not tolerate any grammatical device which makes explicit reference to (code)switching or language(s). The guiding principle for such a program of research might be stated as in (7) (MacSwan, 1999).

(7) Nothing constrains CS apart from the requirements of the mixed grammars.

In the next section, we offer a Minimalist implementation of (7) along with a set of specific analyses intended to serve as proof of concept for it. We will argue that the Minimalist Program (MP) has important advantages for the analysis of CS data, but we note that (7) could in principle be implemented in any number of alternative syntactic frameworks (GPSG, OT, or other).

Implementing a Constraint-free Approach

In important respects, the theoretical contexts in which many influential theories had been formulated within the constraint-based research program did not provide the tools needed to permit the implementation of a true constraint-free research program. An approach to syntax which built structure from the top down, as in the Aspects and later GB models, postponed lexical insertion until well after the word order had been laid out. The approach posed a significant challenge to CS researchers: The structure could not be sensitive to which language contributed a specific lexical item until the end, when lexical insertion occurred, but the language contributing the lexical item appeared to have significant consequences for the syntactic structure at the onset.

The MP provides an opportunity for implementing a rich, constraint-free approach. Within the MP, structures are built from a stock of lexical items, with lexical insertion (formalized as Select) taking place at the outset. This important development permits CS researchers to probe the structural consequences of particular lexical items from specific languages, with no need to keep track of which languages may contribute which specific lexical elements during a final stage of lexical insertion.[1]

The Minimalist Program

According to Chametzky (2003), the "lexical entry driven" approach to syntax was part of the general effort underlying X' reduction, with significant contributions from Stowell (1981) and Speas (1990), among others. X' Theory, which had been introduced in Chomsky (1970), effectively eliminated phrase structure grammar in favor of the view that structures are projected from lexical items; however, remnants remained, with reference to lexical insertion rules reasonably common among GB era syntacticians

(Chomsky, 1981; Stowell, 1981; Lasnik & Uriagereka, 1988). With a return to its derivational roots, Minimalist syntax reduced generation to the simplest possible form – free Merge (Chomsky, 1991, 1994), building structures from the ground (the lexical string) up (the hierarchical phrase structure) based on the specification of lexically-encoded features. Independently, Borer (1984) had suggested an account of language variation in which parameters were also associated with the lexicon, rather than with the system of syntactic rules. Hence, the system of rules could be seen as invariant, with all variation associated with the lexicon, the traditional repository of arbitrariness.

In the MP there are two components of grammar: C_{HL}, a computational system for human language, believed to be invariant across languages; and a lexicon, to which the idiosyncratic differences observed across languages are attributed. An operation called Select picks lexical items from the lexicon and introduces them into a Numeration or Lexical Array (LA), a finite subset of the lexicon used to construct a derivation. Merge takes items from the LA and forms new, hierarchically arranged syntactic objects. Movement operations (Internal Merge) apply to syntactic objects formed by Merge to re-arrange elements within a tree (Chomsky, 1995, 2000). Phrase structure trees are thus built derivationally by the application of the operations Select and Merge, constrained by the condition that lexically encoded features match in the course of a derivation.

Movements are driven by feature valuation, and may be of two types. A head may undergo head movement and adjoin to another head, or a maximal projection may move to the specifier position of a head. In either case, the element moves for the purpose of valuing morphological features of case and ϕ (number, person, and gender). In addition, its movement may be overt or covert. Overt movements are driven by strong

features and are visible at PF (phonetic form, where they are pronounced) and LF (logical form, where they are interpreted). Covert movements, driven by weak features, are visible only at LF.

Principles of Economy select among convergent derivations. One such principle, Full Interpretation (FI), requires that no symbol lacking a sensorimotor interpretation be admitted at PF; applied at LF, FI entails that "every element of the representation have a (language-independent) interpretation" (Chomsky, 1995, p. 27). Thus, uninterpretable features (denoted -Interpretable) must be checked and deleted by LF. The +Interpretable features are categorial features plus ϕ-features of nominals (Chomsky, 1995, p. 278). [+Interpretable] features do not require valuation (checking). A derivation is said to converge at an interface level (PF or LF) if it satisfies FI at that level; it converges if FI is satisfied at both levels. A derivation that does not converge is also referred to as one that crashes. If features are not valued, the derivation crashes; if they mismatch, the derivation is canceled (that is, a different convergent derivation may not be constructed).

