



Hipertexto 20
Verano 2014
pp. 150-161

**Single Morpheme Tendencies in
Spanish English Codeswitching with
Implications for Further Research on the
Differences Between Nouns and Verbs**

Daniel J. Smith
Clemson University

[Hipertexto](#)

Abstract: Myers-Scotton's 4-Morpheme (4-M) model classifies the morphemes of all human languages into four classes: (1) content morphemes and three types of system or grammatical morphemes, (2) early system morphemes, (3) late system bridge morphemes, and (4) late system outsider morphemes. According to Myers-Scotton's Matrix Language Frame (MLF) model, in a language contact setting, the matrix language provides most of the system morphemes, especially the late ones, but content morphemes may freely come from either language. Bilingual data sets in previous research support these models. This paper focuses on singly occurring morpheme insertions of one language into sentences of the other in a Spanish and English bilingual data corpus collected in northeast Georgia. A focus for further research is the number of noun as compared to verb insertions in this and other bilingual data sets.

Introduction

Single morpheme or word insertions from English in a sentence of Spanish in a bilingual conversation or text is the most frequent type of codeswitching (CS) found in the data of speakers in a study of a Northeast Georgia Hispanic community. Discussions of the data will be in the context of Myers-Scotton's Matrix Language Frame (MLF) model and 4-Morpheme (4-M) model. The MLF model accounts for the grammatical configurations of codeswitching. The 4-M model accounts for the morphemes of a language categorized into four major categories of language production, also affecting the grammatical configurations of codeswitching. The two models together help explain the insertion of single word insertions from one language into sentences of another language.

The 4-M model

The 4-M model, a model of language production proposed by Myers-Scotton and Jake (2000), classifies all morphemes of every natural human language into four categories: (1) content morphemes

three types of system morphemes

(2) early system morphemes

(3) late bridge system morphemes

(4) late outsider system morphemes

As seen in example [1], content morphemes carry the basic content of the sentence (e.g. *dog*, *run*). Content morphemes assign or receive thematic roles. System morphemes, in contrast, neither assign nor receive thematic roles. Verbs, nouns, and adjectives, minus their gender, person, and number affixes, for example, are content morphemes.

[1] Dogs run. John's dog gallops.

Dog	-s	run
content	early	content
John	-’s	dog
content	bridge	content
gallop	-s	
content	late outsider	

An early system morpheme modifies the meaning of a single content morpheme within the immediate maximal projection (immediate phrasal unit) of a content morpheme, for example within a Noun Phrase (NP). Example of this is the plural –s on nouns in English: *dog -s*.

A late bridge system morpheme does not modify the meaning of any single content morpheme but must be there for grammatical correctness and sometimes connects two content morphemes into a larger relationship outside the immediate maximal projection of at least one of the morphemes. Example of this is the possessive –’s in English: *John ’s dog*.

A late outsider system morpheme refers to a content morpheme outside its own immediate maximal projection, not to the morpheme of its own immediate maximal projection. The first person singular –s on verbs refers to the subject content morpheme, not to the verb content morpheme of its own immediate maximal projection but to the noun phrase subject of the verb, which is in a phrase outside the immediate phrase of the verb: *The dog run-s*.

Example [2] illustrates the four different types of morphemes in the 4-M model in a Spanish sentence with a noun, and adjective, and a verb.

[2]

C=Content	E=Early	B=Bridge	LO=Late Outsider
los perros negros corren			
l	o	s	perr
B	E	E	C
refers to no content morpheme	refers to 'perr' in same immediate phrase		refers to 'perr' in same immediate phrase
negr	o	s	corr
C	LO	LO	C
	refers to 'perr' outside immediate phrase of 'negr'		refers to 'corr' within immediate phrase
			refers to 'perr' outside immediate phrase of 'corr'

The MLF model of CS

In Myers-Scotton's (1993a: 485, 486) MLF model, embedded language (EL) islands are embedded into the matrix language (ML). In CS, according to the MLF model, the ML is the language that establishes the morpho-syntax and thus the structure for an utterance. At the same time the ML is also the language of the majority of the morphemes in an utterance of at least two sentences in length (pp. 486-487).¹

Myers-Scotton's (1993b, 1995, 1997) MLF Model is based on two major "hierarchies, "the Matrix Language vs. Embedded Language hierarchy," and "the system vs. content morpheme hierarchy" (1995: 235). The first hierarchy assumes that every utterance has a matrix language grammatical frame into which morphemes may be inserted, whether they derive from one or more than one language. The matrix language frame is from one language or is a composite of more than one language, but there is never more than one matrix language frame for any given utterance. The matrix language frame dictates word order; system morphemes come from the ML.

