Bilingualism: Language and Cognition 8 (1), 2005, 23–38 © 2005 Cambridge University Press doi:10.1017/S1366728904002056

Bare forms and lexical insertions in code-switching: A processing-based account*

JONATHAN OWENS
Bayreuth University

Bare forms (or Ø forms), uninflected lexical L2 insertions in contexts where the matrix language expects morphological marking, have been recognized as an anomaly in different approaches to code-switching. Myers-Scotton (1997, 2002) has explained their existence in terms of structural incongruity between the matrix and embedded languages, while Poplack (Budzhak-Jones and Poplack, 1997) looks to distributional symmetries or asymmetries with other anomalous phenomena such as non-standard case marking. In corpus-based studies, bare forms often emerge as clinal in nature, with full matrix language marking appearing alongside bare forms. This suggests that discrete structural constraints are not the only factors, nor necessarily always the dominant ones in play. In this paper it is proposed that on-line processing constraints governed in particular by lexical frequency effects and inherent latencies associated with inflectional attachment may lead to bare forms. The argument is based on a multilingual corpus of native Arabic speakers from the city of Maiduguri in northeastern Nigeria.

1. Borrowing and code-switching

In this paper I propose expanding a typology of insertions from L2 into L1 matrices on the basis of evidence from a code-switching (CS) corpus of native Nigerian Arabic speakers. The debate over lexical insertions in recent years has become orientated around two prominent poles. On the one hand, Poplack and her associates (e.g. Poplack, 1988; Poplack, Wheeler and Westwood, 1990; Adalar and Tagliomonte, 1998; Budzhak-Jones, 1998; Eze, 1998; Poplack and Meecham, 1998) have shown in a number of studies that lexical L2 insertions are often fully integrated into the grammar of the matrix language. Such insertions are interpreted as borrowings. Myers-Scotton (1997, 2002) has also described the grammatical integration of lexical insertions, but considers the phenomenon in terms of a theory of code-switching. Both writers have also noted that so-called bare, or Ø, forms (i.e. Ø inflection), forms which lack expected matrix language grammatical properties, are anomalous with respect to these explanations, and have sought to account for them in different ways. In this paper I suggest that an important aspect of the anomalous status of Ø forms can be explained in terms of findings from lexical processing research. The linguistic basis of the explanation is a corpus collected among native speakers of Arabic in northeastern Nigeria, though it is suggested that the results obtained there are generalizable to other situations as well.

* The author would like to acknowledge the support of the German Research Council (DFG) via the special research project 560 centered at Bayreuth University, as well as the unstinting support of Klaus Wolf in the Computer Center. For the interpretation of the Ø forms, Mr. Jidda Hassan of the Department of Languages and Linguistics, Maiduguri University, a native speaker of Nigerian Arabic, was indispensable. Aspects of this paper were presented at the 32nd NWAVE conference. Three readers provided valuable suggestions for improvement.

Address for correspondence
Arabistik, Bayreuth University, 95440 Bayreuth, Germany
E-mail: *Jonathan.owens@uni-bayreuth.de*

1.1. The Matrix Language Frame Model (MLF) of Myers-Scotton and the problem of bare forms

A well-formulated and well-known approach to the incorporation of L2 has been developed by Myers-Scotton (e.g. 1997, 2002). Expanding on a model of monolingual language production pioneered by Levelt (1989), Myers-Scotton proposed that basic structural elements of a sentence (in her recent work, the Complement Phrase or CP) – morpheme sequence and structural morpheme choice – were determined by a single language, termed the matrix language (ML). The matrix language countenances embedded language (EL) elements of different types. One type are single-word lexical insertions, whose presence in an ML constituent produces a so-called mixed ML-EL constituent, as in (1), where the English verb stem *comment* is integrated into a Swahili verbal frame formed by the system morphemes *si-ku*.

(1) *hata si-ku*-comment
even I.NEG-NEG.PST-comment
"I didn't even comment."
(Myers-Scotton, 2002, p. 89)

Another type are longer EL insertions which obey the internal structure of the EL language, but which are inserted in grammatically appropriate points in the ML, as in *le temps passé* in (2). This constituent has the internal structure of a normal French NP, but is inserted in the regular complement position to the Algerian Arabic genitive marker *taʕ*.

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

(2) *waħed* les histoires *taʕ* le temps passé
one the stories of the time past
"some stories of past times"
(Myers-Scotton, 2002, p. 116; after Boumans and Caubet, 2000, Algerian Arabic–French)

These two cases constitute what Myers-Scotton terms 'classic code-switching' (2002, p. 105), in which the overall sentence frame is set by a single matrix language.

In her recent work, Myers-Scotton (2002) has treated insertional phenomena which are not readily understandable in the classic CS framework. One such set of forms are what I term Ø forms, insertions into ML which lack the morphology mandated by the ML. Myers-Scotton, who terms them 'bare forms' (2002, p. 113–131), tends to look for structural congruity or incongruity between ML and EL as an explanation for them. In an example taken from American Finnish–English switching (after Halmari, 1997; Myers-Scotton 2002, p. 126), Myers-Scotton suggests that the lack of case marking (partitive *-a*) on *olive farm* in (3) below produces a bare form, because the case marker is adequately represented on the determiner element *tä*.

(3) *hoiti* *si-tä* olive farm
took care it-PART
"took care of the olive farm"

In this case, the bare form is sanctioned in the presence of other structures in the NP which bear the necessary case marking. In the following case from Dutch–Moroccan Arabic (MA) switching (after Boumans, 1998), on the other hand, the inserted Dutch noun lacks a determiner which is mandated by the rules of Moroccan Arabic (something like *waħed l-muʃkil* in MA).

(4) *waħed* *Ø*-probleem
a problem
"a problem"

Here Myers-Scotton suggests that "somehow, an Embedded Language feature offers sufficient competition to impede the Matrix Language procedure" (2002, p. 126). The competition from Dutch arises inter alia because Dutch does not have a gender system fully congruent with Arabic (MA has Masculine/Feminine, Dutch has Masculine/Neuter), and Dutch itself has a single determiner, not a bipartite one like MA, consisting of Specifier (*waħed* in example (4)) + Article. Residual influence from Dutch thus prevents the ML from assigning all the expected structural morphemes.

I return to Dutch–Arabic switching in the final two sections. I would note that in this article I use much of the terminology from Myers-Scotton's MLF model (like ML and EL), though in an eclectic manner that does not necessarily conform in all respects to her recent terminology (e.g. 'sentence' instead of 'CP'). In the examples, the matrix language is represented in italics, embedded language in regular font.

1.2. Poplack: Borrowing, lexical integration and bare forms

From a different perspective, Poplack and her associates (e.g. Sankoff, Poplack and Vanniarajan, 1990; Budzhak-Jones and Poplack, 1997; Budzhak-Jones, 1998; Eze 1998; Poplack and Meecham, 1998) contrast borrowing with code-switching. Borrowing, for them, occurs when L2 material is fully integrated into the L1 matrix (not Poplack's terminology), as in the following example from Canadian Ukranian–English switching.

(5) *do* car-*y*
to car-F.SG.GEN
"to the car"
(Buzdhak-Jones and Poplack, 1997, p. 226)

Here the English insertion *car* is marked according to Ukranian grammatical rules by the portmanteau suffix *-y* as required by the preposition *do*. For Poplack, *car* is borrowed, not code-switched, since its grammatical integration in the Ukranian matrix is complete. In contrast, in (6) the complement "friend" of *mav* "had" expects the plural accusative *-iv*, which is missing.

(6) . . . *vin* *ne* *mav* friend
he not had.M.SG friend-Ø
"he didn't have any friends"
(Budzhak-Jones and Poplack, 1997, p. 233)

More will be said about Budzhak-Jones and Poplack's treatment of such Ø forms leter in this section.

Besides these two types of insertions, a further class of insertions has been termed constituent insertions, where longer grammatical stretches are taken from the EL, as the English insertion in (7), comparable to (2) above.

(7) *ale vony*, they won't even go out, *jak kazhut*
"But they, they won't even go out, as they say."
(Budzhak-Jones and Poplack, 1997, p. 231)

Cases like (5) and (7) are unproblematic: (5) is considered a case of borrowing in the sense that the English word *car* is completely integrated into the Ukranian grammar while the internal structure of the insertion in (7) is determined entirely by English. The status of the Ø forms as in (6) is what interests Budzhak-Jones and Poplack. They cannot, ostensibly at least, be considered borrowings, since they lack ML grammatical markings, and they are not constituent insertions, as they have no internal constituency.

