A comparative analysis of Romanian-English and Romanian-Spanish code-switching patterns[*]

Ariana Bancu
University of Michigan
abancu@umich.edu

The present study compares Romanian-Spanish and Romanian-English code-switch patterns at a morpho-syntactic level. Spanish and Romanian have similar grammatical features and share many structural and phonological properties, while Romanian and English differ in most of these aspects. In comparing 240 code-switches from both language pairs, I address whether these differences and similarities in code-switch patterns can be explained by typological factors. The results show that switch points in both language pairs are the same at phrase level. However, different CS patterns occur at morpheme-level, with Romanian-Spanish switches showing more complex combinations between Spanish stems and Romanian inflectional or function morphemes, than in the Romanian-English data. These findings shed light on the role of feature matching between switched categories and the effect of typological similarity.

1. Introduction

Code-switching, a type of language mixing during language contact, has been the focus of extensive research over the past six decades; the phenomenon is defined by the alternate use of more than one language in a conversation by the same individuals (Weinreich 1968), or more specifically "…the alternation of two languages within a single discourse, sentence or constituent" (Poplack & Sankoff 1981:214). Much has been written about code-switching from a sociolinguistic perspective, which has resulted in a closer understanding of why bilinguals code-switch, the circumstances that influence their language use, and what attitudes bilinguals and others have towards code-switching (Myers-Scotton 1988, Romaine 1995). At the same time, there are also questions about the structure of code-switching. This paper will consider the morpho-syntactic properties of intrasentential switches. The structural analysis of mixed utterances has resulted in various attempts to deliver a universal account for the rules that govern code-switching by proposing structural and phonotactic constraints (Poplack 1980, Woolford 1983, DiSciullo et al. 1986). Evidence from various language pairs studied in the 70s and 80s

---

[*] An earlier version of this paper was presented at ILLS 5 (2013), University of Illinois at Urbana-Champaign.

has raised much debate around the most influential constraint theories, and with greater the typological differences between the analyzed languages involved in code-switching, it became clearer that numerous combinations between code-switched constituents were possible. Mahootian (1993) was among the first to show that the same phrase structure principles used in a monolingual context can be used for the code-switching data. Further, Muysken (2000) proposes that code-switching should be associated with the degree of typological similarities between languages, in addition to the sociolinguistic factors that influence it.

The purpose of the present study is to analyze Romanian-Spanish and Romanian-English code-switch patterns at a morpho-syntactic level. I compare the data from the two language pairs to determine the differences and similarities between the code-switch patterns. I then ask whether these differences and similarities in code-switch patterns can be explained solely by typological factors. While the two language pairs do show certain similarities, the differences in switching at the morphological level indicate an effect of typological similarity.

2. Previous research on code-switching

Analyses of the interference of two language systems at syntax level have resulted in various attempts to deliver a universal account for the structure of code-switched utterances. Yet evidence from various code-switching language pairs has contradicted these proposed constraints (Mahootian 1993, Muysken 2000, Chan 2009).

Myers-Scotton (1993) claims that the patterns that arise in code-switching may be shaped by the level of activation of each of the languages the bilingual uses. The language that is activated at a higher level dominates and serves as a base for the code-switched utterances. In her Matrix Language Frame model (MLF), which distinguishes between a Matrix Language (ML), or base-language, and an Embedded Language (EL), restrictions apply to function morphemes, such as definite determiners, such that they usually have to be supplied by the Matrix Language. Additionally, Bentahila and Davies (1983), working with a corpus of naturally occurring Moroccan Arabic/French data, concluded that "switching is freely permitted across all boundaries above that of the word" (p. 329), suggesting an important distinction between syntactic and morphological switches. Woolford (1983) likewise proposed a model which prohibits morphologically mixed lexical items and predicts that terminal nodes of phrase structures can only be filled with lexical items from the grammar of the language that generated them.

