
TOPICS IN LANGUAGE AND LINGUISTICS

Series Editors:
Albert Valdman and Thomas A. Sebeok
Indiana University, Bloomington, Indiana

ISSUES IN INTERNATIONAL BILINGUAL EDUCATION:
The Role of the Vernacular
Edited by Beverly Hartford, Albert Valdman, and Charles R. Foster
Issues in
INTERNATIONAL
BILINGUAL EDUCATION
The Role of the Vernacular

Edited by
BEVERLY HARTFORD
and
ALBERT VALDMAN
Indiana University
Bloomington, Indiana

and
CHARLES R. FOSTER
U.S. Department of Education
Washington, D.C.

PLENUM PRESS • NEW YORK AND LONDON
BILINGUALISM AND THE VERNACULAR

SHANA POPLACK

THE VERNACULAR/STANDARD DISTINCTION

It is well known that any large, socially complex speech community will be linguistically differentiated not only along geographical dimensions, as studied by traditional and modern dialectologists, but also according to sociodemographic characteristics of speakers and the contextual features of situations and activities.

The functioning of a hierarchical social system generates simplifications and stereotypes about the linguistic and other behavior of its members at various levels of the hierarchy. The distinction between vernacular and standard is one such simplification. It obscures a number of distinctions, including those between the spoken and written language, informal and formal speech styles, varieties characteristic of working-class speakers and those of bourgeois speakers, naturally acquired versus educated or literary speech, regionalisms and metropolitan varieties, colloquial and ‘correct’ speech, current usage and that codified in grammars and dictionaries.

The work of Labov (1966, 1968, 1970, 1972a, 1972b) is an important milestone in the demonstration of the systematicity of the vernacular and in debunking the supposed communicative superiority of the standard. To arrive at these conclusions, Labov had to resolve the methodological problem of observing the vernacular freed from the hypercorrectively distorting effects of a normatively imposed standard.

Without denying the importance of the Observer's Paradox and the technology developed by Gumperz (1964), Labov (1970), and others to circumvent it, it is also true that the vernacular, in the sense of ordinary, every-day language, not only differs from speaker to speaker and from
context to context, but the situations in which it is unaffected by influences from one or more standard(s), even in the absence of the observer, may be quite rare. This immediately leads to two problems inherent in the idea of isolating the vernacular. First, there are any number of vernaculars in a single speech community, each situation involving linguistic patterns at least marginally different from the next. Second, there can be no unitary standard for the spoken language, at least in any objective, well-defined sense. The written forms of a language do not coincide with upper class speech, nor with that of prestige centers, nor with the speech behavior of literary or learned persons. Norms of formality, correctness and grammaticality are only occasionally reflected in anyone's performance.

Indeed, the standard may be seen as some idealized set of features which do not completely characterize any real situation, though they may be reflected in different proportions in different linguistic varieties, and whose importance is their contrast with features of one or more vernaculars, informal or colloquial, minority or accented, regional or rural, working-class, etc. Note that these varieties are identified with sociological rather than linguistic labels. It is the ill-defined and idealized nature of the standard which is crucial, since it permits the grouping together of all the diverse varieties of upper-class, learned, conventional speech against all other vernaculars, and perpetuates the notion that the distinction is primarily of linguistic origin when it is really a socially-motivated contrast.

We reformulate the problem of isolating the vernacular, then, as one of describing the speech patterns of a community, as opposed to examining the ideas, intuitions and attitudes which constitute the standard.

METHODOLOGICAL CONSIDERATIONS

Labov (1966) has demonstrated the prevalence of categorical perception in the characterization of linguistic phenomena. Whereas individuals may perform with subtle quantitative differences from situation to situation, both as speakers and as listeners, the subtlety of these differences is not readily accessible to conscious intuition. Many such differences are not remarked at all, while those that are, are exaggerated to the status of categorical occurrences. Indeed, this is a major mechanism leading to the stereotypes mentioned earlier. The problem of categorical perception affects not only methodologies which appeal primarily to native speakers' intuitions, it is also a defect of analyses based on casual observations, even over long periods of time, or collections of examples and anecdotes.

This leaves systematic empirical observation of speech in context as the only reliable way to gather information on the vernacular. Because of the
BILINGUALISM AND THE VERNACULAR

multiplicity of vernaculars, i.e. the variability of the vernacular from one context to the next, and because of the quantity of data necessary to characterize adequately even a few of the important patterns observed, this approach leads to masses of data. To account for these data, an analytic framework is needed which is capable of reflecting the range of differences between contexts in concise and accurate terms. This is the motivation for the development of variation theory.

The internal differentiation of a speech community is not one of some structure or usage being present in one context and absent in a closely related one. Rather it is a case of more or less. In a given situation a speaker uses the available sentence structures in a certain proportion, she chooses appropriate lexical items with specific frequencies, and applies phonological reduction rules at a certain rate. In a somewhat different situation all the frequencies and rates will be slightly different and the differences will tend to be systematic from speaker to speaker. Of course, in very different contexts there may be some more categorical distinctions, but these are relatively rare. A combination of several quantitative changes suffices to make one speech pattern very different from another. To document the differences then, careful counts must be made of various usage patterns, taking into account not only the occurrence of given phenomena, but also every time they did not occur when they might have.

To incorporate the results of these analyses into a theory of linguistic variation, sociolinguists have proposed a number of ways of generalizing the structures of formal linguistics to account for speech performance. These include probabilistic grammars, implicational scales, variable rules, and probabilized lexicons, among others.

While the statistical analysis may be integrated into the linguistic description, it also provides the basis for correlation of community speech behavior with extralinguistic factors. Correlations of frequencies with sociodemographic factors provide clear indices of the linguistic stratification of a community, and the comparison of quantitative results from one situation to another provides an objective taxonomy of linguistically relevant distinctions and similarities among social contexts.