Budzhak-Jones and Poplack (1997, pp. 240–245) note that in their corpus of 399 English noun insertions, Ø-marked nouns are in fact overproportionally

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

represented in comparison with a benchmark set of 481 Ukranian nouns taken from the same corpus. This observation obtains for two sets of speakers, one older Ukranian immigrants into Canada and the other, second-generation immigrants. The percentage of Ø forms increases in the second-generation group, however.

Finally, they note that in the Canadian Ukranian corpus morphological marking can occur in which unexpected suffixes as in (8) appear, where in Standard Ukranian the masculine singular locative inflection *-i* is expected rather than the masculine singular genitive.

(8) *na-* back yard-*a*
on back yard-M.SG.GEN
"in the back yard"

It is noted that even in native Ukranian words (not exemplified in their article, though used as a statistical category), non-standard inflections occur.

Unlike Myers-Scotton, Poplack does not treat bare forms as an independent anomaly in her borrowing paradigm, but rather subsumes them under what she categorizes as non-standard forms. These can be Ø forms, but they can also be non-standard inflectional suffixes, as in (8). Furthermore, much more than Myers-Scotton, Poplack looks at bare forms quantitatively in their variable realization. I discuss her approach in greater detail in the final section.

1.3. The data and a plan of the paper

Bare, or Ø, forms thus emerge as a problem for two well-defined models of CS. Ø forms are also a prevalent category in a code-switching corpus gathered among native speakers of Nigerian Arabic (henceforth NA) in Maiduguri, Nigeria. English lexical insertions in particular are prone to appear bare. This corpus is the focus of the current paper. In the rest of this paper I first give a basic description of the corpus, and the language environment of Nigerian Arabs in Maiduguri (section 2). Two different corpora from Nigerian Arabs are used, one with a good deal of CS in it, the other largely monolingual NA, which serves as a benchmark comparison. I then briefly describe the grammatical properties of three constructions which are heavily used for English lexical insertion, what their normal structure is, and what important deviations are attested (section 3). After this, in section 4, I present a statistical comparison between the English lexical insertions in these three constructions on the one hand and, on the other hand, the corresponding statistics in the benchmark monolingual corpus. The striking differences observed between these two corpora serve as the basis for an explanatory account of Ø forms in terms of processing-based principles, found in section 5. After this explanation, in section 6, typological differences are drawn out between various CS studies with different Arabic dialects as the ML. These typological differences are related not so much to structural features of the contact languages, as to the sociolinguistic situation and to individual language competence. Finally, in section 7, the processing-based account is compared to explanations in terms of congruency (Myers-Scotton) or borrowing (Poplack).

Table 1. *Total words in CS corpus in different languages.*

	NA	Hausa	English	SA	Other
Total	47,230	29,216	3,560	908	2,943
% of total	56%	35%	4%	1%	3%

Note: The "other" category consists of a small number of Kanuri insertions, and proper nouns (= 2,847), which are excluded from the analysis. NA = Nigerian Arabic, SA = Standard Arabic.

2. The data and setting

2.1. A corpus of Nigerian Arabic from Maiduguri

In this section I provide background information to the study and describe English noun insertions from two corpora of native Arabic speakers, recorded in Maiduguri, Nigeria. The main corpus consists of about 87,000 words in ten different texts in which four languages are used: Nigerian Arabic, Hausa, Standard Arabic and English. The four languages have different discourse statuses. Nigerian Arabic and Hausa are frequently used as matrix languages, whereas Standard Arabic and English are used in this way hardly at all; rather, they appear primarily in lexical insertions. The different statuses of the languages can be illustrated in different ways. Since this is not the main theme of the paper, I restrict myself to one basic statistical observation, in Table 1, namely how many words appear in each language.

I distinguish between discourse languages (Nigerian Arabic, Hausa) and lexical donor languages (English, Standard Arabic). In this paper I concentrate exclusively on the properties of English noun insertions in NA matrices. Standard Arabic insertions display rather different integration properties from the English insertions, requiring an independent contrastive study (see Owens, 2002, pp.197–201, for brief comparison of possessive insertion). There are 1,121 English lexical insertions in NA matrices, about a third of the total of all English words. The remaining English words occur as complete sentences (i.e. as an ML), as constituent insertions, analogous to (2) and (7) above, or as lexical insertions (N = 484) in Hausa matrices.

There are a total of 32 active speakers in the texts, active speakers defined as those who use at least 250 words in a given text.

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

The second corpus is termed the 'monolingual corpus'. It consists of six texts, five from Maiduguri,[1] about 27,000 words, in which the only matrices are formed by NA. There is no constituent switching in the texts, and no discourse in anything other than NA. This corpus is thus monolingual in a performance sense: speakers use only NA matrices. Single-word insertions from English comprise only 185 tokens (195 including proper nouns like names of countries), less than 1% of the corpus, and these can be ignored as a separate category here (see footnotes 2, 3 and 10). For the purpose of this study, the monolingual corpus provides a benchmark for a 'normal' distribution of grammatical constructions. Nigerian Arabic, like all Arabic dialects, is not standardized and what detailed studies have been made of the variety (e.g. Owens, 1998) have shown it to be dialectally and sociolectally diverse to a degree. Under these circumstances it was decided that establishing what a 'normal' distribution of a given grammatical phenomenon is will be best effected on a quantitative basis, with the monolingual texts serving as the basic data.

2.2. Background information

In Maiduguri well over 40 languages are spoken natively, and there are three lingua francas. Up until World War II Kanuri was the main lingua franca of the city; thereafter Hausa has spread enormously and is now the primary lingua franca, though it has not completely displaced Kanuri (see Bross, 2002). Besides these two competing lingua francas, English as the de facto national language, language of instruction in the school system, etc. is a further lingua franca, even if oral competence in it is more confined than that of Hausa. A further language (or variety) in the mix in Maiduguri is Standard Arabic (SA). Still, SA cannot be said to have the status of a lingua franca as it is relatively confined both in terms of its population of speakers, and the domains where it is used. The greater spread of English as opposed to SA is directly reflected in the statistics in Table 1, where English is about three times more frequent than SA.

In Maiduguri, all minority linguistic groups will typically use their own language, plus Hausa, English to one degree or another, and perhaps some Kanuri. This is the case with Nigerian Arabs. In addition, segments of the Nigerian Arab population in Maiduguri will have a good knowledge of SA.

Nigerian Arabs are a linguistic minority, both in Maiduguri and in North-Eastern Nigeria as a whole. In each case they constitute at a very rough estimate – hard demographic statistics do not exist – 10% of the population. In Maiduguri this might put their number at up to 50,000 people. They can thus be said to typify a group of fairly large minority language speakers – Bura, Margi, Bade, Mandara, Fulfulde would be other languages in this demographic class – who are multilingual in at least two languages, their own plus Hausa, and will have some knowledge of English as well.

The two lexical donor languages, English, and to a smaller degree Standard Arabic, are particularly important in Maiduguri in that they serve as a source of new lexical concepts in a number of semantic domains, domains such as education, politics, sports, public services, public adminstration, finance or communications (e.g. recently, vocabulary associated with mobile phones), and the like. Examples are found in the next section.

3. The grammatical constructions

According to both Poplack's conceptualization of borrowing and Myers-Scotton's characterization of classic code-switching, EL inserted material should conform to the grammar of the matrix. In this section I describe three grammatical constructions in NA in which inserted English lexical material figures prominently. To the extent that Poplack's and Myers-Scotton's characterizations are met, it would be expected that the English insertions exhibit the same grammar as does the NA. However, deviations from the expected occur, and in this section are exemplified both what is normally expected and what deviates.

3.1. Definiteness

Nigerian Arabic, like most varieties of Arabic, formally has a bimodal system for marking definiteness. An indefinite noun is unmarked, as shown in (9a), while a definite noun is marked by the definite article prefix *al-*, as in (9b). *Al-* is assimilated to a following dental or alveolar consonant, cf. (9c).

(9) a. beet
"a house"
b. al-beet
"the house"
c. ad-dirdir < al-dirdir
"the wall"

The prototypical function of what is termed an indefinite noun is to introduce nouns (first mention). The three most important functions of what is termed a definite noun is to mark previously identified nouns, and to mark unique and generic reference. In the following excerpt from the monolingual corpus, *hettaab* "chaff from grain" is first

[1] The monolingual corpus has 17 active speakers, five of whom also appear in the CS corpus. As with the CS corpus, all are native speakers of Nigerian Arabic. One text was included from a village outside of Maiduguri. The monolingual texts are taken from the sample used in Owens (1998).

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

introduced in the indefinite form, subsequent mention being marked definite by the article.

