

Shana Poplack
Department of Linguistics
University of Ottawa
401-70 Laurier Ave. E.
Ottawa, ON K1N 6N5 Canada

Code-switching (linguistic)

Introduction

Code-switching (CS) refers to the mixing, by bilinguals (or multilinguals), of two or more languages in discourse, often with no change of interlocutor or topic. Such mixing may take place at any level of linguistic structure, but its occurrence within the confines of a single sentence, constituent or even word, has attracted most linguistic attention. This article surveys the linguistic treatment of such *intra-sentential* switching.

In combining languages intra-sententially, various problems of incompatibility may arise. The most obvious derive from word order differences: under what conditions, if any, can the boundary between constituents ordered differently in two languages host a switch? Other potential combinatorial difficulties involve mismatches in grammatical categories, subcategorization patterns, morphology, and idiomatic expressions. Systematic examination of the spontaneous speech of bilinguals resident in a wide range of communities suggests, however, that speakers generally manage to circumvent these difficulties. CS tends not to produce utterances that contain monolingually ungrammatical sentence fragments. Discovery of the mechanisms enabling such "grammatical" CS is the major goal of current research. Central questions include locating permissible switch sites and ascertaining the nature (hierarchical or linear, variable or categorical) of the constraints on switching.

Background

Though CS is apparently a hallmark of bilingual communities world-wide, it has only begun to attract serious scholarly attention in the last few decades. Researchers first dismissed intra-sentential code-switching as random and deviant (e.g., Labov, 1972; Lance, 1975; Weinreich, 1953/1968); but are now unanimous in the conviction that it is grammatically constrained. The basis for this conviction is the empirical observation that bilinguals tend to switch intra-sententially at certain (morpho)syntactic boundaries and not at others. Early efforts to explain these preferences proceeded by proscribing certain switch sites, e.g. between pronominal subjects and verbs (Gumperz, 1976/1982; Timm, 1975) or between conjunctions and their conjuncts (Gumperz, 1976/1982). However, these particular sites were soon reported to figure among the regular CS patterns of some bilingual communities (e.g., Pfaff, 1975; 1976; Poplack, 1978/81).

The first more general account of the distribution of CS stemmed from the observation that CS is favored at the kinds of syntactic boundaries which occur in both languages. The

equivalence constraint (Poplack 1978/81; 1980) states that switched sentences are made up of concatenated fragments of alternating languages, each of which is grammatical in the language of its provenance (see also Lipski, 1977; Muysken, 2000; Pfaff, 1979). The boundary between adjacent fragments occurs between two constituents that are ordered in the same way in both languages, ensuring the linear coherence of sentence structure without omitting or duplicating lexical content.

That general principles, rather than atomistic constraints, govern CS is now widely accepted, though there is little consensus as to what they are or how they should be represented. Much current research assumes unquestioningly that the mechanisms for language switching follow directly from general principles of (monolingual) grammar. Theories based on this assumption tend to appeal to such abstract grammatical properties as inter-constituent relationships (e.g., government, case assignment) and/or language-specific features of lexical categories (i.e., subcategorization of grammatical arguments, inherent morphological features).

Since Klavans' (1985) proposal that CS was constrained by structural relations, the formal linguistic theories successively in vogue have each been extended to encompass the data of CS. Di Sciullo et al. (1986), for example, identified the relevant relations as C-command and government: CS cannot occur where a government relation holds. Replacement of the function of government in standard theory (Chomsky, 1995) by the notion of feature agreement led to a parallel focus on feature matching in CS studies. The *Functional Head Constraint* (Belazi et al., 1994; Rubin and Toribio, 1995) adds language choice to the features instantiated in functional and lexical categories, prohibiting CS where a mismatch occurs. A more recent Minimalist proposal (MacSwan, 1999) restricts CS at structural sites showing cross-language differences in monolingual features.

