

The case of the nonce loan in Tamil

DAVID SANKOFF
Université de Montréal

SHANA POPLACK AND SWATHI VANNIARAJAN
University of Ottawa

ABSTRACT

Nonce borrowings in the speech of bilinguals differ from established loanwords in that they are not necessarily recurrent, widespread, or recognized by host language monolinguals. With established loanwords, however, they share the characteristics of morphological and syntactic integration into the host language and consist of single content words or compounds. Furthermore, both types of loanwords differ from intrasentential code-switching—alternate sentence fragments in the two languages, each of which is grammatical by monolingual standards from the standpoints of appropriate function words, morphology, and syntax. In a large corpus of Tamil–English bilingual speech, many words of English origin are found in objects governed by Tamil verbs and vice versa. The equivalence constraint on intrasentential code-switching predicts that no code-switch should occur between verb and object in an SOV/SVO bilingual situation, and hence that objects whose language differs from that of the verb must be borrowed, if only for the nonce. To verify this prediction, we compare quantitatively the distribution across various syntactic contexts of both native Tamil and English-origin complements of Tamil verbs, and find them to be parallel. But the strongest evidence in favor of the nonce borrowing hypothesis comes from an analysis of variable accusative and dative case marking in these complements, in which the English-origin material is shown, morphologically and syntactically, to be virtually indistinguishable from Tamil (nonpronominal) nouns. In addition, we present supporting evidence from the genitive, locative, and other cases and from nonce borrowings from Tamil into these speakers' English.

The mixing of two languages in bilingual discourse may take on different forms and may be the result of various processes. When fragments from both codes alternate within a single sentence, this is often called intrasentential code-switching, especially if each fragment consists of more than a single

This work was supported by a Social Science and Humanities Research Council of Canada grant to David Sankoff. The interviews and the initial transcription and correction of the corpus were carried out by Meena Balasubramaniam. Construction and management of the computer database were the responsibility of Sali Tagliamonte. We thank Anthony Kroch, William Labov, Magnus Ljung, Gillian Sankoff, and Lisa Travis for discussions and suggestions. We also thank our anonymous informants.

noun or other content word. When a single word etymologically belonging to one code (the donor) appears in a sentence that is otherwise entirely in the other code (the host), this may well be the result of borrowing, particularly if the word is known to monolingual speakers of the host language. Code-switching and borrowing are two very different processes, at least at the level of how they operate. Code-switching within the confines of a single sentence requires access to the syntactic apparatus of both languages, because, as is generally observed, each of the monolingual fragments making up a code-switched sentence is internally grammatical by the rules of its language. Borrowing on the other hand operates independently of the grammar of the donor language, though it may involve lexical items from that language that are not yet incorporated into the monolingual vocabulary of the host language, and these items may retain aspects of the donor language phonology.

Though they are produced by different processes, code-switching and borrowing often have similar outcomes. For example, a code-switch consisting of a single noun in one language within a sentence entirely in the other language may be superficially indistinguishable from a borrowing. As such single-word fragments are common in bilingual sentences, this possibility of confusion is perhaps the most important methodological problem in the empirical study of bilingual syntax, and is the main focus of this article. Our analysis is based on quantitative data drawn from recorded conversations among Tamil-English bilinguals, and we also refer to results of our comparative studies in five other bilingual communities: Puerto Ricans in New York City (Poplack, 1980, 1981), francophones bilingual in English in the Ottawa-Hull region (Poplack, 1985, 1988), Finnish-English bilinguals also resident in Ottawa (Poplack, Wheeler, & Westwood, 1989), as well as Moroccan Arabic-French (Nait M'Barek & Sankoff, 1988) and Tagalog-English bilinguals from immigrant communities in Montréal.

CODE-SWITCHING AND BORROWING

In recent years, many scholars have addressed the problem of where in a sentence a switch could be made from one language into the other (e.g., Bentahila & Davies, 1983; di Sciullo, Muysken, & Singh, 1986; Gumperz, 1982; Hasselmo, 1972; Joshi, 1985; Lipski, 1978; Muysken, 1990a, 1990b; Pfaff, 1979; Prince & Pintzuk, 1986; Romaine, 1989:Ch. 4; Treffers, 1990). Based on studies of Spanish-English bilingualism among Puerto Ricans in New York, Poplack (1981) postulated the equivalence constraint whereby switching is free to occur between two sentence elements if they are normally ordered in the same way by both monolingual grammars and is prohibited elsewhere.¹ Thus, for Puerto Rican bilinguals, switching is permitted between a determiner of one language and a noun in the other but prohibited between noun and adjective, because this would be contrary to English word order, or between (most) adjectives and the noun, because this would be contrary to Spanish order.

Implicit in the claim that code-switching satisfies the equivalence constraint is another, weaker claim: that it always involves alternating stretches of the two languages within the sentence, each fragment being perfectly grammatical according to the speaker's variety of one or the other of the two languages and ideally containing no phonological, morphological, syntactic, or lexical elements specific to the other language. Though there may be the odd exception, and though the lexical and phonological aspects of this latter claim must be formulated in a very specific manner, there is no evidence from our comparative studies of any systematic tendency toward *multiword* sentence fragments where the word order is specific to only one of the languages while the morphology and the lexicon belong to the other (see Bokamba, 1989; Madaki, 1983; Myers-Scotton & Azuma, forthcoming; Pandharipande, 1986, for claims to the contrary). Thus, we do not dwell further on the internal grammaticality of the alternating fragments in the present article.

The equivalence constraint itself is a stronger claim about how these fragments may be assembled into a complete, coherent sentence. Implicit in competing hypotheses about bilingual word order (e.g., Joshi, 1985; Rivas, 1981; Sridhar & Sridhar, 1980) is the same weaker claim about the monolingual grammaticality of monolingual sentence fragments; these hypotheses differ only in how they propose that fragments are put together to form sentences.

As a result of the New York Puerto Rican studies, Poplack (1981) also proposed the free morpheme constraint, which states that *switches* take place only at full word boundaries; two morphemes, one of which is bound to the other, must originate in the same language unless the free morpheme has been linguistically integrated into the language of the bound one, i.e., has been borrowed. For example, this accounted for the observation that English verb stems were not concatenated with Spanish endings unless the verb was borrowed into Spanish. In contrast to code-switching, borrowing involves the grammatical structure of one language only, with the other language playing a solely etymological role. The use of a loanword does not constitute a switch from the host language into the donor language. Indeed, it is ideally impossible to distinguish established loanwords from the native lexicon of the language into which they are borrowed on the basis of phonology, morphology, or syntax.

It might seem to be a straightforward matter, then, to distinguish single-word switches from loanwords depending on whether the lexical item retains the phonological, morphological, and syntactic characteristics of its original donor language while being unattested in the other, or whether it is phonologically, morphologically, and syntactically integrated into a sentence fragment in the host language and has already been attested there. We could then discard the loanwords and use only the switched items to test the code-switching constraints. In practice, however, because of inter- and intra-individual variability, and/or pre-existing similarities in the phonological systems of a bilingual's two speech varieties, phonology is not a reliable criterion for differentiating loanwords from code-switches (see Poplack, Sankoff,

& Miller, 1988, for a study of phonological gradients in usage patterns of several thousand English loanwords, both rare and established, in Canadian French.) In addition, as we make clear in the ensuing sections, the borrowing behavior of bilingual speakers is not confined to the use of established or attested loanwords. This means that we are left with morphological and syntactic properties as the only relatively reliable criteria for determining whether code-switching or borrowing has occurred, and even these may be insufficient in any given sentence: if a single word of one language appears in a sentence in the other at a site where word order coincides and no inflections are required, this may constitute a switch, but it may also be a loanword. Nevertheless, borrowing is a very different process from code-switching, subject to different constraints and conditions. Thus, failure to separate data on the two phenomena for analytical purposes can only lead to confusing results (see, e.g., Berk-Seligson, 1986; Eliasson, 1989; Myers-Scotton & Azuma, forthcoming; Nishimura, 1985).

This problem has prompted a number of studies on the characteristics of loanwords (Mougeon et al., 1985; Poplack, 1980; Poplack & Sankoff, 1984; Poplack, Sankoff, & Miller, 1988; Poplack, Sankoff, Pousada, 1982), especially on the relationships among phonological, morphological, syntactic, and social aspects of their assimilation or integration into usage patterns in the bilingual community. One of the goals of these studies is to develop operational criteria for distinguishing loanwords from code-switches. Thus, for the Puerto Rican data, a working hypothesis was that loanwords from English were phonologically, morphologically, and syntactically integrated into Spanish, were recurrent and widespread, and that an English word not satisfying these criteria could only occur in English monolingual discourse or in code-switches from Spanish to English. In general, however, borrowing is a much more productive process and is not bound by all of these constraints. In particular, phonological integration and the "social" characteristics of recurrence (in the speech of an individual) and distribution (across the community) need not be satisfied. This type of borrowing is sometimes called "*nonce*" borrowing (Poplack, 1988; Poplack, Sankoff, & Miller, 1988; Poplack, Wheeler, & Westwood, 1989; Weinreich, 1953:11). It is doubly difficult to distinguish loanwords from code-switches when this process is prevalent and, hence, to test the code-switching constraints. An inflection from one language on a word from the other could automatically be classified as a nonce loan rather than as a violation of the free morpheme constraint, whereas one of a bilingual pair of words on each side of a prohibited, nonequivalent boundary could also be considered a nonce loan rather than as participating in a violation of the equivalence constraint. Thus, a number of authors (e.g., Eliasson, 1989) have criticized the use of the analytical category of nonce loans as an ad hoc and circular means for explaining away violations of these constraints.