(10) baadeen b-uṣubb-u lee hettaab, hu l-hettaab
then 3-pour-PL on.it chaff it.M ART-chaff
da ma b-ətallif al-qalla
this not 3-spoil ART-grain
"Then they pour chaff on top of it. This chaff doesn't spoil the grain." (2102187, J)

The definite article may occur as expected on an English inserted noun, as in (11).

(11) *fi iid al*-gaman
in hand ART-government
"in the hands of the government"
(504211, SL, in this context, previous mention of 'gaman')

It may happen, however, that sometimes an expected definite article does not appear and what I term a Ø form occurs. Though formally identical to the indefinite form, such forms may be considered to lack an expected definite article. In the example in (12), for instance, *kampeen masta* would be expected to have a definite article on the basis of the unique referent rule.

(12) *hu bas* Ø-kampeen masta *hana*
he just campaign manager of
abaasanjo (expected: *al*-kampeen)
Obasanjo
"*He* is (the) campaign manager
of Obasanjo." (500554, SQ)

3.2. Plural morpheme on English insertions: English, Arabic, Ø

Nigerian Arabic has two main classes of plural nouns, the so-called broken plurals, where the plural is marked by word-internal ablaut, as in *beet/buyuut* "house/houses", and suffix plurals, as in *ʔarabiyya* "Arab.F"/*ʔarabiyy-aat* "Arabs.F.PL".

Inserted English lexemes occur in contexts where a plural noun is required. English inserted nouns may appear in three guises. In such contexts they may be marked by the English plural morpheme *-s* (with different morphophonological realization), by a suffix from NA, or by Ø. The three cases are illustrated below.

(13) *English plural* -s
al-miin-z *hinee-k*
ART-means hana.PL-your.M
"your means" (302875, SI)

(14) *NA plural*
steet-*aat waad-aat bad-o* b-*ikaff -an*
state-PL.F some-PL.F begin-PL.M 3-pay-PL.F
"Some states have begun to pay."
(901785, BK)

(15) *Ø*
steet-Ø *waadaat qassad-an*
state- some-PL.F agreed.PL.F
"Some states agreed." (901830, KR)

3.3. Possessive pronouns

There are two main nominal possessive constructions in NA. In one the possessor nominal is juxtaposed immediately after the possessed. Following the tradition in Arabic linguistics, this is termed the *IDAFA* possessor. In the other it is linked to the possessed via an associative (genitive) morpheme, *hana* (F. *hiil*, PL. *hine*). I term this the indirect or *HANA* possessive construction. The possessor nominal can be either a noun or a pronoun. In this paper I concentrate exclusively on pronominal possessors. (See Owens, 2002, for a comprehensive variationist account of insertions in possessive constructions.) As will be seen in the statistics below, in the monolingual corpus a pronominal possessor nearly always occurs in the *idafa* possessive. The following two examples exemplify English insertions in the two different possessive constructions.

(16) *idafa: N possessed directly by pronoun*
yunifoom-*ak*
uniform-your.M
"your uniform" (204770, AL)

(17) *Indirect possessive, pronoun possessor attached to associative morpheme*
saaykula *hiil-hum di*
circular hana.F-their this.F
"this circular of theirs" (900947, J1)

4. Quantitative results

In this section I present the quantitative results which form the basis of the main argument set out in the following section. The procedure is to compare the behavior of the English lexical insertions in the CS corpus with the corresponding behavior of nouns in the monolingual corpus in respect of each of the three constructions summarized in the previous three sub-sections. The null hypothesis expects that the English insertions should behave in the same way as the nouns in the monolingual corpus. The following statistics will show quite clearly that they do not, and therefore that there is an explanatory case to answer for.

4.1. Marking of definiteness

Table 2 gives a cross-tab summary of how many nouns occur (i) correctly marked for the definite article (as in (9b, c), (10), (11)), (ii) correctly unmarked (as in (9b), (10)), or (iii) incorrectly lacking the definite article (as

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

Table 2. *Nouns in two corpora: definite, indefinite, incorrectly missing definite article.*

			Corpus		
			CS (English items only)	Monolingual	Total
Definiteness	definite	count	374	1575	1949
		% of total	9.7%	41.1%	50.8%
	indefinite	count	310	1258	1568
		% of total	8.1%	32.8%	40.9%
	incorrectly missing definite article	count	242	77	319
		% of total	6.3%	2%	8.3%
Total		count	926	2910	3836
		% of total	24.1%	75.9%	100%

(12)). 'Monolingual' signifies all nouns classified in the monolingual corpus, while 'CS' means only the English insertions in NA matrices in the code-switched corpus. As can be seen by inspection of Table 2, the biggest difference between the monolingual and CS corpus occurs in the 'incorrect' row: the CS is overproportionally represented here, while the monolingual is underrepresented. In fact, the 'incorrect' category represents over 25% of all the CS nouns vs. less than 3% of the nouns in the monolingual corpus.[2]

The basic observation here is that English lexemes inserted in NA matrices are overproportionally likely to lack the definite article in contexts where it is expected, as in (12) above.

4.2. Plural marking

In the monolingual corpus morphological plural marking is always correct, and therefore no statistics are presented at this point. I limit myself here to English lexical insertions in the CS corpus which occur in plural contexts (cf. section 3.2 below). The statistics are as follows: English insertion marked by *-s* = 72 tokens, by NA suffix = 19, by Ø = 39.

In over 25% of the cases the inserted noun lacks an expected marker of plural (whether from English or from NA). It can be noted here that morphologically-marked plurals, whether using an English or NA suffix, are always used appropriately, i.e. there are no instances where a English noun marked for plural is inserted in a singular context.

4.3. Pronominal possessor

In the previous two cases the 'Ø' category can be regarded as 'incorrect' in that the noun is expected to bear an affix which is in fact lacking. In the current case, so far as the suffixation of the possessive pronoun goes, both constructions are correct, according to rules of NA grammar. A pronoun can be suffixed directly to the noun (*idafa*) or to the morpheme *hana*. In practice, in the monolingual corpus a pronoun suffix is almost never suffixed to *hana* rather than to the noun, whereas in the CS corpus, as Table 3 shows, it is the normal way of marking possession with an English origin noun (83/104 cases).[3]

[2] The chi square significance levels for all statistics cited comparing the monolingual corpus and English insertions in the CS corpus are $p < .001$. Regarding the English noun insertions in the monolingual corpus, seven of 122 classified nouns, barely 5% of the insertions lack an expected definite article. This incorrect percentage is well below that of the Ø English insertions in the CS corpus.

[3] Note that in the monolingual corpus three of the six tokens with the *hana* possessor are of English origin, as in *lokal gaman hanaana* "our local government area". In the monolingual corpus, there are also three instances of English-origin nouns with a pronoun possessor (*kompanii-na* "our company"). The numbers are too small to make a comparison with the CS corpus. On the one hand, the *hana* possessor is also used with English possessed nouns; on the other hand, it is proportionally less dominant than in the CS corpus.

There are two structural cases in the data where the *hana*-PRO construction is mandated in NA. One involves a non-specific indefinite head noun, as in *kan fi daliil ha'na* "if there is a reason for it" (2104927 J). Here if the *idafa* possessor were used, *kan fi daliila*, a different meaning would result, something like "if its reason exists". The other involves pronominal uses corresponding to "mine, yours", etc. in English, as in *mumkin bifiil min hinee-hum* "he can take from theirs" (2105980 X). Neither of these is counted in the present statistics. Contexts are counted only where theoretically a grammatical choice between the two constructions is possible.

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

Table 3. *Type of possessor x corpus crosstabulation.*

			Corpus		
			CS	Monolingual	Total
Type of possessor	*hana* + pro possessor	count	83	6	89
		% of total	11.8%	.01%	12.7%
	idafa + pro possessor	count	21	589	610
		% of total	.03%	84%	87.3%
Total		count	104	595	699
		% of total	15%	85%	100%

Moreover, taking other factors into account, the percentage in favor of the *hana* possessive could be considered to be even higher. Eight out of the 21 English nouns which have a possessor attached directly to them (the *idafa*) have the pronoun attached not directly to the English stem, but rather to an NA suffix, either a plural morpheme, as in *tiicay-iin-hum* "teacher-PL-their.M = their teachers" or a so-called feminine singulative suffix -(*iy*)*it*, as in *caamaan-iyit-hum* "chairman-FSG-their = their chairman.F". Since there are no cases where an English noun with a plural or singulative suffix is used with the *hana* possessor (e.g. **tiicayiin hinee-hum*), a case can be made that it is not the NA origin of the noun stem which strongly favors the *idafa* possessor (as suggested in the monolingual statistics), but more specifically that the choice of type of possessive construction with pronoun possessor is determined by the last morpheme of the stem.[4]

Summarizing the three cases, a tendency can be identified of English lexical insertions to avoid (to anthropomorphize) NA inflectional material. This avoidance manifests itself in two ways. First, when there is a choice between two constructions, one requiring affixation of NA material, the other not, then the second alternative will be favored. This is what happens in the two possessive constructions: the indirect *hana* possessor is favored for possessive pronoun suffixation because it allows the suffix to be attached to an NA morpheme, rather than to the English insertion. Secondly, when there is no alternative to affixing directly to the English lexeme, the NA affix will simply not be added in many instances. This is the case with the definite article and the expression of plural. I discuss these two situations at greater length in the final section.