¹ We extend this two-sentence minimum to entire conversations and even beyond to the entire community speech patterns, taking into account social factors along with morpho-syntactic factors.

The second hierarchy assumes that all morphemes are either content or system morphemes and that content morphemes assign and receive thematic roles² while system morphemes do not. Common category names such as ‘pronouns’ may be content morphemes in one language and system morphemes in another. Content morphemes usually include nouns, verb stems, and descriptive adjectives, while system morphemes usually include inflections and determiners. Single morphemes embedded into the ML will be only content morphemes (see example [3]) because all system morphemes come from the ML.

[3] (English content morpheme EL insertion into a Spanish ML)

*todo, el, el, del paseo de kindergarten estaban apagada las luces,
y...*

‘all, the, the, of the kindergarten hall, the lights were off, and...’

(49:62 unsorted)³

This hierarchy is crucial to the MLF model because system morphemes, which come exclusively from the ML, constitute the essential elements in the ‘frame-building’ apparatus of the ML in the formation of sentences. Content morphemes, whether from the ML or the EL, may subsequently be ‘inserted’ into their ‘pre-framed’ slots for the sentence to take on its speaker’s intended ‘content’ (Myers-Scotton 1995: 235).

The data corpus

The data were collected between the years 1996-2001 from 56 Hispanics in Northeast Georgia, mostly in Habersham County, but also some in neighboring Banks and Stephens Counties. Studies on other aspects of this data appear in Smith (2004a,b, 2006, 2007, 2008, 2009a,b)

Patterns of linguistic CS and convergence in the transcribed data for each subject were distinguished. The following language patterns were found in the data: Spanish with no CS; English with no CS; Spanish with single English lexeme insertions; English with single Spanish lexeme insertions; intersentential English/Spanish CS; intrasentential English/Spanish CS. Each turn in the transcripts was analyzed and categorized. Turns

² ‘Thematic roles’ or ‘theta-roles’ are the roles of people or entities in a sentence. For example, in the sentence, ‘Bob hit Bill’, ‘Bob’ fills the thematic role of ‘agent’, and ‘Bill’ fills the thematic role of ‘patient’ (see Radford 1988: 372-392 for a more complete explanation of ‘thematic roles’ and ‘theta-theory’).

³ For this and following short examples, the first number in () is the identification number of the informant, and the second number is the page number of the transcript (type of transcript indicated by “sorted” and “unsorted”).

of multi-sentence length were divided into sentence-length segments and each segment was analyzed and categorized separately. Turns/segments of Spanish with no CS comprised the vast majority of the data. The second largest type category was only English with no CS.

Spanish with single English lexeme insertions (E>Sinsert) is a turn or segment of only Spanish morphemes and grammatical structure, with the exception of one English word or morpheme inserted in a grammatical slot that would be occupied by a **Spanish word or morpheme in an all Spanish (S)** turn/segment. Likewise, **English with single Spanish lexeme insertions (S>Einsert)** is a turn or segment of all English morphemes and grammatical structure with the exception of one Spanish word or morpheme inserted in a grammatical slot that would be occupied by an English word or morpheme in an **all English (E)** segment. E>Sinsert and S>Einsert are both examples of CS, as opposed to convergence.

The other two CS types are not exact opposites like E>Sinsert and S>Einsert, but are similar because both contain instances of multi-word “islands” (Myers-Scotton, 1993b) of one language either embedded in or alternating with word strings of the other language. **Intersentential English/Spanish CS (SEinter)** is a switch at sentence boundaries. **Intrasentential English/Spanish CS (SEintra)** is an instance of sentence-internal embedding of a multi-word string of one language into the other.

Table 1 summarizes the language patterns with numbers of tokens (instances) of each pattern found in the data.