This distinction between lexical and functional categories is not new to CS research. It is a hallmark of theories invoking the complement structure of individual lexical items to characterize permissible CS sites (e.g., Joshi, 1985) and its sequel, the *Null Theory of CS* (Mahootian, 1993; Santorini and Mahootian, 1995); see also Bentahila and Davies' *Subcategorisation Constraint* (1983). Perhaps the most detailed model involving the contrast between lexical properties and functional (or "system") morphemes is the *Matrix Language Frame* model (Azuma, 1993; Myers-Scotton, 1993). Here, structural constraints on CS result from a complex interaction between a dominant matrix language and the prohibition against embedding "system" morphemes from the "embedded" language in matrix language structure.

The assumption that bilingual syntax can be explained by general principles inferred from the study of monolingual grammar has not yet been substantiated. While formal theories of grammar may account well for monolingual language structure, including that of the monolingual fragments in CS discourse, there is no evidence to suggest that the *juxtaposition* of two languages be explained in the same way. Bilingual communities exhibit widely different patterns of adapting monolingual resources in their code-mixing strategies, and

these are not predictable through purely linguistic considerations (Poplack, 1987). The equivalence constraint, as formalized by Sankoff and Mainville (1986) and Sankoff (1998a; 1998b), is a production-based explanation of the facts of CS, which incorporates the notions of structural hierarchy and linear order, and accounts for a number of empirical observations in addition to the equivalent word order characterizing most actual switch sites. These include the well-formedness of the monolingual fragments, the conservation of constituent structure, and the essential unpredictability of CS at any potential CS site. The mechanisms of monolingual and bilingual grammars are not assumed a priori to be identical.

Evaluating CS theories

There has been remarkably little cross-fertilization among CS theories; indeed, each has been greeted with a host of counter-examples. Testing the fit of competing models against the data of CS should be a straightforward matter since they often make competing predictions. But their disparate assumptions, goals and domains of application have hindered such efforts. Assessment of the descriptive adequacy of a theory of CS requires that at least two methodological issues be resolved. One involves classification of other-language phenomena, the other, confronting the predictions of the theory with the data of actual bilingual behavior.

It is uncontroversial that CS differs from the other major manifestation of language contact: *lexical borrowing*. Despite etymological identity with the donor language, established loanwords assume the morphological, syntactic, and often, phonological, identity of the recipient language. They tend to be recurrent in the speech of the individual and widespread across the community. The stock of established loanwords is available to monolingual speakers of the recipient language, who access them normally along with the remainder of the recipient-language lexicon. Loanwords further differ from CS in that there is no involvement of the morphology, syntax or phonology of the lexifier language.

Recent research has shown that borrowing is actually much more productive than implied above (the papers in Poplack and Meechan, 1998b). In particular, the social characteristics of recurrence and diffusion are not always satisfied. This results in what has been called, after Weinreich (1953/1968), *nonce borrowing* (Poplack et al., 1988; Sankoff et al., 1990). Like its established counterpart, the nonce loan tends to involve lone lexical items, generally major-class content words, and to assume the morphological, syntactic, and often, phonological identity of the recipient language. Like CS, on the other hand, nonce borrowing is neither recurrent nor widespread, and necessarily requires a certain level of bilingual competence. Distinguishing nonce borrowings from single-word CS is conceptually easy but methodologically difficult, especially when they surface bare, giving no apparent indication of language membership.

The classification of lone items is at the heart of a fundamental disagreement among CS researchers over 1) whether the distinction between CS and borrowing should be formally recognized in a theory of CS, 2) whether these and other manifestations of language contact can be unambiguously identified in bilingual discourse, and 3) criteria for determining

whether a given item was switched or borrowed. Researchers who consider lone other-language items to be CS tend to posit an *asymmetrical* relationship, in which one language dominates and other-language items are inserted (e.g., Joshi, 1985; Myers-Scotton, 1993; Rivas, 1981). Where the class of CS is (in the first instance) limited to unambiguous *multiword* fragments, both languages are postulated to play a role (Belazi et al., 1994; Sankoff, 1998a; 1998b; Sankoff and Poplack, 1981; Woolford, 1983). Muysken (2000) admits the possibility of both strategies.