[4] Cf. the more extreme condition in Hungarian–English switching (Bolanyi 2003, Moravscik 1975) where English verbs must be 'Hungarianized' via the suffix -*l* before they can be inflected with Hungarian verbal morphology.

4.4. Interacting factors

To this point English insertions in the CS corpus have been globally contrasted with the monolingual corpus. Taking the latter as the benchmark norm for the three constructions, English insertions emerge as overproportionally likely to lack expected NA inflections. Before moving on to the more general discussion, it is worthwhile to look in greater detail at linguistic factors which may, or may not, influence the occurrence of inflections on the English nouns. The occurrence of the definite article and of the plural morpheme will be examined here in terms of possible phonological influence.

Beginning with the definite article, it may be asked whether the lack of the definite article isn't due to phonological factors. Three elements are considered here, length of the English word in syllables, whether or not the first syllable should induce assimilation of the definite article (as in (9c), for example), and whether the insertion is a single word or a multiple-noun insertion.[5] A regression analysis was run with presence/absence of the definite article as the dependent variable (those tokens codified as 'correct' or 'incorrectly lacking definite article' in Table 2). The results in Table 4 show that the combined contribution of these factors, with a multiple R of .18, does not account for a great deal of the overall variation. Looking at the individual factors, only the number of syllables returns a clearly significant result, with assimilable consonant slightly above the .05 significance level.

The status of the insertion as multiple- or single-word plays no role in determining the insertion of the definite article. The number of syllables does. Here a chi square

[5] Multiple nouns are defined strictly by stress criteria. Words with independent stressed syllables were considered multiple, those with a single stress as one word. By this criterion, not only were familiar combinations, such as *ˈfedəral ˈgaman* considered multiple, but also a token such as *ˈhed ˈkota* "headquarter", which was clearly rendered as two separate stress units.

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

Table 4. *Multiple regression results on the effects of phonological factors on the occurrence of the definite article.*

	B	Beta	T	Sig T
Number of syllables	.158	.15	3.85	.00
Assimilable consonant	−.14	−.08	−1.7	.09
Multiple word	.09	.03	.74	.46

distribution shows that monosyllabic insertions like *bool* "ball" and *fakts* "facts" overproportionally accept the definite article and underproportionally lack the definite article as compared to multisyllabic words 2–6 syllables long. As far as assimilation is concerned, the tendency is for words beginning with an assimilable consonant, the alveolar sounds, to lack the definite article (e.g. *Ø-steet* vs. *as-steet* "the state"), though as noted the effect is below significance.

Turning to the phonology of the plurals, here it is the last sound of the word which was looked at. The last sound of the word was initially divided into five classes: single alveolar final (e.g. *studen*, yielding *studen-s*), alveolar final at end of CC cluster (e.g. *fakt-* yielding *fakt-s*), other final C, and V final. The classification followed the surface realizations, hence *studen-* just cited was classed as having a single final alveolar, while the token *student-* was classed as an aveolar final at the end of CC cluster. It would be expected that the Ø-plural could be influenced by final alveolarity in the case of the English plural in particular, a final alveolar favoring the non-occurrence of English *-s*. With certain exceptions, NA does not allow final CC clusters and categorically disallows final CCC clusters. The suffixation of *-s* to a C- or CC-final stem produces such unacceptable clusters. Overall, as seen in Table 5, it is difficult to discern significant tendencies, even if the two variables, plural suffix and final consonant, show a significant association ($chi^2 = 9.41$, $df = 4$, $p < .045$, Fisher's exact test). Note that in Table 5 only three phonological factors are shown. The variable 'final alveolar in CC cluster' produced only one token with a Ø affix and none with an NA plural suffix, so this variable was combined with the 'final alveolar C' category.[6]

A final comparison is particularly relevant to the discussion in the final section below, and detailed comment on it will be reserved until then. This examines the relation between the expression of definiteness and plurality on the English insertions. The results are in Table 6. The association between definiteness and type of plural suffix on the English insertions is not significant ($chi^2 = 6.7$, $df = 4$, $p < .15$, Fisher's exact test). What is noteworthy is that in the 'incorrect' row only one token of an NA plural suffix occurs. This result echoes the observation in 3.3 above that when an English noun receives an NA suffix (e.g. a plural suffix), a possessive pronoun is always attached directly to the noun, never to *hana*. That is, the presence of a NA suffix implies that the word will be treated distributionally like a NA noun.

Summarizing this section, compared to the stark contrast between the monolingual vs. the CS corpus, individual phonological factors investigated relating to the (non-)occurrence of definite article or a plural marker are not particularly striking, accounting for a relatively small amount of overall variation and often returning non-significant results. The two phonological factors which fall within or close to the significance range are discussed briefly in the next section.

5. A processing-based account

The account of the Ø forms which I would like to propose is based on basic findings of language processing. There are two main assumptions and one axiom in the explanation. The two assumptions are as follows.

1. Processing time increases with decreasing lexical frequency.
2. Processing of bound inflectional morphology is more demanding than processing whole words, as measured by reaction time to a given stimulus.

The first assumption is basic in psycholinguistics and needs no further elucidation (cf. e.g. Bybee, 1995, p. 232; Stemberger, 1995, p. 258).[7] The second assumption is well-supported when morphologically complex inflected forms are compared to simple monomorphemic ones. As this is the locus of the current study, I restrict my assumptions to inflectional aspects of morphological processing. To cite but one case, in a study involving Finnish case endings, Laine, Vainio and Hyönä (1999) determined processing costs for morphologically complex words with various types of Finnish case endings as opposed to morphologically simple words. These costs

[6] As a matter of interest, an English plural occurs with nine tokens of the $CC_{aveolar}$ category. Furthermore, regression analyses on type of plural suffix as dependent variable were carried out using the independent variables: presence of definite article, phonology of final segment and syllable length, or the same factors plus sentence function (topic, subject, predicate, object, object of preposition). These returned no significant results and showed a low multiple R.

[7] A third assumption may be necessary as well, namely that L2 processing is slower than L1. This may, however, be subsumable under assumption 1, L2 being a subset of lower frequency. Alegre and Gordon (1999) do indicate that facilitating whole-word frequency effects even in the case of (English) inflections may be observed in words of higher frequency (the authors suggest that more than six tokens per million are facilitating). They do not contrast the relative effects of inflectional vs. derivational affixes in their study, however.

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

Table 5. *Word final phonology and plural suffixation.*

			Plural			
			Ø	NA suffix	English pl	Total
Plural phonology final stem sound:	alveolar-final	count	17	8	29	54
		% of total	13.2%	6.2%	22.5%	41.9%
	other C-final	count	10	1	6	17
		% of total	7.8%	.8%	4.7%	13.2%
	vowel-final	count	11	10	37	58
		% of total	8.5%	7.8%	28.7%	45%
Total		count	38	19	72	129
		% of total	29.5%	14.7%	45%	100%

NA = Nigerian Arabic; 1 cell with expected count less than 5.

Table 6. *Definitess x plural suffix crosstable, English insertions in CS corpus.*

			Plural			
			Ø	NA	English pl	Total
Definiteness	definite	count	13	11	37	61
		% of total	10%	8.5%	28.5%	46.9%
	indefinite	count	16	7	20	43
		% of total	12.3%	5.4%	15.4%	33.1%
	incorrect	count	10	1	15	26
		% of total	7.7%	.8%	11.5%	20%
Total		count	39	19	72	130
		% of total	30%	14.6%	55.4%	100%

NA = Nigerian Arabic; 1 cell with expected count less than 5.

clearly derive from the case suffixes themselves, as whole words ending in the same phonological form (pseudo-inflections, e.g. *nasta* "tack", where the sequence *sta* is part of the stem vs. *auto-sta* "from the car", elative case) showed no more costs than did whole words of completely different form (see also Caramazza, Laudanna and Romani, 1988; Bybee, 1995, p. 231; Caramazza and Chialant, 1995; Levelt, Roelofs and Meyer, 1999, p. 13; Portin and Laine, 2001).