Table 1 Number of tokens by language type

Language Type	Explanation of Language Type	No. of tokens
S	sentence in all Spanish	2111
E	sentence in all English	623
E>Sinsert	sentence with one word from English inserted into a Spanish sentence	160
SEinter	a codeswitch between languages between sentences in the same turn	66
SEintra	a multi-word insertion of one language into the other language	82
S>Einsert	sentence with one word from Spanish inserted into an English sentence	14

Single morpheme English insertions into Spanish MLs (type E>Sinsert)

Single lexeme English insertions into Spanish (E>Sinsert) are the most prevalent of all the types of mixing/convergence, which is explained by the fact that Spanish is the

predominant ML, with or without EL insertions. According to Myers-Scotton (1993c: 119), the single morpheme insertion CS type should require only a minimal amount of competence in a second language. Based on this claim, this kind of CS is exactly the one that one would expect to be the most prevalent in a speech community in which Spanish is by far the dominant language. The fact that Spanish is the ML for the majority of the data also explains why English insertions into a Spanish ML (E>Sinsert) are much more frequent than Spanish insertions into an English ML (S>Einsert, example [4]). Likewise, it probably explains why non-native Spanish patterns (S>Econv) outrank non-native English patterns (E>Sconv), since English on the whole as an ML is used much less than Spanish.

[4] Example of a single Spanish lexeme inserted into an English ML (S>Einsert)

I don't *entendo*. (42:66 sorted) ('*entendo*' is also an example of non-standard Spanish.)

'I don't understand.'

The fact that intersentential (SEinter) CS is less frequent than both E>Sinsert and intrasentential (SEintra) CS supports the claim by Myers-Scotton (1993b: 71, footnote 6) that intersentential CS is the type that requires the greatest amount of proficiency in both languages. This claim contradicts Poplack (1982), who states that intrasentential CS requires more fluency in both languages than intersentential CS. Poplack's claim is based on her data and those of others that contain higher percentages of intersentential than intrasentential switches (p. 249) as compared to our data.⁴ Moreover, she claims that a higher "presumed degree of bilingual proficiency" is needed to produce intrasentential than intersentential switches (pp. 252-253). Evidence from our data lends more credibility to Myers-Scotton's (1993b) position as well as to Weinreich's (1953: 73) earlier characterization of intersentential CS as being more indicative of the "ideal bilingual." We also submit that E>Sinsert CS requires less skill than SEintra CS because E>Sinsert CS does not require well-formed EL islands. This claim is supported by Myers-Scotton's (1993b: 9-16) data,⁵ in which were found 374 instances of single lexeme insertions ⁶ (corresponding to our E>Sinsert), compared to 121 multi-word

⁴ Differences in language type classifications in this and in Poplack's (1982) study make exact comparisons of numbers of language type tokens difficult. This is an example of a problem often encountered when comparing almost any two CS studies from different researchers. Two researchers will often classify the same instance of CS in different ways. After an attempt was made to 'untangle' these differences by reclassifying Poplack's and our data under a common classification system, our data continue to show larger amounts of intrasentential as compared to intersentential CS than in Poplack's data. Poplack also explicitly states that what she calls "extrasentential" as opposed to intrasentential CS "require[s] less knowledge of two grammars" (p. 249).

⁵ These data are from Myers-Scotton's main Nairobi, Kenya corpus of Swahili/English CS, presented in her *Duelling Languages* (1993b).

⁶ This number would be even higher but for the following: in Myers-Scotton's analysis of her main Nairobi data corpus, any single lexeme insertion occurring in more than two conversations was excluded as a CS

insertions (corresponding to our SEintra) and 44 instances of intersentential switching (corresponding to our SEinter). This claim is additionally supported by our data since E>Sinsert ranks the highest in frequency of all the types of CS/convergence in the data.

Myers-Scotton's MLF model accounts for the insertion of single morpheme/lexeme EL insertions into an ML frame. The E>Sinsert type of CS provides the most rigorous test for the MLF model, and E>Sinsert is exactly the type of CS found most frequently in our data. Now we examine the kinds of lexemes inserted into these Spanish frames in this type of CS found in the data.

Myers-Scotton's MLF model predicts the kinds of EL morphemes that will be inserted into an ML frame. Single morpheme insertions into another language ML (example [5]) are part of what Myers-Scotton calls 'classic CS'.