However, examples from typologically different languages showing noun-adjective switches (Mahootian, 1996) have disproved this constraint, and from this opposing perspective, Mahootian (1993, 1996) was the first to propose a Null Theory for code-switching showing that constraints imposed on CS cannot be claimed to be universal. While Mahootian acknowledges the influence socio- and psycholinguistic factors have on code-switching, she argues that "an empirically adequate syntactic account of code-switching is a necessary component of a full account of code-switching behavior" (1996:378) just as it would be the case for a monolingual context. Clearly, with these conflicting viewpoints in previous research, the question of what exactly constrains code-switching is worth further investigation.

Comparisons made between language pairs that share a base language (Clyne, 1987; Kawangamalu, 1997) have primarily intended to provide evidence against proposed universal constraints on code-switching. These studies show that the grammatical rules that operate in CS are specific to the involved languages and use the base language to demonstrate that. However, these studies do not point to morpho-syntactic differences in CS patterns as a result of the typological differences between the languages involved in code-switching.

For this reason, the current study seeks to compare syntactic and morphological switches in two language pairs that differ in typological similarity. The Romanian-English and Romanian-Spanish bilingual communities have received little attention in recent linguistic studies. Furthermore, I know of no study comparing Romanian-English and Romanian-Spanish code-switch patterns at a structural level.

3. Data sources and methodology

The data used for the present study come from various sources. First, I extracted data from a few previous studies analyzing the language use of Romanian-English (Beligan, 1999; Dumitrescu, 2004; Ene, 2007). Second, I recorded and transcribed various conversations between myself and four other first-generation Romanian-English bilinguals who live in the US.[1] They come from different socioeconomic backgrounds and range in age from 26 to 32 years old. They are native speakers of Romanian and have learned English as a second language. All four participants started using English on a regular basis as adults, past what is considered to be the end of the Critical Period for language acquisition (cf. Pinker, 2004) .

---

[1] Unless otherwise indicated, Romanian-English data cited in the paper comes from this source.

The Romanian-Spanish data are from two sources. About 40% of the Romanian-Spanish CS tokens come from Munteanu's (2007) study on the way Romanians in Spain speak. He analyzed the language use of Romanians that have immigrated to several areas in Spain between 1980 and 1995 and presented an inventory of Spanish lexemes or expressions that are used frequently during conversations in Romanian. About 60% of the Romanian-Spanish data comes from a study analyzing the linguistic interferences between Romanian and Spanish in Spain, conducted by Jieanu (2011). The data for her study were provided by 43 Romanian immigrants in Spain from various age groups, with different educational backgrounds. The participants are from two different communities, one in Castellon and the other one in Madrid, both cities registering considerable numbers of Romanian inhabitants.

I extracted all the intrasentential CS utterances for each language pair from the conversation transcripts and the studies mentioned above[2]. The CS tokens were grouped according to types of switched constituents, e.g. determiners and nouns, nouns and adjectives, verbs and complements etc., to obtain a quantitative overview. Then I proceeded in testing the validity of the encountered switches.

An asymmetry between the involved languages can be observed, Romanian being the dominant language in both code-switch patterns. Such an occurrence is typical for the code-switches of first generation bilinguals, such as the ones providing the examples for this analysis, who choose their first language as the base language in a conversation (Grosjean, 2010). Following Muysken (1995), we can classify Romanian as the BL in both language pairs, the base language in code-switching being the language providing the highest number of morphemes. Spanish and English are classified as embedded languages (EL).

4. Results

Switches from the BL to the EL happen at various points in the clause (e.g. between determiners and nouns, prepositions and nouns, verbs and nouns, nouns and adjectives, etc.). Table 1 below delivers an overview of the encountered switches within different phrasal categories for each language pair for the 240 code-switches analyzed in this study. The total number of tokens was counted for each type of switch. The percentage shown next to each phrase represents its ratio in the overall data for each language pair, with Romanian as the base language.

---

[2] For the purpose of consistency, only intrasentential CS examples were included in this analysis. Intersentential CS takes place at clause boundary and therefore does not provide a basis for a morpho-syntactic comparison.