The appropriateness of data is also relevant to evaluating CS theories. The literature on CS is largely characterized by the “rule-and-exception” paradigm. Despite the onslaught of counter-examples provoked by successive CS theories, very few have in fact been systematically tested against the data of spontaneous bilingual usage. Instead, both the theories and tests of their applicability tend to be based on isolated examples, drawn from judgements, informant elicitation and linguist introspection. The relation between such data and actual usage is not known; nor do they permit us to distinguish between the recurrent and systematic patterns of everyday interaction and examples which may be judged “acceptable” in some sense, but which rarely or never occur.

The equivalence constraint has been verified as a general tendency in Spanish-English (Poplack, 1978/81; 1980); Finnish-English (Poplack et al., 1987), Arabic-French (Naït M'Barek and Sankoff, 1988), Tamil-English (Sankoff et al., 1990), Fongbe-French (Meechan and Poplack, 1995), Wolof French (Poplack and Meechan, 1995), Igbo-English (Eze, 1998), French-English (Turpin, 1998) and Ukrainian-English (Budzhak-Jones, 1998) bilingual communities. But most of the voluminous literature on CS, especially of the “insertional” type, is based on data which represents, properly speaking, lexical borrowing. As only the grammar and word order of the recipient language is pertinent to borrowing, attempts to understand the structure of CS based on a mixture of borrowing and true CS (e.g., Myers-Scotton, 1993 and many others) appear unwieldy or descriptively inadequate.

Identifying the results of language contact

Insofar as CS and borrowing are based on some principled combination of elements of the monolingual (i.e. *unmixed*) vernaculars of the bilingual community, it is important to have as explicit an idea as possible of the nature of these vernaculars before concluding that a code-mixed element is behaving like one or the other. The analysis of code-mixing as a discourse mode requires access to the grammars of the contact languages as they are spoken, and spoken language is characterized by structural variability. In confronting, rather than evading this variability, Sankoff et al. (1990) and Poplack and Meechan (1998a) developed a method to compare bilingual structures with the unmixed source languages of the same speakers. Making use of the framework of linguistic variation theory (Labov, 1969; Poplack, 1993), the inherent variability of such forms is used to determine their status. If the rate and distribution of, for example, case-marking of the contentious lone other-language items show quantitative parallels to those of their counterparts in the (unmixed) recipient language, while at the same time differing from relevant patterns in the donor language, the lone other-language items are

inferred to be *borrowed*, since only the grammar of the recipient language is operative. If they pattern with their counterparts in the (unmixed) donor language, while at the same time differing from the patterning in the unmixed recipient language, the lone other-language items must result from CS.

Quantitative analysis of language mixing phenomena in typologically distinct language pairs shows that lone other-language items, especially major-class content words, are by far the most important component of mixed discourse (Berk-Seligson, 1986; Nortier, 1989; Poplack et al., 1988; Treffers-Daller, 1994). These lone items show the same fine details of quantitative conditioning of phonological, morphological and syntactic variability as dictionary-attested loanwords, both of which in turn parallel their unmixed counterparts in the recipient language (Poplack and Meechan, 1998b). This tendency is apparent regardless of the linguistic properties of the language pair. This is evidence that most lone items are borrowed, even if only for the nonce, despite the lack, in some cases, of dictionary attestation or diffusion within the community.

Future directions

Lack of consensus characterizing the discipline is related to a number of methodological problems. Foremost among them is failure to distinguish code-switching from other types of language mixture, which, despite similarities in surface manifestation, are fundamentally different mechanisms for combining languages. The current state of knowledge suggests that borrowing, nonce or established, is the major manifestation of language contact in most bilingual communities. Its linguistic structure is well accounted for in the traditional language contact literature (Haugen, 1950; Weinreich, 1953/1968). Intra-sentential CS involving multiword fragments of two or more languages is also attested in some communities. Achievement of consensus on an empirically verifiable characterization of the rules for juxtaposing these fragments within the sentence remains an important goal for CS research. Fit between theories and data could be improved by a broader empirical base. This would permit researchers to situate bilingual behavior with respect to the monolingual vernaculars implicated in language mixing, account for the disparate CS strategies that have evolved in different bilingual communities, and distinguish among incommensurable manifestations of bilingual language contact.