[5] Example of single morpheme insertion

No sé. No sé. Yo creo que (name) va a ser mi coach. (24:33 sorted)

'I don't know. I don't know. I believe that (name) is going to be my coach.'

Myers-Scotton's (1993b; Myers-Scotton and Jake 2000) MLF Model centers around the distinction between EL and ML as well as the distinction between system and content morphemes. In 'classic CS',⁷ the ML is the frame into which single EL morphemes may be inserted, but only if they are content morphemes.⁸ In example [5] above, 'coach' is a content morpheme because it is not a quantifier and it receives theta role assignment from the copula verb 'va a ser' as equivalent to the subject of the verb '(name)'. As an English content morpheme, it may be freely inserted into the ML frame of the other language, Spanish.

Table 2 shows the lexical categories with number of tokens found in the data for each single English morpheme insertion into a Spanish ML. According to the 1993 MLF model as applied to classic CS, only content morphemes can be inserted into an ML frame. Most of these lexical categories are clearly content morphemes and are

form and classified instead as a BR form. This exclusion, even though not necessary for the claims of the MLF model, was made to allay criticism that the MLF model attempts to account for not only CS but also BR, which some researchers see outside the purview of CS studies (Myers-Scotton 1993b: 15).

⁷ Myers-Scotton and Jake (2000: 2) use the term 'classis CS' to refer to EL morphemes or constituents in ML frames without EL morpho-syntactic influence on the ML morpho-syntax itself, that is, without a composite ML or convergence.

⁸ Content morphemes have [-Quantification] status and have either [+theta role assigner] or [+ theta role receiver] status. All others are system morphemes (Myers-Scotton 1993b).

therefore insertions permitted in the MLF Model. Nouns by far outnumber all other lexical categories among insertions. Other researchers, including Gardner-Chloros (1995: 74), Poplack, Wheeler, and Westwood (1990: 193), Köppe and Meisel (1995: 281-282), and Myers-Scotton (1993b: 15), have found that single noun insertions are the most common single lexeme insertions in their or others' CS data sets. It is well known that, for many languages, nouns are more frequently borrowed than any of the other lexical categories (Finegan 1999: 53). CS and BR are not distinguished in this study because they are viewed as two points on the same continuum (Myers-Scotton 1993b: 15, 182-183). Poplack, Wheeler, and Westwood (1990: 193) also note that CS and BR of nouns is difficult to separate. Therefore, it is not surprising that nouns are the most frequently used single lexemes codeswitched in our data. This observation leads to the consideration of almost all the CS forms as CS instead of BR, if they must be distinguished, because the community under study is recent and Spanish is by far the most frequent language used. Spanish words are readily available for use instead of those English words chosen, and the investigator's queries among members of the community indicate that they did not use the English terms in their country of origin immediately prior to their arrival in the U.S.

The categories that are not clearly accounted for within the MLF Model, according to Myers-Scotton (1993b), are possessive pronouns, interrogative subject pronouns, numbers, definite articles, coordinating conjunctions, subordinating conjunctions, and adverbs. However, in revisions to the MLF model (Myers-Scotton 2002), early system morphemes as well as content morphemes are permitted insertions under the MLF model. Further discussion about these insertions, however, is beyond the scope of this study and thus I leave that to further investigation regarding Spanish English contact and the MLF model.

Table 2 E>Sinsert language type tokens by word class

E>Sinsert	number	of
word class	tokens	
noun	86	
pron.	1	
poss.pron.	1	
inter.sub.pron.	1	
color noun	9	
other adj. noun	1	
verbal noun	2	
verb(pres.part.)	3	
imper.verb	1	
adj.	5	
number	18	
def.art.	1	
coor.conj.	3	
sub.conj.	1	
adv.	3	

adv./interj.	10
interj.	3
vocative	4
locative	3
quote	4
Total	160

Single morpheme Spanish insertions into English MLs (type S>Einsert)

Table 3 shows the lexical categories with number of tokens found in the data for each single Spanish morpheme insertion into an English ML. Similar to type E>Sinsert, no category surpasses nouns in frequency of insertion into English MLs (S>Einsert, example [6]). The total percentage of nouns compared to other lexical categories is not as great as in type E>Sinsert, however.