Phrasal Category	Rom-Sp	Rom-Eng
Total Nº tokens	92	148
DP	44%	45%
VP	29%	20%
PP	19%	25%
AdjP	6%	8%
AdvP	2%	2%

Table 1. Overview of switched phrasal categories

4.1. Similarities in the two language pairs

Overall, we can see that the distribution of code-switches is similar in the two languages when we consider the distribution at the level of syntactic phrases. Switches within DPs occupy the highest rank in the table and are almost the same percentage in each data set. Interestingly, some of the Romanian determiners (e.g. definite, possessive, demonstrative) take a complement to their left, while others (e.g. indefinite, demonstrative) take a complement to their right. Both, Romanian indefinite (examples 1a and 1b) and definite determiners (examples 2a and 2b) can be identified paired up with EL nouns in the data[3]:

(1a) a.m   văzut o   chic.a   (2a) mi-   a   făcut presupost-ul
     have.1S seen DetFS girl.FS   Refl.1S has done estimate-DetMS
     'I saw a girl'   'He gave me the estimate'
          Rom-Sp (Jieanu, 2011:203)   *Rom-Sp* (Munteanu 2007:7)

(1b) a.m   un   roommate   (2b) ei   prim-eau în spaniolă ticket-ul
     have.1S DetMS roommate   they got- 3P in Spanish ticket-DetMS
     'I have a roommate'   'They were getting the ticket in Spanish'
               Rom-Eng   *Rom-Eng*

In switches between a determiner and a noun, determiners were always from the BL and nouns from the EL. Even though EL determiners could take a BL noun complement, no such code-switches occurred in my data or were reported by Munteanu (2007) or Jieanu (2011).

Regardless of the EL, constituents were distributed in the clause following BL rules. Pro-drop was a typical occurrence in code-mixed utterances and where subjects were present, whether they were from an EL or the BL, they either preceded or followed the verb.

---

[3] In code-switching examples, the data in Romanian appears in regular font, while English and Spanish words appear in Italics. Additionally, the data is shown in its original spelling as it was found in the respective sources.

Switches between a BL preposition and an EL NP occur in both language pairs, such as in (3) and (4) below:

(3) m   -am   cazat   la *youth hostel*
   Refl.1S have.1S checked-in to youth hostel
    "I checked-in at a youth hostel."   *Rom-Eng*

(4) m   -am   dus la *cole*
   Refl.1S have.1S went to high-school
   'I went to high-school.'   *Rom-Sp* (Jieanu, 2011:18)

The EL noun phrases in (3) and (4) following the BL preposition '*la*' are missing determiners. The BL phrase structure rules dominate in both of these cases, such that determiners, which are not required by the BL PP structure, are omitted.

All prepositions in PPs are from the BL. Since all three languages are strictly prepositional, all prepositions take a complement to their right, so a switch between an EL preposition and a BL DP should be possible, even though it was not encountered.

Switches within AdjPs and AdvP take place at a low rate. When an EL AdjP is paired with BL noun, the AdjP follows the same syntactic patterns in both language pair, branching to the left or the right of the noun, according to BL rules. Their position is determined and restricted solely by the particular category requirements of BL adjectives, regardless of the language they are from. For example, in (5) we can see the English adjective *underprivileged* following the Romanian noun *copii*.

(5) nişte club-uri pentru copi-i   *underprivileged*
   some club-FP for   child-MP underprivileged
   'some clubs for underprivileged children'   *Rom-Eng*

In this case, the switched AdjP follows BL rules, which require adjectives derived from participles to follow the noun. Furthermore, Romanian nouns and adjectives agree for number and gender, but the lack of such features for English adjectives does not prohibit switching between a Romanian noun and an English adjective.