THE SOCIAL IMPLICATIONS OF THE RESEARCH FRAMEWORK

Studies of minority or stigmatized dialects using qualitative or unsystematic methodologies have done a great deal to contribute to the existing ideologies shared by speakers, observers, and all too often, academics, about vernaculars, in particular urban vernaculars. Thus, in a study of the pronunciation of English in New York City, Hubbell (cited in
Labov 1966:36) concludes that 'the pronunciation of a very large number of New Yorkers exhibits a pattern...that might most accurately be described as the complete absence of any pattern'. Joseph Matluk says of Puerto Rico that there are few norms of standard language, as Puerto Ricans do not recognize the necessity for them. Because of this, there is no linguistic consciousness whatsoever in the schools, without which it is impossible to fight the pressure of English and the progressive deterioration of Spanish syntax on the Island (1961:342).

Examples such as these could be multiplied for all dialects spoken by any population large or visible enough to attract the attention of researchers. More distressing is that these 'findings' about language, particularly those spoken by minorities, are then translated into appreciations shared by the public and the speakers themselves. It is well-known and no cause for surprise that New Yorkers, Blacks, Hispanics, Montreal French speakers, etc. have a high index of linguistic insecurity.

Misdirected educational policies, bias in employment and social situations, which are all too often the lot of monolingual speakers of vernaculars, are compounded in the case of bilingual individuals and communities. Bilinguals have two languages to contend with, and in urban United States settings, they are generally both considered nonstandard varieties. Whereas the monolingual's vernacular may be characterized as degenerate, there is an added dimension to the study of the bilingual's vernacular: the effects of the contact situation. In a well-established tradition of studies in bilingualism, researchers have claimed over and over that the presence of this or that feature is due to contact and convergence with the language of the majority. To exemplify with Spanish speakers in the United States, prime targets for this sort of study, it has been claimed that their Spanish is losing its gender system (Barkin 1980), that it is losing the subjunctive (Floyd 1979) and other tense/mood distinctions (Klein 1976), that through the intermediary of code-switching Spanish syntax is gradually being replaced by English syntax (Urciuoli 1980), that borrowings are impoverishing the language and the culture (Matluk 1961; Varo 1971) —all purportedly due to contact with English.

Systematic quantitative analysis of monolingual vernaculars has succeeded in discovering regularity and stability where other methodologies could only detect randomness and deterioration. Thus Labov's work on Black English in New York City (1968, 1972b) was replicated in other urban centers (e.g. Mitchell-Kernan 1969; Wolfram 1969; Legum et al. 1971; Baugh 1979). Studies on Montreal French (e.g. G. Sankoff and Vincent 1977) are comparable with quantitative studies of other varieties (Pohl 1975; Ashby 1976). Early work on Caribbean dialects of Spanish (Ma and Herasimchuk 1968) has been replicated again and again by later studies (e.g. Cedergren 1973; Terrell 1975, 1979a, b).
BILINGUALISM AND THE VERNACULAR

This tradition of scientific investigation of the spoken language has had relatively little impact on bilingual studies. The kinds of stereotypes and bias which have been at least in part dismantled by monolingual vernacular studies are still widely held with respect to bilingual communities.

A VARIATIONIST APPROACH TO BILINGUALISM

To exemplify the role of systematic investigation of linguistic variation in bilingual studies, we examine in ensuing sections the notion of convergence, as it applies to various aspects of Puerto Rican Spanish spoken in East Harlem, New York. Bilingualism has been said to be the major determinant of linguistic convergence, which usually involves the lexicon, but may also affect the grammar. In fact, language contact can cause such far-reaching changes as to modify the structural type of a language (Weinreich 1953; Gumperz and Wilson 1971).

The studies described below focus on the linguistic aspects of an interdisciplinary study of language use on a block in El Barrio, perhaps the oldest, and until recently, the largest Puerto Rican community outside of Puerto Rico (Language Policy Task Force 1980).

Over three years of participant observation of 102nd Street (Pedraza 1979) has indicated that although there are certain interactional norms which guide language choice, and which are instrumental in maintaining Spanish, neither Spanish nor English is used exclusively in any setting. This lack of functional separation of the two languages is at least in part responsible for the widespread use of code-switching as a norm of interaction on 102nd Street.

In addition to the participant observation, a detailed language attitude questionnaire was administered to 91 speakers representative of the different social networks on the block (Attinasi 1979). Among the results which concern us here was a near consensus on the attitudes that Spanish is not well regarded by the American community at large, but that it should nevertheless be kept alive in the Puerto Rican community in New York.

Most respondents claimed to speak ‘good Spanish’ regardless of their reported language dominance. Indeed, when asked to rate their Spanish competence on a seven point scale, the majority rated themselves as ‘perfect’ or ‘excellent’. ‘Good Spanish’ was described in a variety of ways, with the most frequently recurring characterizations being good vocabulary and pronunciation. Very few speakers pointed to grammatical correctness as an identifying feature of ‘good Spanish’. When asked who could be considered to speak Spanish well, most respondents cited older people, who in this community are mainly Spanish-dominant speakers. This pattern of
responses indicates a strong awareness of a Puerto Rican Spanish norm, distinct from that of Castilian Spanish. As we shall see, however, it is precisely the areas of vocabulary and pronunciation in which Puerto Rican Spanish speech most diverges from other dialects.

Given the particular configuration of use of both Spanish and English in this stable bilingual community just described, the positive attitudes expressed by speakers towards both, and the generally negative attitudes of noncommunity members towards language varieties used by Puerto Ricans in New York, we may now turn to some properly linguistic questions. What sorts of changes, if any, come about when two languages are in contact, a situation which has been hypothesized to accelerate, or even cause, linguistic change? In the case of the 'mother tongue' or Spanish, in particular, it has been claimed that lack of exposure to a 'standard' caused by transplanting this language and culture into a new environment, must with time, cause the language to diverge farther and farther from the standard until communication with monolingual speakers is endangered.

To address these issues, the sociolinguistic component of our research has examined the morphophonological, syntactic and semantic levels of linguistic structure. The studies reported here were drawn from data collected from a sample of 20 speakers interviewed on the block. These were chosen on the basis of their representativeness with regard to several parameters, including age, sex, participation in key social networks, age of arrival to the United States, age of acquisition of English, and most importantly, present reported and observed bilingual ability. About half of the speakers may be characterized as balanced bilinguals; the remainder are Spanish-dominant. The sample was constructed in this way under the assumption that influence from English would be more apparent in the speech of those who report and are observed to use it as much as Spanish.