I furthermore assume the following axiom:

3. When code-switching, speakers basically maintain the same rate of speech as when they are speaking in a monolingual mode.

I will elaborate on these points below. First, however, the basic explanation for the Ø forms. For Nigerian Arabs the processing of English inserted nouns in inflectable matrices will be doubly slow. They will be slower because the lower frequency of the English lexemes themselves will slow access, and they will be slower because of latencies involved with inflectional processing. The rate of speech, however, is constant. Under these circumstances, should lexical access to the L2 English noun be inordinately long, morphological attachment will be dispensed with in order that the flow of speech continue unabated. The result is a Ø form.

Applying this to the descriptive material above, two options need to be specified. For the failure of the definite article or a plural suffix to appear, the description in the previous paragraph is relevant. Alternatively, when speakers have the choice of an equivalent construction which avoids morphological attachment altogether, this

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

Table 7. *Type/token ratios for monolingual corpus English insertions in CS corpus.*

a. Definite/indefinite nouns	
all noun insertions	
529/2825 = .19	452/926 = .49
only definite nouns marked by *al-*	
337/1213 = .28	225/374 = .60
b. Nouns with plural affixes	
57/297 (sample) = .19	83/130 = .64
c. Nouns with pronominal suffix (*idafa*)	
158/589 = .27	16/21 = .76

alternative will be chosen. In this latter case the explanatory points 1–3 above act as a filter, directing the speaker towards the indirect *hana* possessor, described in 3.3 and 4.3.

This processing-based explanation is to be viewed as a hypothesis, which in certain respects can be supported by ancillary observations, but which in other respects will have to be left as a hypothesis pending further research. In this context I would emphasize that the data deals with, essentially, an unresearched linguistic area so far as corpus-based questions are concerned, so the standards of proof which often accompany such approaches cannot be offered. Corroborating evidence does exist, however. Two types can be adduced.

First, to substantiate that English insertions have a lower degree of frequency than do the nouns in the comparable constructions in the monolingual corpus, the following sets of comparisons are offered. The monolingual corpus is compared to the English lexical insertions in the CS corpus in terms of the type/token ratios derived from these two corpora (see Table 7). For the definite/indefinite noun and the N + pronominal suffix comparison, the entire monolingual corpus was analyzed. For the N + plural affix, a sample of the monolingual corpus was scanned consisting of the first 1,000 and the last 500 words of each of the six texts, altogether about a third of the corpus. In all categories the type/token ratios are considerably lower in the monolingual corpus compared to the English insertions in the CS corpus (by differences ranging from. 28 to .49), indicating that the English lexical frequencies are lower than the native vocabulary of the monolingual corpus.[8]

Table 8. *Pauses before object in verb–object sequences: No pause pause.*

Monolingual	869	16
English in CS	324	12

Secondly, in support of the rate of speech axiom, object nouns were examined for whether or not they were preceded by pause.[9] As far as English insertions go, the object position is the most frequent one for lexical insertion. In the vast majority of cases in both the monolingual and the English insertions in the CS corpus there is no pause, as Table 8 indicates. The English insertions, proportionally, are slightly more likely than are object nouns in the monolingual corpus to be preceded by pause, though the difference is below significance ($p < .09$), and in both the absolute numbers for preceding pause are very low. English insertions in this corpus, thus do not appear to inordinately attract pauses before their insertion.[10]

Parameters corroborating assumptions/axiom 1–3 can be extracted from the corpus, and – to the extent that they can be found – the explanatory hypothesis for the high

[8] As one reader points out, it is necessary not only to compare type/token ratios between 'native' words and English insertions in the CS corpus, but also to elucidate the relation between Ø and affixed forms among the English insertions. It could be, for instance, that Ø forms concentrate among frequent English insertions, or alternatively, among infrequent ones. As one test of this hypothesis, I listed all English insertions (CS corpus) of words which have 10 tokens or more in any context. The word *nomads* for instance, occurs not only as a single lexical insertion, whence its inclusion in the statistics in this study, but also in a complete English sentence (e.g. *there are other nomads*). In all there are 28 English words (types), each with 10 tokens or more. These were then inspected for whether they belong in the *al-* or Ø definite noun class (a subset of the total). From this set, 201 tokens are classified as having either *al-* or Ø, divided into 81 tokens (24 types) lacking an expected definite article, 120 (28 types) with the definite article. 4 types are restricted to correctly-marked definite insertions, the other 24 realized by both correctly and incorrectly marked ones. The approximately 3:2 ratio (definite article:lacking expected definite article) closely mirrors the overall ratio of these two categories in the corpus (see Table 2). In English lexical types of higher frequency at least, Ø/*al-* is statistically evenly distributed.

[9] Potentially, each word boundary is followed by pause. Perceptible pauses were marked in the transciptions, but were not measured individually in milliseconds. Besides hiatus, pauses are typically characterized by a falling tone, and will generally coincide with a structural boundary (e.g. sentence), though of course precise generalizations require detailed study. As a further indication that pause does not correlate with Ø definite article, one may consider the global distribution of pauses before English lexical insertions, marked by the definite article *al-* or by Ø (i.e. missing an expected definite article). Beginning with nouns marked by the definite article, the figures (pauses/noun) are 32/372 vs. 20/246, $p < .87$. The two groups are preceded by pause at the same rate.

[10] In this respect the present corpus differs considerably from Poplack's work in Canada, where flagging – marking insertions with pause or with discourse fillers of various sorts – is quite common (Poplack, 1988, p. 238, on English–French switching; Poplack, Wheeler and Westwood, 1990, on Canadian Finnish-English; Budzhak-Jones and Poplack 1997, p. 245, on Canadian Ukranian–English). In the monolingual corpus, incidentally, none of the object nouns before which a pause occurs are of English origin.

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

percentage of Ø forms among the English insertions in the CS corpus can be maintained. Of course more wide-ranging data is desirable, including type/token frequency counts both for NA and for English in North-Eastern Nigeria, for English among NA speakers, experimental evidence, and so on. In the light of these desiderata, the present treatment can be seen as a specific hypothesis for future orientation.

Note that the broad explanation for Ø forms presented in this section does not preclude the effect of further factors. It was noted in the previous section that two phonological factors, monosyllabicity favoring the insertion of the definite article and an assimilable initial consonant disfavoring insertion, could be identified as well. Both of these, moreover, potentially have a processing-based interpretation as well: monosyllabic English insertions are easier to retrieve than are multi-syllabic ones, and assuming an extra cost for assimilation, definite article insertion will be disfavored in contexts where it must be assimilated. However, the magnitude of these two phonological effects are relatively slight in comparison to the large differences obtained between the monolingual and the English insertions in the CS corpora.

6. Comparative studies

It was seen in the statistical section that there is a clear tendency for English noun insertions, as opposed to native NA nouns in the monolingual corpus, to avoid morphological affixation. There are a number of studies on Arabic code-switching which allow, in greater or lesser detail, a direct comparison with the present data. They may be classified into two categories, based on the parameter of how EL insertions are treated morphologically.

On the one hand are various studies carried out on Arabic–French CS in Morocco (Bentahila and Davies, 1983, 1998; Nait M'barek and Sankoff, 1988) and Algeria (Chebchoub, 1985; Boumans and Caubet, 2000), and on Arabic–English bilinguals (of Jordanian, Egyptian, Palestinian origin) living in the USA (Atawneh, 1992; Myers-Scotton, Jake and Okasha, 1996) and Jordan (Mustafa and Al-Khatib, 1994). In all of these cases the degree of bilingual competence is reported to be very high. For present purposes of comparison only constructions directly comparable to those treated here are summarized. These are the occurrence of the definite article in all cases, and the occurrence of the possessor pronoun for the non-North African studies. In western North African Arabic a construction comparable to the indirect *hana* possessor (using the morpheme *dyal*, *di*, *d* or *mtaʕ* instead of *hana*) has become the norm, while the *idafa* is rarely used. Therefore comparison between the two types of possessive constructions is not relevant here.

As far as the three non-North-African studies go, none deal specifically with the question of morphological marking of lexical insertions, and therefore they lack extensive quantitative information. Looking through the individual examples in these studies an impression can be gained, however, and it emerges that in nearly all cases English nouns are marked 'correctly' according to the rules of the relevant Arabic dialect matrix. For instance Mustafa and Al-Khatib (1994, p. 220) cite the example:

(18) *il*-communities *ṣ-ṣaɣiira tʕallam-u*
ART-communities ART -small learned
istiʕmaal l-manure
use ART-manure
"Small communities learned the use of manure."

