

PEER COMMENTARY

What does the Nonce Borrowing Hypothesis hypothesize?

SHANA POPLACK
University of Ottawa

In a climate where many researchers are content to characterize the linguistic manifestations of language contact using ad hoc criteria or anecdotal data, Jonathan Stammers and Margaret Deuchar (S&D) are to be commended for undertaking an empirical study with speakers, data and an accountable analysis. S&D (this issue) are right to say that the classification of lone other-language items (LOLIs) has been dogging the study of code-switching for decades. They use an interesting diagnostic, soft mutation, novel in this field, a typologically distinct language pair, and quantitative methodology to test the Nonce Borrowing Hypothesis (NBH). They find that soft mutation on English-origin verbs increases as a function of dictionary attestation (“listedness”), and frequency of occurrence. They interpret their findings to “unequivocally refute” the NBH (under section heading: “Discussion”), and conclude that the “category of nonce borrowings is redundant” (under section headings: “Discussion” and “Conclusion”). Not surprisingly, we disagree that this nullifies the NBH or disposes of the phenomenon of nonce borrowing itself. On the contrary, their results are consistent with those that have emerged from the nonce borrowing studies they critique, to wit: morphosyntactic integration of BORROWED items is largely abrupt and categorical, whereas phonological integration is often gradual and highly variable.

The Nonce Borrowing Hypothesis: What it is and what it is not

Before outlining the reasons why S&D’s findings do not refute the NBH, it will be useful to summarize what the NBH hypothesizes, and the nature of the evidence that has been brought to bear on it. Poplack, Sankoff and Miller (PSM, 1988) were the first to address the problem of distinguishing borrowing from code-switching of LONE other-language items head-on through analysis of a vast corpus of bilingual French–English discourse. Collected from a representative sample of 120 unrelated members of the bilingual speech community of Ottawa–Hull (Canada),

it contains close to 20,000 naturally-occurring LOLIs. This unprecedented body of data in this line of inquiry (to our knowledge still unmatched in quality or quantity) enhances the confidence we can have in their results.

Working under the accepted twin assumptions of the era that (i) bona fide “loanwords” were fully integrated into recipient-language discourse linguistically and socially, and (ii) code-switches were entirely unintegrated, PSM were intrigued when empirical analysis turned up substantial numbers of (unambiguous) loanwords and code-switches, each with the expected characteristics of the other. This prompted extended investigation of the other-language material, with a view to determining the characteristics of borrowing and code-switching, as actually instantiated in the speech of bilinguals. PSM were the first to document a finding which has since been replicated in many other bilingual situations and language pairs: the very disproportionate distribution of LOLIs ($N = 19,579$), vs. MULTIWORD fragments of the other language ($N = 1,766$). The latter are generally unambiguous code-switches; the former may be code-switches or borrowings. Reasoning that by their numerical preponderance alone, classing these LOLIs A PRIORI as one or the other would bias any analysis of either code-switching or borrowing to reflect only or largely the characteristics of LOLIs, as for example, with the Matrix Language Frame model (Myers-Scotton, 1993), PSM undertook a large-scale, multi-pronged analysis specifically of the behavior of such items. The sheer size of the database enabled them to eschew arbitrary criteria (e.g. the “three-occurrence” metric for the loanword/code-switch divide (Jones, 2005; Myers-Scotton, 1993)) and risky inferences (e.g. that dictionary-attested LOLIs are frequent) in favor of objective quantitative measures of frequency and diffusion – IN THE COMMUNITY ACTUALLY UNDER STUDY. They classified LOLIs according to four basic frequency categories, ranging from “widespread” (= over 10 occurrences by a total of over 10 speakers) at one extreme, and “nonce” (= 1 occurrence, 1 speaker) at the other, and analyzed the behavior of each cohort from four independent perspectives: linguistic integration

Address for correspondence:

Department of Linguistics, University of Ottawa, 401–70 Laurier Avenue East, Ottawa, ON, K1N 6N5 Canada
spoplack@uottawa.ca

into recipient-language grammar, diffusion across the community, dictionary attestation and age of attestation. These had all been invoked by various scholars as measures of (unambiguous) loanword status.