[6] Example of Spanish noun insertion into an English ML (S>Einsert)

thunder and *truenos* (24:32 sorted)

The smaller frequency of total S>Einsert as compared to total E>Sinsert was discussed earlier, as well as the asymmetry between English and Spanish regarding the insertion of single lexemes (Gal 1979). It was observed that this difference may be related to the general nature of Spanish as the community ML and English as the community EL. The categorization of instances of E>Sinsert and S>Einsert by lexical category shows that nouns are not as frequently inserted as compared to other categories in S>Einsert as in E>Sinsert. The smaller number of total S>Einsert tokens may be a reason for this pattern but, given the propensity for nouns to be much more frequently codeswitched than other categories across data sets, the difference may also be due to the asymmetry between the ML and the EL. In other words, since Spanish and English are asymmetrical in their contact relationship, it should not be surprising that differences would also appear in the lexical categories switched.

Myers-Scotton's MLF Model permits insertion of nouns, adjectives, content verbs,⁹ and prepositions because they are content morphemes. Interjections and tags are also permitted insertions and are also assigned content-morpheme status under the MLF Model. Insertions that are not clearly permitted because of their questionable content morpheme status are copula verbs, numbers, and demonstrative pronouns; again, I will leave this topic to a future study.

⁹ The copula (e.g. 'to be' in English, 'ser' and 'estar' in Spanish) and the 'do' auxiliary (e.g. 'to do' in 'Do you study?') are system morphemes. Most other verbs (not copulas or 'do' auxiliaries) are content morphemes or content verbs, since they, unlike system verbs, assign theta-roles such as agent or patient (Myers-Scotton, 1995: 240).

Table 3 S>Einsert language type tokens by word class

S>Einsert word class	Number of tokens
noun	3
other adj. noun	1
verb(finite)	1
copula verb	2
number(ordinal)	1
dem.pron.	1
prep.	1
interjection/tag	1
interjection	3
Total	14

Insertion of Verbs and Nouns and Implications for Further Research

Even though verbs and nouns are both clearly content morphemes in the 4-M model, and therefore both, as content morphemes, may be inserted into a matrix language frame as ELs, in our data and in the majority of the data (Myers-Scotton and Jake 2013), nouns are inserted much more than verbs. Why is this the case? Myers-Scotton and Jake (2013) give reasons why verbs may not be inserted as much as nouns, then give reasons why there is no obstacle to their insertion, since there are several options that verbs have to be inserted as nonfinite forms instead as finite forms. The extreme poverty of EL verbs of any kind, including nonfinite and 'do' verb constructions, is not explained by Myers-Scotton and Jake's analysis of nonfinite verbs in the literatures. All verbs and verbal forms, including infinitives and participles, are inserted much less frequently than nouns (excluding all noun-like verbals) in speech data in the research literature across language contact situations. Myers-Scotton suggests that social factors and prestige level or prescriptive attitudes about their language might have something to do with it (personal communication 2014). Therefore, the issue has not been resolved linguistically, and opens the way for social or extra-linguistics reasons for explanation as to why. EL verbs, however, are so rare in our data as to prevent any association with social factor analysis as compared to EL nouns.

Regarding Spanish English contact in general, we have encountered prescriptive attitudes frequently. But most of the negative prescriptive attitudes toward mixing are centered around a fear of 'corrupting' Spanish, which is usually the ML, rather than 'corrupting' English, the EL. This is likely due to the social pressure to use English and the fear that Spanish will suffer the most due to contact. These attitudes merit more study with closer investigation of related social factors. It is also likely that nouns have a closer connection to objects and ideas in the EL culture/setting than verbs do, thus increasing the likelihood of EL noun insertion over EL verb insertion.

Conclusion

Singly occurring morphemes are the most common form of CS in the Spanish English

language contact data of this study in northeast Georgia. The same pattern is found in most other CS data sets. This study also details the percentages of each kind of morpheme in this type of CS. No one to date has determined why more nouns are inserted as compared to verbs. Researchers have commented that such is the case but with no empirically based explanation. The question remains as to why more nouns from one language are inserted into sentences of another language as compared to so few verb insertions and thus is the direction for future research.