In this article, we refute this critique by characterizing in some detail the process of nonce borrowing among Tamil-English bilinguals and by showing that there are several lines of evidence strongly justifying both this

analytical category and our operational means of distinguishing it from code-switching in our corpus, independent of, and in addition to, the fact that if nonce tokens were incorrectly classified as switches, a great many of them in these data would seem to violate both of these (independent) constraints simultaneously.

THE DATA

Aside from the difficulty of distinguishing borrowing from code-switching, verification of the equivalence and free morpheme constraints in various bilingual communities has not been a straightforward matter. Grammatical intuitions about bilingual sentences, even by fluent bilinguals, although as accessible as monolingual judgments, are notoriously misleading when compared to real performance. In addition, grammatical theory gives us no reason to believe that intuitions about bilingual sentences have the same psycholinguistic status as evidence of grammatical structure as do intuitions about monolingual sentences. (Indeed, there is no evident justification in grammatical theory that all or any principles governing monolingual syntax need be pertinent to language mixing in bilingual production, though this has been postulated by many authors [e.g., di Sciullo et al, 1986; Woolford, 1983].)

Second, natural performance data including code-switching are hard to come by, requiring skilled ethnographic fieldwork techniques, usually by in-group members (Poplack, 1981). Third, performance data are characterized by extrasyntactic material such as hesitations, pauses, repetitions, fillers, and parentheticals. It is relatively easy, being less noticeable for both speaker and hearer, for speakers to violate syntagmatic constraints, either monolingual or bilingual, between the two parts of a sentence thus interrupted. Several studies have identified a concentration of such material at switch points in bilingual speech (Poplack, 1985; Poplack, Wheeler, & Westwood, 1989; Zentella, 1982).

Fourth, the study of unedited linguistic performance has to cope with other irregularities of production, such as utterances that are partially inaudible, difficult to transcribe or interpret, unfinished or atypical compared to the rest of the speaker's output. In the bilingual context, occasional examples of many kinds of language mixing occur occasionally, although the overwhelming majority of instances of mixing usually involve just one or two important mechanisms. The occasional examples, which may be interesting in their own right, and which should be reported, are nonetheless of far less relevance than systematic tendencies in the analysis and description of bilingual performance.

Fifth, the equivalence constraint allows switching where two languages do not obligatorily differ in word order. A pair of languages like Spanish and English that have similar word order provides relatively few kinds of syntac-

tic boundaries at which switching is prohibited, and thus does not constitute a strong test of the constraint.

In the series of comparative community studies just cited, we have developed an effective methodology for overcoming these difficulties. Thus, the data on which our analysis is based were gathered by a highly bilingual speaker of Tamil and English among a network of her acquaintances and their families living in Ottawa, Canada. All are highly educated and proficient bilinguals. The choice of language pair was in large part motivated by the major typological difference between Tamil and English, there being important word-order patterns in one that are prohibited in the other, thus providing a strong test of the equivalence constraint.

Twelve sets of unusually rich, informal conversations, averaging a total of 3 hours each, were taped by this ingroup member. The resulting corpus contains long stretches of monolingual Tamil discourse that we do not analyze systematically here. However, all sentences containing words or structures of English origin (including the surrounding Tamil context and extrasyntactic material) were transcribed and rechecked for accuracy and entered into a computer file for sorting and comparisons. There were over 1,100 such excerpts in all, comprising a total of 15,000 words about evenly divided among native Tamil forms (56%) and English-origin words (44%). We refer to this as the bilingual corpus to distinguish it from the larger monolingual Tamil corpus.

In our analysis of the bilingual material, we took account of *all* the tokens pertaining to each question under study and used quantitative means to assess the importance of the various mechanisms of mixing. When comparing bilingual structures with monolingual ones, we used the speakers' own monolingual vernaculars for quantitative comparison.

THE BORROWING PROCESS

Established loanwords may retain some traces of the phonological patterns of their language of origin. They may be concentrated in certain semantic fields with some cultural, geographical, or technical connections with that language. They do, however, tend to become at least morphologically and syntactically integrated into the host language. Loanwords take the same inflections and occupy the same syntactic slots as native host language words in the same grammatical categories. In the bilingual context, these facts can help distinguish loanwords from their original forms in the donor language, which, of course, take different inflections and may even occupy different slots. In our analysis, we test the stronger hypotheses that (a) loanwords are distributed quantitatively among syntactic slots in the same way as native words, and (b) variability of morphological marking is statistically parallel for borrowings and native words.

Another distinction between the borrowed lexicon and the general lexicon of the donor language is the grammatical category of the words that tend to be borrowed. Generally, content words are borrowed, and function words are resistant. Each bilingual context, of course, has its exceptions, and certain utterance-initial interjections, connectors, and topicalizers (some of which might be considered function words) are rather readily borrowed. Nevertheless, in most studies the mass of established loanwords consists primarily of nouns, with many adjectives and verbs, and a number of adverbs. Pronouns are very seldom borrowed nor are articles, quantifiers, demonstratives, and prepositions.

There is little controversial in these considerations as they apply to established loanwords. If they apply equally well to nonce borrowings, this will be a validation of our classification of all borrowings together – whether accepted or momentary – as distinct from code-switching. We are not directly concerned here with the processes of spread, assimilation, or nativization of borrowings, though we have recently studied these in great detail in another bilingual situation (Poplack, Sankoff, & Miller, 1988). The degree of phonological integration of a loanword, or its knowledge and acceptance by monolinguals, is not pertinent to the present argument. Where we claim in this article that a form is borrowed, it is immaterial for our immediate purposes whether it is a nonce loan or an established borrowing.

Though, as we have discussed, loanwords tend to behave like native host-language lexicon, there are a limited number of morphological devices or syntactic slots that are specialized in the function of accepting and integrating borrowed items. One that is widespread across many languages and language families is the use of a pro-verb (carrier, operator, or dummy verb) postposed or preposed to a borrowed untensed verb (see, e.g., B. Kachru, 1978; Y. Kachru, 1968, for Hindi; Madaki, 1983, for Hausa; Boeschoten & Verhoeven, 1985, for Turkish; Chana & Romaine, 1984, and Romaine, 1986, for Panjabi; Smith-Stark, 1976, for Mayan; among others). The pro-verb carries all tense and aspect marking, agreement markers, and other inflections, while the borrowed form is invariant. This strategy is also very productive in Tamil (Annamalai, 1971; Asher, 1982:208; Zvelebil 1983) based on the pro-verb *paNnu* 'do' and a small set of other verbs postposed to the bare infinitive of an English verb. Though *paNnu* is usually a transitive verb in monolingual Tamil, and occasionally takes an English-origin direct object in bilingual speech, its most frequent role in our bilingual corpus is as a pro-verb. This fact is of great methodological help here, as the easily identified presence of an English form followed by a pro-verb determines unambiguously that we have a Tamil-language verb construction incorporating an English borrowing, and not a code-switch.

Another extremely productive device is the creation of Tamil adverbs from borrowed adjectives, or occasionally nouns, by the addition of the adverbial suffix *-aa*, which is used occasionally in spoken Tamil on native forms but

is much more frequent on loanwords (Annamalai, 1971; Zvelebil, 1983). Thus, we have sincere-*aa*, immediate-*aa*, conscious-*aa*, stifling-*aa*, color-*aa*, routine-*aa*, gang-*aa*, and so on.

THE ACCUSATIVE CASE

The Tamil-English material is especially pertinent to switches between major constituents of the main clause. Much previous work on the equivalence constraint has involved languages such as Spanish and English, where there is relatively little difference in main clause structure. Tamil, on the other hand, especially vernacular spoken Tamil, is almost exclusively verb-final, whereas English direct objects and other verb complements almost always follow the verb.² The equivalence constraint predicts that true code-switching between verb and direct object or between direct object and verb would be rare or absent. Thus, if the constraint in fact holds in these data, whenever the object does not derive from the same language as the verb, we would predict that the former should have the properties of an established loanword or a nonce borrowing, and not of a code-switch.³

Examination of our corpus reveals large numbers of English-origin direct objects preceding Tamil verbs as well as a lesser number of Tamil-origin objects following English verbs, making this a crucial test case for the operation of the equivalence constraint. Note that in these data, there is never any ambiguity about the language of the verb. Tamil stems always take Tamil inflections, never English inflections, and English stems always take English inflections unless grafted onto a Tamil pro-verb, thereby becoming Tamil for the nonce.

To test our predictions, we extracted from our corpus those English-origin nouns and noun phrases governed by Tamil verbs or by English stem + inflected Tamil pro-verb that, according to the rules of spoken Tamil, could be considered to require the accusative case marker *-ei* or *-e*. For controls, we also extracted those *Tamil* nouns from the bilingual corpus that should take the accusative case.⁴ This is an important methodological point. Rather than rely on prescriptive, idealized, or standardized accounts of the monolingual varieties, it is essential in the study of bilingual syntax to take into account the speaker's own vernacular in the two languages. Finally, we examined the objects of all English verbs in the corpus except, of course, those stems attached to Tamil pro-verbs *paNnu* 'do', *aayi* 'happen', or *cheyyi* 'do' which had already been analyzed as Tamil verbs. Insofar as the predicted constraint against code-switching between verb and object is correct, any Tamil-origin objects of English verbs and all English-origin direct objects of Tamil verbs should have the properties of loanwords rather than of code-switches.

