

THE PENN REVIEW
o f
L I N G U I S T I C S

12th PENN LINGUISTICS
COLLOQUIUM

FEBRUARY 1988

**MARKEDNESS IN LANGUAGE MIXING:
LEXICAL AND GRAMMATICAL PROCESSES***

Enric Vallduví
University of Pennsylvania

1. Introduction.

In almost all the works on intrasentential codeswitching the reader finds a preliminary warning about what constitutes a true switch and what does not (i.e. a reminder of the distinction between 'borrowing' and codeswitching in situations of language mixing). However, while the existence of codeswitching seems obvious and unquestionable, both inter- and intrasententially, the existence of synchronic 'borrowing' -a parallel lexical process- is admitted by some, vaguely suggested by others, and even put in doubt by a third group of authors.

2. The Lower Level Model of Language Mixing.

The goal of this paper is to show that a number of facts of language mixing described in the literature can be accounted for with respect to the distinction between 'borrowing' -henceforth labeled lexical switching to avoid any confusions with diachronic borrowing- and codeswitching as a cognitively distinct form of synchronic language mixing. Furthermore, I shall argue that lexical switching is the unmarked form of language mixing, and that the facts of language mixing presented or revisited in this paper are the reflection of the preference for lexical switching over codeswitching, since the former is a low-level lexical interaction and the latter requires a shift in the use of a whole grammar. I shall call this approach the Lower Level Model in Language Mixing.

The Lower Level Model of Language Mixing assumes that lexical switching (LEX-S) is a cognitively unmarked form of language mixing (LMIX), since it is an exclusively lexical phenomenon, independent from grammar, which does not constitute a 'violation' of any of the intervening codes. In contrast, intrasentential codeswitching (COD-S) is a cognitively marked form of LMIX, and it represents an alternation between grammatical systems, i.e. a shift from one grammar to another.¹ A simplistic outline is in [1], while [2] expresses the cognitive markedness condition:

[1] GRAMMAR₁ <- COD-S -> GRAMMAR₂
 | |
 LEXICON₁ <- LEX-S -> LEXICON₂

[2] Speakers prefer LEX-S to COD-S due to the cognitive unmarkedness of the former, eg.

- [a] [-marked]: comieron [LEX-S LETTUCE]
 [b] [+marked]: comieron [COD-S LETTUCE]
 (they) ate

A number of criteria have been used to distinguish between COD-S and LEX-S (and/or borrowing): phonological integration, frequency of use, distribution, etc. However, none of these seems appropriate. I will refer you to Pintzuk & Prince 1984 and Poplack & Sankoff 1984 for further details. Appel & Muysken 1987 arrive at the conclusion that 'all the relevant evidence points to the fact that it is not possible to distinguish individual cases of code-mixing from not-yet-integrated borrowing on the basis of simple diagnostic criteria. The distinction has a theoretical basis [...but] further work on the implications of this difference will need to yield new operational criteria' (:173). This is precisely the goal of this paper.

There is yet another point that should be made clear. Many works that focus on the study of the constraints that rule grammatical codeswitching make a distinction between this and an independent lexical process. However, in many cases LEX-S turns out to be used as a category that takes care of the exceptions to the constraints posited for COD-S, and, even sometimes, exclusively defined by this body of exceptions. The approach here presented tries to make some headway towards some a priori operational criteria that allow a bias-free diagnosis of language-mixed tokens.

3. Frequency and Location of Mixings.

The first empirical fact that can be accounted for by the LEX-S/COD-S distinction in the light of the Lower Level Model as stated in [2] is the different mixing frequencies of different grammatical categories and constituents.

[3]: Language-Mixed Categories: X^0 vs XP .

	N^0	NP	V^0	VP	P^0	PP
Ar. → Eng.	63	13	23	8	1	1
Sp. → Eng.	818	69	71	5	0	36
Jap. → Eng.	296	107	26	10	0	2

(Source: Mohamed 1983, Arabic-English; Pfaff 1975, Spanish-English; Nishimura 1985, Japanese-English)

The distributional evidence shown in table [3] can be understood within our Model.