At some point in the derivation, an operation Spell-Out applies to strip away from the derivation those elements relevant only to PF; what remains is mapped to LF by a subsystem of C_{HL} called the covert component. The elements relevant only to PF are mapped to PF by operations unlike the covert component, operations which comprise the phonological component. The phonological component is also regarded as a subsystem of C_{HL}. The subsystem of C_{HL} which maps the lexicon to Spell-Out is the overt component. Note that the various components (overt, covert, phonological) are all part of C_{HL}, the computational system for human language.

The leading aim of the MP is the elimination of all mechanisms that are not necessary and essential on conceptual grounds alone; thus, only the minimal theoretical assumptions may be made to account for linguistic data, privileging more simplistic and elegant accounts over complex and cumbersome ones. These assumptions would naturally favor accounts of CS which make use of independently motivated principles of grammar over those which posit rules, principles or other constructs specific to it. Hence, implementing (7), which posits that all of the facts of CS may be explained just in terms of principles and requirements of the specific grammars in interaction, the formal claim is that, for G_x a grammar of L*x* and G*y* a grammar of L*y*, CS falls out of the union of the two grammars ($\{Gx \cup Gy\}$) and nothing more (MacSwan, 1999). Thus, while free union of phrase structure grammars led to the complications noted by Sankoff and Poplack (1981) for reasons associated with late lexical insertion, free union of lexically-encoded grammars does not.

MacSwan (2000, 2005) argues that the properties of morphophonology force bilinguals to separately encapsulate distinct lexicons. Phonology is sensitive to inflectional content, and the rules of word formation are presumed to be internal to the lexicon in Minimalist syntax, with words entering the derivation with their morphological content fully specified. We therefore face two alternatives: Either (a) there is a single lexicon, and each lexical item is marked for a specific set of phonological and morphological rules which yield the appearance of one language or another; or (b) the lexical items in a bilingual's repertoire are mentally compartmentalized in some way, with a specific set of phonological and morphological operations associated with each subset. The second alternative appears more economical, since the morphophonology is

associated with sets of elements rather than with individual members. We therefore will assume the latter to be correct, and refer to each subset of lexical items within a bilingual's linguistic repertoire as a lexicon. The model sketched here is represented graphically in Figure 1 (see MacSwan, 2000, 2005). Let us consider some sample analyses pursued within the Minimalist approach to CS under the direction of (7), our constraint-free research program.

[Insert Figure 1 about here]

Asymmetrical Switches in DPs

Moro (to appear) developed an account of asymmetrical switches in Spanish-English DPs in which a Spanish determiner may precede an English N (e.g., *los teachers*, 'the teachers') but an English determiner may not precede a Spanish N (e.g., **the casa*, 'the house'). Moro notes that the composition of the set of φ-features differs for English and Spanish. In Spanish, *gender* and *number* are morphologically marked on determiners and nouns, but in English only *person* and *number* are marked while *gender* is absent. In monolingual contexts, the feature matrices of Ds and Ns are identical ({person, number} for English, {person, number, gender} for Spanish). However, with respect to the bilingual constructions, the DPs are well-formed only if the φ-set of N is included in the φ-set of D.

Moro takes advantage of Chomsky's (2000, 2001) proposal that uninterpretable features (such as the φ-features of D, but not of N) enter the derivation without values

specified. The operation Agree values and deletes these features from the narrow syntax "in one fell swoop" (2000, p. 124), all or none. Thus, in Moro's analysis, expressions like **the casa* crash because the N's φ-features *person, number,* and *gender* attempt to value and delete the D's features *person*, *number* and *gender* in "one fell swoop"; however, the *gender* feature cannot succeed, as there is no corresponding feature in the English determiner. As a result, the derivation fails to converge, as the uninterpretable features of D are not deleted. In the well-formed case, on the other hand, the φ-features of N (person, number), being included in D's φ-set, successfully value and delete D's uninterpretable features "in one fell swoop," as in the monolingual cases.