Of the 136 English-origin direct objects governed by Tamil verbs, all preceded the verb except 5:⁵

- (1) *naan pooyi paaDuvein* Hindi song-*ei*
 I go (inf) sing (1p-sg-fut) (acc)
 I will go and sing a Hindi song. (2341/SI)
- (2) *avanga vantu keeTuTu iruntaanga* questions
 they (filler) ask (inf-cont) be (3p-pl-past)
 They were asking questions. (6922/C)
- (3) *pooTuruvaan* letter
 put (3p-sg-masc-fut)
 He will write a letter (7194/C)
- (4) *Roughna ennanaa appaDi throw paNNaratu* things
 roughness what it means like that do (inf)
 Roughness means throwing things. (7025/C)
- (5) *Indiaala repeat paNNiNDe irupaan* questions
 (loc) do (cont) be (3p-sg-masc-fut)
 In India they (indef) would be repeating the questions. (2114/SI)

Could these cases be considered as violations of the equivalence constraint? To answer this, we first note that object-final constructions occur in Tamil as a result of postposing: "Major sentence constituents can be subjected to left movement (to sentence-initial position) or right movement (to postverbal position) to express emphasis" (Asher, 1982:85). This is especially true of literary Tamil, but even in the spoken language it can occur when emphasis is to be placed on the verb. In fact, our own corpus contains monolingual Tamil sentences that manifest VO order.

- (6) *aanaa enakku piDikka-ille anta poNnu*
 but I (dat) like (inf-neg) that girl
 But I didn't like that girl. (SD 2B/242)
- (7) *yaarum kuupida veeNdiatu-ille avangaLei*
 who call (inf) need to (inf-neg) they (acc)
 Nobody needs to call them. (DK 2B/162)

Such examples are admittedly infrequent in the monolingual part of the corpus, but the same holds for the bilingual sentences: cases (1)–(5) represent less than 4% of the English-origin accusatives.

The fact that in examples (1), (2), (4), (5), and (7), the object occurs after an auxiliary, and not directly after the verb governing it, is clear evidence that these sentences indeed result from postposition and do not represent underlying VO structures (see Pintzuk & Kroch, 1989, for this same argument in relation to Old English). Only examples (3) and (6) are ambiguous with respect to underlying object placement. We can conclude that examples (1)–(5) are not aberrant by Tamil word-order rules and hence do not violate the equivalence constraint.

What of the remaining 96%: 131 cases of apparently English direct objects preceding Tamil verbs? If these represent code-switches, they are all in clear violation of English word order. Our main goal in this section is to

prove that this is not the case: that the nouns in question are borrowings—whether nonce or established—and are being treated morphologically and syntactically exactly as if they were native Tamil nouns. The first piece of evidence is the nature of these English-origin direct objects themselves as exclusively content words. None of them is pronominal, compared to 40% pronominal forms among the native Tamil direct objects. No English articles, quantifiers, or demonstratives are preposed to the English-origin nouns, though the Tamil article *oru* ‘one’, quantifiers *konjam* ‘some’, *jaasti* ‘more’, *ellaam* ‘all’, and demonstratives *anta* ‘that’, *inta* ‘this’ do occur with them, as in examples (8)–(10), as do the interrogative particle *enna* and its variants and the emphatic suffix *thaan*.

- (8) *anta car-ei drive paNNanum*
 that (acc) do must
 We must drive that car. (846/Sh)
- (9) *oru seal pooTTu koDuppaanga*
 one put (inf) give (3p-pl-fut)
 They will put on the seal and give it. (1855/M)
- (10) *oru aambiLLainaaka konjam discretion use paNNuvaan*
 one man if some do (3p-sg-masc-fut)
 If he be a man, he will use some discretion. (4495/V)

The lack of English function words such as pronouns or articles is exactly what we would expect from borrowed vocabulary. There are two minor details that must be clarified in this connection. Borrowings generally occur as single words, but in our data there is a small proportion of English adjective–noun pairs—*snide remarks*, *serious subjects*, *educational system*—and very occasionally longer sequences—*slacks and blouses*, *Government of India scholarship*—constituting the direct object. In other situations, these sequences might be considered full NPs, fragments of English monolingual speech, rather than borrowings. In the present case, however, they must be treated as compound borrowings, because the function words typical of English NPs *never* co-occur with them.⁶ Second, as in (11) and (12), English plural inflection is occasionally present (Asher, 1982:227), though inflections on true borrowings should generally be in the host language. The plural inflection, however, is often cited as being susceptible to borrowing along with the noun (e.g., Bynon 1978:227), especially among educated speakers (Lehmann, 1973:218) and especially when the borrowings in question are not yet established (Weinreich, 1953:31). Furthermore, Tamil plurality is frequently unmarked morphologically and is often inferred from the semantics of the noun, quantifiers, and other aspects of the context, in ways apparently not appropriate to borrowed forms in the spoken language.

- (11) *Indian women-e avaa discriminate paNNa-ille*
 (acc) they do (neg)
 They don't discriminate (against) Indian women. (6078/B)

- (12) Only *kalaimagaLLataan* movies-*e* *patti peece ille*
 (magazine name) (loc) (only) (acc) about talk (emph) (neg)
 Only in *KalaimagaL* there is no talk about the movies. (6605/B)

But the presence of the accusative case marker in Tamil (the suffix *-ei* or *-e*) on English-origin preverbal direct objects is the strongest argument that these are morphologically and syntactically integrated borrowings, if only for the nonce. Not all the English-origin direct objects carry case marking, however. Does this mean that the others are code-switches? Not at all, because native Tamil direct objects are also marked only on a variable basis, as can be seen from examples (1)–(12).

The extent to which the use of *-e* is obligatory may be related to the choice of verb. . . . Inanimate nouns are optionally marked for accusative with all verbs, in the sense that an utterance containing a transitive verb and an inanimate object-NP can be grammatical both when this NP is overtly marked for accusative and when it is not. However, the choice is not entirely random; if the reference is specific and definite, there is a tendency for the accusative suffix to be used, while indefinites tend to select the nominative case form. (Asher, 1982:106–107)

In our data, the animate/inanimate constraint is only quantitative—even object personal pronouns do not always have to be marked, although unmarked tokens occur less frequently than with inanimate nouns. Nevertheless, the existence of variability in monolingual Tamil helps us show that the quantitative conditioning of variable case marking on Tamil and English-origin objects is remarkably parallel. First, we examine the contexts in which accusative case marking seems obligatory and those where it is optional. No simple structural criterion has been formulated for distinguishing between these two kinds of context. Operationally, we relied on native speaker judgments of the marked and unmarked versions of each object, using Tamil equivalents for the English-origin objects. We found that the instances considered to obligatorily require marking occurred when its absence led to incorrect readings due to nominative–accusative ambiguity, such as in certain examples when PRO-drop has occurred, or as in (13), when the verb inflection would not disambiguate between two ways of construing *ivan*—as subject or object—were the accusative marker not present. Conversely, the case-marker is only optional in spoken Tamil when it would be redundant, as in (14)—‘guest’ could only be the object; the inflection on the verb indicates that the subject is first person singular. In this kind of context, case marking can sometimes be associated with stylistic foregrounding of the direct object.

- (13) *anta poNNukku ivane paakaNum*
 that girl (dat) he (acc) see must
 For that girl, she must see him. (7070/Ch)

TABLE 1. *Distribution of English-origin and native Tamil direct objects among obligatory and optional marking contexts^a*

	Obligatory Contexts	Optional Contexts	Total
English origin	28 (20.6%)	108 (79.4%)	136
Native Tamil	58 (37.9%)	95 (62.1%)	153
Native Tamil (pronouns excluded)	18 (26.1%)	51 (73.9%)	69

^aChi-squared test with continuity correction: rows 1 and 2, $\chi^2 = 9.5$ ($p = .002$); rows 1 and 3, $\chi^2 = 0.51$ ($p = .47$).

- (14) Guest *ellaam paattein*
 all see (1p-sg-past)
 I saw guests and all. (336/S)

If English-origin borrowings are being treated identically to native Tamil direct objects, two statistical conditions should hold. First, the two sets of objects should be distributed in the same proportions among obligatory marking contexts and optional (or generally nonmarking) contexts—loanwords should generally be distributed in the same way as other words in the grammatical category into which they are borrowed. At first glance, there is a serious and statistically significant disproportion, as may be seen in the first two rows of Table 1. However, many of the Tamil objects are pronominal, whereas none of the English ones are, which is consistent with our contention that the English objects are borrowings. Now, more Tamil pronouns fall into obligatorily marked contexts than other nouns do. When the pronoun data are removed, we find remarkable congruence with no statistically significant difference in the distribution of the two sets of objects.

Second, and more important, we would expect the actual rates of case marking to be similar for English-origin objects and native Tamil objects, within each class of contexts. Examining first the optional marking contexts, the aggregated data again suggest a statistically significant disproportion between rates of accusative marking on English-origin versus native Tamil nouns, as in the first two rows of Table 2. But, again, when the more frequently inflected Tamil pronoun data are removed (as, consistent with the hypothesis that the English-origin objects are borrowings, English-origin pronouns do not occur in this context), the disproportion in marking rates is sharply reduced and is no longer statistically significant. Note that eight cases of quantifiers (acting as nominals) were also excluded from this tabulation because they usually behave like pronouns, but as only four of these were marked, their exclusion has scarcely affected the percentages.

This lack of English-origin pronouns cannot be explained away on the grounds of morphophonological incompatibility with Tamil verbs, as might be the case if preverbal objects were cliticized to the verb, because the Tamil pronouns in these positions are full forms, as in examples (13) and (44).