A clear pattern is immediately apparent: X^0 categories are mixed with much more frequency than XP constituents (with the exception of PP, to which we shall return below). If a single mechanism is responsible for mixing in language contact, how do we account for this difference? To my knowledge, no interpretation of these facts is found in the literature. Using a Lower Level Model approach, the facts do make sense. We cannot be certain of the status of mixed X^0 : they could be either COD-Ss or LEX-Ss. However, we can be certain about the status of mixed XPs: they have to be COD-Ss. Since there is no reason to believe that codeswitching is more effective with bare constituents than with phrasal constituents, we can assume without too much risk that the difference between X^0 and XP is caused by the intervention of LEX-S. Again we observe how COD-S is a much more restricted phenomenon than LEX-S, as one would expect from a marked, cognitively complicated operation.

In table [3] we also see the divergent behavior of nouns and verbs when mixed, a puzzling fact that has long been observed. Remember that one of the basic differences between LEX-S and COD-S is that of the involvement of the grammar in the latter and the exclusively lexical nature of the former. If we regard verbal inflectional morphology (especially tense) as part of the grammar and nominal morphology as part of the word formation component² we can explain these facts. That tense might be relevant to the point is supported by the fact, illustrated in table [4], that only 5/71 language-mixed verbs in Pfaff's data, only 1/23 language-mixed verbs in Mohamed's corpus, and none in Nishimura's data are tensed verbs, i.e. are language-mixed along with tense.

[4] Language-Mixed Tensed and Tenseless Verbs

	$V^0[+tns]$	$V^0[-tns]$	
A->E	1	22	= 23
S->E	5	66	= 71
J->E	0	26	= 26

If we bear the Lower Level Model in mind, it makes perfect sense that nouns and tenseless verbal forms show a higher rate of mixing, since they can be lexically-switched. In contrast, tensed verbs cannot be lexically switched unless additional operations are used in order to separate the lexically-switchable lexical material from the non-lexically-switchable

grammatical material. These additional strategies make the verbal LEX-S more cognitively complicated and therefore less desirable than a nominal LEX-S. Thus the difference in the amount of verbal and nominal mixes. The Lower Level Model would also predict that language-mixed tensed verbs represent actual COD-S. As we observe, the mixing rate -actually codeswitching rate- of tensed verbs is significantly lower, in accordance with our markedness condition, than the mixing rate of tenseless verbs, which can be lexically-switched.

But, what are these strategies? On the one hand, we have the periphrastic or pro-verb construction strategy described in Pfaff (1979), Joshi (1984) and others, and shown in [5]:

- [5]a [Spanish-English]
 su hija hace TEACH allá en San Jose
 do+tns
 'his daughter teaches there in San Jose' (Pfaff's #27)
- b [Marathi-English]
 to parat jayco DECIDE karto
 he back going do+tns
 'he decides to go back' (Joshi's #3.24)
- c [Japanese-English]
 Boston ni HIT shita toki ga
 do+tns time
 'the time when (we) hit Boston (Nishimura's #4.14)

We see how tense is attached to a dummy verb (generally do) and the lexical information of the verb, i.e. the root, is then lexically switched.

A second strategy is to lexically switch the root of the verb and inflect it with the host language inflectional morphology, as illustrated by [8], resulting in a word-internal lexical switch:

- [6]a [Spanish-English]
 los hombres me TRUSTearon
 +tns
 'the men trusted me' (Pfaff's #21a)
- b [Arabic-English]
 ana GUESS-t-ha
 I +tns+it
 'I guessed it' (Mohamed's #1)

These strategies have generally been seen as a means of avoiding 'switching' the tense of the sentence, which is a closed class item (see below). However, both strategies are viewed by the Lower Level Model as devices to avoid COD-S in favor of LEX-S, which we have assumed is a far less complex process, following [2]. Therefore, if these strategies are used to escape COD-S, the mixed verbal lexical item involved in the strategy has to be a reflection of LEX-S. In other words, tense cannot be lexically switched -it is not lexical- and therefore the mentioned strategies have to be used in order to separate the lexical material from the grammatical material and lexically switch the former.

