# How to mix

## Confronting "mixed" NP models and bilinguals' choices

Rena Torres Cacoullos[1], Nathalie Dion[2], Dora LaCasse[3], and Shana Poplack[2]
[1] Penn State University | [2] University of Ottawa | [3] University of Montana

The widespread occurrence of nouns in one language with a determiner in the other, often referred to as mixed NPs, has generated much theorizing and debate. Since both a syntactic account based on abstract features of the determiner and an account highlighting the notion of a Matrix language yield largely the same predictions, we assess how the tenets of each play out in speaker choices. The data derive from a massive corpus of spontaneous nominal mixes, produced by bilinguals in New Mexico, where bidirectional code-switching is the norm. Bilinguals' choices concern (1) NP status (mixed vs. unmixed); (2) mixing type (limited-item vs. multi-word); and (3) language of the noun (here, English vs. Spanish). Results show that the community preference is for mixed NPs, independent of their theoretical felicity as dictated by determiner language properties. As to mixing type, these NPs are mostly constituted of lone nouns, such that the language of the determiner and any associated verb is perforce that of the surrounding discourse. Finally, the overwhelming choice is for English lone nouns incorporated into Spanish, and hence for a Spanish determiner. The language of the determiner thus proceeds, not from abstract linguistic properties, but instead from straightforward adherence to bilingual speech community conventions.

**Keywords:** mixed NP, classes of nominal mixes, lone other-language nouns, multi-word code-switching, bilingual community norms

### 1. Introduction

There is considerable debate amongst theoreticians of language contact over what motivates the choice of the language of the determiner when it is not realized in the same language as the noun, in configurations variously referred to as, for example,

https://doi.org/10.1075/lab.20017.tor | Published online: 19 February 2021
*Linguistic Approaches to Bilingualism* issn 1879-9264 | e-issn 1879-9272

"determiner/noun code-switches" (Parafita Couto & Stadthagen-Gonzalez, 2019), "codeswitched determiner-noun sequences" (Herring, Deuchar, Parafita Couto & Moro Quintanilla, 2010), or simply MIXED NPS (Jake, Myers-Scotton & Gross, 2002, p. 72), the working term we adopt here.

Proponents of a variety of formal theories have weighed in, some seeking to apply general principles of monolingual grammar, others appealing to models of bilingual language mixture. Prominent within the former approach are generative syntax accounts, which invoke the richness of the nominal agreement systems of the languages involved to argue that, as a functional, or closed-class, item, the determiner must be realized in the language with the stronger (uninterpretable *phi*) features. Accordingly, for English and Spanish, they predict that the determiner in mixed NPs will be Spanish because, in contrast to English determiners, Spanish determiners carry gender (e.g., Liceras, Fuertes, Perales, Pérez-Tattam & Spradlin, 2008; MacSwan, 2005, p. 18; Moro Quintanilla, 2014). Within the second approach, most notable are models that additionally assume an asymmetrical relationship between the languages, regarding one of them as the dominant or Matrix language that supplies the "abstract grammatical frame" in bilingual utterances (e.g., Jake et al. 2002, pp. 72–74). Per the Bilingual NP Hypothesis, the language of the determiner should therefore come from this language. In practical terms, the language of the determiner should agree with that of the finite verb (e.g., Parafita Couto & Stadthagen-Gonzalez, 2019, p. 351).

Thus, in the examples of mixed NPs below, the abstract feature (AF) account predicts (1) and (2), but not (3), because the determiner is English, while the Matrix language (ML) account predicts (1) and (3), but not (2), which has a Spanish determiner despite the English verb.[1]

(1) $DET_{[SP]} + N_{[EN]}$
jalaba **la *sweater*** pa' cá. [14, 45:47–45:48]
'she would pull the sweater this way.'

(2) $DET_{[SP]} + N_{[EN]}$ [DET-V mismatch]
.. *I .. pray t- with her* allá en **el *cemetery***, [23, 46:10–46:12]
'I pray t- with her there in the cemetery,'

1. All examples are from the New Mexico Spanish-English Bilingual (NMSEB) corpus (Torres Cacoullos & Travis, 2018, Chapters 2 and 3). Within brackets following examples is the recording number and the beginning-ending time stamps of the lines reproduced. In the examples, roman type font indicates Spanish speech, and italics, English. Mixed NPs are those in which the determiner [DET] and noun [N] (bolded) are not in the same font style (roman vs. italics).

(3) $DET_{[EN]} + N_{[SP]}$
*… what did* ***the*** **partera** *used to do*? [24, 34:01–34:03]
'what did the midwife used to do?'

As is typical in the field of bilingual language mixing, many other counter-examples have been documented for both approaches. The preference for English determiners in Nicaraguan Creole English-Spanish combinations, despite Spanish having the richer array of features (Blokzijl, Deuchar & Parafita Couto, 2017), or Papiamento determiners in Papiamento-Dutch mixed NPs, where Dutch has them (Parafita Couto & Gullberg, 2019, p. 703), to name but two, contravene the AF account, while examples such as (2) violate ML predictions, unless they are qualified as Embedded Language (EL) Islands or the Matrix language is reidentified midway through the clause.

In spite of their conflicting principles, however, both approaches find a good deal of support in the published data thus far examined. The two camps coincide in the observation that in mixed NPs there tends to be a preferred language for the determiner, even if they do not agree on the reason. For generative syntacticians it is linguistic, an idea that is supported by the preference for Spanish determiners in Spanish-English mixed NPs (e.g., Herring et al., 2010, p. 564; Liceras et al., 2008; Liceras, Spradlin & Fernández Fuertes, 2005; Moro Quintanilla, 2014; Ramírez Urbaneja, 2019; Valdés Kroff, 2016), German determiners in German-English combinations (Eppler, Duescher & Deuchar, 2017) or Welsh determiners in Welsh-English ones (Herring et al., 2010, p. 564). But these facts are also compatible with the ML position (provided that the language of the determiner matches that of the Matrix language). The findings of ML adherents for the Nicaraguan Creole English-Spanish and Papiamento-Dutch cases cited above lead them to posit that the preference is rather socially motivated, stemming from differential community selection of a Matrix language (Blokzijl et al., 2017, pp. 8–9). Explicit comparisons of the two models report that ML has a slight edge in terms of accuracy (Duran Eppler, Luescher & Deuchar, 2017, p. 47; Herring et al., 2010, p. 565; Parafita Couto & Stadthagen-Gonzalez, 2019, p. 357), while the AF account leads in coverage (or, "the extent of the data about which [a] theory makes predictions"; Herring et al., 2010, pp. 558, 565). But these differences are relatively minor. The overall success of both models in accounting for most of the data, while failing to account for the remainder, makes it difficult to choose between them.

Inspection of actual bilingual usage on the ground suggests a number of reasons for this impasse. One is that in most contact situations that have been systematically documented, language mixing tends to be *directional*: one language predominates, and other-language items are occasionally incorporated into

it. The directionality of such mixing -- *whether it coincides with a theoretically optimal configuration or not* -- may be for reasons independent of the linguistic properties of the languages or their determiners. This makes it difficult to know whether the reasons for observed DET+N language imbalances are linguistic as opposed to social or cognitive. A second problem results from disregarding the fact that speakers have a choice of whether or not to mix, as well as a choice of where to do so. As such, they are able (and can be expected) to *avoid* mixed NPs that result in a combination deemed infelicitous, opting for an *un*mixed other-language NP. In (3), for example, the predicted unmixed NP alternative would be *what did* **la partera** *used to do*?. Any test of the coverage and accuracy of a model should take these bilingual choices into account. Third, and perhaps most important, not all language mixing is created equal. Whereas in (1)–(3) the other-language noun is a lone item, in (4), for example, the noun initiates a multi-word code-switch. The literature on mixed NPs has generally regarded all other-language nouns as instances of code-switching (CS), despite general acknowledgement of a difference between using one versus more than one word from another language. Regardless of how they are labelled or construed, these types of language mixing have different properties. In what follows we show that distinguishing between them crucially impacts the accuracy and coverage of proposed models.

(4) .. en **los *weekends*** *they would get together and*, jugaban dados áhi no? [18, 43:55–43:48]
'.. on the weekends [...] they'd play dice there right?'

## 2. Data

As is so often the case in the widely investigated but sparsely actualized data of bilingual language mixing, many of the claims and tests characterizing the literature suffer from scant and/or insufficiently contextualized data, making them difficult to assess. In this study we return to the language pair that spawned the debate, Spanish-English. We draw our data from the New Mexico Spanish-English Bilingual (NMSEB) corpus (Torres Cacoullos & Travis, 2018, Chapters 2 and 3), a community-based compilation of sociolinguistic interviews (Labov, 1984). The bilinguals are Hispanic northern New Mexicans (Bills & Vigil, 2008) in the Southwest of the United States who use English and Spanish regularly in their daily interactions. The corpus comprises 31 recordings with 40 speakers totalling 29 hours, or 300,000 words, transcribed orthographically and prosodically. The prosodic transcription is based on the Intonation Unit (IU) (Du Bois, Schuetze-Coburn, Cumming & Paolino, 1993, p. 47); see Appendix for transcrip-

tion conventions. In drawing on prosodically-based transcription, we can objectively delimit sentences and grammatical units in speech, enabling us to establish the set of NPs most pertinent to the issues at hand (uncontaminated by cases of prosodically separated DETs, and identifying instances of unattached, or syntactically isolated, NPs).

