Ashlyn Winship
LING 6477
Prof. Helena Aparicio
13 Dec. 2023

Effect of DP Structure on Gender Agreement in Spanish-Englished Codeswitched DPs (Experimental Methods in Language Sciences: Final Project)

1 Introduction

While codeswitching has several different interpretations and definitions depending on the framework in which it is being approached, this paper will specifically be concerned with the use of two languages, Spanish and English, within DPs.

In the Matrix Language Frame (MLF) model of codeswitching, the *matrix language*, or main language of the sentence, determines the syntactic structure of the switched phrase, and this can impact word order within the switch (Myers-Scotton 2004:110). While this paper will not focus on exploring the MLF or other syntactic models of codeswitching, the term *matrix language* will be useful for describing the structure of switches. In the example below, the matrix language is English.

(1) Hand me the *libro*.

Valdés Kroff (2016) describes an "overwhelming preference for masculine-marked mixed NPs" in the Bangor Miami corpus (2016:290). Bangor Miami is a Spanish-English bilingual speech database that includes both audio recordings and transcripts, collected from 2008 to 2011 in Miami, Florida (Bangor Miami). Valdés Kroff searched the corpus for "mixed NPs": NPs consisting minimally of a determiner and a noun, each in a different language, and any of the noun's modifiers if present (2016:289). The modifiers could be in either the language of the determiner or the language of the noun. If the determiner was in Spanish but did not inflect for gender, the token was discarded, as were bare nouns, and mixed NPs in which the noun was a phonologically adapted loanword, like *la breca* ('the brake') (Valdés Kroff 2016:288-9). This left the following possible constructions for mixed NPs:

(2)

(a) Sp. Det + Eng. N *el* [(red) dress]
(ex. from VK 2016:288)

(b) Eng. Det + Sp. N some [*agüitas*]
(ex. from Bangor Miami, reported in VK 2016:289)

Valdés Kroff reports the distribution of the 316 mixed NPs he found in the corpus as follows:

(3)

(a) Sp. MASC Det + Eng. NP	93.7%
(b) Sp. FEM Det + Eng. NP	2.5%
(c) Eng. Det + Sp. NP	3.8%

(VK 2016:289)

Cruz (2021) performed a similar investigation into the CESA (Corpus del Español en el Sur de Arizona) database and found similar results that suggest the presence of the MASC default in switches involving English nouns with feminine Spanish translations (2021:31).

Castaño-Nuñez (2019) also investigated the Bangor Miami corpus for similar NPs (or DPs). Her investigation looked more closely at the internal structure of these DPs. She looked at 27 of the 56 transcripts in the database, and found only 68 switches within DPs, to Valdés Kroff's 316 (CN 2019:31). Additionally, because she was not specifically investigating gender like Valdés Kroff was, her criteria for switched DPs did not exclude Spanish bare nouns. Therefore, her tokens included DPs like the following, where Valdés Kroff's did not:

(4) *Mi carro tiene* heated seats

(Bangor Miami, as reported in CN 2019:34)

Castaño-Nuñez also observed switched DPs where the adjective was post-nominal, as in (4) ((4)a was also observed and discussed by Valdés Kroff).

(5)

(a) *la* cheerleader *pesada* Eng. N, post-N Sp. Adj
(b) *un micrófono* built-in Sp. N, post-N Eng. Adj

(Bangor Miami, as reported in CN 2019:34)

Castaño-Nuñez writes that these types of switches, between N and Adj, "are a rarity," but that "the switch between determiners and nouns, however, is one of the most productive one (sic)" (2019:34). These productive switches between Det and N include those like (1)b, which Castaño-Nuñez also reported, as well as those in (5).

(6)

(a) *el* food festival
(b) *un* roommate
(c) I did all my *trabajo*

(Bangor Miami, as reported in CN 2019:33-34)

Because Castaño-Nuñez's investigation only looked at about half of the Bangor Miami corpus, it is possible that switches of this type occur more frequently in the corpus than she reports. However, her claim is supported by data from Poplack (1980), which cites switches that occur at an adjective as occurring in only 0.8% of switches (1980:602). Poplack's data consisted of interviews and casual speech recordings collected from residents of the Barrio, a Puerto Rican neighborhood in New York City (Poplack 1980:590, 595).