I should mention here that Sankoff et al also pay attention to the periphrastic constructions, which they call 'devices or syntactic slots which are specialized in the functions of accepting and integrating borrowed items' (1985:6). However, in their approach, there is no tangible reason for which the verbal root involved in the periphrastic construction must be a lexical switch. They do predict, as the Lower Level Model does, that in a sentence like [7a] COD-S must have occurred, since 'the syntactic and morphological systems change within the sentence' (:5). For the example in [5a] above, they claim the analysis in [8a], rejecting [8b]. The truth is that, without any further conditions on LMIX, both [8a] and [8b] are possible. Only if we assume something like the cognitive markedness condition can we affirm that [8b] is invalid, since the function of the periphrastic construction is precisely avoiding COD-S.

[7]a su hija TEACHES allá en San Jose

[8]a su hija hace [LEX-S TEACH] allá en S.J.
 b su hija hace [COD-S TEACH] allá en S.J.

An interesting and striking case is presented by Prince 1987. In her data all verbal inflection is realized in the language of the sentence. While in Yiddish sentences, English verbs show Yiddish inflection [11a], in English sentences, Yiddish verbs show English inflection [11b]. In contrast, plurals are almost always pluralized with the morphology of the guest language, a fact that would support the lexical nature of nominal morphology, which, therefore, would not be a hindrance for the LEX-S of nouns.

[9]a [Yiddish-English]
 er WATCHt nor andere zoln arbeten
 he +tns only other AUX to-work
 'but he only sees to other people working'

- b [English-Yiddish]
 PATSHed 'slapped' (3sg pret)
 KHALISHed 'fainted' (3pl pret)
 [from Prince (1987)]

In sum, if all these bare-constituent LMIXs are thought of as LEX-Ss, and the theory of cognitive markedness assumed in the Lower Level Model is taken into account, the sharp distributional difference analyzed in this section can be explained.

4. The Lower Level Model and Interference.

There is a third interesting element in table [3]. The cases of P⁰ and PP. Here we should also take into account what we have already mentioned about tense, and also that, for example, in Mohamed's (1983) and Nishimura's (1985) data there are no language-mixed articles, complementizers, quantifiers, or pronouns. This leads us to the question of Closed Class Items (CCI) and Open Class Items (OCI). For some works on intrasentential codeswitching this is an important distinction: Joshi's (1984) model has as one of its two basic constraints the Closed Class Item Constraint, which states that CCI cannot be switched. Counterexamples to this constraint are readily found, as Pintzuk & Prince (1984) (P&P 1984) show. See [12]:

- [10]a we go MIT the bus
 with [P&P's #16c]
- b it was MAYN daughter's house
 my [P&P's #16e]

Their approach to the question is to catalog all CCI LMIX as 'interference'. What encourages them to do so is the fact that all mixed CCIs are found in L2 matrix sentences, i.e. they constitute a switch from non-dominant to dominant language. Since by interference they understand the influence or intrusion of one language -the dominant one- on another -the non-dominant one- as the result of an 'incomplete model' of the latter (:6), their proposal makes apparent sense. Therefore, for them, interference is distinguished from COD-S.

The psycholinguistic evidence for the difference between CCI and OCI is strong and the Lower Level Model should incorporate it. There is, however, a loose end in both Joshi's and P&P's treatment of the CCI/OCI distinction, which we shall

point out in a moment.

The evidence from language acquisition shows that, during the two-word and the telegraphic stages, children acquire lexical items which they incorporate into their OCI lexicon. It has been also claimed, although this remains controversial, that CCIs, along with syntax, come later. On the other hand, the work of several researchers in psycholinguistics (cf. P&P 1984), specially Garrett 1980, suggests not only that CCIs and OCIs are stored in two different sublexicons, but also that CCIs are much closely related to the syntactic planning level than OCIs, which are selected at a second level.

If we regard COD-S as an alternation of two grammars, i.e. 'the interaction between two grammatical systems' (Joshi 1985:190), then we should allow the language-mixing of CCI, due to their intimate relationship with the grammar. In other words, it is more reasonable to argue that a LMIX is a COD-S when the observable mixed item is a CCI than when it is a OCI, which has been shown to be more distant from the grammar. Similarly, in P&P 1984, it seems counterintuitive to argue for the closeness of CCIs and syntax and afterwards conclude that interference of CCI is not COD-S.