TABLE 2. *Variable accusative marking on English-origin and native Tamil direct objects in optional marking contexts^a*

	Marker Present	Marker Absent	Total
English-origin	31 (29%)	77 (71%)	108
Native Tamil	51 (54%)	44 (46%)	95
Native Tamil (pronouns excluded)	20 (39%)	31 (61%)	51

^aChi-squared test with continuity correction: rows 1 and 2, $\chi^2 = 12.1$ ($p = .0005$); rows 1 and 3, $\chi^2 = 1.3$ ($p = .25$).

TABLE 3. *Accusative marking on English-origin and native Tamil direct objects in obligatory marking contexts^a*

	Marker Present	Marker Absent	Total
English-origin	22 (79%)	6 (21%)	28
Native Tamil (pronouns excluded)	18 (100%)	0 (0%)	18

^aChi-squared test with continuity correction: $\chi^2 = 2.7$ ($p = .1$).

Turning to the obligatory marking contexts, Table 3 indicates six apparent anomalies among the English-origin data. One of these six is an artefact of our test for obligatoriness. As stated earlier, if the Tamil equivalent for an English loanword would categorically receive marking, we classified the context as obligatory. In the following example, however, the Tamil equivalent for 'telephone' (*tolei peesi*) occurs only in 'high' Tamil or the literary language, and it is the stylistic clash between this form and the nonexpression of the accusative marker that accounts for our classification of the context as obligatory. In fact, sentence (15) as it stands is perfectly acceptable; indeed, it would even be judged unusual in this particular example to have the *-ei* marker on *phone*.⁷

- (15) Phone pick *paNNareeLaa?*
do (2p-pl) (request)
Will you pick up the telephone? (1166/Sh)

In a second example, the unmarked direct object *divorce* is not in the usual immediately preverbal position in the verb phrase.

- (16) Divorce *inta uurula romba easy-aa paNNiTtatunaale*
this city (loc) very (adv) do (past) because
Divorce—since they have made it very easy in this city. . . (6004/B)

It thus has an alternative reading as a preposed object, in which case it would require nominative (null) marking. Note that if (16) does represent preposing, this could also be a case of code-switching with no violation of the equivalence constraint.

A third example is (4), where the postposition of the object and the lack of marking conspire in foregrounding the verb. In another example, the speaker pauses slightly between the direct object and the rest of the sentence, which may reflect processing difficulties with this aberrant construction.

- (17) *Brahmin. . . ellaam romba hate paNNaraa*
 all very do (3p-pl-pres)
 They hate Brahmins very much. (6160/B)

Depending on how many of these explanations are valid ways of accounting for the anomalous examples, we are left with at most 20%—probably less—of the 26 English-origin direct objects in obligatory marking contexts that received no marking. As with the 10% shortfall in marking rates in optional contexts, this figure is not statistically significant by itself, though we will have reason to return to it after evaluating the dative case.

We now examine the other half of the data where the roles of Tamil and English are reversed. Some 259 clearly English verbs (i.e., not counting English verb + Tamil pro-verb constructions, which we have identified as Tamil) in the corpus governed one or more objects, largely direct objects. Over 90% of these objects contained no Tamil elements whatsoever, either preceding or following the verb. The object phrases showed expected English structure and placement in the sentence, including the presence of English articles (example 18), quantifiers (20), demonstratives (19), and pronouns (21) in the object noun phrase.

- (18) *batilaa immediate-aa* they take the easiest route
 instead (adv)
 Instead, they immediately take the easiest route. (254/M)
 (19) I used to write down this Hindi song. (2340/SI)
 (20) I respect each person's views. (1294/Sh)
 (21) He'll call me to his room and scold me. (930/Sh)

These can all be seen to be ordinary English sentences or fragments of English monolingual speech. There remain 24 Tamil objects of English verbs in a total of 15 sentences. Can these be identified as Tamil loanwords, nonce or established, in the English of these Tamil speakers? If so, they should be treated morphologically and syntactically like English-origin material, that is, they should be situated postverbally; they should involve single nouns or common adjective + noun expressions; they should take English inflections (plural, if applicable), articles, demonstratives, and quantifiers, and should include no Tamil inflections, demonstratives, quantifiers, or pronouns. All of these conditions hold; the objects are largely single nouns and are all post-verbal, no Tamil inflections or function words are present, whereas English inflections and determiners do occur in a number of cases, as in (22)–(25).

- (22) You have *muuDis* for everything. (3068/Su)
lids
- (23) You have *karanDis* for seventy-five cents. (3062/Su)
spoons
- (24) It has got a *muuDi*. (3070/Su)
lid
- (25) So the Lord Jesus Christ is going to come back someday to take the
maNavaaTTi. (5166/D)
bride

Another widely used diagnostic of loanword status would be whether the words designate particular cultural concepts or idiomatic expressions to which there are no satisfactory English equivalents. Indeed, 15 of the 24 occurrences involved Tamil food or cooking terms as in (22)–(24) and (26)–(27).

- (26) My mother had made *taraTTipaal* or something. (2610/Su)
milk sweet
- (27) They still eat the same *rasam caatam* (2805/Su)
rice dish

Three involved Tamil cultural or religious celebrations.

- (28) I like *kacceri*. (1321/Sh)
concert music
- (29) They used to have Friday *puujai*. (2985/Su)
religious services
- (30) I miss all those *diipaavali*, *navaraatri bajanei*, and all that. (3030/Su)
festivals services

And three were high Tamil forms for biblical concepts within some cited Christian discourse, as in (25) and (31).

- (31) Then we will have *santoosham samaataanam*. (5155/D)
happiness peace

Two cases were clearly idiomatic expressions, as in (32).

- (32) Telling them *kaNNu*, *muukku*, *kaatu*, and everything. (375/Su)
eyes nose ears

The remaining case is a *palindromic switch*, sometimes called a copy translation or *portmanteau* construction.

- (33) They make *candai pooDaraanga*.
fight make (3p-pl-pres)
They are fighting. (7340/Ch)

TABLE 4. *Distribution of English-origin and native Tamil datives among three syntactic contexts^a*

	Dative Subjects	Dative Objects	Governed by Postpositions	Total
English-origin	11 (14%)	57 (71%)	12 (15%)	80
Native Tamil	9 (13%)	48 (69%)	13 (19%)	70

^aChi-squared test: $\chi^2 = 0.35$ ($p = .84$).

This is the only violation of the equivalence constraint among these 24 tokens of Tamil-origin objects of English verbs. In fact, it is the only clear code-switch. We return to these palindromic switches later. The other tokens are mostly single noun or idiomatic compounds, not simply translatable without losing some of the connotation. This, together with their lexical nature, morphology, the structure of the noun phrase in which they are found, and postverbal placement all indicate that they are borrowings and not code-switches.

THE DATIVE CASE

The dative marker *-kku* in Tamil is used for a larger variety of grammatical functions than the accusative. Aside from verb complements, it also is used frequently in subject positions for certain verbs, in many postpositional phrases, for nominalized sentences, and certain adverbials.

Of these, only those datives governed by verbs and postpositions are generally in positions prohibited by English word-order patterns. Nevertheless, it is instructive to include dative subjects in our comparison as well. As a first test for loanword status, we note that as with the accusatives, no English pronouns or articles occur among some 91 English-origin tokens in dative slots. This contrasts with the 70% pronominal native Tamil forms occupying dative slots. Example (34) shows the only two arguable exceptions to the content-word status of these English-origin forms.

- (34) *inta five-ukku, each-ukku oru head irukku*
 this (dat) (dat) a be (3p-fut)
 There will be a head to this (group of) five, to each (group of five). (4045/Su)

In Table 4, we compare the grammatical distribution of English-origin forms with native Tamil ones, omitting the pronoun data.⁸ The distributions are essentially identical, there being no statistically significant difference, with the large majority of datives occurring as objects of Tamil verbs, the rest being almost equally divided between subject position and postpositional phrases.

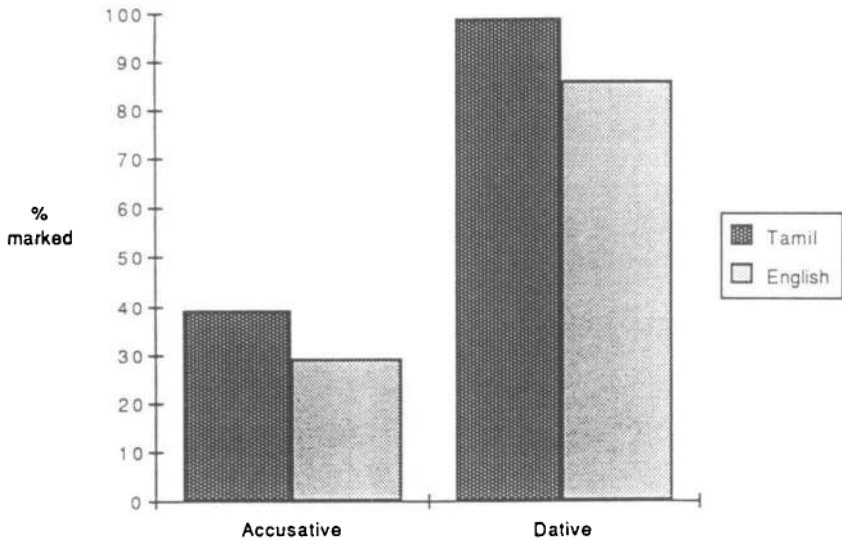


FIGURE 1. Variable marking of English-origin and native Tamil objects.

Tamil datives occur more frequently in postverbal positions than do accusatives. About 5% of the dative objects in our bilingual corpus were postverbal. The distribution was exactly as could be predicted from the rest of the dative data: 18 Tamil pronouns (70%), 4 Tamil content words (15%), and 4 English content words (15%), all marked for the dative.