Castaño-Nuñez also conducted a follow-up experiment that consisted of two tasks. The first provided support for the dispreference of N-Adj switches. It was an acceptability judgment task that asked participants to judge different types of switches on a scale of 1-7, including those between adjective and noun (CN 2019:47). Castaño-Nuñez tracked which participants were heritage language learners (HLLs) and which were non-HLL native Spanish speakers

(SPs), and reported that the average acceptability of N-Adj switches was 2.09 for HLLs, and 2.03 for SPs (2019:49).

The second task was a forced choice task that tested the order and language of the adjective and noun in a DP-internal switch (CN 2019:47). An example question is provided below, where participants would see either the English or Spanish matrix language sentence as the prompt, and choose the best of the 4 options:

(7) *Yo vivo en la* ___ / I live in the ___
  (a) big *casa*
  (b) *casa* big
  (c) *grande* house
  (d) house *grande*

(CN 2019:47)

The results showed that participants displayed a preference for the word order of the language of the sentence, regardless of the language of the adjective and determiner (CN 2019:57). Thus, participants would select the post-nominal order if the sentence was in Spanish, and the pre-nominal order if the sentence was presented in English, as in (7) (CN 2019:57).

(8)
  (a) *Puse los libros en la* table *robusta*. Spanish S: D N Adj
  (b) I put the books on the sturdy *mesa*. English S: D Adj N

(CN 2019:59)

The relevant conclusion from Castaño-Nuñez's study is that switches between adjectives and nouns are infrequent in natural language, and that when people are forced to choose switches of this type, they adhere to the word order of the matrix language.

This paper investigates the effect of the structure of the mixed NP on selection of the MASC default observed by Valdés Kroff. In Castaño-Nuñez's experiment, all of the Spanish modifiers that inflect for gender were correctly modified according to the Spanish translation of the noun (see (8)a above: An option for this sentence was not "table *robusto*"). This paper focuses on sentences in which the matrix language is Spanish, and in which the (first) switch occurs after the determiner, which has been shown by both Valdés Kroff and Castaño-Nuñez to be an acceptable location for a switch. The noun is in English, and the adjective (if present) varies in both language and position pre- and post-nominally, similar to options (c) and (d) in the example question (7) above. The results of this pilot study do not show a significant difference in answer selection depending on either presence of the adjective, language of the adjective, or position of the adjective. However, the DPs in which the adjective is in Spanish and pre-nominal (e.g. [*la roja* t-shirt]) do show a higher rate of FEM responses for nouns with feminine Spanish translations than the other DP structures. A repetition of this experiment with a larger sample size may yield a significant difference for this condition, which would have interesting implications for current accounts of agreement in codeswitching.

2 Experiment

This experiment was previously run in a slightly different format, with 5 participants. This data was discarded after the experiment was reformatted following a discussion with bilingual speakers, who pointed out that there was a flaw in the design of the stimuli. The matrix language of the test items was originally English, with the switch first appearing with the Spanish-language determiner, then the English noun. The bilingual speakers pointed out that, as discussed in the previous section, switches are unlikely to happen at the determiner, and it is more likely that the determiner will be in the matrix language. For that reason, I changed the matrix language of the test stimuli to Spanish, so that the determiner that the participant chose was in the same language (refer to the next section). The experiment was then redistributed to new participants. Everything else about the experiment design remained the same as it is described in the following section.

2.1 Methods

2.1.1 Design and procedure

The experimental stimuli were designed to test 4 different types of DP-internal switch. The types are as follows:

Type	Structure	Sample experimental item
A	Determiner only; No Adj present	Pásame el/la t-shirt
B	English Adj present	Pásame el/la red t-shirt
C	Spanish Adj present, pre-nominal	Pásame el/la rojo/roja t-shirt
D	Spanish Adj present, post-nominal	Pásame el/la t-shirt rojo/roja

The experiment was a forced-choice task between the two gender agreement options, masculine vs. feminine, and it was administered through PC Ibex (demo)[1]. In sentences of type A and B, participants would select the determiner they thought was most appropriate (*el* or *la*). In sentences of type C and D, participants would select both the determiner and adjective they deemed most appropriate. The determiner and adjective were presented together as one choice, both displaying either masculine or feminine agreement. Examples of what a participant would see are shown below.