Basic Subject/Verb Word Order

A striking result of the MP is the account of differences in basic word order in terms of movement requirements associated with feature strength. MacSwan (2004) proposed an analysis of word order facts obtained in CS between languages differing with regard to whether subjects occurred pre-verbally (SV languages) or post-verbally (VS languages). The universal base structure is assumed to be underlyingly SVO with a VP-internal subject for both SV and VS languages. V^0 raises to T^0 (=Agr^0) to value ϕ-features. If the subject overtly checks its EPP feature in the specifier position of T^0, then an SVO order results. If it checks its EPP feature covertly, however, then the resulting word order will be VSO. Thus, the typological distinction between SVO and VSO languages is captured in terms of the strength of the EPP feature in T^0.

Specific properties of the phonological system are responsible for a ban on mixed-language heads in CS contexts; the relevant condition, derived from basic properties of the linguistic system, is known as the PF Interface Condition (PFIC) (MacSwan, 2009;

MacSwan & Colina, 2007, to appear). Therefore, if V^0 raises to T^0 to check ϕ-features, both elements must be of the same language, otherwise the complex so formed would be ill-formed on the PFIC. As a result, the language of the verb will determine the language of T^0, hence the value of its EPP feature. In this way the system developed here predicts that the language of the verb will determine the position of the subject – if the verb is from an SVO language, the subject should occur preverbally, whether it is from an SVO language or not; if the verb is from a VSO language, the subject should occur postverbally, regardless of the requirements of the language of the subject. These facts are attested in a wide range of corpora, as shown in (8) and discussed in MacSwan (2004).

(8a) VS verb (Irish), SV subject (English) (Stenson, 1990)
Beidh *jet lag* an tógáil a pháirt ann
be-FUT taking its part in-it
'Jet lag will be playing its part in it'

(8b) VS verb (Irish), SV subject (English) (Stenson, 1990)
Fuair sé *thousand pounds*
get-PA he
'He got a thousand pounds'

(8c) VS verb (Breton), SV subject (French) (Pensel, 1979)
Oa ket *des armes*
be-3S IMP NEG of-the arms
'There were no arms'

(8d) VS verb (Breton), SV subject (French) (Troadec, 1983)
Setu oa *l'état-major* du-se barzh ti Lanserot
There be-imp the military-staff down-there in house Lanserot
'There was the military staff down there in Lanserot's house'

(8e) VS verb (SLQ Zapotec), SV subject (Spanish) (MacSwan, 2004)
S-to'oh *mi esposa* el coche
DEF-sell my wife the car
'My wife will definitely sell the car'

Pronominal and Lexical Subjects

As a final example, consider the well known ban on CS in pronominal subjects illustrated in (2) in contrast to (1), repeated here.

(1) Mi hermano bought some ice cream.
My brother bought some ice cream

(2) *El bought some ice cream.
He bought some ice cream

Van Gelderen and MacSwan (2008), following a line of work proposing that pronouns are heads rather than phrases (Abney, 1987; Longobardi, 1994; Cardinaletti, 1994; Cardinaletti and Starke, 1996; Carnie, 2000), argued that subject extraction from a VP-internal shell proceeds according to a principle of economy proposed in Chomsky (1995): "F carries along just enough material for convergence," where F is the target of movement. As such, van Gelderen and MacSwan (2008) argue that Ds may check features with T via head movement, while lexical DPs must move as phrases, pied-piping their phonetically filled complements. Since D-to-T movement would result in a mixed-language complex head, ruled out by the PFIC, the pronominal subject case (as in (2)) crashes at PF. However, in (1), the lexical DP moves as a phrase to the specifier of T with no ill-formedness resulting. Corroborating evidence involving coordination of pronouns with lexical DPs in CS contexts further support the analysis.

We have offered some specific examples of analyses of CS conducted within a constraint-free research program to support the claim that such analyses are both possible and fruitful. The analyses give rise to new families of questions about other language pairs with related and distinct properties, and about general properties of linguistic theory. Other illustrations may be observed in other work referenced here.