In contrast to many other bilingual corpora, NMSEB is exceptionally rich in language mixing, not only of the expected single other-language nouns (Examples (1)–(3)), but also of the generally far sparser multi-word strings (4) (Torres Cacoullos & Travis, 2018, pp. 44–46, 181–182). Though CS has been described as occurring more frequently in some communities than others (e.g., Lipski, 2005, p. 2), interpretable comparisons require some reproducible measure of CS frequency, which has been hitherto lacking. Such a measure is established on the basis of the prosodic boundary as a major delimitor of CS. For the NMSEB corpus, the frequency of intra-sentential CS across IUs (as between the lines in (4)) is approximately 8% (out of 36,000 IUs); the rate of within-IU CS (as in the first line in (4)) is 2.5% (out of 38,000 IUs, where the the universe is the total number of IUs eligible to host CS) (Torres Cacoullos & Travis, 2020, p. 263).

It comes as no surprise then that NMSEB features the largest number of mixed NPs studied thus far in this connection ($n = 1195$; Figure 3) – approximately four times more than the corpus-based analyses of determiner language choice to date. More important for present purposes, not only are these data copious in number and diverse in classes of mixes, they are further exceptional in their bi-directionality, or balance. CS goes from English to Spanish and Spanish to English in about even proportions within sentences (as in main-and-complement clause complexes with CS at clause boundary [Torres Cacoullos, 2020, p. 8]). It is important that there are also nearly equal numbers of Spanish clauses as there are English (based on the language of the verb; Torres Cacoullos & Travis, 2018, p. 67) in which to potentially introduce mixed NPs. This contrasts with (what we infer to be) the case in the previous studies known to us.[2]

Bi-directionality is crucial to establishing whether the language of the determiner proceeds from linguistic properties or community norms, if only because, as noted above, the unidirectional mixing characterizing most bilingual datasets, in conjunction with the practice of conflating classes of mixes, may coincidentally yield the theoretically "optimal" configuration. That is, there simply is no opportunity to assess the likelihood of predicted structures in datasets in which mixing is generally unidirectional. Considering different mixes both from Spanish to

**2.** In a study of gender assignment and semantic domains in a large ($n = 4746$) data set of mixed NPs of English and Spanish in Northern Belize, "English/Spanish DPs" (e.g. the *hoja* 'the leaf') constitute a mere 0.4% of the "overall switched data set" (Balam, 2016, p. 420).

English *and* English to Spanish means that our data provide a more balanced starting point from which to assess the relative frequency of structures that are alternately felicitous and infelicitous for one or both of the frameworks under investigation. This in turn allows us to establish whether speakers avoid the non-predicted structures, an element that has heretofore been missing from the discussion.

## 3. Method

### 3.1 Framework

Our approach here is broadly variationist (Labov, 1969; Poplack, 2020). We construe the decision of *whether* to mix nouns from one language into another, as well *where* in the NP to do so, as *choices* speakers make in the course of their bilingual discourse. To investigate the factors contributing to this choice, we apply the Principle of Accountability, to count all instances in which the phenomenon of interest occurred as well as those where it *could have* occurred (Labov, 1969, p. 738, n. 20), which here involves contextualizing mixed NPs with respect to unmixed alternatives.

The focus for this study is DET+N sequences that are the locus of mixing in bilingual utterances. NOMINAL MIXES include *all nouns preceded by a determiner at the border between the two languages, regardless of whether they result in mixed NPs or not*. To best capture the various elements of the claims of these models, we refine the concept of nominal mixes into the following four classes, sorted according to status of the NP (mixed vs. unmixed) and mixing type (limited-item vs. multi-word).

### 3.2 Nominal mixes

The four-way distinction among NPs is illustrated in Table 1. NP status has to do with the choice between MIXED vs. UNMIXED NPs, according to whether the determiner and noun are realized in (a) different languages or (b) the same language, preserving linguistic cohesion within the NP. This is considered in two separate contexts, by what we label mixing type, according to the extent of the mixed element. For NOMINAL-LIMITED ITEMS, we distinguish between incorporations of (a) a LONE N of one language into a stretch of another language, thereby creating a mixed NP (5–6), or (b) the determiner along with the noun in a DET+N item, thereby producing an unmixed NP (7–8). The second mixing type targets NOMINAL-LED MULTI-WORD CODE-SWITCHES (CS), insofar as they are either (a) initiated by a noun, resulting in a mixed NP in N-INITIAL CS (9–10), or (b) occur

at the determiner, yielding an unmixed NP in DET+N-INITIAL CS (11–12). The four configurations apply to English as well as Spanish nouns.

**Table 1.** Nominal mixes

| | **Mixed NPs** | **Unmixed NPs** |
|---|---|---|
| | Nominal-limited items | |
| | Lone N | DET+N item |
| English N | (5) tenían **unos *desks*** muy grandes, [53:25–53:27] 'they had some very big desks' | (7) si serían ***four lanes*** aquí, [06, 2:31–2:33] 'if it would be four lanes here' |
| Spanish N | (6) *from* ***a*** **pollito** *that we would raise.* [04, 21:30–21:33] '… chick …' | (8) *he's* **el vaquero** *and,* [16, 38:38–38:30] '… the cowboy …' |
| | Nominal-led multi-word code-switches (CS) | |
| | N-initial CS | DET+N-initial CS |
| English N | (9) cruzando **la *road*** *once in a while.* [17, 37:00–37:03] 'crossing the …' | (11) porque ella quería ***that wood stove*** *for sure.* [04, 45:25–45:27] 'because she wanted …' |
| Spanish N | (10) *because* ***the*** **pelo** se hace pa' allá. [17, 39:44–39:46] '… hair goes one way' | (12) *and of course* **el vino** que sirven. [28, 33:39–33:41] '… the wine that they serve' |

Note that there are two types of mixed NPs. Lone Ns (5–6) are operationally defined as single words (or source-language dictionary-listed compounds) that (a) are surrounded by unilingual discourse of the other language by the same speaker on both sides, or (b) adjoin unilingual discourse of the other language on one side and occur at a prosodic boundary on the other, as in (13), where "*appointment*" adjoins Spanish on one side only, but still counts as a lone N because it occurs at the end of a prosodic sentence.[3]

(13) *let's say* que me daban **un *appointment.***
'[…] that they gave me an appointment.'
*so,*
*.. the day before,* [06, 53:53–53:57]

**3.** Lone English nouns are prosodically connected with Spanish on both sides most of the time (66%, $n = 634/959$); Spanish follows final intonation 27% the time ($n = 257/959$) (plus 7 cases of truncation) and English (as in (13)) less than 10% of the time ($n = 61/959$). The *prosodic sentence* is an IU or a series of IUs that contains a finite verb and ends in a final or appeal intonation contour (marked by a period or question mark, respectively) (cf. Chafe, 1994, p. 139).

In (9) and (10), on the other hand, we have N-initial multi-word CS. Here the determiner and noun are each part of juxtaposed multi-word strings in different languages. In (9), the first string ends in a Spanish determiner, the second begins with an English noun; in (10) the direction is the opposite, with a determiner-final English string first, and a noun-initial Spanish string following.

### 3.3 Predictions

The first prediction concerns NP status, that is, the choice between producing mixed over unmixed NPs. At stake is whether preferences diverge for Spanish as opposed to English nouns. From the perspective of an AF account, if abstract features are relevant, bilinguals should strongly prefer *un*mixed NPs when the noun is Spanish, as these would avoid the creation of NPs constituted of (the relatively impoverished) English determiners and Spanish nouns. When the noun is English, no such preference is predicted. This directionality is similarly important in a (more recent) recasting of the ML framework according to which EL islands are predicted to be "few" for English nouns in a Spanish ML, whereas "when English is the ML and Spanish nouns occur, they are likely to occur with Spanish determiners in Spanish EL islands" (Myers-Scotton & Jake, 2017, p. 356–357). In this version, the ML framework converges with the AF account (as also noted by Parafita Couto & Stadthagen-Gonzalez, 2019, p. 351), both predicting unmixed rather than mixed NPs for *Spanish*, as opposed to English, nouns.

Thus, in the context of nominal-limited mixing, we ascertain whether speakers tend to produce DET+N items rather than lone Ns when the noun is Spanish, as predicted by an AF -- and amended ML (Myers-Scotton & Jake, 2017, p. 356–357) -- account. That is, (8) "el vaquero"$_{[\text{DET+N}]}$ should be far more frequent than (6) "*a* pollito"$_{[\text{LONE N}]}$, whereas no such preference is predicted for "*four lanes*"$_{[\text{DET+N}]}$ (7) relative to (5) "unos *desks*"$_{[\text{LONE N}]}$. In nominal-led CS, again the prediction is for DET+N-initial to be more frequent than N-initial when the noun is Spanish, i.e., more "el vino"$_{[\text{DET+N-INITIAL CS}]}$ (12) than "*the* pelo"$_{[\text{N-INITIAL CS}]}$ (10), unlike when the noun is English, i.e., "*that wood stove*"$_{[\text{DET+N-INITIAL CS}]}$ (11) relative to "la *road*"$_{[\text{N-INITIAL CS}]}$) (9).

