A milestone study: Structured variability as the key to unraveling (contact-induced) language change

RENA TORRES CACOULLOS
Department of Spanish, Italian and Portuguese,
The Pennsylvania State University
rena@psu.edu

Despite increasing attention to bilingualism – conferences, publications, grants – linguists are as far as ever from reaching consensus. Is code-switching the alternation between two equally activated languages or is it the insertion of elements from a source language into a recipient language? Can and should we distinguish borrowing and code-switching of single words? Is there grammatical convergence between bilinguals' two languages and does code-switching promote it? Since the first accounts of the structure of code-switching in the 1970s, the same questions have been readdressed with astoundingly little, if any, cumulative advances. Scientific progress has been obstructed by polemic debate, often fueled by elicited judgments, which may display random error (Labov, 1996), or reports of the behavior of stray individuals, which are uninterpretable in the absence of knowledge of the systematic community pattern (Labov, 2006/1966, p. 5).

Poplack, Zentz and Dion (PZD; Poplack, Zentz & Dion, 2011, this issue) put forward a scientific approach to the synchronic study of code-switching and convergence, grounded in solid data and objective analyses. With quantitative reasoning resting on community-based speech data and the variationist comparative method, PZD deflate the widely held conjecture that code-switching entails grammatical change. This work should henceforth be seen as a paradigm for ascertaining change in the grammar(s) of bilingual speakers and testing contact-induced change in the community.

The methodological steps are:

1. to locate a community of bilinguals who do in fact code-switch, construct a principled sample of speakers, and record unreflecting speech, reducing observation effects;

2. to delimit a linguistic variable that involves a candidate for grammatical convergence, that is, a form or construction with respect to which the languages in contact have some overlap – which would enable bilinguals' "interlingual identifications" (Weinreich, 1963, p. 7) or "grammatical replication" (Heine & Kuteva, 2005, p. 2) – but also some differences or "conflict sites" (Poplack &

Meechan, 1998, p. 132) – which would enable the analyst to unambiguously discern the provenance of the forms of interest, ruling out cross-linguistic tendencies;

3. to define constraints on the selection of variant forms or constructions, based on quantitative measures of goodness of fit between the theoretical model and observed distributions.

Vital to the test of convergence here are comparisons of bilingual patterns with non-contact benchmarks: pre-contact and contemporaneous non-contact varieties, and, most critically, comparison with the presumed source (or model) language.

The bilingual data examined by PZD are contextualized with respect to a well-defined community. This is the first step distinguishing this study from most code-switching research, and an essential one for cumulative scientific progress. While it is becoming harder to publish papers based on one or two subjects, the imperative for meaningful sample sizes is still often not adhered to. But even requiring larger speaker samples is not sufficient. Because contact situations and patterns of code-switching are community-particular, even for the same language pair (see Poplack, 1987), research based on bilinguals of unspecified background is neither replicable nor revealing.

Establishing convergence, let alone whether it is promoted by code-switching, means first ascertaining change. Why have linguists been susceptible to hasty diagnoses of (contact-induced) change? One reason has been the equation of analysts' judgments of departures from a prescriptive or idealized norm with change. But while all change involves variability, "not all variability and heterogeneity in language structure involves change" (Weinreich, Labov & Herzog, 1968, p. 188).

We know that variation can be stable for quite a long time, for example, in the case of proscribed double negation in English (Labov, 2001, pp. 85–92). Linguists have also not been sufficiently wary of categorical perception, which might make a few overheard examples seem like an overall tendency. Reflections about speech are no substitute for unreflecting speech as data, especially

for nonstandard varieties (Sankoff, 1988, pp. 145–146).

Besides the *Ottawa-Hull French Corpus*, the assembling of benchmarks to ascertain change is exemplary in PZD's study, with painstaking comparisons both in time and across contact and non-contact varieties of French. Crucial, too, is PZD's examination of preposition placement in a variety of Canadian English that would actually be a likely target model for the bilingual francophones. Many a claim of convergence has been made in the absence of accountable comparison with the presumed source language, another deficiency which has diverted research on convergence.

How do PZD gauge the attainment of structural similarity between the language varieties in contact? Having contextualized the data with respect to the bilingual community, step two on the list above is to contextualize the form/construction that is the candidate for convergence with respect to the grammatical system(s) in which it is embedded. In delimiting two linguistic variables – preposition "stranding" and "orphaning" – PZD situate phrase-final prepositions in Quebec French against their apparent counterparts in English but also against associated French constructions. By counting ALL relative clauses involving prepositions, stranded as well as the variants with which they alternate in discourse, PZD show that "stranding" is actually of low frequency relative to the other variants (scientifically answering categorical perception). And, by also considering "orphaning", or the construction of verb + preposition with a "complement" in the preceding discourse, they show that phrase-final prepositions are an extension to the relative clause context of this French construction rather than a replica of superficially similar English stranded prepositions.