References

- Azuma, S. 1993. The frame-content hypothesis in speech production: evidence from intrasentential code switching. *Linguistics* 31: 1071-93.
- Belazi, H.M., E.J. Rubin, and A.J. Toribio. 1994. Code switching and X-bar theory: The functional head constraint. *Linguistic Inquiry* 25 (2): 221-37.
- Bentahila, A., and E.E. Davies. 1983. The syntax of Arabic-French code-switching. *Lingua* 59 (4): 301-30.
- Berk-Seligson, S. 1986. Linguistic constraints on intrasentential code-switching: A study of Spanish/Hebrew bilingualism. *Language in Society* 15 (3): 313-48.

- Budzhak-Jones, S. 1998. Against word-internal code-switching: Evidence from Ukrainian-English bilingualism. *International Journal of Bilingualism* 2 (2): 161-82.
- Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: The M.I.T. Press.
- di Sciullo, A.-M., P. Muysken, and R. Singh. 1986. Government and code-mixing. *Journal of Linguistics* 22 (1): 1-24.
- Eze, E. 1998. Lending credence to a borrowing analysis: Lone English-origin incorporations in Igbo discourse. *International Journal of Bilingualism* 2 (2): 183-201.
- Gumperz, J.J. 1976/1982. Conversational code-switching. In J. J. Gumperz (ed.), *Discourse Strategies*. Cambridge: Cambridge University Press, 59-99.
- Haugen, E. 1950. The analysis of linguistic borrowing. *Language* 26 (2): 210-31.
- Joshi, A.K. 1985. Processing of sentences with intrasentential code-switching. In D. R. Dowty, L. Karttunen, and A. M. Zwicky (eds), *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*. Cambridge: Cambridge University Press, 190-204.
- Klavans, J.L. 1985. The syntax of code-switching: Spanish and English. In L. D. King, and C. A. Maley (eds), *Selected Papers From the XIIIth Linguistic Symposium on Romance Languages*. Amsterdam: Benjamins, 213-21.
- Labov, W. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45 (4): 715-62.
- Labov, W. 1972. *Language in the Inner City*. Philadelphia: University of Pennsylvania Press.
- Lance, D. 1975. Spanish-English code-switching. In E. Hernandez-Chavez, A. D. Cohen, and A. F. Beltramo (eds), *El Lenguaje de los Chicanos*. Washington: Center for Applied Linguistics, 138-53.
- Lipski, J. 1977. Code-switching and the problem of bilingual competence. In M. Paradis (ed.), *Aspects of Bilingualism*. Columbia, SC: Hornbeam Press, 250-64.
- MacSwan, J. 1999. *A Minimalist Approach to Intrasentential Code Switching*. New York: Garland Press.
- Mahootian, S. 1993. *A Null Theory of Codeswitching*. Ph.D. dissertation, Northwestern University.
- Meechan, M., and S. Poplack. 1995. Orphan categories in bilingual discourse: Adjectivization strategies in Wolof-French and Fongbe-French bilingual discourse. *Language Variation and Change* 7 (2): 169-94.
- Muysken, P.C. 2000. *Bilingual Speech: A typology of Code-Mixing*. Cambridge, MA: Cambridge University Press.
- Myers-Scotton, C. 1993. *Duelling Languages*. Oxford: Clarendon Press.
- Naït M'Barek, M., and D. Sankoff. 1988. Le discours mixte arabe/français: des emprunts ou des alternances de langue? *Revue Canadienne de Linguistique* 33 (2): 143-54.
- Nortier, J. 1989. *Dutch and Moroccan Arabic in Contact: Code-Switching Among Moroccans in the Netherlands*. Ph.D. dissertation, University of Amsterdam.
- Pfaff, C. 1975. Syntactic constraints on code-switching: A quantitative study of Spanish/English. Paper presented at 50th Annual Meeting of Linguistic Society of America.
- Pfaff, C. 1979. Constraints on language mixing. *Language* 55 (2): 291-318.