[1] As of 12/13/2023, the demo of this experiment does not function properly. Following the PC Ibex downgrade, the experiment is throwing an error when it tries to run. The JavaScript console in my browser does not recognize any of the commands in the experiment. The main.js file that I submit on GitHub will contain all of the code, which worked properly when I was gathering data from the 15 participants.

Type A:
Pásame ___ t-shirt

1. el
2. la

Type B:
Pásame ___ red t-shirt

1. el
2. la

Type C:
Pásame ___ t-shirt

1. el rojo
2. la roja

Type D:
Pásame ___ t-shirt ___

1. el, rojo
2. la, roja

The experimental items consisted of 128 sentences: 32 different English nouns (such as t-shirt, above) were used to construct 4 sentences of the types A-D above, and the experimental items were organized in a Latin square design. All of the English nouns used were inanimate nouns whose Spanish translation is grammatically feminine (e.g. t-shirt: *la camiseta*), to test whether the MASC default was surfacing. Each participant encountered 32 experimental items and 32 filler items. The experimental items were randomized for each participant, but each received the same filler items.

The 32 fillers were split into 4 types:

- Number agreement in Spanish
  - Ex. Ella tiene ojos ___ (azul, azules)
  - Correct answer: azules (agreeing in number)
- Switch with English matrix sentence
  - Ex. I want to read the ___ interesante (book, libro)
  - No “correct” answer; participants may strongly prefer Spanish N
  - Switches occur more frequently at the object NP level (the *libro interesante*) than at the Adj (the book *interesante*) (Poplack 1980:602)
- Switch at infinitive VP
  - Ex. Ella va a ___ ingredients at the store (buscar, look for)
  - No “correct” answer; switch should be allowed at either point (i.e. ‘*va a buscar* ingredients’ or ‘*va a* look for ingredients’) (cf. Poplack 1980)
- Switch at *estar* + progressive V
  - Ex. Mi amigo ___ singing a song tomorrow (está, is)
  - No “correct” answer; participants may prefer *estar* over *be*
  - *estar* + V_{Prog} switch has been observed in literature (cf. Balam et al. 2020)

Before the experiment began, participants answered background questions and training questions. The background questions asked the participant’s age and whether they identify as bilingual, and also asked the participant to describe the dialects of Spanish and English they speak, the age at which they learned each language, and the contexts in which they use each language in their daily lives (at home, at school, etc.). Each participant saw 4 training questions

designed to introduce them to the forced choice task and to seeing sentences presented with words in both Spanish and English. One training question included the comma-separated answers of the kind seen in test sentences of Type D, with instructions explaining that the first word fills the first blank and the second word fills the second blank, so participants were familiar with this format when the experiment began.

2.1.2 Participants

15 subjects participated in the experiment. For 2 participants, the experiment did not actually present them with the experimental trials, so they only completed the training questions. One of these participants also selected “no” for the background question “Do you identify as a Spanish-English bilingual speaker?”, so their data would have been omitted regardless. The other 13 participants selected “yes” for this question and verified that they were over the age of 18, so for this analysis, n=13.

The subjects were between the ages of 19-25. They reported mainly speaking Mexican Spanish, with 1 speaker reporting Venezuelan, 1 Castilian, and 1 español bogotano (from the capital of Colombia). They all reported speaking American English, with a few specifying the region (Western, Californian, etc.). The majority described learning Spanish as their L1 and learning English at a young age, from school and/or media. 2 of the speakers described themselves as late L2 Spanish learners (both over the age of 11). The majority of the speakers reported using Spanish at home and with family, and English at school and with friends.