The basic question the methodology of PZD addresses is: How do we measure grammatical similarity or difference (and thereby gauge grammatical convergence)? In place of casual observation of apparently similar forms and/or functions, or even ascription of underlying structure according to some theory of autonomous syntax, evaluating convergence requires quantitative argumentation. Convincing reports of language change must include a quantitative component, because change is by nature quantitative, as alternative forms or constructions increase in frequency. Moreover, variationist diachronic studies indicate that increasing rate of use is accompanied by changes over time in the configuration of constraints contributing to variant choice (Poplack & Malvar, 2007; Torres Cacoullos, 2009). This brings us to step three of the list above.

Implementing and advancing the variationist comparative method, PZD demonstrate that superficial similarities between languages in contact may belie "deeper differences" (abstract). Rather than relying on the calculation of overall rates of use – itself an advancement beyond merely registering candidates for convergence – the variationist comparative method incorporates inherent variability into the traditional comparative method of historical linguistics, by examining distribution and co-occurrence patterns (Poplack & Meechan, 1998; Sankoff, Poplack & Vanniarajan, 1990). Structured heterogeneity, which constitutes an inherent part of the grammatical system, becomes the tool for measuring grammatical similarity or difference. This structure lies in the constraints on speakers' choices among variants, or the LINGUISTIC CONDITIONING. Overall rates – average relative frequencies of forms – may misrepresent the grammar, because they fluctuate according to data collection procedures and situational factors of channel, genre, topic or style, but the linguistic conditioning – direction and magnitude of effect of language-internal factors (dis)favoring the occurrence of forms – holds across extra-linguistic circumstances.

An example of the fineness of linguistic conditioning as a gauge of grammatical systems is seen in Spanish subject-pronoun expression, where despite DISPARATE RATES, we find PARALLEL LINGUISTIC CONDITIONING across genres. Travis (2007) found a rate of expression nearly one-and-a-half times greater in (Colombian) conversation than in (New Mexican) narratives, but shared morphosyntactic and discourse constraints (verb class and tense, realization of previous coreferential subject, distance in clauses from previous coreferential subject). The narratives had a higher degree of subject continuity than the interactive conversations, which had more shifting of topics. Given the subject continuity constraint, such that expression is favored when the subject of the preceding clause is non-coreferential, Travis (2007) concluded that the overall rate difference was due to genre-driven differences in distribution rather than to different grammars.

Just as divergent overall rates may mask shared linguistic conditioning (grammatical similarity), close rates may mask differences. An example is the case of bare (determinerless) nouns in the Spanish and English portions of interviews with New Mexican bilinguals. Torres Cacoullos and Aaron (2003) found SIMILAR RATES of bare nouns, but DISPARATE LINGUISTIC CONDITIONING (effects of semantic class, syntactic role and specificity of the NP). This confirms the language-particular usage of superficially similar determiners in the two varieties in contact. In New Mexican Spanish, predicate nominals designating occupations or social status favor bare nouns, but in New Mexican English, the direction of effect is reversed. These language-particular occupation-predicate nominal constructions are maintained in code-switching, as illustrated below. In the English sequence in (1), the speaker uses indefinite article *a*, but a

bare form (marked by Ø) in the immediately following Spanish.

(1) he was **a** teacher, you know,
 fue   *Ø maestro y   todo*
 be.PFV.3SG   teacher and everything
 "he was a teacher, you know, he was a teacher and all"
 (NMCOSS 144.4A)[1]

The comparison of multivariate models of variation, step three above in this methodology, allows for the specification of "conflict sites" between the language varieties in contact. For theories of bilingual representations and processing, especially those seeking to integrate the facts of production into models of cognition, perhaps most pertinent is PZD's conclusion that individual bilinguals adhere to different patterns of preposition placement in their French and English. Similarly, the differences in linguistic conditioning between the same New Mexican bilingual speakers' Spanish and English provide the strongest kind of evidence against convergence between their two language varieties, as PZD underline.

Particularly intriguing in this study is the identification of conflict sites involving lexical effects. PZD find that lexical identity of the preposition does not shape English use, but accounts for the majority of the variance of the French data. Most decisively, the lexical effects operate the same way in relative clauses with phrase-final prepositions ("stranding", the candidate for convergence with English) and in the more general construction of verb + preposition with a complement in the preceding discourse ("orphaning", which has no counterpart in English). The mechanism for the spreading of the existing "orphaning" construction to the relative clause context is thus not convergence with English but analogical extension within French. This raises the question of how the authors would answer the notion that contact triggers and/or accelerates internal change processes (e.g., Heine & Kuteva, 2005).

The strong lexical effects found by PZD suggest that future studies of contact-induced change may profit from consideration of lexical effects in grammatical variation. In contrast with the symbolic rules posited by generative linguists, analogy makes reference to lexical items (Bybee, 2010, pp. 69–74). For the study of bilingualism, particularly noteworthy is PZD's finding that the translation counterparts are the site not of convergence but of divergence – there is no distinction in English, for purposes of stranding, between *to* and *from* on one hand and *with* and *for* on the other, in contrast with the French distinction between *à* and *de* on one hand and *avec* and *pour* on the other, which strengthens the case for analogy.