We now review the findings of a series of quantitative studies of language use in a bilingual setting, which all point most strikingly to a single result—the integrity of the Spanish language at its deeper levels despite constant contact with English and purported lack of exposure to the 'standard'.

Variable Concord in Plural Marking

The first problem area we discuss is the variable deletion of plural markers (s) and (n).

Standard Spanish marks the plural redundantly across the noun phrase onto each determiner, noun and adjective in the constituent. Plurality is then repeated in the verb phrase, where the verb must agree with its subject in person and number, as in (1).
1. *Tu sabes, los doctores trataron*—*sin operarla, a ver si la podían arreglar.* 'You know, the doctors tried to—without operating, to see if they could fix her up.' (02/12)

This results in a great deal of redundancy in the sentence. Moreover, almost every sentence or its context also contains non-inflectional indicators of plurality—morphological, lexical, syntactic and semantic—so that redundancy is even greater than would appear from the surface.

In a study of some 2,500 sentences consisting of an overt or deleted noun phrase and a verb phrase (Poplack 1980a), we sought to answer two types of questions. The first concerns the nature of patterns of plural marking at the sentence level, and the second, the effects on these patterns of long-term contact with English.

Although English also has a concord rule for plural marking, it is characterized by a lesser amount of surface inflection than the Spanish rule. As may be seen in the translation of (1), English only marks plurality inflectionally on the noun. The standard English and Spanish marking sequences may be schematized as in (2). But because both (s) and (n) are subject to deletion, we could also theoretically obtain, for a Spanish noun phrase with two components, any of the eight sequences listed in (2c), after deletion has applied.

2a. Standard Spanish marking sequence: SSN* (Los doctores trataron.)
2b. Standard English marking sequence: OSØ (The doctors tried.)
2c. Possible Spanish marking sequences after deletion has applied:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S S N</td>
<td>S S Ø</td>
</tr>
<tr>
<td>S Ø N</td>
<td>S Ø Ø</td>
</tr>
<tr>
<td>Ø Ø N</td>
<td>Ø Ø Ø</td>
</tr>
<tr>
<td>Ø S N</td>
<td>Ø S Ø</td>
</tr>
</tbody>
</table>

In contrast to other studies which have examined these segments in isolation (e.g. Cedergren 1973; Terrell 1979a, b; López-Morales 1980; Alba 1980), we decided to study the interaction between (s) and (n) deletion processes in the same sentence, since plural concord in Spanish is really a sentence level phenomenon. The study was undertaken to see whether marker deletion in the noun phrase and marker deletion in the verb phrase specifically constrained each other, or whether the noun phrase and the verb phrase acted independently, subject only to an overall functional constraint against ambiguity.

Given the differences just mentioned between Spanish and English in both amount and place of plural marking, were we to find that balanced bilinguals deleted markers overall more frequently than Spanish-dominant
Table 1. Proportion of deleted and retained plural markers on determiners, nouns and verbs for bilingual and Spanish-dominant speakers.

<table>
<thead>
<tr>
<th>BILINGUAL</th>
<th>SPANISH-DOMINANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determiner</td>
<td>Noun</td>
</tr>
<tr>
<td>Ø</td>
<td>S</td>
</tr>
<tr>
<td>.492</td>
<td>.508</td>
</tr>
</tbody>
</table>

speakers and/or that the sequence of such marking as is present in surface structure resembles the English schema (ØØO), we might hypothesize that these results are due to influence from and convergence with English.

As Table 1 shows, however, this is not the case. Though bilinguals delete more from determiners (.492 vs. .294), they also delete more from nouns, and less from verbs, compared to the Spanish-dominant speakers. In both of these latter cases, this is the opposite of what we would expect from English influence.

We also note from Table 1 that in the noun phrase most plural inflections are deleted. Indeed, in only 39% of all noun phrases was plurality conveyed inflectionally at any site: determiner, adjective, or noun. Although these deletion rates might seem rather high, they are in fact fairly common for Caribbean dialects of Spanish, when compared with the quantitative studies of Cedergren in Panama (1973), of Terrell in Cuba (1979a) and the Dominican Republic (1979b). They are actually lower than the rates found for a functionally monolingual Puerto Rican community in Philadelphia (Poplack 1979a). Verbs, on the other hand, are inflected with one or another phonetic variant of (n) 93% of the time, showing a tendency in Puerto Rican Spanish to mark the plural on the verb and not in the noun phrase, the opposite of what is required for English.

We may next ask whether the verbal plural markers are retained in the same sentences in which the nominal plural markers are lost. Table 2 shows that the opposite effect prevails. It is true that a large proportion of verbal markers is retained (about 86-88%) when markers are deleted from the noun phrase, but this simply reflects the fact that (n) deletion is not very advanced as a phonological process. Indeed, a significantly greater percentage of verbal (n) is retained when the markers are also retained in the noun phrase.  

BILINGUALISM AND THE VERNACULAR

Table 2. Percentage of deleted verbal (n) for different marking patterns in the noun phrase.

<table>
<thead>
<tr>
<th>Noun Phrase Marking Pattern</th>
<th>% Deleted Verbal (n)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>ØO</td>
<td>14%</td>
<td>(13/91)</td>
</tr>
<tr>
<td>Ø</td>
<td>12</td>
<td>(27/233)</td>
</tr>
<tr>
<td>Total for uninflected NPs</td>
<td>12</td>
<td>(40/324)</td>
</tr>
<tr>
<td>S Ø</td>
<td>7</td>
<td>(8/114)</td>
</tr>
<tr>
<td>SS</td>
<td>7</td>
<td>(5/67)</td>
</tr>
<tr>
<td>S</td>
<td>6</td>
<td>(5/90)</td>
</tr>
<tr>
<td>ØS</td>
<td>23</td>
<td>(3/13)</td>
</tr>
<tr>
<td>Total for inflected NPs</td>
<td>7</td>
<td>(21/284)</td>
</tr>
</tbody>
</table>

We may term this a concord effect. It is a weak quantitative version of the categorical normative agreement rule. Moreover, the rule of concord works independently of the fact that a sentence is semantically plural, i.e. either to delete markers or to retain them.