Evaluating Theories and Programs

It is important to note that a research program is not itself a theory, but rather constitutes an analytic framework which in turn spawns theories and proposals, as Boeckx (2006) has discussed in the context of a more general discussion of linguistic research programs. Lakatos (1970), a philosopher of mathematics and science best known for his defining work on scientific research programs, noted that the "typical descriptive unit" of scientific achievement was not a particular theory or proposal, but indeed a research program. As Boeckx (2006) elaborates, a research program is made up of a core with a "logico-empirical character."

> … whatever the constitutive elements of a program's core may be, that core is rightly characterized by Lakatos as 'irrefutable,' and 'stubbornly defended.' Wittgenstein would have called it 'unassailable and definitive.' The core is tenaciously protected from refutation by a vast protective belt of auxiliary hypotheses (Boeckx, 2006, p. 89).

Research programs, according to Lakatos, are not to be distinguished as right or wrong but rather as progressive or degenerating – that is, as fertile or sterile. A progressive research program is one which gives rise to new families of questions and leads to the discovery of new facts. A program is degenerating or sterile, on the other hand, if the theories it spawns are conceived only to account for known facts. Boeckx further notes,

> As Lakatos points out, one must treat budding programs leniently: programs may take decades before they take off and become empirically progressive. As Lakatos observes, 'criticism is not a Popperian quick kill, by refutation. Important criticism is always constructive: there is no refutation without a better theory' (p. 91).

Note that (7), the agenda for a genuine constraint-free approach to CS, is similarly not a particular theory about CS but a program for CS research. It gives rise to particular theories or proposals, each formulated with a commitment to tolerate no CS-specific

device. Because (7) is not itself a theory, it is not subject to falsification; rather, it should be abandoned as an agenda for CS research if it proves sterile or unsuccessful.

We might reasonably argue that CS research conducted in the constraint-free program constitutes a dramatic departure from theories articulated within the constraint-based program. A review of the research literature shows that CS constraints are proposed fundamentally to account for known facts, with each successive proposal offering criticisms of previous with no significant or new insights provided. The effort to identify over-arching constraints on CS has been exhausted, and the program now appears sterile (or regressive, in Lakatos' terms).

Furthermore, it is not meaningful or informative to make empirical comparisons between theories formulated within a constraint-free approach and those formulated within a constraint-based approach, as has been done. If relevant counter-evidence is presented regarding a proposed constraint on CS, constraint-oriented researchers either attempt to disprove the counter-evidence or augment the system of constraints in some way. One sees this notably among proponents of the MLF Model, whose system of constraints has admitted increasingly numerous add-on constraints, principles, and hypotheses which are successively intended to inoculate the model from falsification.

Conversely, if a particular theory proposed within the constraint-free program is shown to be false, it does not lead researchers within the program to turn to CS-specific constraints but rather to adjust the analysis to newly discovered data. Suppose, for instance, that evidence from additional language pairs suggested that Moro's (to appear) theory of DP-internal checking, or van Gelderen and MacSwan's (2008) theory about subject licensing in CS context, was incorrect. Researchers within the constraint-free

program would be expected to discard the previously analyses in favor of improved alternatives covering a greater range of empirical ground, as linguists do generally. But such data would not be interpreted as evidence that the very "logico-empirical character" of the program should be rejected. That conclusion would follow only if continued research appeared fruitless in light of insurmountable barriers, and the project grew sterile – a result that is not likely to obtain soon, given the enormous promise of ongoing research.

The excellent session on CS convened at the Chicago Linguistics Society 44 included several papers focused on CS, including one paper besides the current one which focused on the grammatical character of CS. Because of the relevance of that paper to this one, we turn to a few specific comments on Schindler, Legendre and Mbaye's contribution to the present volume volume.

Comments on Schindler, Legendre and Mbaye (2009 [this volume])

The paper by Schindler, Legendre and Mbaye (2009 [this volume]) documents the substance of their presentation at CLS 44, which focused on my work on phonologically-mediated restrictions on language mixing, originally termed the PF Disjunction Theorem (MacSwan, 1999) and subsequently the PF Interface Condition (PFIC) (MacSwan, 2009; MacSwan & Colina, 2007, to appear).