Further, when comparing data from code-switches between Romanian-English and Romanian-Spanish, we see certain specific structural similarities, such as in (6) and (7). Both English and Spanish nouns appear with Romanian gender and number inflections and enclitic determiners, indicated in bold font in the examples below:

177

(6) (…) a.i   de învăţat cu   *finals*- **uri- le**
  (…) have.2S to learn   with finals- FP-DetFP
  'You have to study with the finals (coming up).'   *Rom-Eng*

(7) mâine   încep *rebahas*-**uri- le**
  tomorrow   start.3P sales-   FP-DetFP
  'The sales start tomorrow.'   *Rom-Sp* (Munteanu 2007:8)

Both tokens reveal an instance of a double-plural, visible in the morpheme-by-morpheme gloss. The data illustrate that Romanian rules are being applied to English and Spanish nouns alike, resulting in a similar combination. There are variable degrees of typological similarity in the two language pairs: Spanish and Romanian have similar grammatical gender features, inflectional verb morphology, and share many phonological properties, while Romanian and English differ in most of these aspects. Despite these differences, many of the patterns that occur in the Romanian-Spanish data are also present in the Romanian-English data. Even more, the same Romanian inflectional morphemes attached to English or Spanish noun stems occur in Romanian-Spanish and Romanian-English mixed sentences, whether they can fully be integrated phonologically to the stems or not.

In summary, there are many similarities between the switches in the two language pairs, and the distribution by syntactic form is about the same in both pairs. There are, however, differences to be found in the data.

4.2. Differences in the two language pairs

Where we find differences in the data is in the verbal morphology. Romanian verbs inflect for person, number, and tense, and pro-drop is common in Romanian. Both sets of data show EL verbs with BL inflections. Furthermore, such tokens show pro-drop and other occurrences typical for BL clausal architecture.

Table 2 delivers an overview of encountered BL inflections occurring attached to EL verb roots, with an example for each encountered BL inflection shown in **bold**. The X indicates the absence in the data of a combination for the respective language pair.

Pers.	Nº	Rom-Sp[4]	Rom-Eng
1st		*engord*-**esc** (*engordar* +**esc**)	*gradu*-**ez** (*graduate*+**ez**)
2nd	**Sg**	*guard*-**ezi** (*guardar*+**ezi**)	X
3rd		*mehor*-**eaza** (*mehorar* + **eaza**)	*choke*-**ăneşte**[5] (*choke*+**ăneşte**)
1st		*actu*-**ăm** (*actuar* + **ăm**)	*hang*-uim (*hang* + **uim**)
2nd	**Pl**	*divert*-**iţi** (*divertir*+**iţi**)	X
3rd		X	X

Table 2. EL verb roots with BL inflections

The majority of the EL verbs occur with BL auxiliaries, so they often keep their infinitive or participle form. However, differences between language pairs can be observed in such cases as well. Examples (8) and (9) show a similar switch between a BL auxiliary and an EL participle verb. The difference can be observed in the EL verb morphology. The Spanish verb in (9), *acercar* – 'to approach', carries a Romanian participle ending '-*at*'. The English participle verb, *linked,* on the other hand does not carry any Romanian inflectional morphology.

(8) e *linked* la cont-ul   de Amazon
    is linked to account-DetMSg of Amazon
    'It is linked to the Amazon account.'   *Rom-Eng*

(9) A   fost *aserc-*   at
    has been approach-ed
    'He has been approached'   *Rom-Sp* (Jieanu, 2008:13)

Most English verbs in the Romanian-English data appear either in the subjunctive mood[6] or with the auxiliaries 'to be' and 'to have'. (10) shows a sequence of Romanian subjunctives followed by an English infinitive, where Romanian verbs inflect and the English verb remains unchanged:

(10) trebe să te   apuc-i   să scr- i,   să cit-eşti, să *summarize*
    must to Refl.2S begin-2S to write-2S to read-2S to summarize
    'You have to start to write, to read, to summarize'   *Rom-Eng*

The Romanian-Spanish data, on the other hand, show more complex combinations between the Romanian subjunctive *să* followed by a Spanish verb stem carrying a Romanian verb inflection: să *envas*-ăm 'to bottle' (*envasar* + 1sP), să *disfrut*-e 'to enjoy' (*disfrutar* + 3P), să *guard*-ezi 'to save' (*guardar* + 2S).

---

[4] Examples taken from Munteanu (2007) and Jieanu (2008).   [5] (Ene, 2007, p. 53).
[6] Soare (2005) argues that the subjunctive particle *să* and infinitive marker *a* behave in a similar way with regards to the verb and should be treated as mood marking particles.