Most relevant to S&D's claim is that on a variety of measures of morphological and syntactic integration, the overwhelming majority of NONCE LOLIs, OPERATIONALLY defined as uttered once by a single speaker (of 120), featured most of the characteristics of widespread and/or dictionary-attested loanwords, which in turn behaved like their counterparts in the recipient language. This led PSM to formulate the NBH, which captures the empirical observation that speakers not only code-switch spontaneously, but may also BORROW spontaneously, and these spontaneous borrowings assume the morphological and syntactic identity of the recipient language even PRIOR to achieving the social characteristics of established loanwords (recurrence in the speech of the individual, and dispersion across the community). The same quantitative analysis also showed that although morphosyntactic integration occurs at, or soon after, the stage of spontaneous borrowing, PHONOLOGICAL integration is gradual, proceeding as a function of degree of diffusion and age of attestation of the LOLI (PSM, 1988, pp. 72ff.). Importantly, however, the phonology remains VARIABLE at both end points of the elapsed interval, the time of first attestation and the time of the study.

PSM also observed that while the overwhelming majority of LOLIs were integrated in this way, a small minority was not: "Not all of the single-item forms can be expected to represent established loanwords, nor even nonce borrowings, since single-word code-switches are theoretically possible" (p. 53). Furthermore, "some of these nonce tokens may properly belong to the conceptually distinct category of single-word code-switches" (p. 95). This fact was explicitly reiterated in Poplack and Meechan (1998): "Some lone other-language items are not borrowings" (p. 135), and "Lone other-language items are not necessarily code-switches" (p. 133), as well as in a number of subsequent studies. This should have lain to rest the persistent misapprehension (also espoused by S&D) that the NBH predicts or requires that ALL LOLIs be considered nonce borrowings. It does not. Nor does it rule out single-word code-switches.

Which LOLIs are code-switches and which ones are borrowings? This is an empirical question. Sankoff, Poplack and Vanniarajan (1990) and subsequent studies of a variety of typologically different language pairs developed a highly ramified methodology to help make this determination in as many cases as possible. Essentially, it is based on triangulation of the quantitative patterning of LOLIs on some diagnostic with the patterns displayed by other types of language mixing whose status is not in question: unambiguous (multiword) code-switches, dictionary-attested loanwords, and most

tellingly, the monolingual, i.e. UNMIXED vernaculars entering into the contact situation.

S&D deploy this methodology to compare the application of soft mutation in unlisted LOLIs to dictionary-attested loanwords and native Welsh verbs. But their analysis does not test (nor by extension, refute) the NBH. It tests their own REFORMULATION of the NBH: "There is no difference between frequent and infrequent donor-language items in terms of their degree of integration" (under section heading: "The nonce borrowing hypothesis"). Their findings – a difference in rates of mutation on frequent and infrequent verbs – do refute this reformulation, but they are not pertinent to the NBH. The following sections detail why.

The role of frequency in the Nonce Borrowing Hypothesis

We of course endorse S&D's characterization of our position that nonce borrowings and established loans can be distinguished by frequency of use – indeed, they were operationally DEFINED by their differing frequencies of use. But this is not tantamount to saying that "there is no difference between frequent and infrequent DONOR-LANGUAGE ITEMS in terms of their degree of integration". That formulation implies, erroneously, that the NBH considers ALL LOLIs to be borrowings, without operationally distinguishing nonce borrowings from widespread and/or dictionary-attested loanwords, and none to be code-switches. This is not the case. Code-switches – single- OR multi-word – by definition are NOT integrated into a recipient language, unlike established loanwords (and nonce borrowings). Like nonce borrowings, however, specific code-switches are generally non-recurrent (and IPSO FACTO infrequent) and unlisted. By virtue of their diametrically opposed behavior in terms of linguistic integration alone, contamination of the unlisted/infrequent categories of LOLIs by unintegrated single-word code-switches (since these were not distinguished) could by itself account for some of the lower rates of soft mutation (henceforth SM) in these categories. Indeed, S&D entertain this possibility themselves (under section heading: "Discussion"), but are still not averse to treating code-switching and borrowing as one in order to refute their version of the NBH.