Schindler, Legendre and Mbaye (2009) present counter-examples to the descriptive generalization underlying the PFIC based on Wolof affixation on a French stem, an example of word-internal switching. In discussion at CLS 44, we noted that unlike many other cases where the word-internal switching ban is well-attested, Wolof affixes, and *–woon* in particular, can be stranded . Torrence (2003) posits two versions of

the affix, one that is free and another bound; however, one can imagine a plausible account whereby the relevant Wolof affixes are treated as syntactic rather than morphological elements, the difference between English genitive *–s* and English third-person single *–s*. The authors assert that the evidence points to the conclusion that "Wolof has two types of *-woon*, identifiable both syntactically and phonologically." However, the evidence reviewed is in fact quite narrow, related to phonological processes that affect the element in some contexts and not in others, and that an utterance may contain two instances of *–woon*. One could draw similar conclusions about Romance clitics, based largely on parallel arguments, alleging that head-climbing clitics are pre-syntactic affixes and other instances are syntactic words. In the absence of compelling evidence to the contrary, an analysis of *–woon* which attributes the same lexical status to the element in all instances should be preferred. If *–woon* may be displaced, it is almost certainly an independent syntactic object, probably an XP, and is not a morpheme attached pre-syntactically in the lexicon. Because the proposed restriction on word-internal mixing applies to heads and not phrases, the *-woon* cases do not appear to contradict the ban on word-internal CS.

In a second set of proposed counter-examples, the authors note that a French word may trigger Wolof vowel harmony in an adjacent Wolof element. The authors argue that the fact suggests phonological interaction of a sort not permitted by the PFIC. Note, however, that although an adjacent word cues a phonological change (vowel harmony) in Wolof, the changes are restricted to Wolof, so far as is evident. If we were to see vowel harmony on the French word, along with distinctive phonological characteristics of French, that would indeed be amazing and highly relevant. However, so long as the

internal modification of sound structure is restricted around word boundaries, cases such as these do not constitute counter-evidence. In fact, the findings parallel MacSwan and Colina (2007, to appear) in which it was found that adjacent English words could trigger phonological operations on Spanish, but that neither Spanish nor English could trigger these operations on *English* words. In this sense, the phonologies appear to remain separate.

While the data are indeed fascinating and warrant further exploration, even if they should be interpreted in the way the authors suggest, they do not lead to the conclusion that data from other language pairs are incorrect. There is robust evidence which has a much clearer status that the Wolof-French cases reviewed by Schindler, Legendre and Mbaye, such as the prohibition against adding Spanish tense affixes to English stems (*love*-ó 'I loved', for instance). Even if the Wolof-French counter-examples stood on firm ground, the facts should lead us to ask why CS could occur in the Wolof-French cases but not in other word-internal contexts (Spanish-English, for instance). Such questions could lead us to abandon the PFIC in favor of a better proposal, or to modify it in some respects. Or the questions might lead us to probe Wolof to discover the true nature of affixes like *–woon* in these specific contexts, to permit us to reconcile our understanding of Wolof syntax with the CS facts. In the end, the authors' proposal of "co-phonologies" or of "morpheme-specific phonologies" does not appear to be radically different from the notion of "separately encapsulated phonologies" embedded within the PFIC.

More germane to the present discussion, we note that falsification of the PFIC, which is a specific theory about interface conditions governing CS, does not constitute the falsification of the constraint-free research program. Rather, it invites us to explore

other options, based on new questions arising from the discussion, which further the mission of a constraint-free approach to CS.

However, the authors suggest a retreat to a constraint-oriented paradigm, offering an OT implementation of the MLF Model CS-specific constraints. Such proposals generally do little more than restate the descriptive facts observed in CS, and the specific proposal of Schindler, Legendre and Mbaye (2009) offers little or no convincing treatment of new empirical ground at the cost of admitting a substantial catalog of linguistic machinery uniquely suited to account for their CS data. While I see the French-Wolof facts as worthy of further investigation which might help us refine our understanding of CS, they should not compel us to abandon the promise of the constraint-free research program for CS.

References

Abney, S. P. (1987). The English noun phrase in its sentential aspect. Ph.D. dissertation, MIT.

Azuma, S. (1991). Processing and intrasentential code-switching. Ph.D. diss. Univ. of Texas, Austin.

Azuma, S. (1993). The frame-content hypothesis in speech production: Evidence from intrasentential code switching. Linguistics, 31, 1071-1093.