Example (18) has two insertions from English, *communities* and *manure*, both marked by a definite article. Here the definite article is sanctioned by the 'generic' reference condition.[11] In the three studies I count 33 tokens of nouns in these non-North-African studies correctly marked (i.e. correctly marked with the definite article or correctly lacking it if indefinite). There is only one token where it appears that the definite article is expected but missing:

(19) *ħalla n-ʃuuf* decomposition rate *tabaʕ-ha*
now we-see of-it.F
"Now we see its decomposition rate."
(Mustafa and Al-Khatib, 1994, p. 222)

The pronoun on *tabaʕ-ha* marks the English insertion as definite, yet the definite article is missing. In the same study one finds *saturation* 'correctly' marked as definite and the definite article correctly assimilated to *-s*.

(20) *is*-saturation *tabaʕ-ha*
ART-saturation of-it
"its saturation"
(Mustafa and Al-Khatib, 1994, p. 221)

As far as occurrence with pronominal suffixes go, data comparable to that in section 3.3 is found, though only in three tokens, one exemplified in (19) and (20), where *tabaʕ* is structurally equivalent to *hana*). Note that as with NA it is theoretically possible in Jordanian Arabic to attach the possessor pronoun directly to the noun (*idafa* possessive). In this respect the relatively little data found is comparable to the NA, with the indirect possessor used with English insertions.

[11] In fact, the translation is given as "The small communities learned the use of manure". Since the example is not cited in a larger context, it is not clear whether or not the translation "small communities" should include the definite article. In my interpretation it is a generic statement, which in English would not take the definite article. Mustafa and Al-Khatib have used the definite article "the" in their translation, implying previous mention, and therefore non-generic. However, the definiteness marking on "manure" can only be construed as generic, which leads me to suspect that "small communities" is generic as well.

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

The North African Arabic data can be dealt with more perfunctorily, as the study by Nait M'barek and Sankoff (1988) provides unambiguous quantitative evidence, and is commensurate with the non-quantitatively based evidence in Bentahila and Davies (1983) for Moroccan Arabic and Boumans and Caubet (2000, p. 154) for Algerian. As far as the definite article is concerned, in all instances (N = 110) where a French noun is inserted in an Arabic matrix, the rules of Arabic are followed; (21) illustrates.

(21) *fi* *waħid* *l*-paysage *dyal* *al-xawoir*
exist a ART-passage of ART-extraordinary
"There is an extraordinary landscape."
(Nait M'barek and Sankoff, 1988, p. 149)

In (21) the French noun is inserted into a frame in which North African Arabic dictates that the noun be marked by a definite article. In general in North African Arabic demonstrative adjectives and the specifier *waħid* require that the head noun be marked by a definite article. In fact, the definite article can come from Arabic, as in this example, or, and this is more frequent, from French (e.g. *duk les avions* lit. "those the planes = those planes", Nait M'barek and Sankoff 1988, p. 149). The explanation for the use of the French rather than Moroccan Arabic article has generated a considerable degree of debate (see Heath, 1989; Boumans, 1999; Muysken, 2000, pp. 81–87; Myers-Scotton, 2002, p. 123) though the matter is not directly relevant to the basic observation that examples like (21) with the NP structure Specifier + Art–Noun are mandated by North African Arabic grammar, rather than by French grammar. What one does not find, and this is crucial to the current summary, are examples lacking any article, examples like *fi waħid Ø paysage* "in a landscape". That is, in North African Arabic–French switching, Ø forms are rarely reported in the literature.

Summarizing this aspect of the comparative data, there is considerable evidence, some based on extensive quantitative material, some gleaned from examples illustrating other issues, that EL items inserted into Arabic conform completely to the rules of Arabic. It is worth reflecting briefly on the situations where this type of CS is attested (see Muysken, 2000, chapter 9, for a summary of broad correlations between switching type and social setting). Foremost is the situation in North Africa, where, as seen, French lexemes are completely integrated into North African Arabic frames. In North Africa French and Arabic have been in close contact for over a century, and the degree of bilingualism is high. In Morocco, French continues to have an officially-recognized status in many domains (e.g. education), and even in Algeria, where it was officially displaced by the Arabicization policy of the early 1990s, it remains an important language, both practically and symbolically (see Brahimi and Owens, 2000). There has been ample time for area-specific CS strategies to develop (e.g. use of an etymological French determiner in place of Arabic *al-* in an NP frame mandated by Arabic). Likewise, in the Arabic–English studies cited, it appears that the degree of bilingualism is high. Mustafa and Al-Khatib (1994, p. 216) explicitly state that their corpus comes from fully bilingual speakers. Provisionally, then, it may be suggested that a high degree of bilingualism favors full morphological integration of EL nouns.[12]

The counterpole, to turn to the second type of situation, is offered not only by the NA data set out above, where Ø forms are prevalent, but also by Arabic–Dutch CS in Holland. Most Arab immigrants to Holland are of North African origin, so their Arabic only offers a basis of comparison in the definite article (see above). Nortier (1990, pp. 197–205; also Boumans, 1998, pp. 187–191 in confirmation) discusses the realization of the construction at length. In her data, from Utrecht, there are 285 single Dutch nouns inserted in Arabic matrices (what she terms the 'host language') and of these, she determines, 82 occur in contexts where, according to the rules of Moroccan Arabic, a determiner is expected but missing (Nortier, 1990, p. 199). The expected grammatical marking in Moroccan Arabic is similar, though not identical, to that in NA, described above. One such example is (22).

(22) *waħed* *Ø*-bejaardencentrum
one old people's home
"an old people's home" (Nortier, 1990, p. 199)

This is structurally the same construction as (21) above, where the specifier *waħed* must be followed by an article.

Strikingly, the token count of 82/285 = 29%, the number of tokens where a definite article is expected but missing on the inserted Dutch noun divided by the total of all noun tokens, is very close to the 'incorrect' Ø marking of English noun in the NA data (see Table 2 above). Based on this observation, a second typological situation can be established. In contrast to the studies summarized above, here Ø forms on inserted nouns are

[12] Noun insertion is not the only insertional phenomenon which can be adduced here. It is surely no coincidence that precisely in the same corpora a high degree of morphologization of inserted verbal material is attested as well, i.e. English or French insertions are rendered not by a "do + object" construction, common in many parts of the Arabic world, including Nigeria and Nortier's and Bouman's Dutch material, but are adapted to Arabic morphological patterns. Examples are *t-at-gratter* "you scratch" from Morocco (Bentahila and Davies, 1983, p. 315) and from Arabic CS in the USA, *nudris-ing* with an Arabic imperfect "we are studying" amplified by an English imperfect suffix, and *cancel-t-uh* "I canceled it", with an Arabic perfect suffix + object suffix attached to the verb (Myers-Scotton et al., 1996, pp. 27, 33). Apparently, the Moroccan examples are routinized to a high degree whereas the American ones, so far as the literature allows a general interpretation, are highly creative but also spontaneous and not conventionalized.

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

common. Just as the global social situation helps inform the maintenance of Arabic inflectional material in North Africa and elsewhere, so too does it help understand the relative prevalence of Ø forms in the NA and Dutch situations. In Holland, Arabic is in a state of flux. Moving from first to second and third generation speakers, there is a gradual shift from Arabic dominance to Dutch dominance (see El Aissati and Boumans, 2001). The conditions for stable bilingualism as in North Africa, or even in Jordan for that matter, where English is an important language of education, do not exist. Similarly in Maiduguri, Hausa, not English, is the main lingua franca, and the quality of English language teaching in schools is poor. Thus, the explanation for Ø forms or lack thereof based on individual language processing has a corollary in the global social status of English (or Dutch as far as Holland goes) in the relevant community of speakers. Ø forms will tend to occur where, even if important, one of the contact languages is still socially restricted (English in Maiduguri), or where the overall statuses of the languages are in a state of flux, with one of the languages regressing.[13]

It was seen in this section that studies on CS with Arabic as the matrix language reveal two different treatments of lexical insertions. In one group, the insertions are fully integrated grammatically into the Arabic ML, whereas in the other group, a large percentage of insertions surface as Ø forms. The problem of Ø forms exemplified in detail for NA CS in the statistical section thus has broad sociolinguistic typological dimensions as well.