- Poplack, S. 1978/81. Syntactic structure and social function of code-switching. In R. P. Duran (ed.), *Latino Discourse and Communicative Behaviour*. New Jersey: Ablex Publishing Corporation, 169-84.
- Poplack, S. 1980. Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching. *Linguistics* 18 (7/8): 581-618.
- Poplack, S. 1987. Contrasting patterns of code-switching in two communities. In E. Wande, J. Anward, B. Nordberg, L. Steensland, and M. Thelander (eds), *Aspects of Bilingualism: Proceedings from the Fourth Nordic Symposium on Bilingualism, 1984*. Uppsala: Borgström, Motala, 51-76.
- Poplack, S. 1993. Variation theory and language contact. In D. Preston (ed.), *American Dialect Research*. Amsterdam: Benjamins, 251-86.
- Poplack, S., and M. Meechan. 1995. Patterns of language mixture: Nominal structure in Wolof-French and Fongbe-French bilingual discourse. In L. Milroy, and P. Muysken (eds), *One Speaker, Two Languages*. Cambridge: Cambridge University Press,
- Poplack, S., and M. Meechan. 1998a. How languages fit together in codemixing. *International Journal of Bilingualism* 2 (2): 127-38.
- Poplack, S., and M. Meechan (eds). 1998b. *Instant Loans, Easy Conditions: The Productivity of Bilingual Borrowing; Special Issue of the International Journal of Bilingualism*. London: Kingston Press.
- Poplack, S., D. Sankoff, and C. Miller. 1988. The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26 (1): 47-104.
- Poplack, S., S. Wheeler, and A. Westwood. 1987. Distinguishing language contact phenomena: Evidence from Finnish-English bilingualism. In P. Lilius, and M. Saari (eds), *The Nordic Languages and Modern Linguistics*. Helsinki: University of Helsinki Press, 33-56.
- Rivas, A. 1981. *On the Application of Transformations to Bilingual Sentences*. Amherst: University of Massachusetts: Department of Spanish and Portuguese.
- Rubin, E.J., and A.J. Toribio. 1995. Feature-checking and the syntax of language contact. In J. Amastae, G. Goodall, M. Montalbetti, and M. Phinney (eds), *Contemporary Research in Romance Linguistics*. Amsterdam and Philadelphia: John Benjamins, ??
- Sankoff, D. 1998a. A formal production-based explanation of the facts of code-switching. *Bilingualism: Language and Cognition* 1 (1): 39-50.
- Sankoff, D. 1998b. The production model of code-mixed discourse. In *COLING-ACL '98: Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics*. Montreal: University of Montreal Press, 8-21.
- Sankoff, D., and S. Mainville. 1986. Un modèle pour l'alternance de langue sous la contrainte d'équivalence. *Revue Québécoise de Linguistique* 15: 233-46.
- Sankoff, D., and S. Poplack. 1981. A formal grammar for code-switching. *Papers in Linguistics: International Journal of Human Communication* 14 (1): 3-45.
- Sankoff, D., S. Poplack, and S. Vanniarajan. 1990. The case of the nonce loan in Tamil. *Language Variation and Change* 2 (1): 71-101.
- Santorini, B., and S. Mahootian. 1995. Code-switching and the syntactic status of adnominal adjectives. *Lingua* 96: 1-27.

- Timm, L.A. 1975. Spanish-English codeswitching: El porqué y how-not-to. *Romance Philology* 28 (4): 473-82.
- Treffers-Daller, J. 1994. *Mixing Two Languages: French-Dutch Contact in a Comparative Perspective*. Berlin: Mouton de Gruyter.
- Turpin, D. 1998. "Le français c'est le last frontier": The status of English-origin nouns in Acadian French. *International Journal of Bilingualism* 2 (2): 221-33.
- Weinreich, U. 1953/1968. *Languages in contact*. The Hague: Mouton.
- Woolford, E. 1983. Bilingual code-switching and syntactic theory. *Linguistic Inquiry* 14 (3): 519-36.

Shana Poplack
University of Ottawa

This research was supported by the Social Sciences and Humanities Research Council of Canada. The author is a Canada Research Chair in Linguistics.