2.2 Predictions

The comparisons between sentence type where differences in the rate of the participants’ FEM selection are expected are as follows:

- Presence of Adj: A ↔ B & A ↔ C — A has no Adj, B has Eng. Adj, C has Sp. Adj
- Adj language: B ↔ C — B has Eng. Adj, C has Sp. Adj
- Adj position: C ↔ D — C has pre-N Sp. Adj, D has post-N Sp. Adj

In consultation with a Spanish speaker, sentences of type A and B are predicted to show predominantly MASC answers:

A: Pásame el t-shirt
B: Pásame el red t-shirt

Additionally, selection of FEM answers is predicted to increase for types C and D, with type D showing the highest rate of FEM selection:

C: Pásame la roja t-shirt
D: Pásame la t-shirt roja

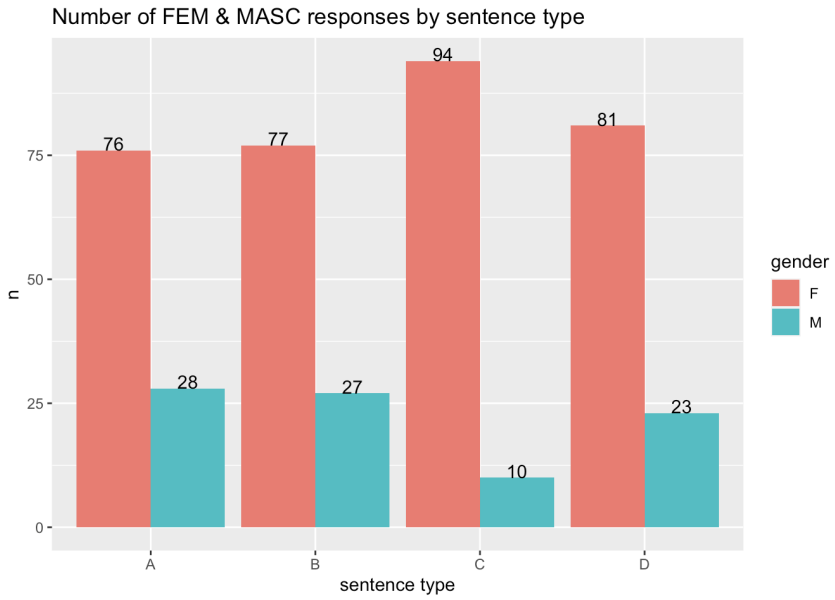
The assumption is that in C and D, the fact that there are now two modifiers that inflect for gender (the adjective in addition to the determiner) will trigger the FEM gender agreement over the MASC default. This will be especially true for D, in which the order of the DP is Spanish word order (Det N Adj), and Spanish is the matrix language of the sentence. This structure aligns with the results of Castaño-Nuñez’s experiment, in D would be the preferred word order for a Spanish sentence. If Spanish word order is preferred, Spanish agreement may be as well, so it

is predicted that participants will be especially likely to choose the FEM option, which agrees with the Spanish translation of the noun.

2.3 Results

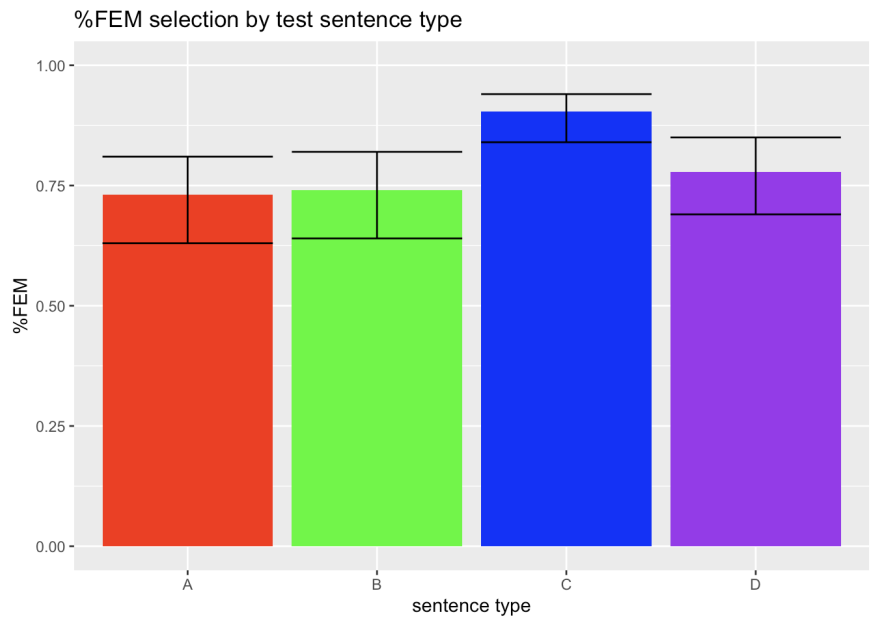
Participants saw 32 experimental items (8 from each sentence type A-D). In total, 416 data points were collected (104 for each sentence type).
The graph below shows the counts for selection of MASC vs FEM response by sentence type.