Frequency has been integral to the formulation of the NBH from the outset. It is measured in one of two ways: frequency of the WORD (token frequency) to which some process (e.g. mutation) applies, and frequency (rate) of APPLICATION of that process. PSM used both, as do S&D. Determining word frequency is by no means as simple as S&D imply. A word categorized as "frequent" by S&D's criteria could have simply been "idiosyncratic" by PSM's: used by only ONE speaker, even if recurrently. A listed word could show up as a nonce form in a corpus, while many unlisted forms are both frequent and

widespread (PSM, 1988, pp. 57–59). Token frequency is most meaningful when combined with DIFFUSION. PSM employed the rigorous requirement that a LOLI be used spontaneously more than 10 times BY MORE THAN 10 SPEAKERS to qualify as “widespread”, while NONCE LOLIs were uttered once by a single speaker. Based on these exacting metrics, PSM were able to arrive at the very robust and significant finding (nicely confirmed by the present study) that word frequency and diffusion were relevant, but only to phonological, and NOT to morphosyntactic, integration. This explains why subsequent studies made use of morphological, syntactic and lexical diagnostics, but did not explicitly analyze either word frequency or phonology.

Frequency of APPLICATION, on the other hand, is the dependent variable in every study carried out by Poplack and associates, as well as in S&D’s. But despite S&D’s interpretation, this is neither the only, nor even the main, diagnostic of language MEMBERSHIP used in Poplack and Meechan (1998) and other studies testing the NBH.

Measuring integration: Conditioning vs. rates

Poplack and Meechan (1998, p. 130) observe that to the extent that rates of application are language-specific, they may be useful in assessing language membership. But these must be interpreted judiciously, they caution, since differences in the rate of variant selection (here SM vs. no SM) may be masking disproportionate, but unrelated, distributions in context, linguistic (e.g. eligible environments) or extra-linguistic (e.g. speech style, bilingual ability). Among potential confounding factors in S&D’s study, the distribution of consonants eligible for SM (since /k/ is much more favorable than /b/ (under section heading: “Syntactically triggered mutation”)) and the language dominance of the speaker (since their goal was to obtain a wide range (under section heading: “The nonce borrowing hypothesis”)) could be problematic. Indeed, PSM (1988, p. 74) documented a clear effect of bilingual proficiency on level of (phonological) integration.

The CONDITIONING, or the factors CONTRIBUTING to variant selection, are deeper constraints that remain constant regardless of the extra-linguistic circumstances. To the extent that they are language-specific, they suggest which grammar most likely gave rise to the LOLIs. To take but one example, Sankoff, Poplack and Vanniarajan’s (1990) study of integration of lone English-origin nouns in Tamil showed that only a third of them are case-marked in the accusative, but almost all are case-marked in the dative. Why should so many more English-origin accusatives remain ostensibly unintegrated than English-origin datives, when English supplies no rules for case-marking either? By analyzing the unmixed Tamil of THE SAME speakers, Sankoff, Poplack and Vanniarajan discovered that this is the NATIVE pattern of variable case-

marking. The case constraint shows that Tamil grammar is operating on BOTH English-origin and native Tamil nouns. THIS is the evidence underlying their conclusion that (some of) the English-origin ones were borrowed, if only for the nonce. Thus, identity of the case is one of the factors implicated in variant selection – whether a LOLI will receive Tamil case-marking. Rather than simply assume, as do S&D and many other scholars of language contact, that lone English-origin nouns with zero morphology FAILED to undergo integration into the recipient language, examination of the CONDITIONING of variable case-marking in donor and recipient languages revealed that even the selection of morphological ZERO was constrained in the same way in both native Tamil and English-origin material. Although (striking) parallels in rate were also observed, this is much stronger evidence of integration.

Now, SM is also highly variable in Welsh; S&D’s Table 5 shows that it occurs only 73% of the time overall on native verbs, and even less when those verbs are infrequent. What motivates a speaker to apply SM – to English-origin OR native Welsh verbs? The answer resides in the variable conditioning of variant choice. If, as in the Tamil case and many others, the factors that contribute to the selection of SM in the former parallel those operating in the latter, the English-origin verbs are being treated grammatically as Welsh, EVEN IF AT A LOWER RATE.

