DE GRUYTER MOUTON
DOI 10.1515/ling-2013-0039 — Linguistics 2013; 51(Jubilee): 11–14

The second decade (1973–1983)

Shana Poplack, *"Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL": Toward a typology of code-switching. Linguistics* 1980, Volume 18, issue 7/8, 581–618. DOI 10.1515/ling.1980.18.7-8.581

Introductory comments by the author

What qualifies *Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL* as the most heavily cited paper in the 50-year history of *Linguistics*? It did inspire a substantial and productive research tradition, but it has also generated recurrent and ongoing attacks. I'd like to reflect on why its main proposals have been so controversial, and what their current status is today. A remarkable fact is that despite 33 years of intense research activity since *Sometimes* was published, there is still no consensus on the nature or identity of even the major manifestations of language contact (codeswitching [CS] and borrowing [B]), let alone the linguistic conditions governing their use. This discord, so characteristic of the field of contact linguistics, arises not so much from the recalcitrance of its subject as from the incommensurable perspectives of its practitioners on language, the conduct of research, the nature of "fact" and evidence, and the principles of scientific proof.

Sometimes represents the first application of the variationist paradigm to the study of CS. Foremost among the theoretical and methodological tenets that distinguish this framework is the recognition that bilingual speech, just like its monolingual congeners, is inherently variable, i.e. involves choices during speech production which incorporate some degree of unpredictability. Key aspects of a methodology capable of handling such variability include (i) consideration of the data of actual bilingual interactions situated in the context of the speech community in which they were produced, (ii) distinction among different manifestations of language contact, (iii) contextualization of mixing strategies with respect to each other and to the monolingual benchmark varieties (standard or not) involved, and iv) systematic quantitative analyses of the data. Taken together, these methodological imperatives can help us determine when (and how) other-language items may come to assume the linguistic structure of a recipient language (L_r) into which they are incorporated, or alternatively, retain their donor-language (L_d) identity. More important, they reveal *patterns* of language mixing within and across communities, enabling us to tease out the major strategies from idiosyncratic outliers.

Sometimes was a nascent attempt to delineate, for quantitative analytical purposes, the fundamental distinction between CS and B. This distinction is manifest in the contrast between those L_d items that function like their L_r counterparts (i.e., are *borrowed*), and those retaining L_d grammar (*switched*). It was introduced in this article in terms of its corollary, the "Free Morpheme Constraint" (FMC), essentially a preliminary formulation of the claim that B (established or nonce) and CS (single-word or multiword) could be identified and characterized. The reasoning behind the FMC was that any apparent *word-internal* change in language could be traced to the conversion – phonological, morphological and syntactic – of an L_d free morpheme into an L_r counterpart, following well-documented processes of loanword integration. Switching *between* words intra-sententially, on the other hand, was found to be mainly conditioned by the "Equivalence Constraint" (EC), a formalization of the widespread avoidance by bilinguals of word order conflicts at switch points. Subsequent work in the variationist framework has clarified that "nonce" borrowing in running speech is not restricted to previously attested forms, and that the gradience of phonological integration renders it a generally poor indicator of loanword (or CS) status.

What have we learned since 1980?

The last several decades have witnessed the accumulation of a coherent body of variationist studies on a large number of typologically distinct and similar language pairs, using a rich variety of morphosyntactic diagnostics involving such key linguistic phenomena as determiner expression, adjective placement, case-marking, word order and verb formation, among many others. These have confirmed beyond a doubt the original insight of *Sometimes* that CS and B are indeed two distinct processes, which are governed by different rules, and which – using the appropriate methodology – may be operationally distinguished as such (except in inherently language-neutral constructions). They have also converged on two groundbreaking discoveries, correcting assumptions implicit in the article (as in the field more generally), while refining the broad lines of the proposals sketched there.

(1) Phonological and morphosyntactic integration are independent. Phonology, of both CS and B, is variable, and thus cannot reliably be used to distinguish between them.
(2) Morphosyntactic integration is not contingent on frequency and diffusion (i.e., achievement of loanword status), but rather occurs abruptly at the nonce borrowing stage.

Only hinted at in *Sometimes* was what we now know from subsequent community-based studies to be the quantitatively preponderant categories of nonce and more frequent, though unattested, borrowings. Other recurrent findings to emerge from systematic analyses of tens of thousands of tokens of spontaneous bilingual production data are:

(3) The vast majority of all instances of language mixing are *single* other-language items, and
(4) Nearly all of them are immediately integrated morphologically and syntactically into L_r.

This means that single L_d items tend to display the linguistic characteristics of established loanwords, independent of their frequency or recurrence, while single-word CS are exceedingly rare. (Notwithstanding, single L_d incorporations cannot simply be *assumed* to be B; the variationist literature offers clear discovery procedures to determine their status.) This disproportion was already clearly in evidence in *Sometimes*, despite the fact that many single L_d items were misidentified there as CS (largely because the principle in (1) had not yet been recognized). Crucially, the linguistic properties of CS – of whatever length – have by now been shown repeatedly to contrast diametrically with those of Bs: both the internal constituency and positioning of Bs come from $L_{r;}$ the internal constituency of CS is that of L_d, but its placement in the sentence tends to respect the word order requirements of *both* L_d and L_r. This is the basic insight behind the EC.

These differences, coupled with the severe quantitative disproportions among single- and multiword L_d items, conspire to render any study treating them as one and the same in effect a study of B. (They also explain the appeal of insertion models for CS. They do in fact account for the bulk of the language mixing data, but not for the data of the relatively rare multiword CS.)

Counterexamples?

Over the years, many counterexamples have been offered to the original constraints proposed in the article. In assessing the arguments based on them, we must consider that relatively few were ever actually produced by bilingual speakers. Others arise from failure to distinguish purported violations from replications of (variable) L_r structure, often including null-marking. The bulk of them concern the EC prediction that (unambiguous) intra-sentential CS will occur at sites sanctioned by the grammars of both languages. Remarkably, the lion's share of these "counterexamples" involves *single* L_d words. These of course do not respect the EC, but instead pattern syntactically and morphologically with L_r, just

as predicted by the FMC and its more recent formulation, the Nonce Borrowing Hypothesis.

Crucially, only rarely have any of these counterexamples been demonstrated empirically to represent quantitatively meaningful *patterns* of language use in any well-defined speech community. Once the data of language mixing have been accurately circumscribed, and the analytical and methodological imperatives outlined above applied, (nonce) borrowing and CS under equivalence emerge as the quantitatively major strategies employed in bilingual communities studied using accountable methodology.

The preoccupation with isolated counterexamples characteristic of the field of language contact highlights the disappointing fact that even where *Sometimes* and other quantitative empirical work like it is enthusiastically received, it only rarely inspires the type of research required to replicate it. To be sure, many scholars of language contact have elicited bilingual production data of varying degrees of naturalness, and some have even constructed sizeable corpora. But exploitation of this material is too often restricted to a few selected examples, or declarations in lieu of demonstrations that all the data satisfy some constraint. The real value of a corpus can be realized only through the *principle of accountability*, where every token of all constructions carrying out a certain role is counted and placed in statistical context. Only then are patterns discernable and general trends emergent from the complex production of spontaneous discourse.

Shana Poplack: University of Ottawa. E-mail: spoplack@uottawa.ca