Is there any difference between Spanish-dominant and bilingual speakers? Table 3 gives the observed frequencies and proportions of plural marking sequences in Spanish, along with expected proportions for each group. The expected values are calculated from the information in Table 2, under the hypothesis that deletion at any given position is statistically independent of deletion in any other.

We note first that the differences between observed and expected values depend on whether or not local concord obtains in the noun phrase. For those sequences containing local concord (marked with an asterisk), we observe more sentences than would be expected under the hypothesis that marking in each slot proceeds independently of any other. Where there is no local concord we observe less than would be expected. This indicates that marking of the two components in the noun phrase does not proceed independently.

While there are distinctions between Spanish-dominant and bilingual speakers, these are consequences of differences in predicted values for the two groups, i.e. differences in overall deletion rates for each grammatical category. What are the consequences of the dissimilarities between the two groups? We saw from Table 2 that bilingual speakers show a greater overall tendency to delete markers from both determiners and nouns than the Spanish-dominant speakers and a lesser tendency to delete from verbs. This
Table 3. Proportions of various plural marking patterns on the sentence level, as observed and as predicted under the null hypothesis of no concord and no functional compensation.

<table>
<thead>
<tr>
<th>Marking Patterns for: 2-slot Noun Phrase</th>
<th>BILINGUAL</th>
<th>SPANISH-DOMINANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>VERB</td>
<td>N</td>
</tr>
<tr>
<td>*ØØ</td>
<td>N</td>
<td>54</td>
</tr>
<tr>
<td>S Ø</td>
<td>N</td>
<td>49</td>
</tr>
<tr>
<td>*SS</td>
<td>N</td>
<td>18</td>
</tr>
<tr>
<td>*ØØ</td>
<td>Ø</td>
<td>9</td>
</tr>
<tr>
<td>ØS</td>
<td>N</td>
<td>4</td>
</tr>
<tr>
<td>S Ø</td>
<td>Ø</td>
<td>2</td>
</tr>
<tr>
<td>*SS</td>
<td>Ø</td>
<td>1</td>
</tr>
<tr>
<td>ØS</td>
<td>Ø</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marking Patterns for: 1-slot Noun Phrase</th>
<th>BILINGUAL</th>
<th>SPANISH-DOMINANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>N</td>
<td>98</td>
</tr>
<tr>
<td>S</td>
<td>N</td>
<td>32</td>
</tr>
<tr>
<td>Ø</td>
<td>Ø</td>
<td>6</td>
</tr>
<tr>
<td>S</td>
<td>Ø</td>
<td>3</td>
</tr>
</tbody>
</table>

leads to different preferences between the two groups for the three most important sequences at the top of the Table. Spanish-dominant speakers prefer marking on the determiner, deleting from the noun and marking on the verb (SØN), followed by the standard marking sequence SSN, and then by marking on the verb alone (ØØN). Bilinguals favor verbal marking alone, then marking on the determiner and the verb, and finally, the standard full concord form.

Now, if one result has emerged consistently from different quantitative studies of (s) deletion in Caribbean dialects of Spanish and Brazilian Portuguese, it is that determiners, or elements in the first position in the string, are most conservative with regard to marker retention (Ma and Herasimchuk 1968; Cedergren 1973; Terrell 1975; Guy and Braga 1976; Scherre 1978), although by no means does retention operate categorically here, even in functionally monolingual communities (Poplack 1980b). The behavior of the bilinguals may be explained by the fact that they have generalized the deletion rule, which operates most frequently in nouns and adjectives, to a category, determiner, where its operation has been comparatively infrequent. The differences between expected and observed frequencies are the same for Spanish-dominant and bilingual speakers,
BILINGUALISM AND THE VERNACULAR

indicating that aside from overall deletion rates the two groups behave similarly. In particular, the differences in pattern frequencies cannot be attributed to influence from English; none of the sequences involved is cognate with English. Patterns which would resemble English sequences ØSØ (The doctorš tried.) and SØ (Doctorš tried.) are almost nonexistent, representing less than 1% of the Spanish data for both groups.

Of course, standard English is not the only influence on Puerto Rican speech. It has been observed in this community (Pedraza 1979) and shown in others (Wolfram 1974; Poplack 1978a), that Black English may also influence Puerto Rican speech. The concord rule in Black English is also variable, largely due to -s deletion and hypercorrect -s reinsertion on both nouns and verbs. However, the outcomes of these processes are not directly comparable. Black English hypercorrect -s insertion does not necessarily function as a plural marker, but rather results in neutralization to third person singular on verbs, and neutralization of the singular/plural distinctions in nouns. The only attested instances of hypercorrect -s insertion in Puerto Rican Spanish verbs, which may be considered analogous to Black English -s insertion, serve, in contrast, to reinforce rather than neutralize the singular/plural distinction.

Thus we see that while reorganization has clearly taken place with regard to the standard Spanish rule for plural marking, we have no reason to attribute it to influence from English. It is due rather to an old and widespread process of phonological weakening and deletion of syllable-final (s)—first attested in Spain in the 16th century (Lapesa 1965), and to the elimination of inflectional, but not other types of redundancy from the sentence. This is made possible because 1) the Spanish verb (unlike the English verb) carries information as to person, number and tense, 2) inflections are generally retained on these verbs, and 3) a large amount of noninflectional redundancy remains in the discourse.

Social and Syntactic Functions of Code-Switching

Another area of study concerns code-switching, or the alternation of two languages within a single discourse, sentence or constituent, as in (3).