Turning to the rates of dative case marking in all dative slots, we find that despite the acceptability of its absence to informants in some contexts, marker presence is almost categorical (99%) for both Tamil pronouns and other nouns. It is statistically less so ($p = .01$), but still overwhelmingly present (86%) for the English-origin material. This parallel between Tamil and English-origin material is particularly striking when these rates are contrasted with nonpronominal accusative rates of 39% and 29%, respectively, as depicted in Figure 1. Grammatical conditioning of variable case marking thus follows the same general pattern for English-origin and native Tamil verb complements. It is this pattern that we find to be the strongest evidence for the nonce loan hypothesis.⁹

Except for the locative, data on the other Tamil cases were generally too sparse to allow for the type of quantitative analysis effected for the accusative and dative. It is also sometimes more difficult than with accusatives and datives to unambiguously identify unmarked tokens that potentially could be marked for these other cases. Systematic examination, however, confirms the results of the two cases studied in detail. Locative case marking (*-la*) is virtually categorical (even more so than the dative) for both Tamil and English-origin material. Another parallel is that the locative occurs postverbally more frequently than the dative or accusative, for words of both ori-

gins. Genitive marking (*-oDa* or *-uDaya*) is more variable for words of both origins, but there are not enough (nonpronominal) data to measure it with any degree of accuracy. Nevertheless, the genitive—and the even rarer instrumental (*-aale*), ablative (*-lentu*), and comitative (*-oDa*)—case markers are all used on nonce borrowings in our corpus in the same way as on native Tamil nouns. Thus, insofar as cases requiring *overt* marking are concerned, English-origin nouns behave much like native Tamil ones.

THE NOMINATIVE CASE

The nominative case requires quite a different analysis than the other cases studied. First, Tamil nominatives receive null marking, so that there are no morphological indices helping us to identify single English-origin nouns as nonce loans with this case. Second, recall that the equivalence constraint predicts not only that there should be no code-switches between constituents that are ordered differently in the two languages, as in the cases of the verb complements already discussed. Other nonsyntactic (pragmatic, social, etc.) conditions permitting, we should also be able to find bona fide switches between any constituents that are ordered in the *same* way in both languages. Thus, switches immediately after the subject and before the verb phrase should occur freely in Tamil–English bilingual discourse. This means that in contrast to the accusative and the dative objects, we cannot classify English-origin subjects of Tamil verbs or Tamil subjects of English verbs as necessarily nonce borrowings, though they may well be, in some cases. Indeed, we have no way of positively identifying nonce loans in subject position. In some tokens, however, we can rule this possibility out, namely when a pronoun or a multiword noun phrase including function words is involved. In these instances, we are clearly dealing with code-switches.

Examples (35)–(39) are unmistakable cases of switching between Tamil subjects and English predicates.

- (35) *en niydi* is very clear.
 my principle
 My principle is very clear. (4511/MS)
- (36) *anta koRantai* goes through a lot of hell.
 that child
 That child goes through a lot of hell. (4511/SI)
- (37) *naan vantu* unfortunately get into chemistry.
 I (filler)
 I unfortunately get into chemistry. (7057/Ch)
- (38) *avaa* trying to keep it within the Indian range.
 they
 They are trying to keep it within the Indian range. (6178/B)
- (39) *atu vantu* doesn't bother me.
 that (filler)
 That doesn't bother me. (282/M)

There are also many relatively clear switches in the other direction.

- (40) Especially young widows *kalyaaNam paNNikka paDaatu*.
 marriage do (refl) should (neg)
 Especially young widows should not marry. (3161/RR)
- (41) That originality *illave-ille*.
 (neg-emph-neg)
 That originality is just not there. (1064/Sh)
- (42) Even the lowest middle class *ellaarum pooyiduvaa*.
 all go (3p-pl-fut)
 Even the lowest middle class will all go. (5283/B)

Much of the data, however, are less clear. Many subjects of otherwise Tamil sentences are single English-origin nouns or other noun phrases that could easily be nonce borrowings, according to the same criteria we have used in the previous sections. For example, it would be interesting to find more English subjects concatenated with Tamil direct objects, as in (40). These would confirm that the equivalence constraint operates on the level of adjacent constituents (considering O + V as a single constituent) and not simply on the level of adjacent lexical categories (Sankoff & Mainville, 1986a, 1986b). There are indeed examples such as:

- (43) Professor *enna* mark-e *eRuti* *koDutaar*
 what (acc) write (inf) give (3p-sg-past)
 What marks did the professor write and give? (3860/Su)
- (44) Boyfriends *unne* *date-ukku* *kuupaDaraanaaka* . . .
 you (acc) (dat) call (3p-pl-pres) if
 If boyfriends call you for a date. . . (542/B)

These could be postsubject, preobject code-switches. Though the article *the* is absent in (43), this is common in Indian English, especially in sentence initial position. On the other hand, the subjects could just as well be nonce borrowings – they are not pronouns, they are not accompanied by demonstratives, articles, or other indices that they are (single-word) fragments of English discourse. Even the presence of the plural marker -s can be consistent with borrowing, as we have seen. That there is no Tamil inflection is not pertinent, because the nominative is a null-marked case. The same sort of problem was encountered in example (16).

The search for English subjects concatenated with Tamil objects is hampered by the PRO-drop properties of Tamil as well. Subject indications are incorporated in the Tamil verbal inflections so that it is natural and frequent (40%) for sentences with transitive verbs to be subjectless. This drastically reduces the possibility of finding Tamil object + verb combinations with English pronominal subjects, which would be unequivocal code-switches.

Despite this conspiracy of factors mitigating against this one particular type of sentence, it is clear that, in general, postsubject switches occur in both

directions, as in examples (35)–(42) and as predicted by the equivalence constraint.

The discussion in this section highlights an important advantage of quantitative analysis. The question of whether the subjects of (43) and (44) are single-word English fragments or null-marked borrowings cannot be answered for each and every token. Their status cannot be determined; it is inherently ambiguous. The closest we can come is at the aggregate level, by placing the ambiguous cases in the context of the entire system. We know that nouns are borrowed in all the other cases, and so it is safe to say that some proportion of the single-noun subjects must have been produced by the same process, though which particular tokens cannot be singled out. Using an appropriate sampling scheme (see note 8, for problems with the present corpus), we could even estimate what proportion.

PROPOSITIONAL COMPLEMENTS

The equivalence constraint is not the only logically possible way of assembling alternating monolingual fragments from two languages into coherent sentences. It is basically a linearization constraint that can be seen to operate on the level of production, allowing speakers to concatenate successive fragments without compromising the internal grammaticality of each fragment, without having to repeat the same information in both fragments (see the section on palindromic switches), and without skipping any essential information. An alternative way of constraining code-switching is through expanding the phrase structure rules of one or both languages to permit a change of code in one or both directions at some subset of the nonterminal nodes of the phrase structure.¹⁰ This proposal, which is implicit in many of the articles cited at the beginning of this article, has been clearly formalized by Joshi (1985). In applying this formulation to the data on English-origin nouns in Tamil object position, the change of code would have to be restricted to the lowest NP nodes. One of the difficulties is that the English NP thus created would still have to take Tamil inflections (see note 6). Another is that it would not account for code-switches between full or pronominal subject noun phrases and the verb; these would require code changes at higher level NP nodes, changes that somehow have to be prohibited for object NPs. In any case, introducing this single type of code change in phrase structure is but an alternate notation for the nonce loan process as it affects Tamil object position, unless it can be shown to occur at a variety of other phrase structure nodes.

In the previous sections, we have discussed how the equivalence constraint acts to prohibit or allow switches at the level of main clause constituents. We will now deal with a specific type of switch at the interclausal level where this constraint clearly does not apply in our data. Indeed, it appears to be a situation better described by a code change at the level of an embedded sentence node in phrase structure. Direct and indirect quotations in Tamil and

other propositional verb complements are marked by the obligatory complementizer particle *-nu* affixed to the end of the complement and preceding the head verb. In Tamil-English bilingual syntax, the equivalence constraint would predict no switching between the verb and the complement, because both the verb and any complementizer in spoken English precede the complement. Nevertheless, we find cases like (45) and (46), which contravene Tamil patterns, both with respect to word order and the lack of the particle.

- (45) They say that if the boy's side say "*ceri, paakka veeNdaam*"
 okay see (inf) want (fut-neg)

then we will say "*ceri, paakka veeNdaam*"
 okay see (inf) want (fut-neg)

They say that if the boy's side says, "okay, we don't want to see," then we will say, "okay, we won't see." (2752/Su)

- (46) She asked me "*ettanaavatu rank?*"
 which
 She asked me, "which rank?" (2924/Su)

Perhaps even more interesting, we find many cases where it is the English sentence that is embedded and Tamil, not English, rules are followed.

- (47) Even there, I am really lucky-*nu collaNum*
 (that) say must

Even there, one must say that I am really lucky. (817/S)

- (48) The system has completely changed-*nu enakku tooNaratu.*
 (that) I (dat) feel

I feel that the system has completely changed. (743/SD)

- (49) The *iLaniir* is best-*nu Tu, I was sticking onto iLaniir.*
 coconut water that (quot) coconut water

Thinking that the coconut water is best, I was sticking with it. (458/SD)

- (50) It corrodes your confidence-*nu enakku oru* feeling.
 that I (dat) a

I have a feeling that it corrodes your confidence (246/M)

- (51) As long as you are much better than other people-*nu*
 that

collaraaL.

say (3p-sg-fem-pres)

She says that as long as you are much better than the others. (487/SD)

Examples (47)–(51) would seem to be consistent with a constraint often postulated to the effect that a complementizer must be in the language of the matrix verb (di Sciullo et al., 1986; B. Kachru, 1978). However, in systematic studies of language pairs that share the same word order with respect to complementizer placement, it has been shown that no such constraint holds categorically (e.g., Bentahila & Davies, 1983; Sankoff & Poplack, 1981). What is of interest in the given examples is not so much the language or position of the complementizer, which could not be otherwise without de-

stroying the coherence of the sentence, but the very fact that a switch has occurred.