S&D do in fact invoke two potential conditioning factors (constraints) – type of trigger and identity of the consonant to be mutated. They report that only the latter is statistically significant, with substantially higher SM rates when the consonant is /k/ as opposed to /b/. But because they treat all of native Welsh, English listed and English unlisted verbs TOGETHER in determining significance (S&D’s Table 6, Figure 4), we cannot recover, and certainly not compare, the effect these conditioning factors might have had on SM in each.

A third conditioning factor is of course word frequency. Here S&D DO separate out the different cohorts of verbs. But in interpreting their finding that the incidence of SM on lone English-origin verbs rises as a function of word frequency, we must factor in the crucial information that word frequency conditions the occurrence of SM in NATIVE Welsh verbs in the same way (S&D’s Table 5, Figure 3). The obvious alternative to their interpretation is that the correlation between increase in SM as a function of frequency of the LOLI is simply a reflection of this same effect in Welsh. This would be a sign that not only those LOLIs that underwent SM but also those that did not were INTEGRATED into the recipient language, rather than the opposite, as S&D claim.

Type of diagnostic: Morphosyntax vs. phonology

Independent of the above considerations, perhaps the key reason why S&D’s finding does not refute the NBH

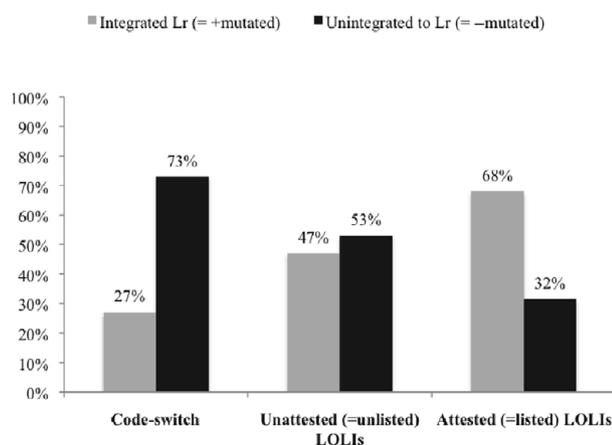
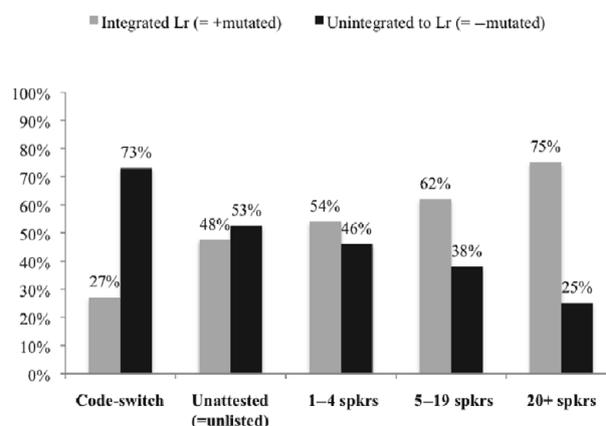


Figure 1. Average phonological integration of lone other-language items (LOLIs) according to LISTEDNESS in the Ottawa-Hull French corpus (adapted from Figure 7 in Poplack et al., 1988; Lr = recipient language).

resides in the nature of the evidence they bring to bear, and in particular, in the nature of SM as a diagnostic. S&D assert that SM is not “merely a phonological phenomenon” (under section heading: “English verbs in Welsh: Integration by soft mutation”), but part of the grammar of Welsh. Elsewhere they refer to it as a morphophonological process. Some scholars regard SM as phonetic, others as phonemic. Whatever the characterization, it is clear that although mutation may be TRIGGERED by the relevant morphological, lexical and/or syntactic environments, it manifests PHONETICALLY, as lenition of eligible initial consonants, and does not involve alteration of the morphology or syntax.¹ S&D’s finding that SM increases as a function of the frequency and listedness of the LOLI is exactly what PSM (1988) found with respect to phonological integration more generally, on the basis of a variety of phonological diagnostics, almost 40 times more data, and objectively-established frequency and attestation categories. Compare Figures 1 and 2, adapted from Figures 7 and 8 in PSM (1988, pp. 73–74), with S&D’s Figures 1 and 3: phonological integration is variable, increasing with frequency and listedness of the LOLI.