3. But I used to eat the bofe, the brain. And then they stopped selling it because tenian, este, le encontraron que tenía [they had, uh, they found out that it had] worms. I used to make some bofe! Después yo hacía uno d'esos [then I would make one of those] concoctions: The garlic con cebolla, y hacia un mojo, y yo dejaba que se curara eso [with onion, and I'd make a sauce, and I'd let that sit] for a couple of hours. Then you be drinking and eating that shit. Wooh! It's like eating anchovies when you're drinking. Delicious! (04/10i)
Code-switching has been observed by Pedraza (1979) to function as a widespread mode of communication on the block. Educators and intellectuals (Varo 1971; de Granda 1968; LaFontaine 1975) have seen such language mixture to constitute evidence of the disintegration of Puerto Rican Spanish language and culture, although community members themselves recognize its existence and do not view it negatively.

An early pilot study of a single balanced bilingual speaker observed to be one of the most prolific code-switchers on the block (Poplack 1978b), provided quantitative empirical evidence that first, code-switching is governed by functional or pragmatic constraints, as had already been noted anecdotally in the literature (Gumperz 1971, 1976; McClure 1977; Valdés-Fallis 1976, 1978). We found that the occurrence of code-switching is affected by both the ethnicity of the interlocutor and the formality of the speech situation, such that it is more frequent when speaking with another Puerto Rican, and in informal speech styles.

Aside from the extralinguistic constraints, we wanted to account for the syntactic distribution of switch points. Most of the literature on code-switching is only tenuously related to empirical research, and almost none of it is based on systematic analysis of a large corpus of natural speech. Nevertheless many theories of syntactic constraints on code-switching have been presented. Scholars who have proposed such constraints include Gumperz (1976), Hasselmo (1972, 1979), Timm (1975, 1978), Wentz (1977), McClure (n.d.), and Gingrás (1974). Most of the constraints derived in an ad hoc way from the linguist's own intuitions or from acceptability judgements.

In the 400 switches analyzed from the balanced bilingual speaker, we found exceptions to every one, which led us to posit two more general constraints which hold very strictly for these data as well as the data published in other studies, though they had not been formally enunciated as such in the literature.*

The first, the *free morpheme constraint*, prohibits a switch from occurring between a bound morpheme and a lexical form unless the latter has been phonologically integrated into the language of the former. This excludes switches like (4a), which in fact are not attested, but not forms like (4b), although we consider this an instance of monolingual Spanish speech, and not a code-switch.

4a. *RUN-ando* [oʊn-e'ando] ‘running
4b. *janguwando* [hæŋge'ando] ‘hanging out’

The second, the *equivalence constraint*, states that the order of sentence constituents immediately adjacent to and on both sides of the switch point
must be grammatical with respect to both languages involved simultaneously. This requires some specification: the local cogrammaticality or equivalence of the two languages in the vicinity of the switch holds as long as the order of two sentence elements, one before and one after the switch, is not excluded in either language. This is illustrated in Figure 1, where the dotted lines indicate permissible switch points and the arrows indicate ways in which constituents from two languages map onto each other. The speaker's actual utterance is reproduced in (C).

A. Eng I ; told him ; that ; so ; that ; he ; would bring it ; fast.
B. Sp (Yo); le ; dije ; eso ; pa' ; que ; (él); la ; trajera ; ligero.
C. CS I told him that I pa' que la trajera ligero. (04/73)

Figure 1. Permissible code-switching points.

Operation of the syntactic constraints permits only code-switched utterances which, when translated into either language, are grammatical by both L₁ and L₂ standards, and indicate a large degree of competence in both.

To ascertain whether these results were merely the artifact of having studied a balanced bilingual, a larger study was undertaken including both Spanish-dominant and balanced bilingual speakers (Poplack 1979). Analysis of almost 2,000 switches confirmed that these constraints were upheld by even the nonfluent bilinguals in the sample.

Speakers were found to engage in at least three types of code-switching, each one characterized by switches of different levels of constituents, and each one reflecting different degrees of bilingual ability. These are exemplified in (5).

5a. 'Tag'-like Switch:
Vendía arroz 'n shit. 'He sold rice and all.' (03/24)
Ave Maria, which English? 'Oh God, which English?' (08/436)

5b. Sentential Switch:
It's on the radio. A mi se me olvida la estación. I'm gonna serve you another one, right? 'I forget which station.' (08/187)

5c. Intra-sentential Switch:
Si tu eres puertorriqueño, your father's a Puerto Rican, you should at least de vez en cuando, you know, hablar español. 'If you're Puerto Rican, your father's a Puerto Rican, you should at least sometimes, you know, speak Spanish.' (34/25)
(5a) shows a switch of a tag or interjection, freely moveable constituents which can be inserted almost anywhere in discourse without fear of violating a grammatical rule of either language. This type of segment may be switched into L₂ with only minimal knowledge of the grammar of that language. Next on the scale are full sentences or larger segments, which require much more knowledge of L₂ to produce (5b), although hypothetically, not as much as is required by the third category, intra-sentential switches (5c). In order to produce this latter sort of switch, the speaker must know enough about the grammar of each language and the way they interact to avoid ungrammatical utterances.

Figure 2 displays our finding that nonfluent bilinguals are able to code-switch frequently, and still maintain grammaticality in both languages by favoring tag-switching, the type requiring least skill, while the balanced bilinguals favor the sentential or intra-sentential type, which we had hypothesized to require most skill. (Each line on the graph represents a speaker.) The figure shows that reported language ability (which in all cases but four corresponds to observed ability) is an excellent indicator of code-switching patterns. Figure (2a) shows that most of those who report that they know, feel more comfortable in, and use more Spanish than English, tend to switch into L₂ by means of tag-like constructions, sometimes to the practical exclusion of sentential or intra-sentential switches. Those who claim to be bilingual, on the other hand, show a reversal (2b). They favor large amounts of the switches hypothesized to require most knowledge of both languages, sentential and intra-sentential switches.

The few exceptions to these patterns are represented by the dotted lines on the graphs. These are precisely the cases where ethnographic observation and linguistic analysis were previously found to conflict with self-report, because the speakers either underrated or overrated their ability in English. The studies provide evidence that rather than representing a debasement of linguistic skill, code-switching is actually a suggestive indicator of bilingual competence.

Linguistic performance constrained in this way must be based on simultaneous access to the grammatical rules of both languages, raising the question of the existence and nature of code-switching grammar.