References

- Finegan, E. (1999). *Language: Its structure and use*, 3rd ed. Fort Worth, Texas: Harcourt Brace.
- Gal, S. (1979). *Language shift: Social determinants of linguistic change in bilingual Austria*. New York: Academic Press.
- Gardner-Chloros, P. (1995). Code-switching in community, regional, and national repertoires: The myth of the discreteness of linguistic systems. In L. Milroy & P. Muysken (Eds.), *One speaker, two languages: Cross-disciplinary perspectives on codeswitching* (pp. 68-89). Cambridge: Cambridge University Press.
- Köppe, R., & J. M. Meisel (1995). Code-switching in bilingual first language acquisition. In L. Milroy & P. Muysken (Eds.), *One speaker, two languages: Cross-disciplinary perspectives on code-switching* (pp. 276-301). Cambridge: Cambridge University Press.
- Myers-Scotton, C. (1993a). Common and uncommon ground: Social and structural factors in codeswitching. *Language in Society*, 22, 475-503.
- Myers-Scotton, C. (1993b). *Duelling Languages: Grammatical structure in codeswitching*. Oxford: Clarendon Press.
- Myers-Scotton, C. (1993c). *Social motivations for codeswitching: Evidence from Africa*. Oxford: Clarendon Press.
- Myers-Scotton, C. (1995). A lexically-based model of codeswitching. In L. Milroy & P. Muysken (Eds.), *One speaker, two languages: Cross-disciplinary perspectives on codeswitching* (pp. 233-256). Cambridge: Cambridge University Press.
- Myers-Scotton, C. (1997). New afterward to update *Duelling Languages*. Unpublished manuscript, University of South Carolina, Columbia.
- Myers-Scotton, Carol. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

- Myers-Scotton, C. & J. L. Jake. (2000). Testing the 4-M model: An introduction. *International Journal of Bilingualism* 4(1): 1-8.
- Myers-Scotton, Carol and Janice L. Jake. (2013). Nonfinite verbs and negotiating bilingualism in codeswitching: Implications for a language production model. *Bilingualism: Language and Cognition*. Cambridge: Cambridge University Press.
- Poplack, S. (1982). Sometimes I'll start a sentence in Spanish *y termino en español*: Toward a typology of code-switching. *Spanish in the United States: Sociolinguistic aspects*. Eds. J. Amastae & L. Elías-Olivares. Cambridge: Cambridge University Press. 230-263.
- Poplack, S., S. Wheeler, & A. Westwood (1990). Distinguishing language contact phenomena: Evidence from Finnish-English bilingualism. In R. Jacobson (Ed.), *Codeswitching as a worldwide phenomenon*. Ed. R. Jacobson. New York: Peter Lang. 185-218.
- Radford, A. (1988). *Transformational grammar: A first course*. Cambridge: Cambridge University Press.
- Smith, Daniel J. (2004a). Intra-speaker variation of standard and nonstandard grammatical patterns: Spanish influenced by English in Georgia. *Southwest Journal of Linguistics* 23(2): 163-185.
- Smith, Daniel J. (2004b). Social and grammatical patterns in Spanish/English bilingualism in Georgia, USA. *Southern Journal of Linguistics* 28(1/2): 1-15.
- Smith, Daniel J. (2006). Thresholds leading to shift: Spanish/English codeswitching and convergence in Georgia, USA. *The International Journal of Bilingualism* 10(2): 207-240.
- Smith, Daniel J. (2007). Spanish/English bilingual children in the Southeastern USA: Convergence and codeswitching. *The Bilingual Review* 23(3): 99-108.
- Smith, Daniel J. (2008). Family, friends, and language preference influence on Spanish and English use in the U.S. *Southern Journal of Linguistics* 32(2): 26-77.
- Smith, Daniel J. (2009a). Asymmetrical mixing patterns of Spanish and English in a new immigrant community. *Bilingual Research Journal*, 32(2): 188-206.
- Smith, Daniel J. (2009b). Spanish and English gender/age variation in Georgia. *Hipertexto, Online Journal* v.10(summer), 90-111. Retrieved from <http://www.utpa.edu/dept/modlang/hipertexto/>.
- Weinreich, U. (1953). *Languages in contact: Findings and problems*, (1968 ed.). The Hague: Mouton.