It matters that the unambiguous code-switches to English included in both calculations ALSO show variable phonological “integration”. In other words, even when speaking English, words are sometimes rendered in

¹ In this sense, SM can be likened to plural marking on LOLIs in otherwise French discourse, to which S&D also refer. Although triggered semantically, it manifests itself phonetically. Plural marking also showed gradual integration according to frequency (PSM, 1988) and parenthetically, the LOLIs examined by Turpin (1998) also contained some (independently identified) overtly plural-marked single-word code-switches to English.



Note: Although Figure 2 is couched in terms of diffusion, i.e. the number of SPEAKERS who said the LOLI, it can be interpreted in terms of frequency: a word said by a given number of speakers had to have been said at least that number of times (and more if the word was recurrent in the speech of an individual), while a word said once by one person is just that.

Figure 2. Average phonological integration of lone other-language items (LOLIs) according to FREQUENCY in the Ottawa-Hull French corpus (adapted from Figure 7 in Poplack et al., 1988; Lr = recipient language).

French phonology. EACH of the categories – listed, unlisted, frequent, nonce, AS WELL AS CODE-SWITCHES TO ENGLISH – show some phonological integration and some failure to integrate. Phonological integration is GRADIENT. S&D conclude that SM is gradient too, but extrapolate from that finding that nonce borrowing is itself redundant. This does not follow. Rather, as stated in PSM (1988), and many times thereafter, phonological integration is a poor diagnostic for code-switching or borrowing status, precisely BECAUSE it occurs variably in all of them. Therefore the finding that phonological integration increases with frequency does not refute the NBH, only S&D’s reformulation of it.

In contrast, morphosyntactic integration is a nearly fail-safe indicator that a LOLI has been borrowed. Interestingly, this is exactly what S&D found too: 97.3% of the verbs in Table 2 were ALREADY morphologically integrated, since they are inflected with *-io*, and this, regardless of frequency. This is the MORPHOLOGICAL criterion, and it appears to apply to virtually all English-origin verbs in Welsh, listed and unlisted, just as PSM (1988) reported on a variety of grammatical diagnostics for LOLIs in French. This contradicts S&D’s assertion, “[s]ince we do not find infrequent items that pattern like frequent ones, we also do not find evidence for the category of nonce borrowings” (under section heading: “Discussion”).

Nonce borrowings were never intended as, nor claimed by us to be, a linguistically distinct category from other types of lexical borrowings. On the contrary, we have noted repeatedly that in terms of grammatical

integration, nonce borrowings and established loanwords are virtually the SAME, any distinctions between them being social, in terms of diffusion across the community, and phonological, in terms of increasing integration as a function of frequency, diffusion and age of attestation. If anything, the nonce borrowing is a reinforcement of the distinction between CODE-SWITCHING and borrowing, at the level of lone other-language items. It may be redundant in S&D's reformulation of the NBH, but since we never considered it a distinct category, this does not apply to the original version of the NBH.

In closing, perhaps the most remarkable aspect of S&D's study is that it turns up the same sorts of results as other empirical studies, even if S&D interpret them otherwise. Morphological integration of BORROWED elements is abrupt, code-switched elements remain unintegrated. In both cases, phonological integration is variable. We stress that it is S&D's use of transparent empirical methodology that has enabled dialogue in this contentious area, regardless of whether or not we agree on the interpretation. We can only applaud them for making that effort and hope that their example will encourage more scholars to follow suit by applying accountable tests

to the NBH and other areas of language mixing using different diagnostics and language pairs.

References

- Jones, M. C. (2005). Some structural and social correlates of single word intrasentential code-switching in Jersey Norman French. *French Language Studies*, 15, 1–23.
- Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon Press.
- Poplack, S., & Meechan, M. (1998). How languages fit together in code-mixing. *International Journal of Bilingualism*, 2 (2), 127–138.
- Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, 26 (1), 47–104.
- Sankoff, D., Poplack, S., & Vanniarajan, S. (1990). The case of the nonce loan in Tamil. *Language Variation and Change*, 2 (1), 71–101.
- Stammers, J., & Deuchar, M. Testing the nonce borrowing hypothesis. *Bilingualism: Language and Cognition*, doi:10.1017/S1366728911000381.
- Turpin, D. (1998). "Le français, c'est le last frontier": la structure nominale dans le discours bilingue français/anglais. MA thesis, University of Ottawa.