We thus attempted to describe formally how the code-switching constraints determine the way the two monolingual grammars may be combined in generating discourse containing code-switches (Sankoff and Poplack 1980). As the code-switching constraints operate on the surface syntax of a sentence, we adopted a formalism based on direct generation of surface phrase structures by a context-free grammar (having shown that switched items are in no way constrained to be of the same language as the
elements which may be adjacent to them in deep structure, but rather, if at all, by their surface neighbors. We then probabilized the rules of the grammar so that it could account for the quantitative patterning evident in the data. An important conclusion drawn from this part of the exercise is that even in portions of discourse in close proximity to one or more switches, the speaker strictly maintained both qualitative and quantitative distinctions between Spanish and English grammars. Whenever a stretch of discourse, e.g. a sentence or constituent, could be clearly identified as monolingual, the rules of the appropriate monolingual grammar, and their associated probabilities, were exclusively at play. The analysis of code-switching has deep implications for grammatical theory, as it points to ways in which two languages can reconcile, but not change, their differences, to result in a mode of communication widely functional as monolingual speech.

The evidence we have presented for the syntactic integrity of Spanish and English grammars, even when they are being used sequentially and simultaneously, bolsters other arguments for nonconvergence of Spanish and English in the Puerto Rican speech community.
Verb Usage

In a quantitative semantic analysis of tense, mood, and aspect in the entire Puerto Rican Spanish verbal paradigm, and again involving both Spanish-dominant and balanced bilingual speakers (Pousada and Poplack 1979), we compared the ranges of meaning covered by verb forms for the two groups, with those prescribed in normative Spanish grammars. Our goal was to examine the extent, if any, to which the Puerto Rican Spanish verb system has diverged from the standard, by ascertaining whether the semantic fields, or ranges of meaning, covered by verb forms were being extended or restricted, and in what direction: specifically, were some forms being extended to cover semantic fields of others which have fallen into disuse within the Puerto Rican Spanish system, or is there adaptation to specifically English semantic fields? To do this we first distinguished absolute from extended uses of verb forms. An example of the former would be use of the present progressive to convey ongoing action in present time. Examples of extended uses are given in (6). A construction line (6a) was assumed to result from transference from English, because the present progressive occurred with a verb of perception, which in standard Spanish categorically requires the simple present. On the other hand, (6b), exemplifying use of the present to convey action in future time, is a perfectly acceptable extension of the uses of the Spanish simple present.

6a. Yo no estoy viendo eso. 'I'm not seeing that.' (002/314)
6b. Mañana voy a Caguas. ‘Tomorrow I'm going to Caguas.’

We then compared the rank distribution of the verbal forms in the speech of our sample with the Puerto Rican Spanish standard as represented by the written speech of a prominent Puerto Rican author, then with a sample of 15th century Spanish, with a sample of modern Andalusian Spanish, and finally, with the New York City English of a monolingual.

The results indicated, first, that there were virtually no instances of verbs being used in an ungrammatical or idiosyncratic way. Extended uses not attested in prescriptive grammars constituted less than 1% of the 9,000 verbs studied. Table 4 shows that of the 26 verb forms originally studied, there are only three inflected forms which are quantitatively important: the present, which represents half the data, followed by the preterite, and then the imperfect. All other inflected forms individually represent 3% or less of the data.

When we examine how these verb forms are distributed to convey the semantic fields of {PRESENT}, {PAST} and {FUTURE}, we find that the preferred way of expressing present reference is through use of the simple present, while preterite and imperfect forms are generally used to express
### Table 2. Verb distribution in vernacular and standard Puerto Rican Spanish.