The idea that code-switching may lead to convergence is not, prima facie, an outlandish one. Despite its attraction, it has not been supported, however. To date the methodology pioneered by Poplack and colleagues has been largely ignored by students of convergence, in part because it is easier to make than to prove claims. But when we apply the same method to another group of code-switchers, we find that the Quebec bilinguals are not a lone bastion of resistance to convergence.

Torres Cacoullos and Travis (2011) investigated variable Spanish first-person singular subject expression, illustrated in (2), in conversations of New Mexican speakers of Spanish and English, where code-switching is frequent – occurring essentially unpredictably – and smooth – occurring both across and within constituents.

(2) a. ***Yo tenía***   *una* wringer type machine,
   I have.IPFV.1SG a
  b. *so **yo** me   les   vinía*
   so I   REFL.1SG them.DAT.3PL come.IPFV.1SG
   por la ropa   once a week, or twice a week *y*,
   for the laundry   and
  c. (*Ø*) *la   llevaba   pa' mi casa,*
    it.ACC.SG take.IPFV.1SG to my house
  a. "I had a wringer type machine,
  b. so I would come to them for the laundry
   once a week, or twice a week and,
  c. (Ø) would take it to my house,"
 (NMCOSS 117.1A)

The hypothesis of grammatical convergence via code-switching was directly tested by comparing constraints on subject expression across the bilingual vs. monolingual "modes" of copious code-switchers (here, those who produced at least one-fifth of their first-person singular tokens in the presence of code-switching). Bilingual vs. monolingual "mode" (Grosjean, 1998) was operationalized as code-switching by the same speaker within the preceding three clauses. We found the same linguistic conditioning in operation in the presence vs. the absence of code-switching, and that the slightly higher rate of subject expression in the presence of code-switching was due not to the code-switching per se, but to priming. The linguistic constraints, including priming, were parallel to those observed in non-contact varieties of Spanish.

In the end, need code-switching – the alternating use of other-language material – perforce produce infiltration of other-language grammatical properties? Under which circumstances, if any, does code-switching promote grammatical convergence? Only empirical studies will tell. While historical linguists have sought to identify convergence post-facto, tests of the role of code-switching must be synchronic, since code-switching is a real-time discourse mode. Thanks to Poplack and colleagues, we have a model for testing convergence synchronically.

---

[1] Examples from New Mexico Colorado Spanish Survey (Bills & Vigil, 2008) are cited by interview number and cassette number/side.

References

Bills, G., & Vigil, N. (2008). *The Spanish language of New Mexico and southern Colorado: A linguistic atlas*. Albuquerque, NM: University of New Mexico Press.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1,* 131–149.

Heine, B., & Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Labov, W. (2006/1966). *The social stratification of English in New York City* (2nd edn.). Cambridge: Cambridge University Press. [1st edition 1966]

Labov, W. (1996). When intuitions fail. *Chicago Linguistic Society, 32,* 77–106.

Labov, W. (2001). *Principles of linguistic change* (vol. 2): *Social factors*. Malden, MA: Blackwell.

Poplack, S. (1987). Contrasting patterns of code-switching in two communities. In M. Heller (ed.), *Aspects of multilingualism: Proceedings from The Fourth Nordic Symposium on Bilingualism, 1984*, pp. 51–77. Uppsala: Borgströms, Motala.

Poplack, S., & Malvar, E. (2007). Elucidating the transition period in linguistic change: The expression of the future in Brazilian Portuguese. *Probus, 19,* 121–169.

Poplack, S., & Meechan, M. (1998). Introduction: How languages fit together in codemixing. *International Journal of Bilingualism, 2 (2),* 127–138.

Poplack, S., Zentz, L., & Dion, N. (2011). Phrase-final prepositions in Quebec French: An empirical study of contact, code-switching and resistance to convergence. *Bilingualism: Language and Cognition*, doi:10.1017/S1366728911000204. Published by Cambridge University Press, 11 August 2011.

Sankoff, D. (1988). Sociolinguistics and syntactic variation. In F. J. Newmeyer (ed.), *Linguistics: The Cambridge survey* (vol. 4), pp. 140–161. Cambridge: Cambridge University Press.

Sankoff, D., Poplack, S., & Vanniarajan, S. (1990). The case of the nonce loan in Tamil. *Language Variation and Change, 2 (1),* 71–101.

Torres Cacoullos, R. (2009). Variation and grammaticisation: The emergence of an aspectual opposition. In S. Tsiplakou, M. Karyolemou & P. Pavlou (eds.), *Language variation – European perspectives II*, pp. 215–224. Amsterdam: John Benjamins.

Torres Cacoullos, R., & Aaron, J. (2003). Bare English-origin nouns in Spanish: Rates, constraints and discourse functions. *Language Variation and Change, 15 (3),* 289–328.

Torres Cacoullos, R., & Travis, C. (2011). Using structural variability to evaluate convergence via code-switching: Priming and the structure of variable subject expression. *International Journal of Bilingualism, 15 (3),* 241–267.

Travis, C. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change, 19 (2),* 101–135.

Weinreich, U. (1963). *Languages in contact*. The Hague: Mouton.

Weinreich, U., Labov, W., & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann & Y. Malkiel (eds.), *Directions for historical linguistics*, pp. 95–188. Austin, TX: University of Texas Press.