The second prediction concerns nominal mixing type, that is, bilinguals' choice to produce nominal-led multi-word CS over nominal-limited items. At stake here is whether preferences diverge for nominal mixes beyond single-word noun items surrounded by the other language, namely lone Ns. The choice according to mixing type is most pertinent to the ML account. This is because, to the extent that ML claims are independent of mixing direction, hinging as they do on the language of the verb, there would be no reason to expect (for either

nominal-limited items or nominal-led switches) a predominance of unmixed over mixed NPs depending on the language of the noun. The choice according to mixing type, however, provides a pivotal test for the ML framework, as we will see. The Bilingual NP hypothesis (Jake et al., 2002) states that there is an ML that dictates the language of the determiner in mixed NPs. The ML account should fare equally well for nominal mixes other than lone Ns, in particular, there should be no difference between the two types of mixed NPs according to the extent of the mixed element.

### 3.4 Extraction

Guided by the above considerations, rather than target only mixed NPs, as is characteristic of most of the literature, we (manually) extracted *all* tokens of nominal mixes -- DET+N sequences at the border of the two languages, whether the NPs contained within them were mixed or not. Excluded from these calculations are other-language nouns used in monolingual varieties, as determined by recipient-language dictionary attestation.[4] This is because established loanwords, like "*rifle*" in (14), which by definition are fully integrated into (i.e., indistinguishable from) the recipient language, do not result in *synchronically* mixed NPs when occurring with recipient-language determiners.

(14) parecía uno de esos de México y luego *with that big hat*,
y con el **rifle**$_{[\text{EST. LOANWORD}]}$ también. [20, 26:04–26:08]
'he looked like one of those guys from Mexico and then with that big hat,
and with the rifle too.'

Nor did we consider proper nouns, which may be ambiguous or neutral as to language and as such have been hypothesized to trigger CS (e.g., Clyne, 1991, p.193), including names of institutions (*La Cueva* high school), businesses (*Walmart*), or radio stations (*KDCE*), though we do count names of products (*iPod*), medical

4. We used the *Diccionario de la Lengua Española* (Dictionary of the Spanish language, http://www.rae.es/recursos/diccionarios/drae), the *Diccionario del Español de México* (Dictionary of the Spanish of Mexico, https://dem.colmex.mx/) and Merriam Webster (https://www.merriam-webster.com/). As operationally defined, none of the lone English Ns is listed in monolingual Spanish dictionaries; approximately three-quarters of the tokens are of words unattested even in dictionaries of New Mexican or US Southwest varieties of Spanish, which list, for example, *grandma, truck* or *troca, weekend* (Cobos, 2003; Galván & Teschner, 1989). In addition, most are infrequent: approximately half of the tokens are of a lexical type produced by only one speaker, while only one-fifth involve a noun produced by at least 10 different speakers (of the 40 participants). See Aaron (2015, p.464) for a description of a sub-sample of lone Ns in the NMSEB corpus.

conditions (*PTSD*) or calendar units (*Monday*), so long as these are dictionary-attested in the source language (here, English). Excluded, too, are metalinguistic uses, as in "una pesa es un *scale*." 'a scale (lit. 'weight') is a scale.' [16.2, 1:56]. Also excluded are instances where the determiner is prosodically separated from the noun, occurring in the preceding IU, as in (15), to ensure that the NPs considered are unbroken (i.e., true) units. These protocols yield a total of 1,280 tokens of spontaneously produced nominal mixes.[5]

| | | |
|---|---|---|
| (15) | .. tengo que ponerme **los** -- | '.. I have to put on the -- |
| | ***long johns*** otra vez, [27, 13:41–13:44] | long johns again,' |

## 4. Results

### 4.1 Testing the abstract feature-based directionality hypothesis

Proponents of the AF model claim that the preponderance of mixed NPs consisting of $DET_{[SP]}$ + $N_{[EN]}$ reflects a preference for assigning the language with the richest set of agreement features to the determiner. Yet the Spanish-English bilinguals in the northern New Mexico Hispanic community freely produce configurations lacking such conditions (as in (3), (6) and (10)). We capitalize on the co-existence of felicitous and infelicitous mixed NPs in the data to ascertain whether at least a quantitative avoidance of the latter can be detected. If so, this should be reflected in bilinguals' choices amongst classes of nominal mixes, decisively, the choice between mixed and unmixed NPs for Spanish vs. English nouns.

Recall that inserting a lone N necessarily results in a mixed NP (providing, of course that there is a determiner). For proponents of AF, this becomes problematic if the noun is Spanish, because it yields mixed NPs in the "wrong" configuration. As discussed above, this may be circumvented by inserting the Spanish noun along with its determiner. If abstract features are relevant, then, while bilin-

5. In accordance with our ethical commitment to the participants in this study, and the consent form authorized by the University of New Mexico Institutional Review Board and the funding agency, the data on which this study is based have not been placed in the public domain. Considered protocols for shareability are required for sociolinguistically-constructed corpora recording the spontaneous vernacular of close-knit minority language communities, especially where this has been widely stigmatized. The NMSEB corpus data, which derive from interactions with in-group fieldworkers, are sometimes of a highly personal nature. Protocols for access safeguard the privacy of participants, while protecting them against unintentional publication of stereotyping examples by those unfamiliar with the speech community. For further information, contact Rena Torres Cacoullos.

guals may choose freely between mixed and unmixed NPs for English nouns, as "la *facing*", "*the facing*" in (16), they should eschew mixed NPs and opt as a rule for unmixed NPs (i.e., DET+N items) with Spanish nouns, as with "la gallina" in (17). The same should apply to nominal-led multi-word CS. Spanish nouns should again overwhelmingly occur in unmixed NPs (12), so as to avoid mixed NPs featuring English determiners (10).

(16) es la *facing*, 'it's the facing,
y luego, and then,
.. cosí ***the facing***, [06, 30:42–30:44] .. I sewed the facing,'

(17) .. *I didn't want her to take* **la gallina**. [04, 20:31–20:35] '[…] the hen.'

Figure 1 depicts the relative proportion of mixed versus unmixed NPs in nominal-limited mixes. Among the English nouns incorporated into Spanish, 96% are lone, as in (5) "unos *desks*", as opposed to DET+N items, as in "*the facing*" (16) and "*four lanes*" (7); Spanish nouns in English also occur overwhelmingly -- 92% -- lone, as with "*a* pollito" (6), rather than in DET+N items like "la gallina" (17) and "el vaquero" (8). (The apparent discrepancy in proportions between English and Spanish nouns does not achieve significance, $p = 0.0575$ by Fisher's exact test.) It is clear that bilinguals prefer mixed over unmixed NPs by far, regardless of direction.

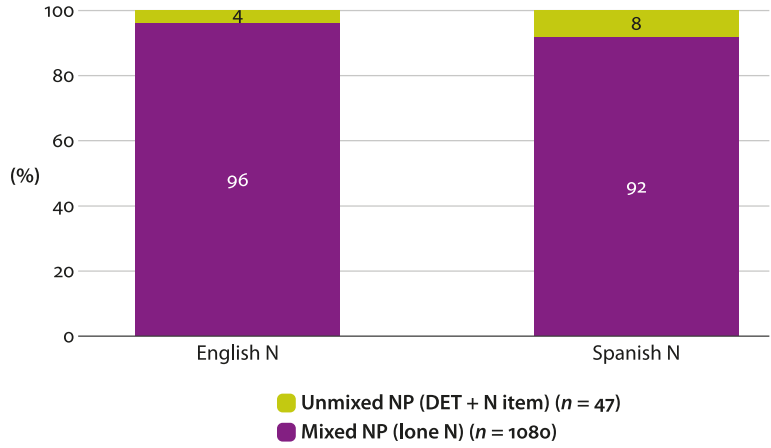


**Figure 1.** Nominal-limited mixes: Relative frequency of mixed vs. unmixed NPs according to the language of the noun ($n = 1{,}127$)

We now examine whether the asymmetry between Spanish and English in the inherent features of their determiners is echoed by an asymmetry in the proportion of mixed to unmixed NPs in nominal-led multi-word CS. Figure 2 shows that

as with nominal-limited items, speakers fail to differentiate Spanish and English nouns, the ratio of mixed to unmixed NPs being the same regardless of direction. The rates again reveal a marked preference for mixed NPs involving N-initial CS, at 75%, as in (9–10), over unmixed NPs involving DET+N-initial CS, at 25%, as in (11–12), even though this contravenes AF predictions when the noun is Spanish.