At this point, looking at individuals in the NA corpus, I would like to add one more piece of evidence in support of the idea that different competences in the contact languages result in different insertional patterns. I would note first of all that language tests were not carried out among the corpus participants, as it would have been difficult, given the spontaneous nature of the recordings, and in any case would not necessarily have been of great reliability (see Owens, 1998, p. 208 for difficulty in administering standardized tests in a culture where such are virtually unknown). Instead, given the author's personal knowledge of the participants, two participants clearly stand out as having a better knowledge of English than the others.[14] One is a lady (identified as FM) who lived in England for four years and holds an MA from the University of Maiduguri in Business. The other (identified as J1 in the tables below) is a lecturer in the Linguistic Department, the only participant in the corpus who must use English extensively in his daily work. From the processing-based perspective, it would be expected that these two speakers would have a lower degree of Ø forms for English insertions, since their better knowledge of English should facilitate lexical access. This expectation is borne out, though in some cases their scores are only slightly above those of other participants. In Table 9 'correctness' indices are given for the insertion of English nouns in the definite and plural contexts. The percentages give the number of cases an English noun is correctly inserted into a given context, according to the rules of the NA matrix language. For the definiteness column the list is limited to the first ten speakers in terms of tokens, while the plural context is limited to participants with at least 5 tokens of an English insertion in a plural context. FM and J1 (marked in bold in Table 9) have the highest 'correctness' indices for both constructions.

Table 9. *Individual 'correctness' indices: Definiteness, plural.*

Definiteness		Plural	
FM	79/102 = 78%	**FM**	9/10 = 90%
J1	232/304 = 77%	**J1**	48/60 = 80%
SK	14/18 = 77%		
KR	29/38 = 76%	KR	4/5 = 80%
SL	53/77 = 69%	SL	8/11 = 72%
ED	55/82 = 67%	ED	7/19 = 37%
SQ	14/22 = 64%	SQ	3/5 = 60%
SI	23/39 = 59%	SI	1/5 = 20%
BK	21/40 = 52%	BK	7/11 = 64%
AB	19/39 = 49%		

7. Processing difficulties and other explanations

In this final section I situate the processing-based account of the Ø forms in the two broad approaches to code-switching summarized at the beginning of the paper. A basic observation is that the Ø forms conform neither to Poplack' characterization of borrowing nor to classic code-switching in Myers-Scotton's MLF framework. For Poplack, borrowed items are completely integrated into the matrix language. As shown above, however, English insertions in the CS corpus overproportionally lack

[13] Similarly, Halmari's (1997, pp. 145–155) main explanation for Ø forms in Finnish–American English switching. She notes that although Ø forms are prevalent (162/550 or 29% of English nouns lack expected Finnish inflections), the great bulk of these are found among less fluent, second-generation Finnish speakers, and/or are produced 'non-fluently'. That is, they are marked by pause and repetition.

[14] Halmari (1997) similarly bases her assessments on personal knowledge of the speakers. Moreover, as Poplack and Meecham (1998, p. 128) point out, the ultimate test of language form is embodied in a spoken corpus, not abstract language tests. However, the present researcher simply did/does not have the time or resources to conduct a corpus-based survey of English in Maiduguri with which to establish an English baseline of comparison.

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

expected inflections, and they overproportionally occur in the indirect *hana* possessive construction. Formally and distributionally they diverge from native NA nouns. By the same token, they do not meet the expectations of classic code-switching, where it would be expected that in mixed embedded language islands, the English insertions would appear with NA inflections. Of course, in many cases they do, but there is a statistically significant set of insertions which do not, which gives rise to the current processing-based explanation.

If they do not conform to the borrowing or mixed EL island criterion, neither do they represent constituent insertions. These have the internal structure determined by the EL. The data considered in this paper has no internal syntactic structure, however, and certainly not the internal syntactic structure of English.

Both Poplack and Myers-Scotton have recognized the problematic status of Ø forms. Their solutions, however, are different from the approach developed here. Poplack notes that the majority of English insertions in the Canadian Ukranian data described in section 1.2 above appear as bare forms, 52% for first-generation speakers, 73% for second-generation speakers. However, Budzhak-Jones and Poplack (1997) do not directly address the reason for the large number of bare forms among English insertions, and while they give bar-graph statistics clearly indicating that compared to native Ukranian nouns (1997, p. 240), English nouns surface much more frequently bare, they do not carry out statistical comparisons between the two classes. A direct comparison with the current data is thus impossible. Their comparison does yield the interesting result that non-standard (usually bare) English insertions mimic the distribution of non-standard (usually non-standardly-inflected) Ukranian nouns among the first generation of speakers, but not among the second.[15] However, even if the criterion, patterning like non-standardly inflected Ukranian nouns, can be said to be a characteristic of borrowing, the comparison still says little about the origin of the Ø forms themselves in their data. In fact, it ignores the obvious difference between the English insertions and the Ukranian nouns, namely that the former are usually bare and the latter usually inflected, even if non-standardly. 'Morphological integration' is factored out of their definition of 'borrowing'. What is interesting for comparative purposes, however, is the authors' basic observation not only that in the majority of cases English nouns surface bare, but also that this tendency increases considerably in the second-generation Ukrainian speakers. This shows further that language knowledge plays a role in predicting a greater occurrence of bare forms, as outlined in the previous section. It was observed there that CS studies on Arabic divide into two broad types with respect to bare nouns. In one type, integration into ML grammar is nearly perfect, whereas in the other, bare forms often occur. It was further suggested that these two situations roughly typologize along the basis of sociolinguistic settings where L1 and L2 knowledge is very high (a large degree of integration) vs. those where one of the two may be less than perfect (Ø forms).[16] It is precisely in the latter case where the processing-based explanation of Ø forms proposed in the previous section would be operative.

In contrast to Poplack, Myers-Scotton does offer a specific explanation for bare forms, indeed from two related perspectives. First, referring to the Arabic data from Utrecht (see section 6 above), she suggests that a general lack of structural congruity between the ML and EL will result in bare forms (2002, p. 126). The trouble with this explanation from the present perspective is that the current data compares the same ML/EL language sets and shows that different outcomes occur. In Arabic/English in Maiduguri, Ø forms are frequent whereas in Arabic/English in Jordan or America they are, apparently, not. Language structure alone cannot be the sole conditioning factor. In a second case, discussed in section 1.1 above, it was seen that Myers-Scotton suggests that a cause for the high degree of Ø forms on Dutch nouns in Nortier's Dutch–Moroccan Arabic CS data is due to a residual projection of incongruent Dutch features into a Moroccan Arabic frame.

This latter hypothesis can be tested indirectly with one set of data presented in Table 6 above. There a cross-tabulation is presented for the expression of

[15] The issue which they address is somewhat complicated, involving comparisons between two generations of speakers and comparisons between English and Ukrainian vocabularies in different grammatical contexts (Budzhak-Jones and Poplack, 1997, p. 248). Their multivariate analysis makes the following comparison. They compare the behavior of what they term 'non-standard marking' in Ukranian nouns with non-standard marking in nouns of English origin. Note that the characterization of 'non-standard' includes two parameters. A noun is non-standard if it (i) has a non-standard affix, as in (7) above, or (ii) surfaces bare in a context where an affix is expected. ("non-standard marking, here defined as any inflection, null or overt, which is at variance with the requirements of standard Ukranian grammar" (1997, p. 238)). The category 'non-standard' thus collapses the very variable which the current study is concentrating on, namely lack of an expected morphological mark, with a completely different variable, namely the presence of a morphological mark, but in the wrong (or "non-standard") form.

[16] I must hedge here, specifying 'one of the two' (ML or EL) in consideration of the Dutch–Moroccan Arabic situation, where in the Dutch studies (Nortier 1990, chapter 6; Boumans 1998, p. 162) Moroccan Arabic/Dutch dominance varies for different group members. Since Nortier gives only a global count of Ø forms, undifferentiated for which group produces them (dominant Dutch or dominant Moroccan Arabic), it can only be surmized that the Dutch–Moroccan Arabic contact situation in general produces linguistic effects similar to those in Maiduguri. These conclusions, of course, may need to be revised in the light of a closer analysis of the Dutch data.

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

plurality on inserted English nouns with the expression of definiteness, marked by the definite article *al-*. A reasonable interpretation of Myers-Scotton's residual influence theory would be that English insertions which carry the English plural *-s* would negatively correlate with the presence of the definite article. That is, the English plural would increase the 'Englishness' of the insertion, which should inhibit the appearance of the definite article from NA. Note that the presence of an NA suffix (either plural or feminine singulative as described in section 4.3 above) on the English insertion does command, almost categorically, the presence of the definite article when the context mandates it.