There is strong evidence, however, that these constructions, although not infrequent, are to some degree problematic for the speakers. For it is here where we find many of the instances of palindromic switches. Not being able to satisfy either Tamil or English word orders, the speaker repeats the "offending" material in both languages at the appropriate places.

(52) I would say it is betterment for India-*nu taan colluveen*.

that only say (1p-sg-fut)

I would strongly say that it will be better for India (527/SD)

(53) And he'll tell me then and there, you are no good-*nu colluvaa*.

that tell (3p-pl-fut)

And he'll tell me then and there that you are no good. (930/Sh)

(54) I think it's the European influence-*nu ninaikiren*.

that think (1p-sg-pres)

I think that it's the European influence. (937/Sh)

Nevertheless, it is with propositional complements that we find the only recurrent type of language mixing in our corpus not accounted for by the equivalence constraint or nonce borrowing, but seeming to arise from a process of inserting a constituent from one language into a sentence of the other, where the location of this insertion obeys the rules of the matrix language only. We take up the implications of this finding in the discussion.

PALINDROMIC SWITCHES

Palindromic switches, also known as portmanteau, copy translation, or mirror-image constructions, are widely attested but are inevitably found to occur rarely in quantitative studies. Thus, these seem to constitute an occasional ad hoc production strategy rather than a systematic approach to bilingual sentence construction. In a strict sense, these constructions violate the equivalence constraint in that they involve a syntactic boundary of form AB in language 1 and BA in language 2. From another point of view, however, it seems as if the speakers' strategy is to circumvent the constraint by producing the correct word order according to both grammars and accepting the duplication of structure as the lesser of evils, as in one of the following four configurations:

	<u>AB</u>	A	A	<u>BA</u>	<u>BA</u>	B	B	<u>AB</u>
language	1	2	1	2	2	1	2	1

In our data, speakers used an average of three or four of these constructions over the course of long interviews containing much borrowing and code-

switching. Almost half of the cases involved embedded clauses with the post-posed *-nu* complementizer as discussed in the previous section.

Another recurrent type of palindromic switching involves the copular verb.

- (55) They don't want to be steady-*aa* strong-*aa* *irukkaratu-ille*.
 (adv) (adv) be (inf-neg)
 They don't want to be steady and strong. (493/SD)

The other cases were scattered among a variety of constructions.

- (56) According to the schedule *paDi* *oNNutaan irukkaNum*.
 according to one only be must
 According to the schedule, there must be only one. (3943/Su)
- (57) They gave me a research grant *koDutaa*.
 gave (3p-pl-past)
 They gave me a research grant. (2854/Sh)
- (58) I was talking to *oru orutanooDa peesinDu iruntan*.
 one person (com) talk (cont) be (1p-sg-past)
 I was talking to a person. (2854/Sum)
- (59) Just because *avaa innoru* color and race *engindratunaale*.
 they different of-because
 Just because they are of different color and race. (6221/B)
- (60) Even MBA-*kku kuuDa irukku*.
 (dat) even be (3p-neut-past)
 Even in the MBA, it is there. (718/SD)

DISCUSSION

The results of any study of the constraints on code-switching or on the characteristics of nonce loans are extremely sensitive to the criteria used to distinguish these two phenomena. We cannot falsify the prohibitions against switches at certain points nor can we substantiate positive predictions about switches at other points, unless we can prove that the attesting examples are not borrowings. Conversely, we cannot generalize about loanwords and how they are integrated into the host language if our data do not strictly exclude code-switched material.

We have proposed and tested a consistent framework for categorizing code-switches and borrowings, making use of morphological and syntactic criteria and quantitative methods. To operationalize these general criteria, we draw on the particular properties of Tamil and English. These include the case inflections of Tamil diagnostic of borrowing when affixed to an English-origin word; the demonstratives, quantifiers, articles, and pronouns of both languages, which are not borrowed and which thus enable us to determine the language of the noun phrase in many instances; and the use of Tamil proverbs whose presence or absence allows us to determine for every English-

origin verb whether it is acting as a nonce borrowing into Tamil or as part of a code-switch or a monolingual stretch in English.

These indices, however, are not always sufficient to distinguish between loans and switches. Often a particular noun phrase in the data could have been produced by either process, such as for English-origin nominatives or the variably null-marked accusatives or genitives. It is to handle these cases that we resort to quantitative methods, comparing variable inflection on candidates for nonce loan status with that on native Tamil nouns. Thus, if the distribution of absent case marking on English-origin noun phrases quantitatively parallels that on native Tamil noun phrases, we are compelled to recognize even the unmarked tokens as nonce loans. We could not of course do this with any individual token if we did not have the weight of statistical evidence pointing in this direction.

In this article, we have made no operational effort to distinguish between nonce borrowings and established loans, because our point is that there is no difference between them with respect to their morphological and syntactic integration into host language contexts. Let us enumerate the independent lines of evidence bearing on the nonce loan hypothesis.

By itself, the fact that this hypothesis accounts for the apparent counterexamples to the free morpheme constraint – words with “mixed morphology,” where the free morpheme is not necessarily phonologically integrated into the other language or widely known – in this article and in most of the literature on the subject might be considered an ad hoc and circular attempt to “rescue” this constraint (e.g., Eliasson, 1989). The category of nonce loans, however, simultaneously satisfying morphological and syntactic rules of the host language, has been well established as one extreme of a gradient (with well-attested loans at the other extreme) in a quantitative study of a massive corpus where intrasentential *multiword* code-switching is very rare (Poplack, Sankoff, & Miller, 1988). Moreover, it is the most economical way of accounting for the striking regularity of Tamil morphology on *preverbal* English-origin objects and English morphology, to the complete exclusion of Tamil inflections, on *postverbal* ones.

Independent of morphology, the category of nonce loans also accounts for the numerous English-origin object + Tamil verb and English verb + Tamil-origin object sequences in our corpus, which would otherwise contradict the equivalence constraint.

The third line of evidence is that whereas there are postverbal Tamil objects of English verbs, as would be expected through nonce borrowing from Tamil into English, there are no preverbal ones, because this is not a slot for English objects and hence not for borrowings, either. Note that where borrowings cannot contribute apparent counterexamples, the equivalence constraint applies unequivocally. Similarly, postverbal English-origin objects of Tamil verbs occur rarely and, for accusatives, apparently only as the result of postposition, the same as for native Tamil postverbal direct objects. Again, the frequency of English-origin object + Tamil verb and the absence of underlying Tamil verb + English-origin direct object constructions are ev-

idence that there is no code-switching between verb and object, just as predicted by the equivalence constraint, only borrowing into the predicted slots and at the predicted rates.

The remaining lines of evidence do not pertain directly to the code-switching constraints, but rather, support the loanword character of the English-origin objects. For example, the fourth type of evidence is the lexical nature of the English-origin material in Tamil object position. This is almost exclusively nominal, with a few idiomatic adjective plus noun forms, and is never pronominal, in contrast with Tamil objects, which are 55% pronominal for direct objects and 70% pronominal in dative contexts. This distributional contrast parallels what we would expect from established loanwords, as no English pronouns are attested as borrowings into Tamil; indeed pronominal borrowings are extremely rare in most bilingual situations.

Fifth, the English-origin direct objects are single nouns without English determiners, demonstratives, or modifiers preceding them and are followed neither by English relative clauses nor by English prepositional noun complements. They participate in Tamil constituent structure in the same way as do native Tamil nouns and established loanwords and differ from the unambiguous code-switches in our data, which typically contain several English words, including the full range of function words, modifiers, relatives, and complements.

Sixth, the distribution of English-origin material among syntactic classes—optional versus obligatory marking contexts in the accusative; and subject, verb object, and governed by postpositions for the dative—quantitatively parallels that of native Tamil forms. This is to be expected of borrowings, but were these code-switches, only very speculative universal considerations could account for the quantitative parallelism.

Seventh, the quantitative rate of case marking of the English-origin objects closely parallels that of the native Tamil objects, both with respect to the optional versus obligatory accusative contexts and also when the accusative is compared with the dative.

In summary, not only does the nonce loan hypothesis, when combined with the equivalence constraint, economically account for the Tamil-English data, but many independent lines of evidence confirm that the verb objects we have studied behave morphologically and syntactically exactly as do established borrowings and native Tamil forms. The only possible differences between them concern degree of phonological integration and assimilation into the monolingual lexicon. These are basically irrelevant to their status as loanwords: bilingual speakers may manifest phonological variability in both nonce and established loanwords, and their lexical access is limited to a closed set of accepted loanwords.

A few anomalous tokens do not impeach these results. Rare English-origin postverbal direct objects of Tamil verbs are compatible with the same construction occurring occasionally in our monolingual Tamil data. English plural inflections on some nouns that we count as borrowings have counterparts in other bilingual contexts and can be understood in terms of the highly

divergent ways of expressing plurality in English and Tamil. Unmarked English-origin direct objects in obligatory marking contexts occur at a low rate as does the palindromic strategy of circumventing the equivalence constraint.