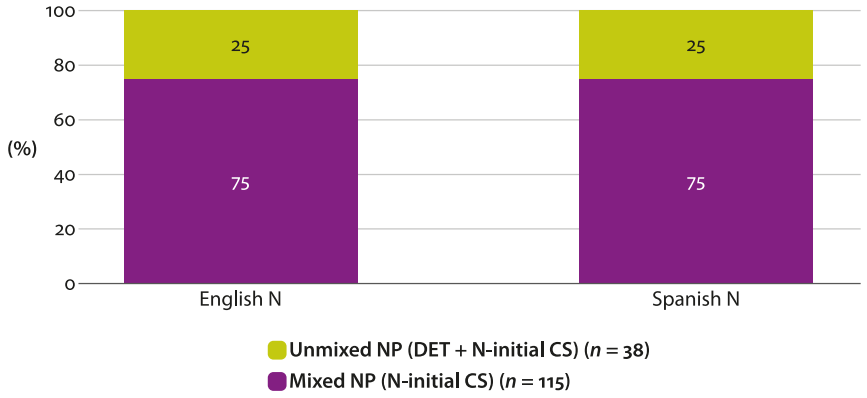


**Figure 2.** Nominal-led multi-word CS: Relative frequency of mixed vs. unmixed NPs according to the language of the noun ($n = 153$)

When we combine all nominal mixes, we confirm that Spanish nouns pattern like the unproblematic English ones, even though the result is that 90% of the tokens involving the former violate AF principles (Figure 3) (The difference in proportions between English and Spanish nouns again fails to achieve significance, $p = 0.1143$ by Fisher's exact test). The driving force behind these speakers' mixing choices appears to be independent of directionality of mixing.

In sum, the AF account leads to diametrically opposed predictions according to language directionality, generally sanctioning $DET_{[SP]} + N_{[EN]}$ and invalidating $DET_{[EN]} + N_{[SP]}$ combinations. Yet bilingual *usage* shows no such constraint, since bilinguals not only do not eschew a Spanish noun following an English determiner for fear of creating an illicit mixed NP, but actually *prefer* mixed NPs in the purportedly wrong configuration over unmixed NPs (Figures 1–3). The indisputable preference for mixed versus unmixed NPs across the board reveals that speakers are unconstrained by language-particular abstract features when choosing where in the NP to mix. Since the option to resolve feature asymmetry between Spanish and English is manifestly not exploited, we conclude that the AF approach fails to provide an adequate account for the language of the determiner in Mixed NPs.

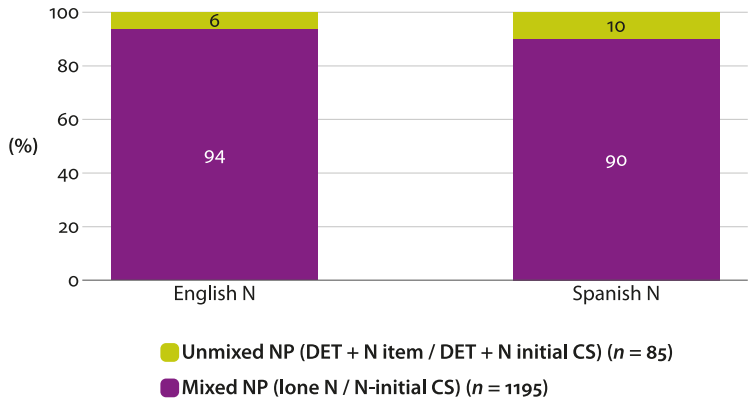


**Figure 3.** Relative frequency of mixed and unmixed NPs by language of the noun ($n = 1{,}280$)

## 4.2 Testing the Matrix language construct beyond lone nouns

As we have just seen, if abstract features matter, there should be no $DET_{[EN]}$ + $N_{[SP]}$; bilinguals will either simply never mix Spanish nouns in with English or they will opt for unmixed NPs for Spanish nouns at the border between the two languages. AF claims about the directionality of the mixed element are echoed in an ML account via the Variable Election Hypothesis, according to which Spanish determiners dominate not only by virtue of Spanish being the Matrix language but also because they carry phi-features and are elected differently from English ones (Myers-Scotton & Jake, 2017, pp. 355–357) (see 3.3 above). Here we address the more widely cited ML framework generalization, that the determiner in mixed NPs is set by the ML. This Bilingual NP Hypothesis specifies that "full EL NPs are dispreferred" (Jake et al., 2002 p. 78), that is, mixed NPs predominate over unmixed NPs (other-language nouns together with their determiners) in accordance with the axiom that the determiner matches the language of the verb (ML).

Pivotal for the efficacy of the ML, however, is extent of the mixed element. Indeed, the account is consequential only for nominal mixes other than lone Ns. When a lone N is inserted into discourse of a recipient, the language of the determiner will by definition match the language of the verb (if there is one), not to mention all of the other elements in its vicinity. The exception is when a lone N itself appears in a multi-word switch within the clause, as in (18) (also (2) above), but this is relatively uncommon. Accordingly, the accuracy of the ML is virtually perfect when it comes to lone Ns (97%, 928/955).

(18) *you could see*$_{[EN]}$ *the white rag*,
también aquí en **un**$_{[SP]}$ ***post***. [17, 43:07–43:10] 'also here on a post.'

Where the ML account is less successful is when mixing involves more than a single word (Table 2). For example, many of the published exceptions (e.g., Blokzijl et al., 2017, p.7) are made up of EL islands, corresponding to the class of DET+N items (unmixed nominal-limited NPs) (19). Where these are produced in discourse of another language, the language of the determiner is unlikely to match the language of the verb, or by extension, the Matrix language. Indeed, all ($n$=43/43) of the DET+N items in NMSEB for which the Matrix language could be identified (i.e., those realized in a clause containing a finite verb) contravene ML predictions for these reasons. Unmixed NPs resulting from DET+N-initial multi-word CS (20) likewise nearly always lead to failure: in 83% ($n$=30/36) of the instances, the language of the determiner is not that of the Matrix language.[6, 7] Mixed NPs resulting from N-initial multi-word CS (21) are far more successful in meeting ML expectations, but still contain DET-V mismatches in 7% of cases ($n$=7/105). Taken together, 43% of extended nominal mixes (DET+N items and both CS classes) are shown to counter the ML contention that the language of the determiner should match that of the Matrix.

(19) … en un bote de esos de manteca que tenían$_{[SP]}$ ***the***$_{[EN]}$ ***lid***? [03, 40;38–40:41]
'…in one of those lard cans that had the lid?'

(20) … se me hace que era$_{[SP]}$ ***four***$_{[EN]}$ ***years*** *ago*. [20, 30:10–30:11]
'…I think it was four years ago.'

(21) en **el**$_{[SP]}$ ***winter***, 'in the winter,
*.. you are doing*$_{[EN]}$ *it for a little while,*
y sí está frío. [06, 39:10–39:13] and it really is cold.'

The ML has also been faulted for its limitations in terms of *coverage* relative to the AF account. When the clause contains no finite verb or no verb at all, "the

**6.** An unmixed NP in DET+N-initial CS whose determiner is in the same language as the associated finite verb can be a subject preceded by an adverbial expression (se me hace que -- como que ***your body*** *is more stronger*, 'I think that -- like that […]' [06, 1:15:23–1:15:25]) or a head noun (una hermana de los ~Pacheco? .. mhm. que tenían la tienda, qu- ***the one*** *that got the store over here*? 'a sister of the Pachecos? .. mhm. who had the store, wh- […]' [25, 40:32–40:35]; see also (12)).

**7.** A reviewer suggests that, if switches tend to be post-verbal in English-Spanish CS (as in, for example, (11) vs. (12)), then it is to be expected that other-language DET+N items would not be in the language of the verb, and thus that a propensity for determiner and verb to be in the same language would be tested by nominal mixing data from OV language pairs.

assignment of a Matrix language is not possible" (Herring et al., 2010, p. 565). For example, in (22), the target NPs are unattached, occurring in "increments" (Ford, Fox & Thompson, 2002) outside the prosodic sentence containing the verb, leaving no means by which to identify a Matrix language.

(22) a. *.. I own the properties of the ~Meza,*
*... and the properties from the,*
*.. ~Arreola.*
con **el *grandpa*** de la ~Hilda? 'with the grandpa of Hilda?
... los dos. [25, 52:21–52:28] ... both of them.'

b. *nothing.*
ni *scribbling*, 'not even scribbling,
*much less* **su nombre**. [22, 25:10–25:14] much less her name.'

c. *well that's good to know*,
*no.*
.. *like **the other*** **vecino**, '.. like the other neighbor,
.. de allá abajo. [23, 41:39–41:43] .. from down there.'

In the NMSEB corpus, the coverage of the ML account is 92% for extended nominal mixes (Table 2) and, overall, 89% ($n = 1139/1280$) of all nominal mix tokens (Table 3).[8] The 11% ($n = 141/1280$) lacking a (finite) verb are comparable to the approximately 10% ($n = 18/163$) "with ambiguous matrix language" reported for a Welsh-English dataset (Deuchar, 2006, p. 2002). When the cases featuring a mismatch are combined with those where it was not possible to identify the Matrix language (and therefore to verify whether there was a match), we see that only slightly more than half (52%) of the nominal mixes greater than a single word are found to be accounted for by the ML approach (Table 2).