Bandi-Rao, S. & M. den Dikken. (To appear). Light switches: On v as a pivot in codeswitching, and the nature of the ban on word-internal switches. In J. MacSwan (ed.) Grammatical Theory and Bilingual Codeswitching. Cambridge, MA: MIT Press.

Belazi, H. M., Rubin, E. J. & Toribio, A. J. (1994). Code switching and X-Bar Theory: The Functional Head Constraint. Linguistic Inquiry, 25(2), 221-237.

Boeckx, C. (2006). Linguistic minimalism: Origins, concepts, methods, and aims. Oxford: Oxford University Press.

Borer, H. (1984). Parametric syntax: Case studies in Semitic and Romance languages. Dordrecht: Foris

Cardinaletti, A. (1994). On the internal structure of pronominal DPs. Linguistic Review, 1994, 11, 3-4, 195-219.

Cardinaletti, A., & Starke, M. (1996). Deficient pronouns: A view from Germanic: A study in the unified description of Germanic and Romance, pp 21-65. In Thrainsson, Hoskuldur, Epstein, Samuel David, & Peter, Steve (Eds.), Studies in comparative Germanic syntax. Volume 2. Dordrecht, Netherlands: Kluwer.

Carnie, A, (2000). On the definition of X^0 and XP. Syntax, 3(2), 59-106.

Chomsky, N. (1964). Current issues in linguistic theory. The Hague: Mouton.

Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge: MIT Press.

Chomsky, N. (1970). Remarks on nominalization. In R. Jacobs and P. Rosenbaum (eds.), Readings in English transformational grammar. Waltham, MA: Ginn.

Chomsky, N. (1981). Lectures on government and binding. New York: Mouton de Gruyter.

Chomsky, N. (1991). Linguistics and cognitive science: Problems and mysteries. In A. Kasher (ed.), The Chomskyan turn. Cambridge: Blackwell.

Chomsky, N. (1994). Bare phrase structure. MIT occasional papers in linguistics 5. Also published in G. Webelhuth (ed.) (1995), Government and Binding Theory and the Minimalist Program. Oxford: Blackwell.

Chomsky, N. (1995). The minimalist program. Cambridge: MIT Press.

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels & J. Uriagereka (eds.), Step by step: Essays on Minimalist syntax in honor of Howard Lasnik, pp. 89-155. Cambridge: MIT Press.

Chomsky, N. (2001). Derivation by phase. In M. Kenstowics (ed.), Ken Hale: A life in language, pp. 1-51. Cambridge: MIT Press.

de Bot, K. (1992). A bilingual production model: Levelt's "speaking" model adapted. Applied Linguistics, 13(1).

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

Espinosa, A. M. (1911). Studies in New Mexican Spanish, Part II: Morphology. Revue de Dialectologie Romane 3, 251-268.

van Gelderen, E. & MacSwan, J. (2008). Interface conditions and codeswitching: An F-movement analysis of pronouns and lexical DPs. Lingua, 118(6), 765-776.

Gumperz, J. (1967). On the linguistic markers of bilingual communication. Journal of Social Issues 28(2), 48-57.

Gumperz, J. (1970). Verbal strategies and multilingual communication. In ed. J. E. Alatis (ed.), Georgetown round table on language and linguistics. Washington, DC: Georgetown University Press.

Gumperz, J., & E. Hernandez-Chavez. (1970). Cognitive aspects of bilingual communication. In W. H. Whitely (ed.), Language and social change. Oxford: Oxford University Press.

Hasselmo, N. (1972). Code-switching as ordered selection, pp. 261-280. In E. S. Firchow (ed.) Studies for Einar Haugen. The Hague: Mouton.

Jake, J., Myers-Scotton, C. & Gross, S. (2002). Making a minimalist approach to codeswitching work: Adding the Matrix Language. Bilingualism: Language and Cognition, 5(1), 69-91.

Joshi, A. (1985). Processing of sentences with intrasentential code switching. In D. R. Dowty, L. Kattunen, & A. M. Zwicky (eds.), Natural language parsing: psychological, computational and theoretical perspectives, pp. 190-205. Cambridge: Cambridge University Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programs, pp. 91-195. In I. Lakatos & A. Musgrave (eds.) Criticism and the Growth of Knowledge. Cambridge: Cambridge University Press.