The facts, however, could be interpreted in the following way. First, since LEX-S is an interaction between two lexicons of OCI, it has no access whatsoever to the CCI set -this is why P⁰ in table [3], tense in table [4] above and other CCIs cannot be lexically switched. It follows from this that any language-mixed CCI represents a case of COD-S and never a case of LEX-S.

On the other hand, CCI LMIX, in the Lower Level Model, is an instance of COD-S, in which the 'incomplete' speakers of L2 switch grammars to L1 because of special difficulties they might have accessing their L2 CCIs (cf P&P 1984:9). Therefore, 'interference' is understood as a case of COD-S with a special characteristic: it is asymmetrical (L2 -> L1). The number of language-mixed CCIs is much lower than the number of language-mixed OCI (more dramatically in balanced bilinguals than in unbalanced bilinguals), since language-mixing the former requires a marked cognitive process -COD-S-, while language-mixing the latter only requires a low-level LEX-S.

5. A Sociolinguistic Argument.

As stated in the introduction to this paper, the distinction LEX-S/COD-S is useful in a number of ways. Another fact that should be considered is a sociolinguistic one. It is well-known that there are different degrees of LMIX in

different bilingual communities. Frequency of LMIX has been seen as a continuum with some speech communities at the top, others at the bottom.

It has been pointed out that French-Canadian speakers are very reluctant to use LMIX and that flagging of English elements is extremely common (Poplack 1985). Another example is the case of Catalan. In a sociolinguistic study of a neighborhood in Barcelona, Calsamiglia & Tuson 1980 report 4 Spanish utterances out of a total of 577 recorded utterances among a group of young Catalan-dominant bilingual speakers. There is no data regarding LEX-S, but in my own corpus of 1 1/2 hours of speech of Catalan-Spanish bilinguals the language-mixed items into Spanish amount to 16. 13 of the Spanish switches are single words, and the other 3 are direct quotes, conveniently flagged.

Scotton 1987 displays a thorough theory of sociolinguistic markedness in codeswitching seen as a status affirmator and negotiator. As she says, linguistic behavior is a clear marker of collective identity. She points out that in order for codeswitching to occur 'both varieties must be indexical of social identities which are positively evaluated for the exchange' (:21). In other words, mixing of a code -a grammar- does not take place unless both languages are equally valued in a particular situation. This would explain why, as pointed out before, certain speech communities (francophones in Quebec, Catalan-speakers in Catalonia, etc.) do not show any -or almost any- (intrasentential) COD-S.

But, why do these speech communities show LEX-S? The answer is straightforward and it follows from the Lower Level Model of Language Mixing. Within the Lower Level Model the continuum mentioned above is reanalysed as a discrete distinction between speech communities using less marked LMIX, LEX-S (and intersentential codeswitching), and others using more marked styles of LMIX (COD-S). One reason for avoiding COD-S is to avoid violating the integrity of a code, of a grammar. However, LEX-S does not violate the code, and therefore can be used by all bilingual communities in higher or lower degree, whether both languages involved are positively valued or negatively valued.

This amounts to saying that two Catalan-dominant Catalan-Spanish bilinguals engaged in conversation will seldom codeswitch into Spanish, but will feel free to use LEX-S without feeling as if they were intruding in their Catalan grammar. Our model of cognitive markedness, given Scotton's theory of sociolinguistic markedness, is supported by the behavior of these non-codeswitching bilingual speech

communities.

6. Conclusion.

The Lower Level Model of Language Mixing states that, given two cognitively different processes of LMIX, one lexical -LEX-S- and the other grammatical -COD-S-, the former is the unmarked, cognitively simpler option, and that LEX-S is to be preferred by the speakers over COD-S for this reason.

When considering the Lower Level Model one might think that it does not constitute a fruitful avenue of research. It is true that the distinction COD-S/LEX-S, for the most part, cannot be made at the X^0 level, but the Model does suggest some important conclusions:

- [11]a Language-mixing of closed-class items (including tensed verbs, i.e. tense) is the result of COD-S (caused by interference or otherwise), but it cannot be the result of LEX-S.
i.e. *we go [LEX-S MIT] the bus.
- b Language-mixing of other X^0 constituents might be either COD-S or LEX-S, but the unmarked option is LEX-S (in fact, it must be LEX-S in cases of V^0 which are part of a periphrastic construction or carry tense in the host language morphology).
i.e. *mi hija hace [COD-S TEACH] en S.J.