The number and rate of exceptions cited in the literature pales in comparison to the results reported here. Indeed, in one analysis of Spanish-English mixed NPs in a Miami corpus, mismatches appear just 2% of the time (Blokzijl et al., 2017, Table 3). Why should this be? For one thing, most studies restrict their focus to mixed NPs only, omitting other nominal mixes involving determiners and nouns (DET+N items, DET+N-initial CS). These are precisely the classes that fare worst in accuracy checks (Table 2). But more revealing, the accuracy of the ML account is *bolstered* in studies that fail to differentiate between lone

**8.** Relying on the verb to identify a Matrix language is even more fraught for multi-word CS more generally: examination of switches occurring at the sentence boundary in the wider NMSEB corpus reveals that a full one-fifth ($n = 280/1343$) are increments, containing no finite verb; the proportion of increments is the same in a unilingual sample from the corpus (LaCasse & Trawick, 2019).

**Table 2.** Accuracy, coverage and overall success of the ML account in predicting the language of the determiner for extended nominal mixes (other than lone Ns)

| | Accuracy (DET-V match) | | | Coverage (Presence of finite V) | | | Overall success* (Accuracy + Coverage) | | |
|---|---|---|---|---|---|---|---|---|---|
| **Nominal mix** | **%** | ***n*** | **TOT** | **%** | ***n*** | **TOT** | **%** | ***n*** | **TOT** |
| N-initial CS | 93% | 98 | 105 | 91% | 105 | 115 | 84% | 98 | 115 |
| DET+N-initial CS | 17% | 6 | 36 | 95% | 36 | 38 | 16% | 6 | 38 |
| **Total CS** | **74%** | **104** | **141** | **92%** | **141** | **153** | **67%** | **104** | **153** |
| DET+N item | 0% | 0 | 43 | 91% | 43 | 47 | 0% | 0 | 47 |
| **Total extended** | **57%** | **104** | **184** | **92%** | **184** | **200** | **52%** | **104** | **200** |

* Accuracy TOT is the total possible, excluding tokens lacking a finite V. Overall success is calculated as accuracy/total (all tokens).

nouns vs. mixing involving more than a single word. As we will see just ahead (Section 5.1), given the ease with which lone nouns satisfy ML predictions, their relative frequency conspires in inflating the capacity of ML to account for the language of the determiner.

The volume of mismatches in language between the determiner and the associated verb, then, will depend on the frequency of DET+N items and multi-word CS: the more mixing beyond the lone N, the more likely the exceptions to ML predictions. The actual rate may differ across corpora, due either to genuine community differences or as an artefact of data collection. Nevertheless, the preponderance of lone Ns vis-à-vis other classes of mixing applies cross-linguistically (Poplack, 2018).

## 5. Discussion

### 5.1 The role of the lone noun in the controversy

We noted earlier that the literature has made no effort to distinguish lone Ns from multi-word CS. We suggest that this is a key, if unacknowledged, driver of the controversy over the language of the determiner in mixed NPs. In this section, we show how this move particularly *benefits* the ML account.

A first important consideration is that lone Ns are by far the major class of mixing in just about every language pair systematically studied (cf., Backus, 1992; Halmari, 1997; Jake et al., 2002, p.79; Myers-Scotton, 1993 [1997]; Nortier, 1990; Pfaff, 1979, p.303–308; Poplack, 2018, p.120–121). The New Mexico community also displays a marked preference for inserting lone Ns into other-language

discourse, far outweighing nominal-led CS and DET+N items. This is the case regardless of directionality, and purported (in)felicitousness of the resulting combinations. Given the multitude of typological similarities (equivalence sites, cf. Poplack, 1980) between Spanish and English in particular, there is little to prevent many more instances of extended nominal mixes. Yet, these are nonetheless far outweighed by the incorporation of lone Ns, which constitute approximately four-fifths of all nominal mixes, as shown in Figure 4.

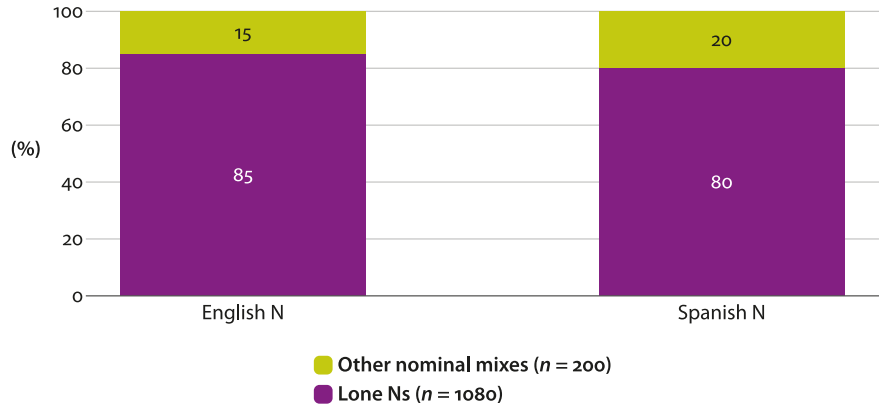


**Figure 4.** Frequency of lone Ns relative to extended nominal mixes (DET+N items, N-initial CS, DET+N-initial CS) ($n$ = 1,280)

What are the repercussions of this fact? For the ML account, the huge disproportion of lone Ns is responsible for its apparent overall accuracy level. Table 3 confirms that accuracy is heavily boosted when lone Ns are included in the count (91% vs. 57% [Table 2] when they are not). The same is true of overall success rate, where coverage is taken into account – 81% including lone Ns vs. 52% for other classes of nominal mixes.

As noted above, because lone Ns are incorporated into discourse of another language, the language of the determiner will de facto match the language of the verb (as well as other linguistic elements in the surrounding discourse). But conflating lone Ns with multi-word CS and DET+N items obscures the fact that 43% of the latter classes show a DET-V language mismatch; the number of contraventions rises to 48% when lack of coverage resulting from a missing verb is factored in (Table 3). These facts must be considered in evaluating the model.

For the AF camp, in contrast, this same preponderance of lone Ns constitutes a major problem when the community under study happens to prefer the "wrong" direction, as was reported for the Nicaraguan Creole English-Spanish and Papiamento-Dutch contexts (Blokzijl et al., 2017; Parafita Couto & Gullberg 2019,

**Table 3.** Accuracy, coverage and overall success of the ML account in predicting the language of the determiner by nominal mix class: Lone Ns vs. extended (DET+N items, N-initial CS, DET+N-initial CS)

| | Accuracy (DET-V match) | | | Coverage (Presence of finite V) | | | Overall success* (Accuracy + Coverage) | | |
|---|---|---|---|---|---|---|---|---|---|
| **Nominal mix** | **%** | ***n*** | **TOT** | **%** | ***n*** | **TOT** | **%** | ***n*** | **TOT** |
| Lone Ns | 97% | 928 | 955 | 88% | 955 | 1080 | 86% | 928 | 1080 |
| Total extended | 57% | 104 | 184 | 92% | 184 | 200 | 52% | 104 | 200 |
| **Total (all data)** | **91%** | **1032** | **1139** | **89%** | **1139** | **1280** | **81%** | **1032** | **1280** |

* Accuracy TOT is the total possible, excluding tokens lacking a finite V. Overall success is calculated as accuracy/total (all tokens).

p. 703); see also the propensity of anglo-Quebeckers to insert French nouns into English (Poplack, Walker & Malcolmson, 2006). Even in the language pair that prompted their proposal, lone Ns account for 80% of nominal mixes from English to Spanish in the NMSEB corpus, all of which contravene the AF model (Figure 4, right column).

In some of the literature on AF, exceptions such as those described above have been explained (away) as borrowings. NONCE BORROWINGS (Weinreich, 1953, p. 11) are lone other-language incorporations that feature the linguistic characteristics of established loanwords, assuming the morphological and syntactic identity of the recipient language, even though they lack the social characteristics of their established counterparts, namely recurrence in the speech of the individual and dispersion across the community (Poplack, 2012, 2018, pp. 122–157). Were those lone Ns that so qualify to be recognized as borrowed, the mixed NPs would not really be "mixed". What are the consequences of such a move?

For the AF proposal, the number of counterexamples would be vastly reduced, since its proponents explicitly exclude loanwords (which lose the properties – phi-features included – of their lexifier language) from their purview. Nevertheless, the unambiguous code-switches, i.e., the nominal mixes initiating multi-word CS, remain as a refutation, since the percentage of mixed NPs is identical for the two languages (Section 4.1). At the same time, removal of the lone nouns, while getting rid of a good number of $DET_{[EN]} + N_{[SP]}$ exceptions, would also remove a great number of the felicitous $DET_{[SP]} + N_{[EN]}$.

This very move would also hurt the ML model, since, as we have shown above, lone Ns radically inflate its accuracy (Table 3). Regardless of how lone Ns are characterized, once they are incorporated into another language, the language of the determiner will automatically correspond to that of the surrounding dis-

course and likely that of the verb (or the matrix language, in ML terms), *regardless* of its linguistic properties.

Where the possibility of choosing the language of the DET -- and whether or not to mix within the confines of the NP -- arises is precisely when the other-language material extends beyond a lone N. Thus a proper evaluation of the language of the DET in mixed NPs should come from their behavior in nominal mixes greater than a single noun. Since neither camp distinguishes between lone Ns and multi-word CS, we have no idea how many of the latter are contained in their data. But when we evaluate their predictions in the nominal-led CS occurring in NMSEB, we find that they are subverted (75% for AF [Figure 2] and 16% (100–84) for MLF [Table 2]).