<table>
<thead>
<tr>
<th>Tense/Mode</th>
<th>Spanish Dominants</th>
<th>Bilinguals</th>
<th>All Speakers</th>
<th>Standard PPS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N %</td>
<td>N %</td>
<td>N %</td>
<td>N %</td>
</tr>
<tr>
<td>present</td>
<td>3231 49.5</td>
<td>1078 50.2</td>
<td>4309 49.6</td>
<td>133 49.3</td>
</tr>
<tr>
<td>preterite</td>
<td>904 13.8</td>
<td>324 15.1</td>
<td>1228 14.1</td>
<td>25 9.3</td>
</tr>
<tr>
<td>imperfect</td>
<td>543 8.3</td>
<td>148 6.9</td>
<td>691 8.0</td>
<td>15 5.6</td>
</tr>
<tr>
<td>present perfect</td>
<td>143 2.2</td>
<td>43 2.0</td>
<td>186 2.1</td>
<td>6 1.1</td>
</tr>
<tr>
<td>conditional</td>
<td>49 0.8</td>
<td>14 0.7</td>
<td>63 0.7</td>
<td>10 3.7</td>
</tr>
<tr>
<td>preterite perfect</td>
<td>22 0.3</td>
<td>5 0.2</td>
<td>27 0.3</td>
<td>2 0.7</td>
</tr>
<tr>
<td>future</td>
<td>12 0.2</td>
<td>3 0.1</td>
<td>15 0.2</td>
<td>5 1.9</td>
</tr>
<tr>
<td>preterite anterior</td>
<td>0 0.0</td>
<td>1 0.0</td>
<td>1 0.0</td>
<td>0 0.0</td>
</tr>
<tr>
<td>future perfect</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>preterite conditional</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>present modal + infinitive</td>
<td>245 3.8</td>
<td>78 3.6</td>
<td>323 3.7</td>
<td>12 4.4</td>
</tr>
<tr>
<td>present periphrastic future</td>
<td>158 2.4</td>
<td>43 2.0</td>
<td>201 2.3</td>
<td>4 1.5</td>
</tr>
<tr>
<td>imperfect modal + infinitive</td>
<td>23 0.4</td>
<td>16 0.7</td>
<td>39 0.4</td>
<td>0 0.0</td>
</tr>
<tr>
<td>preterite modal + infinitive</td>
<td>15 0.2</td>
<td>6 0.3</td>
<td>21 0.2</td>
<td>1 0.4</td>
</tr>
<tr>
<td>imperfect periphrastic future</td>
<td>4 0.1</td>
<td>4 0.2</td>
<td>8 0.1</td>
<td>0 0.0</td>
</tr>
<tr>
<td>present progressive</td>
<td>135 2.1</td>
<td>54 2.5</td>
<td>189 2.2</td>
<td>2 0.7</td>
</tr>
<tr>
<td>imperfect progressive</td>
<td>26 0.4</td>
<td>5 0.2</td>
<td>31 0.4</td>
<td>0 0.0</td>
</tr>
<tr>
<td>preterite progressive</td>
<td>6 0.1</td>
<td>3 0.1</td>
<td>9 0.1</td>
<td>0 0.0</td>
</tr>
<tr>
<td>present subjunctive</td>
<td>257 3.9</td>
<td>51 2.4</td>
<td>309 3.5</td>
<td>7 2.6</td>
</tr>
<tr>
<td>imperfect subjunctive</td>
<td>71 2.0</td>
<td>10 0.5</td>
<td>81 0.9</td>
<td>2 0.7</td>
</tr>
<tr>
<td>preterite perfect subjunctive</td>
<td>4 0.1</td>
<td>2 0.1</td>
<td>6 0.1</td>
<td>0 0.0</td>
</tr>
<tr>
<td>present perfect subjunctive</td>
<td>1 0.0</td>
<td>1 0.0</td>
<td>2 0.0</td>
<td>1 0.4</td>
</tr>
<tr>
<td>future subjunctive</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>future perfect subjunctive</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>imperative</td>
<td>173 2.6</td>
<td>82 3.8</td>
<td>255 2.9</td>
<td>0 0.0</td>
</tr>
<tr>
<td>infinitive</td>
<td>510 7.8</td>
<td>176 8.2</td>
<td>686 7.9</td>
<td>45 16.7</td>
</tr>
<tr>
<td><strong>TOTALS</strong></td>
<td><strong>6532</strong></td>
<td><strong>2147</strong></td>
<td><strong>8679</strong></td>
<td><strong>270</strong></td>
</tr>
</tbody>
</table>
reference, and that the periphrastic and simple present are also the preferred ways of expressing future time. This explains how Puerto Rican Spanish expresses past, present and future reference by means of a basic present–past tense distinction: the present has been extended to cover the semantic fields of the future form, which is used only rarely, and not necessarily to convey futurity. (In fact, one third of the attested future forms were used for present reference.) These findings are in keeping with studies on Spanish in the Southwest (Floyd 1978), which although not quantitative, repeatedly indicate that these three verb forms are the most productive, maintaining their usual functions as well as expanding to include those of other forms.

Table 4 also shows strikingly little difference between what we are calling ‘standard’ Puerto Rican Spanish and that of the East Harlem sample, despite the fact that the standard material was extracted from a written text. Statistical tests on these figures showed that the most significant differences between the two data sources are in the area of past tense forms—the East Harlem speakers used more due to their more informal speech style, which contained many narratives of personal experience requiring verb forms in the past tense. The standard also shows significantly more of the inflected conditional and future forms. Use of the conditional is probably an aspect of learned or academic speech, and the future, as in other languages such as French and English, may be largely reserved as a marker of formal speech performance.

The Table also shows that there is remarkably little difference between the Spanish-dominant and bilingual speakers, despite the fact that the latter could be hypothesized to be more under the influence of English. In fact, the most startling aspect of these findings is their great regularity. Log-likelihood tests based on these figures show that the only significant area of difference is in use of the subjunctive, which the bilinguals use somewhat less than the Spanish-dominant speakers. Now just this tendency has been hypothesized (e.g. de Granda 1968) to be due to influence from English, which has lexicalized or lost most of the distinctions expressed by the subjunctive. Although this possibility cannot be overruled, no conclusive evidence in its favor has yet been presented. In fact, the Table shows that ‘standard’ Puerto Rican Spanish is characterized by subjunctive usage closer to that of the bilinguals than to the Spanish-dominant speakers. These results, then, cannot be considered evidence for any significant degree of convergence of vernacular Puerto Rican Spanish towards English.

Figure 3 correlates the rank order of inflected verb form frequencies among the East Harlem speakers and *La Celestina*, a 15th century Spanish picaresque novel.
Figure 3. Rank order of inflected verb form frequencies in East Harlem data versus order in La Celestina.

Points lying near the diagonal represent forms of relatively equal importance in each corpus. Strikingly enough, Figure 3 shows that the relative ranking of verb form usage has remained basically unchanged since the 15th century.

The only real exception is in use of the inflected future, which had much greater importance in the 15th century than today. Indeed, this form has practically been replaced by the periphrastic future in vernacular Puerto Rican Spanish. As may be seen in (7), the rank correlation of these figures is .85, indicating a strong similarity in the distribution of verb forms.
7. Rank correlations of verb form usage.
   East Harlem — 15th century Spanish  .85
   East Harlem — modern Andalusian Spanish  .79
   East Harlem — English  .53
   Andalusian — English  .57

When compared with a corpus of English verb forms, in contrast, we find that their distribution is significantly different for every verb form but one; that is, influence from English does not appear to have affected these core areas of the Spanish language. The correlations in (7) summarize this lack of influence.

Indeed, in our study of East Harlem, the only factor which was found to differentiate verb usage in any significant way was the speech situation in which the form was uttered. Different speech situations were shown to favor different proportions of verbal forms, providing yet another example of the inherent stylistic variation which characterizes natural language.

Thus we see that the only significant change in the system, decrease in use of the inflected future, is intrasystemic: it is the result of the extension of the present periphrastic and simple present forms to convey futurity. This process is widespread in other Romance languages as well as in English.

The example given in (6a), Yo no estoy viendo eso, is one of only three in the entire corpus. The evidence thus is against convergence with English. Such conclusions were not drawn from qualitative studies of verb usage in other varieties of United States Spanish (cited in Floyd 1978; Bills 1975). However those studies have concentrated on supposed deviations from the standard, without quantitative study of the standard itself.