Therefore, from my point of view, intra-sentential codeswitching should be studied, from a grammatical point of view, only in those cases in which more than a bare lexical category is involved. Some generalizations may be missed, but, when LEX-S is incorporated into the studies of COD-S, two completely different phenomena are subsumed under the same label and any results stemming from such an approach are flawed to a certain extent.

In sum, the theory examined in this paper calls for opening (or re-opening) a way of looking at language mixing facts, taking lexical switching and codeswitching as two different phenomena with different observable behavior. The existence of lexical switching is guaranteed by the need of the same (or a similar) process in other aspects of language and in the existence of other lexically confined contact phenomena. In addition, the Lower Level Model of Language Mixing is supported by different effects observable in areas of language contact: the behavior of different bilingual communities with respect to lexical switching and codeswitching, the different patterning of V^0 vs. N^0 , and also X^0 vs. XP (traditional codeswitching analyses have no account for this fact), and, finally,

incorporates interference within codeswitching, thus reflecting appropriately the intimacy of the CCI sublexicon and the grammar.

NOTES.

*I would like to thank Ellen Prince, for drawing my attention to the study of intrasentential codeswitching, and the other members of the bilingual codeswitching seminar at the University of Pennsylvania (Spring '87) for much helpful and stimulating discussion. I am also indebted to many other members of the department of linguistics at Penn, especially Claudia Mazzie, who made very valuable comments on earlier drafts of this paper.

1. Codeswitching above the sentence level will not be dealt with here, since it deserves attention on its own.

2. This has been claimed by a number of studies. Noun morphology would be located in the word-formation component in the lexicon (cf Kiparsky 1982, among others).

REFERENCES.

- Appel, René & Peter Muysken. 1987. *Language Contact and Bilingualism*. London: E. Arnold.
- Calsamiglia, Helena & Empar Tuson. 1980. *Us i alternança de llengües en grups de joves d'un barri de Barcelona: St Andreu del Palomar*. *Treballs de Sociolinguística* 3.11-82.
- Garrett, Merrill F. 1980. *Levels of Processing in Sentence Production*. *Language Production I*, ed. by B. Butterworth, 177- 220. New York: Academic Press.
- Joshi, Aravind K. 1984. *Processing of Sentences with Intra-Sentential Code-Switching*. *Natural Language Parsing: Psycholinguistic, Computational and Theoretical Perspectives*, ed. by D. Dowty, 190-205. Cambridge: CUP
- Kiparsky, Paul. 1982. *Explanation in Phonology*. Dordrecht: Foris.
- Mohamed, Fatma. 1983. *Arabic-English Code-Switching in the Speech of a Six-Year Old*. Ms.
- Nishimura, Miwa. 1985. *Intrasentential Codeswitching in Japanese and English*. Dissertation. University of

Pennsylvania.

- Pfaff, Carol W. 1979. Constraints on Language Mixing. *Language* 55. 291-319.
- Pintzuk, Susan & Ellen F. Prince. 1984. Bilingual Code-Switching and the Open/Closed Class Distinction. Paper presented at the LSA Annual Meeting. Minneapolis.
- Poplack, Shana. 1985. Contrasting Patterns of Code-Switching in Two Communities. *Papers From the 5th International Conference on Methods in Dialectology*, ed. by H. Wakentyne, 363-85. Victoria: University of Victoria.
- Poplack, Shana & David Sankoff. 1984. Borrowing: the Synchrony of Integration. *Linguistics* 22.99-135.
- Prince, Ellen F. 1987. Language Mixing and Lexical Categories. Paper presented at the 1987 GURT on Languages and Linguistics. Georgetown: March 12, 1987.
- Sankoff, David, Shana Poplack & S. Vanniarajan. 1985. The Case of the Nonce Loan in Tamil. Paper presented at the XIV NWAWE Conference. Georgetown University.
- Scotton, Carol M. 1987. Codeswitching and Types of Multilingual Communities. Paper presented at the 1987 GURT on Languages and Linguistics. Georgetown: March 12, 1987.