Lasnik, H. & J. Uriagereka. (1988). A course in GB syntax: Lectures on binding and empty categories. Cambridge: MIT Press.

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

Longobardi, G. (1994). Reference and proper names: A theory of N-movement in syntax and logical form. Linguistic Inquiry, 25(4), 609-665.

MacSwan, J. & S. Colina. (2007). Interface conditions on language mixing. Paper presented at the 6th International Symposium on Bilingualism. . Hamburg, Germany, May 30-June 2.

MacSwan, J. & S. Colina. (To appear). Some consequences of language design: Codeswitching and the PF Interface. In J. MacSwan (ed.) Grammatical Theory and Bilingual Codeswitching. Cambridge, MA: MIT Press.

MacSwan, J. (1999). A minimalist approach to intrasentential code switching. New York: Garland Press.

MacSwan, J. (2000). The architecture of the bilingual language faculty: Evidence from codeswitching. Bilingualism: Language and Cognition, 3(1), 37-54.

MacSwan, J. (2004). Code switching and linguistic theory. In T. K. Bhatia & W. Ritchie (eds.), Handbook of bilingualism. Oxford: Blackwell.

MacSwan, J. (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.

MacSwan, J. (2009). Generative approaches to codeswitching. In A. J. Toribio & B. E. Bullock (eds.), Cambridge Handbook of Linguistic Codeswitching. Cambridge University Press.

Moro Quintanilla, M. (To appear). The semantic interpretation and syntactic distribution of determiner phrases in Spanish-English codeswitching. In J. MacSwan (ed.) Grammatical Theory and Bilingual Codeswitching. Cambridge, MA: MIT Press.

Muysken, P. & de Rooij, V. (1995). Review of Social motivations for code-switching: Evidence from Africa and Duelling languages: Grammatical structure in codeswitching. Linguistics, 33, 1043-1066.

Muysken, P. (2000). Bilingual speech: A typology of code-mixing. Cambridge: Cambridge University Press.

Myers-Scotton, C. (1993). Dueling languages: Grammatical structure in code switching. Oxford: Clarendon Press.

Nishimura, M. (1997). Japanese/English code-switching: Syntax and pragmatics. In I. Rauch (ed.), Berkeley insights in linguistics and semiotics, vol. 24. New York: Peter Lang.

Pensel, I. (1979). Testeni. Hor Yezh, 126, 47–73.

Ross, J. (1967). Constraints on variables in syntax. Ph.D. dissertation, MIT.

Sankoff, D. & S. Poplack. (1981). A formal grammar for code-switching. Papers in Linguistics, 14, 3-45

Schindler, M., Legendre, G. & Mbaye, A. (2009). Violations of the PF Interface Condition in Urban Wolof. CLS 44 Proceedings [this volume].

Speas, M. (1990). Phrase structure in natural language. Dordrecht: Kluwer.

Stenson, N. (1990). Phrase structure congruence, government, and Irish-English code switching. Syntax and Semantics, 23.

Stowell, T. (1981). Origins of phrase structure. Ph.D. dissertation, MIT.

Timm, L. A. (1975). Spanish-English code-switching: El porqué and how-not-to. Romance Philology 28, 473-82.

Torrence, H. 2003. Verb Movement in Wolof. UCLA Working Papers in Linguistics, 9.