In fact, as far as the English *-s* is concerned, no such correlation exists. The English plural *-s* is slightly overproportional both for co-occurrence with the definite article, and with its lack (Ø form). Of course, the lack of evidence in the present data for the residual influence hypothesis, as it may be called, does not mean that it should be rejected in all cases. The issue is an empirical one, and within a processing perspective a reasonable hypothesis can be formulated: at a certain point of bilinguality EL knowledge could transfer into ML frames, even when the EL is a single lexical item.[17]

In any case, the overall conclusion points to support of the hypothesis that Ø forms can emerge as a response to processing difficulties. Besides complete integration of lexical material into the ML and complete maintenance of EL structure in constituent insertions, a third broad category of switching phenomenon needs to be recognized, Ø forms, which arise from processing difficulties. Moreover, the processing-based account offers an explanation for a phenomenon which is hard to explain in the borrowing or the congruency-based perspective. This is the situation where there are two comparable constructions, one requiring morphological inflection, the other not. In the current data, the *idafa* possessive requires that a possessive pronoun be suffixed to the noun, whereas the *hana* indirect possessor suffixes the pronoun to *hana*. As seen in Table 3 above, given a choice between the two, it is the *hana* construction which is preferred for English lexical insertions. From a structural perspective, there is no a priori reason for this, as can be seen in the fact that direct suffixation on an English insertion does occur (see (16)). The preference for the indirect *hana* possessor, however, is explained by the general propensity to avoid inflection on English inserted nouns.[18]

As a matter of caution I should add that it is not claimed here that all Ø forms in all CS corpora will be explicable in terms of processing difficulties. It is maintained, however, that this can be an important factor, and is the dominant one in respect of the Maiduguri Arabic CS corpus, very likely in the Dutch (Utrecht)–Arabic CS corpora and other corpora as well.

[17] In fact this hypothesis is compatible with another observation from this data, namely that an English plural is never followed by a suffix pronoun. The presence of *-s* might be interpreted as blocking a further suffix. Two other observations are relevant here, however, which make the interpretation less than clear-cut. First, as seen in Table 3, pronoun suffixes are rare in general on English insertions, and when they occur, they are often attached not directly to the English stem, but rather to an NA plural marker. Secondly, there are also no pronoun suffixes on Ø form plurals. In this instance a global interpretation may be left open.

[18] Moreover, this propensity can probably be related to Muysken's observation (2000, p. 94) that, as he terms them, 'dummy' elements may appear in CS structures whose function it is to bear ML matrix structural elements, allowing EL items to appear bare.

References

Adalar, N. & Tagliomonte, S. (1998). Borrowed nouns; bilingual people: The case of the 'Londrali' in Northern Cyprus. *International Journal of Bilingualism*, 2, 139–159.

Alegre, M. & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, 40, 41–61.

Atawneh, A. (1992). Code-mixing in the speech of Arabic–English bilinguals. In E. Broselow, M. Eid & J. McCarthy (eds.), *Perspectives on Arabic linguistics* IV, pp. 219–244. Amsterdam: Benjamins.

Bentahila, A. & Davies, E. (1983). The syntax of Arabic–French code-switching. *Lingua*, 59, 301–330.

Bentahila, A. & Davies, E. (1998). Code-switching, an unequal partnership. In R. Jacobsen (ed.), *Code-switching worldwide*, pp. 25–49. Berlin: De Gruyter.

Bolanyi, A. (2003). Verbs don't come easy: Bilingual verb formation in language contact and acquisition. Paper presented at 32nd NWAVE conference, October 2003.

Boumans, L. (1998). *The syntax of code-switching: Analysing Moroccan Arabic/Dutch conversation*. Tilburg: Tilburg University Press.

Boumans, L. (1999). Code-switching and the organisation of the mental lexicon. In G. Extra & L. Verhoeven (eds.), *Bilingualism and migration*, pp. 281–299. Berlin: Mouton de Gruyter.

Boumans, L. & Caubet, D. (2000). Modelling intrasentential code-switching: A comparative study of Algerian/French in Algeria and Moroccan/Dutch in the Netherlands. In Owens (ed.), pp. 113–180.

Brahimi, F. & Owens, J. (2000). Language legitimization: Arabic in multiethnic contexts. In Owens (ed.), pp. 405–446.

Bross, M. (2002). State and perspective of languages in Maiduguri. In R. Kawka (ed.), *From Bulamari to Yerwa to metropolitan Maiduguri*, pp. 103–116. Köln: Köppe.

Budzhak-Jones, S. (1998). Against word-internal code-switching: Evidence from Ukranian–English bilingualism. *International Journal of Bilingualism*, 2, 161–182.

Budzhak-Jones, S. & Poplack, S. (1997). Two generations, two strategies: The fate of bare English-origin nouns in Ukranian. *Journal of Sociolinguistics*, 1, 225–258.

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press

Bybee, J. (1995). Diachronic and typological properties of morphology and their implications for represention. In Feldman (ed.), pp. 225–246.

Caramazza, A. & Chialant, D. (1995). Where is morphology and how is it processed: The case of written word recognition. In Feldman (ed.), pp. 55–78.

Caramazza, A., Laudanna, A. & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28, 297–332.

Chebchoub, Z. (1985). A sociolinguistic study of the use of Arabic and French in Algiers. Ph.D. thesis, Edinburgh University.

El Aissati, A. & Boumans, L. (2001). Moroccan languages in a multi-ethnic neighborhood in Utrecht. Paper presented at "Language, Migration and the City" conference, Bayreuth Germany, November 2001.

Eze, E. (1998). Lending credence to a borrowing analysis: Lone English-origin incorporation in Igbo discourse. *International Journal of Bilingualism*, 2, 183–201.

Feldman, L. (1995) (ed.). *Morphological aspects of language processing*. Hillsdale, NJ: Lawrence Erlbaum.

Halmari, H. (1997). *Government and code-switching: Explaining American Finnish*. Amsterdam: Benjamins.

Heath, J. (1989). *From code-switching to borrowing*. London: Kegan Paul.

Laine, M., Vainio, S. & Hyönä, J. (1999). Lexical access routes to nouns in a morphologically rich language. *Journal of Memory and Language*, 40, 109–135.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge: Cambridge University Press.

Levelt, W., Roelofs, A. & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.

Moravscik, E. (1975). Borrowed verbs. *Wiener Linguistische Gazette*, 8, 3–30.

Mustafa, Z. & Al-Khatib, M. (1994). Code-mixing of Arabic and English in teaching science. *World Englishes*, 13, 215–224.

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

Myers-Scotton, C. (1997). *Duelling languages: Grammatical structure in code-switching*. Oxford: Clarendon.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

Myers-Scotton, C., Jake, J. & Okasha, M. (1996). Arabic and constraints on code-switching. In M. Eid & D. Parkinson (eds.), *Perspectives on Arabic linguistics* V, pp. 9–43. Amsterdam: Benjamins.

Nait M'barek, M. & Sankoff, D. (1988). Le discours mixte arabe/français: Emprunts ou alternance de langue? *Canadian Journal of Linguistics*, 33, 143–154.

Nortier, J. (1990). *Dutch and Moroccan Arabic in contact: Code switching among Moroccans in the Netherlands*. Dordrecht: Foris.

Owens, J. (1998). *Neighborhood and ancestry: Variation in the spoken Arabic of Maiduguri, Nigeria*. Amsterdam: Benjamins.

Owens, J. (2000) (ed.). *Arabic as a minority language*. Berlin: De Gruyter.

Owens, J. (2002). Processing the world piece by piece: Iconicity, lexical insertion and possessives in Nigerian Arabic code-switching. *Language Variation and Change*, 17, 173–209.

Poplack, S. (1988). Contrasting patterns of code-switching in two communities. In M. Heller (ed.). *Code-switching: Anthropological and sociolinguistic perspectives*, pp. 215–244. Berlin: Mouton de Gruyter.

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

Poplack, S., Wheeler, S. & Westwood, A. (1990). Distinguishing language contact phenomena: evidence from Finnish–English bilingualism. In R. Jacobsen (ed.). *Code-switching as a worldwide phenomenon*, pp. 187–218. Frankfurt: Lang.

Portin, M. & Laine, M. (2001). Processing cost associated with inflectional morphology in bilingual speakers. *Bilingualism: Language and Cognition*, 4, 55–62.

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

Stemberger, J. (1995). Phonological and lexical constraints on morphological processing. In Feldman (ed.), pp. 247–270.

Received November 18, 2003
Revision received February 2, 2004
Revision accepted March 29, 2004

https://doi.org/10.1017/S1366728904002056 Published online by Cambridge University Press