Although we have been focusing on word order in the main clause, our methods allow us to characterize borrowing more generally. Thus, the following sentences can be seen to consist entirely of nonce borrowing of content words from English, with Tamil morphology and syntax:

- (61) Religion-*uDaya* main purpose *vantu oru* supernatural being-*la oru*
 (gen) (filler) a (loc) a
 belief create *paNNaratu*.
 do (inf)

Religion's main purpose is to create a belief in a supernatural being. (242/M)

- (62) Normal-*aave* people *vantu* generalize *paNNuvaa*.
 (adv) (emph) (filler) do (3p-pl-fut)

Normally people will generalize. (287/M)

- (63) Parents-*nuDaya* support *irukkum* arranged marriage-*la*.
 (gen) be (3p-neut-fut) (loc)

In the arranged marriage, there will be parents' support. (840/SD)

Note that when nonce borrowings from a donor language can be detected, then the sentence or constituent into which they are borrowed must necessarily be identified as belonging to the host language. This may explain why students of bilingual syntax in SVO-SOV situations such as Marathi-English (Joshi, 1985), Japanese-English (Nishimura, 1985), and Kannada-English (Sridhar & Sridhar, 1980) have been led to claim that each "code-switched" sentence is basically in one language or another. In fact, most of the L2 attestations are not unambiguous switches but, most probably, nonce borrowings. In contrast, in the Spanish-English (SVO-SVO) literature, where true code-switches can occur at many more syntactic boundaries, it has become clear that such sentences need not be assignable to one language or the other (Sankoff & Poplack, 1981; Woolford, 1983).¹¹

Recognition of nonce borrowing as a well-structured process in sentences like these not only removes a great deal of confusion from the data on bilingual syntax but belies notions of a "non-systematic and improvisatory. . . mixed language . . . a very makeshift kind of device that defies analysis . . . a humiliating reminder of the plight of Tamil during colonial times" (Kandiah, 1978:66).

Having made this case, we emphasize that the goal in this study has not been to rescue the code-switching constraints. Indeed, being accountable to our data has required documenting such phenomena as palindromic switches and the insertion of propositional complements before *-nu* + head verb, neither of which is predicted by the equivalence constraint. Moreover, the nonce loan hypothesis is not a "patch" on the free morpheme constraint. In the light of new studies on the nature of loanword integration and the availability to

bilinguals of the entire content-word lexicon for purposes of borrowing, we cannot rely only on phonological and sociological characteristics of borrowing to verify the loanword status of a free morpheme bound to a morpheme of the other language. The nonce loan hypothesis adds to these characteristics a wide variety of diagnostics involving morphological, syntactic, lexical, semantic, and distributional properties of loanwords. Our basic point has been rather that the single-word examples offered up repeatedly in the literature as counterexamples to one or the other constraint are simply irrelevant without systematic corpus-based assessments of whether they are, at least in the aggregate, borrowings or switches.

We can conclude that most of the evidence that purports to refute the equivalence or the free morpheme constraints on the basis of single-word switches does not necessarily bear on these claims. In order for them to be pertinent, it would be necessary to first establish, using the type of independent criteria we have presented, that they are not nonce loans. Barring this demonstration, the most that can be said of such data is that they are ambiguous. Indeed, the free morpheme constraint should be seen as a preliminary formulation of the claim that nonce loans and single-word switches may be distinguished, if not for every token, at least statistically by their distribution and by whether or not they tend to be syntactically and morphologically integrated, proving that they are produced by different processes. The nonce loan hypothesis, which basically states no more than that borrowing, whether nonce or established, is a phenomenon of language mixture distinct from code-switching and is operationally distinguishable as such, at least at the aggregate level, adds to the predictions of the free morpheme constraint in its most important context, the occurrence of single words from one language in a sentence or multiword fragment of the other language.

There is another, less important, aspect to the free morpheme constraint as originally formulated, namely that switches should not occur across a morpheme boundary where one of the morphemes is at the end of a clearly monolingual multiword fragment in one language and is bound to another at the beginning of a clearly monolingual multiword fragment in the other language. Were many such switches to occur, given the multiword context of the free morpheme in its original language, there would be less justification in regarding it as a borrowing, and the free morpheme constraint would be falsified. Even though we are aware of no context where counterexamples occur systematically, we do not consider this to be a major finding. The constraint was basically intended to account for the frequent occurrence of single lexical items from one language preceded and followed by material from the other.

On the other hand, the well-motivated removal of nonce borrowings from consideration as code-switches permits us to focus the study of the validity or limitations of the equivalence constraint on those constructions that are directly pertinent. Are there contexts in which palindromic constructions characterize a significant number of switches? Are the equivalence-violating

switches of propositional complements representative of some more general phenomenon? As we have mentioned, recourse to palindromic switching has never been shown to be a quantitatively important effect. On the other hand, there is a well-documented situation, Arabic-French bilingualism in Morocco, in which an entire French NP can be inserted in any slot for NPs in an Arabic sentence, including some that do not occur in French (Nait M'Barek & Sankoff, 1988). This insertion of constituents, which is the quantitatively dominant aspect of bilingual syntax among the Moroccan speakers of French and Arabic, bears some resemblance to the embedding of an English propositional complement in a Tamil matrix. It is the only other recurrent kind of language mixing in our series of studies that is best modeled as a change of code at a node in phrase structure, in this instance, at a high-level NP node. It is extremely specific, however, not occurring at appreciable rates in Arabic-French bilingual communities originating outside the Maghreb, nor among bilinguals in Moroccan Arabic and Spanish, English (Sankoff & Nait M'Barek, 1990), or Dutch (Nortier, 1989).

Whatever the sociolinguistic and psycholinguistic forces that give rise to the mixing of elements of both codes in bilingual discourse, it is clear that the problem of word-order discrepancies between the two languages must be resolved at some level. The palindromic-switch strategy would be one solution, but as we have found, speakers do not use this on a regular basis. The equivalence constraint is another solution, one which seems to hold in the data sets that have been examined systematically. In cases such as Tamil-English bilingualism, however, the constraint prohibits switches from occurring at most syntactic boundaries. Constituent insertion or the change of code at a node in phrase structure has been empirically well documented in only the Tamil-English and the Arabic-French studies, for only one type of constituent in each case. Thus, most of the mixing of the two codes in this type of situation, which may be considerable (as in 61-63), could well be expected to, and in fact does, come about through nonce borrowing.

Although, as we have seen, it is often difficult to distinguish methodologically between different bilingual phenomena, our results bolster the conceptual distinction between code-switching and borrowing. That nonce borrowing has access to all the content words in the lexicon of the donor language, in contrast to the restricted stock of established loans, does not mean that it is an intermediate process situated somewhere between code-switching and borrowing in the more traditional sense. The morphological and syntactic roles of nonce loans are identical to those of established loans, and as processes, the two of them contrast sharply with code-switching. Indeed, if there is a conceptual distinction that cannot be sharply drawn, it is that between the two types of borrowing. From the use of English technical vocabulary in the discourse of educated bilinguals, to the speech of monolinguals incorporating well-established loanwords, there is a continuum of possibilities, distinguished only by gradations in dialectal distribution patterns, usage frequencies, and acceptability.

NOTES

1. Variants of this principle were enunciated as well by Pfaff (1979), Lipski (1978), and others. For a fully formalized version, see Sankoff and Mainville (1986a, 1986b).

2. See Sridhar and Sridhar (1980), B. Kachru (1978), and Nishimura (1985) for other studies of language mixing involving English and SOV languages.

3. Conversely, the equivalence constraint allows code-switching between the subject and the verb because both languages are subject-initial. Partly because the nominative case takes null marking, however, there are more severe difficulties for subjects than for objects in distinguishing borrowing from switching. We return to this problem later.

4. A small number (15) of direct object constructions, containing either Tamil or English-origin words, that *cannot* take the marker (cf. Asher 1982:105–106) were also noted but are not discussed here as they provide no evidence for or against our hypothesis.

5. The transcription system is based on that of Annamalai (1975): t: +anterior; T: –anterior; L: –anterior; R: +retroflex; C: –voice, +affricate; j: +voice, +affricate; N: –anterior; sh: palatal fricative. Other abbreviations: inf, fut, pres, cont, neg, refl: infinitive, future, present, continuative, negative, reflexive; 1, 2, or 3p-sg or pl, masc, fem, or neut: (first, second, or third) person singular or plural, masculine, feminine, neuter; acc: accusative, dat: dative, loc: locative, gen: genitive, com: comitative; emph: emphatic; adv: adverb; quot: quotative.

Citations contain line numbers in the bilingual corpus, followed by speakers' initials.

6. This absence is not restricted to the variable omission of initial determiner characteristic of varieties of Indian English but precludes all English function words preceding or following the noun, as well as English relative clauses or prepositional phrases complementing the noun. If there had been more elaborate English NPs (without determiners) acting as preverbal direct objects in the corpus, their presence together with objects containing Tamil determiners such as in (8)–(10) might have suggested that both types of object, together with all the single-noun English-origin direct objects, represent code-switches *within* the NP, between the Tamil determiner or demonstrative (usually, but not always—as in examples 8–10—a null determiner) and the English noun. (Note that this would assume that any case marker postposed to the English-origin fragment would have to have been generated in the phrase structure.) In this case, there would be no violation of the equivalence constraint, because the determiner + noun order is the same in both languages. We do not pursue this line of explanation here, however, because it does not account for the highly restricted nature of English-origin objects—largely single nouns—that the nonce borrowing hypothesis explains well.

7. Compounding of verb and object can occur in Tamil, in which case there is no accusative marking. Certain of the unmarked English-origin noun + English-origin verb + *paNnu* sequences might be explained in this way as compound borrowings. However, there is no reason to believe that this would account for a larger proportion of unmarked English-origin forms than would Tamil noun + Tamil verb compounding.