## **5.2** The role of the community in the controversy

As conflation of classes of mixes muddies the waters, so too does disregard for tacit community norms. We have detailed above why we reject the linguistic properties of the determiner, the noun, or any other typological considerations involving the language pair in question as explanatory. We found no evidence that speakers are *avoiding* mixed NPs that are "undesirable" according to the language of the determiner (Figures 1–3). Instead, we suggest that the tendency toward unidirectionality reported in the literature simply reflects community norms.

We know from previous research that such norms are a staple of bilingual communities, and we also know that there is generally no linguistic motivation for them. Thus, despite typological similarity between their language pairs, Puerto Rican Spanish/English bilinguals in New York and French/English bilinguals in Ottawa-Gatineau employ different CS strategies, the former seamlessly alternating between languages as a discourse mode, the latter tending to switch to fulfill certain rhetorical functions (Poplack, 1985). Tunisian Arabic tends to mark possession through nominal inflection, but when bilinguals incorporate French nouns into it, they integrate them via a less common option (postposing an Arabic periphrastic particle to the noun and inflecting *it*), leaving the French noun morphologically intact (Poplack, Sayahi, Mourad & Dion, 2015). The most compelling evidence that community norms are independent of typological or structural properties of the languages involved comes from examining different dialects of the same language pair. How else to explain differential gender assignment to the same English-origin nouns in France and Canada ("$le_{[\text{MASC}]}$ *job*", "$le_{[\text{MASC}]}$ *gang*" in the former; "$la_{[\text{FEM}]}$ *job*", "$la_{[\text{FEM}]}$ *gang*" in the latter [Poplack, 2017]), or the preference for incorporating English-origin verbs via light verb *hacer* 'do' + bare infinitive in New Mexican Spanish, as in (23), in contrast to verbal inflection, as in (24), for Puerto Rican speakers (cf. Wilson & Dumont, 2015)?

(23) … lo **hic-ieron** [do-PFV.3PL] ***hire*** pa' eso. [04, 53:28–53:29]
'… they hired him for that.'

(24) … nosotros te ***baque*-amos.** [back- PRS.1PL] [23,32:08–32:10]
'… we'll back you up.'

The results presented here offer further support for the primacy of community norms over typological expectations. Two kinds of norms are in evidence, one universal, one specific. The first involves the overwhelming predilection for lone Ns, discussed in the preceding section. We have already explained the impact of this fact on the accuracy of the models.

A second community norm is less well-documented, and may even be specific to Northern NM bilinguals (though it appears to capture the situation in the published literature on Spanish-English mixed NPs elsewhere as well). This concerns the aforementioned directionality of nominal mixes. As shown in Figure 5, there is disproportionately more mixing involving English nouns than Spanish nouns.

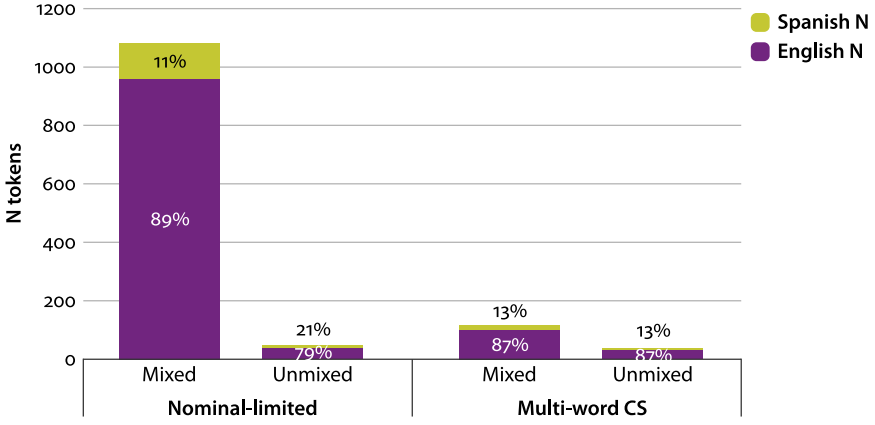


**Figure 5.** Directionality of nominal mixes: Language of noun by mixing type and NP status ($n = 1280$; see Appendix 2)

The stock explanation for the situation depicted in Figure 5 is that English is the prestige language (e.g., Blokzijl et al., 2017, pp. 8–9), and accordingly, bilinguals draw from it when speaking Spanish (though the same directionality is observed in Northern Belize, where English apparently does not enjoy the same high status as in the United States [Balam & Parafita Couto, 2019, p. 213]). Another kind of explanation invokes pragmatic considerations (e.g., emphasis; Myers Scotton & Jake, 2017, p. 357). Such claims are difficult to confirm in spontaneous production data, especially when specific mixing sites are at stake.

These speakers simply prefer to resort to English when incorporating other-language *nouns*. Indeed of all lone English words incorporated into Spanish discourse in the NMSEB corpus, nouns make up nearly three-quarters, proportionally three times more than for Spanish lone words incorporated into English (71%, $n = 1541/2155$ vs. 24%, $n = 203/836$). Idiosyncratic distributions by semantic domain and lexical type additionally highlight the precedence of *community* norms over generalizations such as the notion of cultural vs. core borrowing (contra, e.g., Haspelmath, 2008, pp. 48–51, Myers-Scotton 2002, p. 239, among many others). Most propitious to English lone Ns are kin terms; among these the community preference is for "*grandma*" and "*grandpa*" (rather than "abuela" and "abuelo") but "hermano" and "hermana" (rather than "*brother*" and "*sister*") (Aaron, 2015, pp. 466–467; Torres Cacoullos & Aaron, 2003, p. 300).

Yet another explanation for such asymmetries involves language "dominance" (e.g., Liceras, Fernández Fuertes & Klassen [2016]; see also discussion in Parafita Couto & Gullberg [2019, pp. 703–704], and references therein). In this community, however, the preference for inserting English nouns into Spanish cannot be attributed to language dominance. First, the regular use of *both* languages is evidenced by the even distribution of clauses according to language (of the verb): 51% English, 49% Spanish (Figure 6, left).

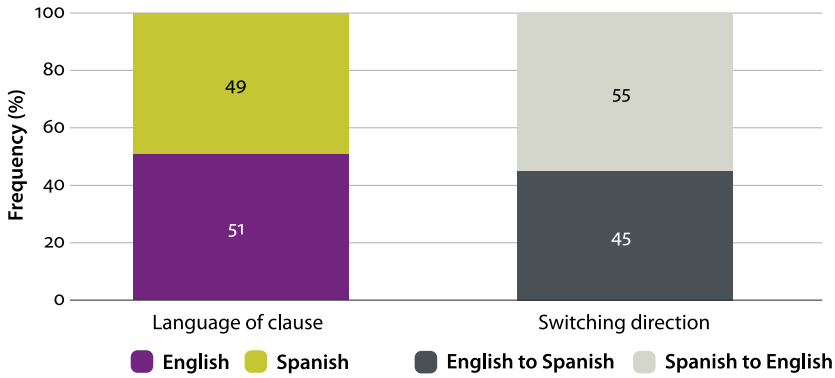


**Figure 6.** Even distribution of languages by language of clause ($n = 36{,}011$, from Torres Cacoullos and Travis [2018, p. 67]) (left) and bidirectionality of intra-sentential multi-word CS ($n = 407$, from Steuck [2018, Table 4-1]) (right)

Another piece of evidence comes from code-switching direction. In a wider sample of intra-sentential multi-word CS not confined to nominal-led switches, the proportions according to language direction are nearly equal, at 55% Spanish to English switches, as in (25), and 45% English to Spanish switches, as in (26) (Figure 6, right) (cf. Steuck & Torres Cacoullos, 2019). Such balanced distribution

(in contrast to previous studies of mixed NPs) precludes any argument that one of the two languages is the overall Matrix language (cf. Myers-Scotton, 1993). For these speakers, regularly using both languages and alternating between them is simply a fact of daily life, with no local triggers motivating CS (Torres Cacoullos & Travis, 2018, pp. 58–71).

(25) como el muchito ese, 'like that boy,'
*they used to take care of this summer.* [09, 45:12–45:15]

(26) .. that little bottle,
.. *qué tanto vale*? [09, 20:35–20:37] '.. how much does it cost?'

This pair of results -- (a) near-equal numbers of English clauses into which lone Spanish nouns could have been inserted and Spanish clauses into which English nouns can be inserted, and (b) near-balanced bidirectionality in intra-sentential multi-word CS -- indicates that these bilinguals have real choices, not imposed by language dominance. The pertinent point here about these approximately even distributions (English clauses and multi-word CS, Spanish clauses and multi-word CS), is that these bilinguals have about the same opportunities to use an English noun as a Spanish noun in a nominal mix. Yet they disproportionately prefer English nouns -- as is evident by juxtaposing the near unidirectionality in Figure 5 and bidirectionality in Figure 6. We conclude that the unidirectional tendency to choose English rather than Spanish nouns in nominal mixes must be recognized as a "conventional production pattern" (Valdés-Kroff 2016, p. 281) for the northern New Mexico bilingual community, and likely others as well.[9] In this we concur with Blokzijl et al.'s (2017, p. 1) conclusion that […] "social rather than grammatical factors may be at play" in the selection of language of the noun and "the matrix language itself", though our analyses counter their further claim that "the matrix language of the clause provides the language of the determiner".