**Lexical Borrowing**

The lexicon is one area in which language contact has had clear influence on Puerto Rican Spanish. Systematic study of the introduction and incorporation of English loanwords into Puerto Rican Spanish (Poplack, Sankoff, and Pousada 1980) indicates first, that borrowings from English are used largely in areas of modern urban culture, for which no traditional Spanish usage existed—though there are now, of course, neologisms and calques in modern metropolitan and other prestigious Spanish varieties. In this sense, borrowings represent an expansion of the lexical resources of the community, rather than a suppression of Spanish vocabulary.

Second, despite the fact that many speakers command an English phonology which is quite distinct from their Spanish, the loanwords which are used with any frequency are rapidly Hispanicized, so that no systematic perturbation of phonological processes or syllabic patterns internal to Spanish can be detected. Indeed, statistical tests prove that frequency of use
BILINGUALISM AND THE VERNACULAR

is clearly a determinant of the degree of phonological integration of a loanword.

Third, words of English origin tend to take on very specific and normal Spanish grammatical functions. In examining several hundred borrowed nouns in natural speech, we found that in virtually every case where Spanish syntax required the presence of a determiner, the loanword was so accompanied (Poplack and Pousada 1980). In those contexts from which a determiner is usually omitted, the loanword had none.

Fourth, the rules for gender assignment to Spanish nouns apply rigorously to loanwords. The well-defined criteria depending on word-final segments operate consistently, even for words which show two alternative paths of phonological integration (e.g. el suéter, la suera 'sweater'; el hambérguer, la hambérga 'hamburger').

In summary, the addition of new words from English to the Puerto Rican vocabulary is seen to have little effect on the existing lexical stock, and no effect on the phonological, morphological and syntactic patterns of Spanish.

DISCUSSION

We have argued that the vernacular/standard distinction contrasts two epistemologically incommensurable concepts, neither of which have been well characterized. The standard is an idealization not tied to systematically observed behavior, and the vernacular has been attested to by anecdote. In the reality studied within the variationist paradigm, there are only vernaculars, upper-class vernacular vs. lower-class vernacular, formal vs. colloquial, etc. The abstract notion of the standard has often been used to show that the language of socially dominated groups is incorrect and inferior. In the case of many bilingual minorities in the United States the vernacular/standard distinction cuts two ways. The English of these minorities, insofar as it may be characterized by community-wide features, is considered deviant. And their language of origin differs, of course, from any abstract standard, so that it too is considered aberrant. The role of the linguist in this situation has often been to invoke the easiest type of explanation for developments in local varieties (i.e. convergence with English), and thus lend scientific authority to negative attitudes toward the minority language.

A systematic variationist study of the speech community is an effective way to avoid stereotypes and assumptions about the standard. The deviant forms cited by researchers on Chicano and Puerto Rican Spanish are also present in our data, but the studies reported here show that when apparent
deviations are placed within the context of the entire system, they are seen to constitute only a minuscule portion of the total verbal output. This leads us to confirm that emphasis on anomalies in multilingual situations on the part of researchers, educators, and intellectuals, is merely stereotyping due to the phenomenon of categorical perception, whereby deviation from a norm may be seen as far more prominent than its negligible frequency would warrant.

These studies raise serious questions about the theory of language convergence in contact situations. It is clear that the Puerto Rican Spanish spoken in El Barrio has been influenced by English, but this influence is most evident in the lexicon. However, the grammar of Spanish, which has been shown by our ethnographic and attitudinal studies to serve a wide range of communicative functions, has been extraordinarily resistant to influence from the grammar of English; this despite the economic and political dominance of the English-speaking community.

The evidence of Gumperz and Wilson (1971) for the convergence of Urdu, Marathi and Kanada, in Kupwar, India, must be due to a greater time scale, or to very different historical and cultural contexts, perhaps reflecting a restricted functioning of one or more of the codes involved, which as has been shown in preceding sections, is not the case here. Other cases of putative language convergence may simply reflect an uninformed purism, or nonquantitative methodology, or a preoccupation with borrowing of noncore vocabulary. More dangerous than the linguistic outcomes of contact to a vigorous and thriving Spanish, at least in this instance, are those who would restrict its usage from schools, government, and other institutions, and purists who stigmatize the expanded expressive repertoire of bilingual speakers in a linguistically complex milieu.

NOTES

*The analyses reported here are part of a research project on Intergenerational Perspectives on Bilingualism supported by the National Institute of Education under NIE-G-78-0091 and the Ford Foundation.

1 This characterization was made on the basis of observed frequency of usage combined with self-report of the language that 'feels most comfortable' and is used most frequently in a series of domains. We shall see that the resulting assessment generally corresponds to the linguistic assessment of language capability.

2 Of course, influence from English may be expected to be most apparent in the speech of those who are 'English-dominant'. These speakers were not included in the present sample because they simply did not produce enough Spanish to compare quantitatively with the others. Two possible outcomes of the contact situation may be hypothesized for their Spanish:
BILINGUALISM AND THE VERNACULAR

(1) such Spanish as they do speak may undergo restructuring so as to approximate English structures; that is, it may converge with English, (2) the Spanish language in toto may gradually disappear from the repertoire of these speakers, and be retained only for emblematic purposes (e.g. through insertion of Spanish words or idiomatic expressions into English), in which case the grammar of the language would remain unaffected. Our ethnographic observations of language use in this community may be interpreted in favor of the second possibility, but we have no linguistic evidence with which to substantiate either hypothesis at the moment.

¹Numbers in parentheses are speaker identification codes.

*‘S’ and ‘N’ refer to any phonetic realization of the variables (s) and (n) respectively, other than phonetic zero.

¹Note that this behavior holds when the subject noun phrase precedes the verb, the canonical sentence structure in Spanish, but not when it follows. There is no significant difference in verbal deletion rates between inflected and uninflected noun phrases when these are postponed.

 Differences in these constraints has been observed independently in the literature by e.g. Jacobson (1976), Lipski (1978), and Pfaff (1979). However, none of these scholars has specifically incorporated them into their analyses.

The data on early modern Castilian Spanish were provided by a frequency analysis of verb usage in La Celestina by Criado del Val (1966).