8. Note that English-origin material occurs somewhat less often in dative than in accusative contexts—53% versus 67% when Tamil pronouns are not counted in the totals. However, the way our bilingual corpus was sampled from the larger corpus (which contains mostly monolingual discourse) precludes any interpretation of this difference in distribution. Sentences sampled because they contain an English-origin object in one case often contain a Tamil object in the other case. Since there are more English-origin accusatives than datives, this reduces the overall proportion of English-origin datives.

9. Though the differences are not significant, the marking on English-origin material does seem to occur at a slightly lower rate in all the contexts we have examined quantitatively. We do not attempt here to account for these small discrepancies, which may well be associated with production phenomena such as hesitation, repetition, or flagging in the vicinity of the English-origin material. A detailed analysis of this type has been carried out in the context of Finnish-English bilingualism by Poplack, Wheeler, and Westwood (1989).

10. The equivalence constraint may also be formulated in this general way, but this requires expanding the set of grammatical categories through the use of superscripts that encode information about linearly adjacent sister constituents (Sankoff & Mainville 1986a, 1986b).

11. Chana and Romaine (1984) also recognized the difficulty of assigning a base language to Panjabi-English code-switched discourse.

REFERENCES

- Annamalai, E. (1971). Lexical insertion in a mixed language. *Papers from the Seventh Regional Meeting, Chicago Linguistic Society*. 20-27.
- (1975). Phonology of Tamil nouns. In H. R. Schiffman & C. M. Eastman (eds.), *Dravidian phonological systems*. Seattle: University of Washington Press. 122-179.
- Asher, R. E. (1982). *Tamil* (Lingua Descriptive Studies, Volume 7). Amsterdam: North-Holland.
- Bentahila, A., & Davies, E. (1983). The syntax of Arabic-French code-switching. *Lingua* 59:301-330.
- Berk-Seligson, S. (1986). Linguistic constraints on intrasentential code-switching: A study of Spanish/Hebrew bilingualism. *Language* 15:313-348.
- Boeschoten, H., & Verhoeven, L. (1985). Integration niederlandischer lexikalischer Elemente ins Türkische: Sprachmischung bei Immigranten der ersten und zweiten Generation. *Linguistische Berichte* 98:347-364.
- Bokamba, E. G. (1989). Are there syntactic constraints on code-mixing? *World Englishes* 8:277-293.
- Bynon, T. (1978). *Historical linguistics*. Cambridge: Cambridge University Press.
- Chana, V., & Romaine, S. (1984). Evaluative reactions to Panjabi/English code-switching. *Journal of Multilingual and Multicultural Development* 5:447-473.
- di Sciullo, A.-M., Muysken, P., & Singh, R. (1986). Government and code-mixing. *Journal of Linguistics* 22:1-24.
- Eliasson, S. (1989). English-Maori language contact: Code-switching and the free-morpheme constraint. *Reports from Uppsala University Department of Linguistics* 18.
- Gumperz, J. (1982). Conversational code-switching. In J. Gumperz (ed.), *Discourse strategies*. Cambridge: Cambridge University Press. 59-99.
- Hasselmo, N. (1972). Code-switching as ordered selection. In E. Finchow et al. (eds.), *Studies for Einar Haugen*. The Hague: Mouton. 261-280.
- Joshi, A. K. (1985). Processing of sentences with intrasentential code-switching. In D. Dowty et al. (eds.), *Natural language processing: Psychological, computational, and theoretical perspectives*. Cambridge: Cambridge University Press. 190-205.
- Kachru, B. (1978). Toward structuring code-mixing: An Indian perspective. *International Journal of the Sociology of Language* 16:28-46.
- Kachru, Y. (1968). *An introduction to Hindi syntax*. Urbana: University of Illinois, Department of Linguistics.
- Kandiah, T. (1978). Standard language and socio-historical parameters: Standard Lankan Tamil. *International Journal of the Sociology of Language* 16:59-75.
- Lehmann, W. (1973). *Historical linguistics*. New York: Holt, Rinehart and Winston.
- Lipski, J. M. (1978). Code-switching and the problem of bilingual competence. In M. Paradis (ed.), *Aspects of bilingualism*. Columbia, SC: Hornbeam. 250-264.
- Madaki, R. O. (1983). *A linguistic and pragmatic analysis of Hausa-English code-switching*. Doctoral dissertation, University of Michigan.
- Mougeon, R., Beniak, E., & Valois, D. (1985). *Variation in the phonological integration of loanwords in a bilingual speech community* [Working paper]. Toronto: Centre for Franco-Ontarian Studies.
- Muysken, P. (1990a). Concepts, methodology and data in language contact research: Ten remarks from the perspective of grammatical theory. In *Papers for the European Science Foundation Network on concepts, methodology and data*. Strasbourg: European Science Foundation. 15-30.
- (1990b). *The intersection of the equivalence and government constraints*. Paper presented at the University of Amsterdam workshop on code-switching, May 1990.
- Myers-Scotton, C., & Azuma, S. (forthcoming). A frame-based model of code-switching. In K. Deaton et al. (eds.), *Papers from the 26th Chicago Linguistic Society Annual Meeting*. Chicago: Chicago Linguistic Society.
- Nait M'Barek, M., & Sankoff, D. (1988). Le discours mixte arabe/français: des emprunts ou des alternances de langue? *Revue canadienne de linguistique* 33:143-154.
- Nishimura, M. (1985). *Intrasentential code-switching in Japanese and English*. Doctoral dissertation, University of Pennsylvania.
- Nortier, J. (1989). *Dutch and Moroccan Arabic in contact: Code-switching among Moroccans in the Netherlands*. Doctoral dissertation, University of Amsterdam.

- Pandharipande, R. (1986). *Formal and functional constraints on code-mixing*. New Delhi: International Congress of Sociology.
- Pfaff, C. (1979). Constraints on language mixing. *Language* 55:291-318.
- Pintzuk, S. and Kroch, A. (1989). The rightward movement of complements and adjuncts in the Old English of *Beowulf*. *Language Variation and Change* 1:115-143.
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish *y termino en español*. *Linguistics* 18:581-618.
- _____. (1981). Syntactic structure and social function of code-switching. In R. Duran (ed.), *Latino discourse and communicative behavior*. Norwood, NJ: Ablex. 169-184.
- _____. (1985). Contrasting patterns of code-switching in two communities. In H. Warkentyne (ed.), *Methods V: Papers from the Fifth International Conference on Methods in Dialectology*. Victoria: University of Victoria Press. 363-386.
- _____. (1988). Language status and language accommodation along a linguistic border. In P. Lowenberg (ed.), *Georgetown University Round Table on Languages and Linguistics 1987*. Washington, DC: Georgetown University Press. 90-118.
- Poplack, S., & Sankoff, D. (1984). Borrowing: The synchrony of integration. *Linguistics* 22:99-135.
- Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26:47-104.
- Poplack, S., Sankoff, D., & Pousada, A. (1982). Competing influences on gender assignment: Variable process, stable outcome. *Lingua* 57:1-28.
- Poplack, S., Wheeler, S., & Westwood, A. (1989). Distinguishing language contact phenomena: Evidence from Finnish-English bilingualism. *World Englishes* 8:389-406.
- Prince, E. F., & Pintzuk, S. (1986). *Bilingual code-switching and the open/closed class distinction*. Unpublished manuscript.
- Rivas, A. (1981). *On the application of transformations to bilingual sentences* (UMass Working Papers on Hispanic Linguistics and Bilingualism). Amherst: University of Massachusetts, Department of Spanish and Portuguese.
- Romaine, S. (1986). The syntax and semantics of the code-mixed compound verb in Panjabi/English bilingual discourse. In D. Tannen & J. Alatis (eds.), *Languages and linguistics: The interdependence of theory, data and applications*. Washington, DC: Georgetown University Press. 35-50.
- _____. (1989). *Bilingualism*. Oxford: Basil Blackwell.
- Sankoff, D., & Mainville, S. (1986a). Code-switching of context-free grammars. *Theoretical Linguistics* 13:75-90.
- _____. (1986b). Un modèle pour l'alternance de langue sous la contrainte d'équivalence. *Revue Québécoise de Linguistique* 15:233-246.
- Sankoff, D., & Nait M'Barek, M. (1990). *Dramatically contrasting language mixture strategies in two communities of fluent Arabic-French bilinguals*. Paper presented at NWAVE 19, Philadelphia, October 1990.
- Sankoff, D., & Poplack, S. (1981). A formal grammar for code-switching. *Papers in Linguistics* 14:13-46.
- Smith-Stark, T. (1976). Jilotepequeño Pocomam texts. In L. Furbee-Losee (ed.), *Mayan Texts. International Journal of American Linguistics* (Native American Text Series 1:72-87).
- Sridhar, S., & Sridhar, K. (1980). The syntax and psycholinguistics of bilingual code-mixing. *Canadian Journal of Psychology* 34:407-416.
- Treffers, J. (1990). *Towards a uniform approach to code-switching and borrowing*. Paper presented at the European Science Foundation workshop on constraints, conditions and models in code-switching. London, September 1990.
- Weinreich, U. (1953). *Languages in contact*. The Hague: Mouton.
- Woolford, E. (1983). Bilingual code-switching and syntactic theory. *Linguistic Inquiry* 14:520-536.
- Zentella, Ana C. (1982). Spanish and English in contact in the United States: The Puerto Rican experience. *Word* 33:41-57.
- Zvelebil, K. V. (1983). Word-borrowing and word-making in modern South Asian languages: Tamil. In I. Fodor & C. Hagège (eds.), *Language reform, vol. III*. Hamburg: Buske Verlag. 431-440.