## 6. Conclusion

The widespread occurrence in bilingual speech of nouns in one language with a determiner in the other, often referred to as mixed NPs or DPs, has generated much theorizing and many counterarguments. The facts of bilingual usage, however, hardly justify the intricate theoretical apparatus marshalled by the compet-

**9.** This is certainly the case for the Canadian French-English speakers of Ottawa-Gatineau (Poplack, 1989) and the Canadian English-French bilinguals of anglophone Quebec (Poplack et al., 2006).

ing accounts. In the present dataset, as in others, the predominant combination is that of $DET_{[SP]}$ + $N_{[EN]}$. Since both an account based on an abstract (uninterpretable Gender) feature on the determiner and an account highlighting the notion of a Matrix language yield largely the same predicted outcomes, exactly so when the Matrix language is also the language with more abstract features, in this paper we have assessed how their tenets play out in speakers' choices. Our test is based on choices along three dimensions:

a. where to mix, that is, whether to produce a mixed NP ("$a_{[EN]}$ pollito$_{[SP]}$") or an unmixed one ("el vaquero$_{[SP]}$");
b. which language to draw the other-language noun from, here, English vs. Spanish ("unos *desks*$_{[EN]}$", "*a* pollito$_{[SP]}$"); and
c. whether to incorporate an other-language noun on its own as opposed to initiating a sequence of other-language words with the noun ("cruzando la *road once in a while*"; "*because the* pelo se hace pa' allá").

First, we contextualized the mixed NPs with respect to their unmixed counterparts in bilingual utterances. This enabled us to examine empirically the inference that speakers are *avoiding* the purportedly infelicitous $DET_{[EN]}$ + $N_{[SP]}$ combination. By applying the variationist Principle of Accountability (to consider not only cases where the phenomenon of interest occurred, but also all those where it could have occurred even if it did not), we could access the *alternative* choice speakers would have made in the same context: to resort to the other-language determiner along with the noun. As we have seen (Section 4.1), speakers rarely avail themselves of this option. Instead, the community preference is to incorporate the other-language noun *without* its determiner, resulting in a mixed NP, and this, for nouns of either language. This means that the preference for mixed over unmixed NPs is independent of the (theoretical) felicity of the mixed NP. This first finding could not have emerged in the absence of the bi-directionality that characterizes NMSEB, which allowed us to apply accountable quantitative reasoning.

Second, though the corpus is remarkably balanced according to language and intra-sentential CS direction (Figure 6), we nevertheless observe more recourse to English nouns when speaking Spanish than to Spanish nouns in English (Figure 5). Absent any evidence implicating features of the determiner, this unidirectionality is revealed to derive from a community convention. Whether socio-historically determined or accidental, it must be stressed that such norms do not derive from linguistic properties. Whatever the extra-linguistic explanation -- if there is one beyond plain recognition of a component of powerful, if implicit, speech community norms -- this convention does have important

linguistic consequences, however. Here, the preference for English vs. Spanish other-language nouns is what accounts for the predominance of the $DET_{[SP]}$ + $N_{[EN]}$ combination.

Third, we sought data from bilinguals who have options other than the predominant mixed NP combination. Herein lies the value of spontaneous bilingual speech, enabling us to capitalize on the bi-directionality of switching, on the one hand, and the choice of other-language material longer than the lone noun, on the other. Corpora with less data overall, fewer English clauses into which Spanish nouns could have been incorporated and fewer nominal-led switches have not served the controversy well. Adoption of the distinction between nonce borrowing and multi-word CS is not required to resolve debates on mixed NPs. Yet the distinction between lone other-language nouns and other nominal mixes is incontrovertible (cf., for example, Muysken's [2000, 2015: 251–254] distinction between CS of the insertional and alternational kind); this turns out to be key to unraveling the issues. While these speakers avail themselves of all three options in both directions – lone Ns, DET+N items and nominal-led CS – the former is by far preferred (Figures 4 and 5).

In this overwhelming incorporation of lone Ns, a documented cross-linguistic tendency is upheld. Simply put, nominal mixes tend to arise from insertion of lone Ns. This undisputed fact underlies the successes of the competing theories. As a result, the determiner (and any associated verb) will perforce be in the language into which the lone noun is incorporated. And the disproportion of specifically *English* lone Ns (Figure 5) yields the inexorable result that the determiner in mixed NPs will be Spanish.

In sum, the general predominance of lone Ns, coupled with community-specific preferences for the language of those nouns, may have prompted the inference that some configurations are ungrammatical, and encouraged the proliferation of theories to account for the presumed ungrammaticality. These, however, do not stand the test of bilingual behavior, which provides its own answers for linguistic patterns.

## Funding

The work reported here has benefitted from the support of the National Science Foundation (BCS-1019112/1019122 to R. Torres Cacoullos and C. Travis; 1624966 to R. Torres Cacoullos and S. Poplack) and a Canada Research Chair to Poplack.

## Acknowledgements

We thank Trevor Bero, Grant Berry, Deyanira S. Moya-Chaves, Jonathan Steuck and Sonya Trawick for their contributions to data extraction, coding and analysis, as well as three LAB referees for thoughtful questions and helpful suggestions.

## References

Aaron, J. E. (2015). Lone English-origin nouns in Spanish: The precedence of community norms. *International Journal of Bilingualism*, *19*, 459–480. https://doi.org/10.1177/1367006913516021

Backus, A. (1992). *Patterns of language mixing. A study in Turkish-Dutch Bilingualism*. Otto Harrassowitz.

Balam, O. (2016). Semantic categories and gender assignment in contact Spanish: Type of code-switching and its relevance to linguistic outcomes. *Journal of Language Contact*, *9*, 405–439. https://doi.org/10.1163/19552629-00903001

Balam, O., & Parafita Couto, M. (2019). Adjectives in Spanish/English code-switching. Avoidance of grammatical gender in bi/multilingual speech. *Spanish in Context*, *16*, 194–216. https://doi.org/10.1075/sic.00034.bal

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

Blokzijl, J., Deuchar, M., & Parafita Couto, M. (2017). Determiner asymmetry in mixed nominal constructions: The role of grammatical factors in data from Miami and Nicaragua. *Languages*, *2*, article 20. Published online by MDPI, October 6 2017. https://doi.org/10.3390/languages2040020

Chafe, W. (1994). *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.

Clyne, M. (1991). *Community languages: The Australian experience*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511597084

Cobos, R. (2003). *A dictionary of New Mexico and Southern Colorado Spanish*. Museum of New Mexico Press.

Deuchar, M. (2006). Welsh-English code-switching and the Matrix Language Frame model. *Lingua*, *116*, 1986–2011. https://doi.org/10.1016/j.lingua.2004.10.001

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. Edwards & M. Lampert (Eds.), *Talking data: Transcription and coding in discourse* (pp. 45–89). Lawrence Erlbaum Associates.

Duran Eppler, E., Luescher, A., & Deuchar, M. (2017). Evaluating the predictions of three syntactic frameworks for mixed determiner–noun constructions. *Corpus Linguistics and Linguistic Theory*, *13*, 27–63. https://doi.org/10.1515/cllt-2015-0006

Ford, C. E., Fox, B. A., & Thompson, S. A. (2002). Constituency and the grammar of turn increments. In C. E. Ford, B. A. Fox, & S. A. Thompson (Eds.), *The language of turn and sequence* (pp. 14–38). Oxford University Press.

Galvan, R. & Teschner, R. (1989). *The Dictionary of Chicano Spanish / El diccionario del español chicano*. National Textbook Company.

Halmari, H. (1997). *Government and codeswitching. Explaining American Finnish*. John Benjamins. https://doi.org/10.1075/sibil.12

Haspelmath, M. (2008). Loanword typology: Steps toward a systematic cross-linguistic study of lexical borrowability. In T. Stolz, D. Bakker, & R. Salas Palomo (Eds.), *Aspects of language contact: New theoretical, methodological and empirical findings with special focus on Romancisation processes* (pp. 43–62). Mouton de Gruyter.

Herring, J., Deuchar, M., Parafita Couto, M. C., & Moro Quintanilla, M. (2010). 'I saw the madre': Evaluating two theories' predictions about codeswitching between determiners and nouns using Spanish-English and Welsh-English bilingual corpora. *International Journal of Bilingual Education and Bilingualism*, *13*, 553–573. https://doi.org/10.1080/13670050.2010.488286

Jake, J. L., Myers-Scotton, C., & Gross, S. (2002). Making a minimalist approach to codeswitching work: Adding the matrix language. *Bilingualism: Language and Cognition*, 569–591.

Labov, W. (1969). Contraction, deletion and inherent variability of the English copula. *Language*, *45*, 715–762. https://doi.org/10.2307/412333

Labov, William. (1984). Field methods of the project on linguistic change and variation. In J. Baugh & J. Sherzer (Eds.), *Language in use: Readings in sociolinguistics* (pp. 28–53). Prentice Hall.

LaCasse, D., & Trawick, S. (2019). Quantity and quality: Using prosodic units to apply accountability to code-switching. Poster presented at New Ways of Analyzing Variation (NWAV) 48, University of Oregon.