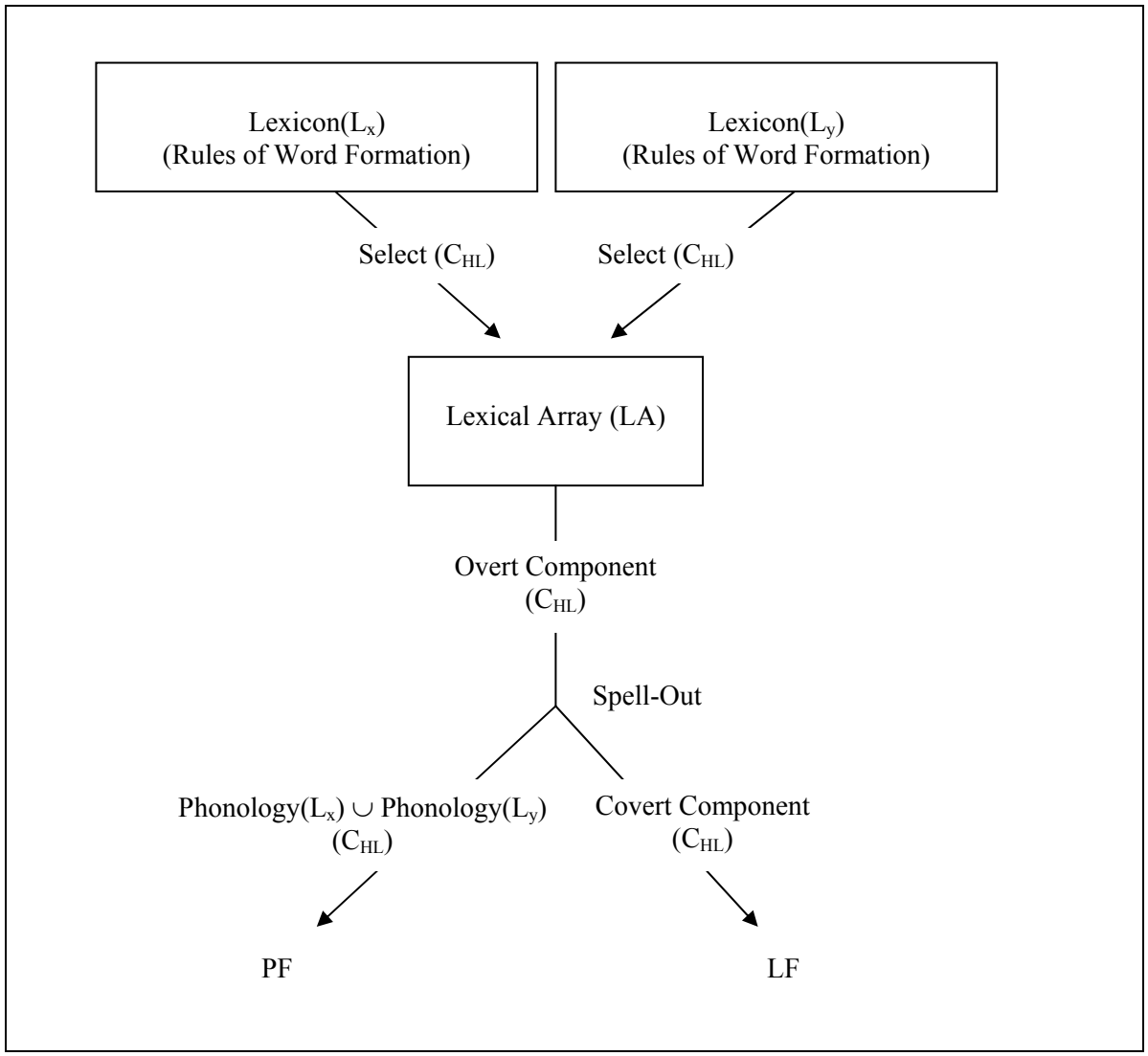
Troadec, B. (1983). Tri fennad-kaoz gant an Itron Bernadette Troadec. Hor Yezh, 71–127.

Vogt, H. (1954). Review of Languages in Contact. Word 10, 79-82.

Weinreich, U. (1993). Languages in contact. The Hague: Mouton.

Wentz, J. (1977). Some considerations in the development of a syntactic description of code-switching. Ph.D. dissertation, University of Illinois, Urbana-Champaign.

Figure 1. A Minimalist Approach to Codeswitching



Endnotes

[*]An extended version of this paper is forthcoming as the introduction to MacSwan (ed.), Grammatical Theory and Bilingual Codeswitching (MIT Press). I am indebted to Yaron McNabb and Juan Bueno Holle and other members of the CLS 44 organizing committee who invited me to speak at the para-session on CS and organized and hosted a magnificent meeting. Many thanks, too, to the editors of the current volume for their patience and kindness.

1. Bandi-Rao and den Dikken (To appear) explore the possibility of a pursuing a constraint-free approach using Distributed Morphology, a late insertion model.