EL verbs are more frequent in Romanian-Spanish CS data, than in the Romanian-English. There are two possible explanations why Spanish verbs show more variation in Romanian-Spanish code-switching, than in Romanian-English code-switching. First, there is a higher congruence between Romanian and Spanish verbs, both inflecting for person, number, and tense. Second, Romanian infinitive verbs always end in a vowel, *-a*, *-e* or *-i* and Spanish infinitive verb endings, e.g. *-ar*, *-er*, *-ir*, also contain the same vowels (*a, e* or *i*). These similarities between Romanian and Spanish verb endings facilitate the selection and attachment of Romanian inflections. The Spanish infinitive ending *-ar* or *-er* is dropped facilitating the attachment of the Romanian participle ending *-at* or *-it*.

Another difference can be seen in gender agreement in switches between nouns and determiners, or nouns and adjectives. Such agreement is often accompanied by Romanian gender marking morphology in the Romanian-Spanish data, but not in the Romanian-English data. For example, there is only one occurrence of a Romanian feminine determiner paired with an English noun in the Romanian-English data, shown in (11). In Romanian-Spanish on the other hand, we encounter various combinations between definite and indefinite Romanian determines and Spanish nouns.

(11) *advisor*-a   me-a mi -a   propus   că …
      advisor-DetFS my-FS me has proposed that …
'My advisor proposed to me that …'   *Rom-Eng*

Since English nouns do not carry a gender feature, such pairings may rely on phonological harmonization between determiner and noun in Romanian-English CS. The most encountered Romanian determiners paired with English nouns are masculine determiners, which do not require any additional inflectional morphology on the noun or stress shift. Feminine determiners, on the other hand, require the agreement with a feminine ending, which is most often a stressed vowel. When agreement is difficult to realize, prepositions are often used as a strategy to incorporate EL nouns without using BL pro-/ enclitics or additional inflections, such as we have seen in (3) above.

It appears that in Romanian-English noun-adjective switches each category maintains its language-specific features. The [-number] and [-gender] features of English adjectives do not hinder switching. In Romanian-Spanish noun-adjective switches on the other hand, number and gender agreement is present, marked predominantly with BL inflections. For example, in (12) we see the Spanish adjective *agradable* – 'pleasant' with the Romanian feminine inflection '-.ă':

(12) am   avut o   discuți.e *agrdabil*-ă
     have.1S had DetFS talk.FS pleasant-FS
     'We had a pleasant talk.'   *Rom-Sp* (Jieanu, 2011:17)

The Romanian-English data does not contain any English nouns carrying Romanian gender marking inflections.

5. Conclusions

At phrase level, each analyzed phrase maintains its BL syntactic properties. When a whole sequence of EL phrases is switched, it follows EL rules, as shown in the case of English adjectives always preceding English nouns. The typological differences between the languages involved do not seem to constrain intersentential CS, even in cases where switched elements do not share the same feature set. Every switch that occurs in one language pair is also present in the other, thus leaving no avenues open for imposing any syntactic constraints in this case.

However, the typological similarities between Romanian and Spanish lead to more harmonious combinations at morpheme level. Spanish nouns or adjectives easily adopt Romanian inflectional and derivational morphology, often with changes in the root form of the lexemes. Both Romanian and Spanish have gender and number features and they are similar to a large extent. Thus, BL determiners, and gender and number markers that are present on Spanish nouns and adjectives in the Romanian-Spanish data are more diverse than the ones in the Romanian-English data. Furthermore, Spanish verbs occur with a wider range of Romanian person, number and tense inflections than English verbs, even in constructions involving BL auxiliaries or subjunctives, where English verbs appear in their participle or infinitive forms. This is again the result of similarities between Romanian and Spanish verb features. In the case of nominal agreement, a similar pattern is encountered, where structural similarity between Spanish and Romanian facilitates integrating in code-switching in contrast to English where distinct phonological systems make such morphological switches difficult.

For the present analysis I have relied on CS data involving two language pairs that share one base language and on known typological and syntactic properties of the languages involved. While the syntactic CS patterns are very similar in both language pairs, differences can be found at a morphological level. In terms of diversity, there are less Romanian determiners, verb infections, gender and number inflectional morphemes that combine with English stems, than with their Spanish counter parts.