Liceras, J. M., Fernández Fuertes, R., & Klassen, R. (2016). Language dominance and language nativeness: The view from English–Spanish codeswitching. In R. E. Guzzardo Tamargo, C. M. Mazak, & M. C. Parafita Couto (Eds.), *Spanish-English codeswitching in the Caribbean and the US* (pp. 107–138). John Benjamins. https://doi.org/10.1075/ihll.11.05lic

Liceras, J. M., Fernández Fuertes, R., Perales, S., Pérez-Tattam, R., & Spradlin, K. T. (2008). Gender and gender agreement in bilingual native and non-native grammars: A view from child and adult functional-lexical mixings. *Lingua*, *118*, 827–851. https://doi.org/10.1016/j.lingua.2007.05.006

Liceras, J. M., Spradlin, K. T., & Fernandez Fuertes, R. (2005). Bilingual early functional-lexical mixing and the activation of formal features. *International Journal of Bilingualism*, *9*, 227–252. https://doi.org/10.1177/13670069050090020601

Lipski, J. (2005). Code-switching or Borrowing? No sé *so* no puedo decir, *you know*. In L. Sayahi & M. Westmoreland (Eds.), *Selected Proceedings of the Second Workshop on Spanish Sociolinguistics* (pp. 1–15). Cascadilla Proceedings Project.

MacSwan, J. (2005). Codeswitching and generative grammar: A critique of the MLF model and some remarks on "modified minimalism". *Bilingualism: Language and Cognition*, *8*, 1–22. https://doi.org/10.1017/S1366728904002068

Moro Quintanilla, M. C. (2014). The semantic interpretation and syntactic distribution of determiner phrases in Spanish/English codeswitching. In J. MacSwan (Ed.), *Grammatical theory and bilingual codeswitching* (pp. 213–227). MIT Press.

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge University Press.

Muysken, P. (2015). Déjà voodoo or new trails ahead? Re-evaluating the Mixing Typology Model. In R. Torres Cacoullos, N. Dion, & A. Lapierre (Eds.), *Linguistic variation: Confronting fact and theory* (pp. 242–261). Routledge.

Myers-Scotton, C. (1993). *Duelling languages*. Oxford University Press.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198299530.001.0001

Myers-Scotton, C., & Jake, J.L. (2017). Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. *International Journal of Bilingualism*, *21*, 340–366. https://doi.org/10.1177/1367006915626588

Nortier, J. (1990). *Dutch-Moroccan Arabic code-switching among Moroccans in the Netherlands*. Foris Publications. https://doi.org/10.1515/9783110877182

Parafita Couto, M.C., & Gullberg, M. (2019). Code-switching within the noun phrase: Evidence from three corpora. *International Journal of Bilingualism*, *23*, 695–714. https://doi.org/10.1177/1367006917729543

Parafita Couto, M.C., & Stadthagen-Gonzalez, H. (2019). El book or the libro? Insights from acceptability judgments into determiner/noun code-switches. *International Journal of Bilingualism*, *23*, 349–360. https://doi.org/10.1177/1367006917728392

Pfaff, C.W. (1979). Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language*, *55*, 291–318. https://doi.org/10.2307/412586

Poplack, S. (1980). "Sometimes I'll start a sentence in Spanish y termino en español": Toward a typology of code-switching. *Linguistics*, *18*, 581–618. https://doi.org/10.1515/ling.1980.18.7-8.581

Poplack, S. (1985). Contrasting patterns of code-switching in two communities. In H.J. Warkentyne (ed.), *Methods V: Papers from the V International Conference on Methods in Dialectology* (pp. 363–385). University of Victoria.

Poplack, S. (1989). The care and handling of a mega-corpus. The Ottawa-Hull French project. In R.W. Fasold & D. Schiffrin (eds.), *Language change and variation* (pp. 411–451). John Benjamins. https://doi.org/10.1075/cilt.52.25pop

Poplack, S. (2012). What does the Nonce Borrowing Hypothesis hypothesize? *Bilingualism: Language and Cognition*, *15*, 644–648. https://doi.org/10.1017/S1366728911000496

Poplack, S. (2017). L'anglicisme à domicile: un portrait (socio)linguistique. In *Les anglicismes: des emprunts à intérêt variable? Recueil des actes du Colloque du réseau des organismes francophones de politique et d'aménagement linguistiques (OPALE), 18 et 19 octobre 2016* (pp. 375–403). Office québécois de la langue française.

Poplack, S. (2018). *Borrowing: Loanwords in the speech community and in the grammar*. Oxford University Press.

Poplack, S. (2020). A variationist perspective on language contact. In E. Adamou & Y. Matras (Eds.), *The Routledge Handbook of Language Contact* (pp. 46–62). Routledge. https://doi.org/10.4324/9781351109154-5

Poplack, S., Sayahi, L., Mourad, N., & Dion, N. (2015). An exception to the rule? Lone French nouns in Tunisian Arabic. *Penn Working Papers in Linguistics*, *21*, 178–186.

Poplack, S., Walker, J.A., & Malcolmson, R. (2006). An English "like no other"? Language contact and change in Quebec. *Canadian Journal of Linguistics 51*, 185–213.

Ramírez Urbaneja, D. (2019). "¿Tú tienes una little pumpkin?" Mixed noun phrases in Spanish-English bilingual children and adults. *International Journal of Bilingualism*, *24*, 824–839. https://doi.org/10.1177/1367006919888580

Steuck, J.W. (2018). The prosodic-syntactic structure of intra-sentential multi-word code-switching in the New Mexico Spanish-English bilingual community. PhD dissertation, Penn State University.

Steuck, J.W., & Torres Cacoullos, R. (2019). Complementing in another language: Prosody and code-switching. In J.-A. Villena-Ponsoda, F. Díaz-Montesinos, A.-M. Ávila-Muñoz, & M. Vida-Castro (eds.), *Language variation: European Perspectives VII* (pp. 219–231). John Benjamins. https://doi.org/10.1075/silv.22.14ste

Torres Cacoullos, R. (2020). Code-switching strategies: Prosody and syntax. *Frontiers in Psychology, Volume 11, Article 2130*.

Torres Cacoullos, R., & Aaron, J.E. (2003). Bare English-origin nouns in Spanish: Rates, constraints, and discourse functions. *Language Variation and Change*, *15*, 287–326.

Torres Cacoullos, R., & Travis, C.E. (2018). *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge University Press. https://doi.org/10.1017/9781108235259

Torres Cacoullos, R., & Travis, C.E. (2020). Code-switching and bilinguals' grammars. In E. Adamou & Y. Matras (Eds.), *The Routledge Handbook of Language Contact* (pp. 252–274). Routledge. https://doi.org/10.4324/9781351109154-18

Valdés-Kroff, J.R. (2016). Mixed NPs in Spanish-English bilingual speech: Using a corpus-based approach to inform models of sentence processing. In R. Guzzardo Tamargo, C. Mazak, & M.C. Parafita Couto (Eds.), *Spanish-English Codeswitching in the Caribbean and the US*. (pp. 281–300). John Benjamins. https://doi.org/10.1075/ihll.11.12val

Weinreich, U. (1953). *Languages in contact*. Mouton De Gruyter.

Wilson, D.V., & Dumont, J. (2015). The emergent grammar of bilinguals: The Spanish verb hacer 'do' with a bare English infinitive. *International Journal of Bilingualism*, *19*, 444–458. https://doi.org/10.1177/1367006913516047

## Appendix 1

**Transcription Conventions** (see Du Bois et al., 1993; Torres Cacoullos & Travis, 2018, p. 211)[10]

Carriage return New Intonation Unit

| | | | |
|---|---|---|---|
| . | final intonation contour | - | truncated word |
| , | continuing intonation contour | .. | short pause (0.2 secs) |
| ? | appeal intonation contour | ... | medium pause (0.3–0.6 secs) |
| -- | truncated intonation contour | ...( ) | timed pause (0.7 secs or longer) |
| ~ | pseudonymized proper noun | | |

10. For the purposes of readability, we have removed symbols for vocal noises, laughter and vowel lengthening.

## Appendix 2

Distribution of Nominal Mixes (see Table 1; example numbers within parentheses)

| | **Lone N** | **DET+N item** | |
|---|---|---|---|
| English N | 959 (5) | 37 (7) | 996 |
| Spanish N | 121 (6) | 10 (8) | 131 |
| | 1080 | 47 | |
| | **N-initial CS** | **DET+N-initial CS** | |
| English N | 100 (9) | 33 (11) | 133 |
| Spanish N | 15 (10) | 5 (12) | 20 |
| | 115 | 38 | |

## Address for correspondence

Rena Torres Cacoullos
Penn State University
Burrowes Building
University Park, PA, 16802
United States
rena@psu.edu

## Co-author information

Nathalie Dion
University of Ottawa
nathalie.dion@gmail.com

Dora LaCasse
University of Montana
dora.lacasse@mso.umt.edu

Shana Poplack
University of Ottawa
shana.poplack@uottawa.ca

## Publication history

Date received: 3 April 2020
Date accepted: 3 November 2020
Published online: 19 February 2021