The more similarities exist between the languages being switched the more complex the code-switching patterns. The genetically related systems of Romanian and Spanish facilitate the combinations of morphemes from the two languages in code-switching.

REFERENCES

Beligan, A. 1999. *Code-switching patterns in the speech of romanian-english bilinguals*. Monash University Linguistics Papers. 2(1). 3-9.

Bentahila, A., & Davies, E. (1983). *The syntax of Arabic-French code-switching*. Lingua 59(4). 301-330.

Chan, BH. 2009. *Code-switching between typologically distinct languages*. In B.E. Bullock & A.J. Toribio (ed.), The Cambridge handbook of linguistic code-switching (182-198). Cambridge: Cambridge University Press.

Clyne, M. 1987. *Constraints on code-switching – how universal are they?* Linguistics 25. 739-764.

DiSciullo, A., Muysken, P., & Singh, R. 1986. *Government and code-mixing*. Journal of Linguistics 22(1). 1-24.

Dumitrescu, D. 1993. *A preliminary approach to the contact phenomena found in the Romanian spoken by Romanian-Americans of the first generation.* Journal of the American Romanian Academy of Arts and Sciences. 18. 161 – 186.

Ene, E. 2007. *Romanian-English code-switching: A preliminary study*. Arizona working papers in Slat. 8. 45-55.

Grosjean, F. 2010. *Bilingual life and reality*. Cambridge, Mass.: Harvard University Press.

Jieanu, I. 2011. *Interferenţe lingvistice Româno-Spaniole* (Doctoral dissertation). Constanţa, Romania: Universitatea "Ovidius" Constanţa

Kamwangamalu, N.,M. 1997. *Language contact, code-switching, and I-languages. Evidence from Africa*. South African Journal of Linguistics. 15(2). 45-51.

Mahootian, S. 1993. *A null theory of code-switching* (Doctoral dissertation). Evanston, IL: Northwestern University

Mahootian, S. 1996. *Code-switching and universal constraints: Evidence from Farsi/English.* World Englishes. Vol. 15. 377-384. Malden, MA: Blackwell Pub.

Munteanu, D. 2007. *Româna vorbită în Spania poate deveni o nouă modalitate lingvistică?* Proccedings of the Conference: Diaspora în Cercetarea Ştiinţifică Românească. Bucharest, Romania.

Muysken, P. 2007. *Code-switching and grammatical theory*. In L. Wei, (ed.). The bilingualism reader. 124-138. London: Routledge. (Original work published 1995)

Muysken, P. 2000. *Bilingual speech: A typology of code-mixing.* England: Cambridge U Press.

Myers-Scotton, C. 1993. *Dueling Languages - Grammatical Structure in Code – Switching*, Oxford, England: Oxford U Press..

Myers-Scotton, Carol. 2007. *Code-switching as indexical of social negotiations*. In L. Wei (ed.). The bilingualism reader. London: Routledge. (Original work published 1988)

Pinker, S. 1994. *The Language Instinct*. New York: Morrow.

Poplack, S. 2007. *Sometimes I'll start a sentence in Spanish Y termino en espanol: Toward a typology of code-switching*. In L. Wei (ed.). The bilingualism reader. 213-243. London: Routledge. (Original work published 1980).

Romaine, S. 1995. *Bilingualism* (2nd Ed.) Oxford, England: Blackwell.

Sankoff, D., & Poplack, S. 1981. *A formal grammar for code-switching*. Papers in Linguistics. 14(1-4)/ 3-45.

Soare, G. 2005. *Romanian Syntax for FiPS* (Technical report). Laboratoire d'Analyse et de Traitement du Langage, University of Geneva.

Weinreich, U. 1968. *Languages in Contact*. London, The Hague, Paris: Mouton & Co.

Woolford, E. 1983. *Bilingual Code-Switching and Syntactic Theory*. Linguistic Inquiry, Vol. 14, 520-536. Cambridge: MIT