Figure 1



The following graph shows the mean response selection by sentence type (coded as MASC=0, FEM=1). Error bars represent adjusted binomial 95% CIs.

Figure 2



At first glance, these graphs support our predictions to some extent. Type C (and, marginally, D) show more FEM responses than A and B, which aligns with our predictions. However, type C shows the highest rate of FEM selection, and it was predicted that type D would show the highest rate.

A repeated-measures one-way ANOVA was performed to compare the effect of sentence type (A-D) on mean response choice (coded as MASC=0, FEM=1). The one-way ANOVA revealed that there was not a statistically significant difference in response choice between at least two groups ($F(3, 36) = 1.861$, $p = 0.154$).

2.4 Discussion

The data initially supported the predictions that sentence types C and D would produce more FEM responses than A and B. However, these differences were not statistically significant. Additionally, the relative rates of FEM selection of C and D were the opposite of those that were predicted, with C showing the highest overall rate of FEM selection. This does not support the hypothesis that the Spanish word order of the DP, which should be the favored order because of the Spanish matrix language of the sentence, would trigger feminine Spanish gender agreement.

One interesting point is that overall, the response selection for sentences of any type tends toward FEM over MASC, which conflicts with the MASC default theory. Liceras et al. (2008) provide data that is relevant to this observation. They ran an experiment in which participants provided acceptability judgments on mixed DPs consisting of just a determiner and a noun; half of the experimental items were structured as Eng. Det + Sp. N, and the other half were Sp. Det + Eng. N (2008:842). There were 3 groups of participants: 2 groups were non-native Spanish speakers (one group was English L1; one group was French L1, English L2; both were learning Spanish as adults), and 1 group was Spanish L1 speakers learning English as adults (2008:841). The results showed that for the mixed DPs, “masculine is the preferred option for

non-native speakers," while "the native Spanish speakers prefer to somehow 'force' agreement [...] upon the English N" (2008:844). The experiment in this paper involved participants who were mainly child simultaneous bilinguals or early bilinguals. Liceras et al. did not test child/early bilinguals in their agreement experiment (they did, as it turns out, test child and adult simultaneous bilinguals in a separate experiment that did not involve gender agreement), but the results in this paper suggest that they would prefer to assign Spanish gender to the English nouns as well. If this is the case, then the MASC default theory should be reexamined with closer attention paid to speakers' individual bilingual profiles.

3 Conclusion

Although the results of this experiment did not show any significant differences in FEM selection between sentence type, a larger sample size may yield a significant difference between sentences of type C and the other types. The current trends of the data, with pre-nominal Spanish adjectives showing the most FEM responses and all sentence types showing more FEM responses than MASC responses, have interesting implications for current theories of agreement in codeswitching if they are indeed significant with a larger sample size. If the overall greater rate of FEM response than MASC response bears out in future investigation, this poses an issue for the idea of MASC default. It may be that the MASC default is produced in actual speech, but when bilinguals are confronted with a choice, they select the gender of the Spanish noun. In fact, one of the participants suggested as much in the "Feedback" section of the survey, writing: "The responses between 1 or 2 varied depending on whether I approached it as something I would say that other 'spanglish' speakers would understand, and then secondly what made sense grammatically as feminine or masculine."

Additionally, given the insights from Liceras et al. (2008), a future research direction with this experimental design could be to administer the survey to different populations of bilingual speakers (e.g. simultaneous bilinguals, Spanish L1 speakers, English L1 speakers, etc.), to investigate whether FEM selection is more frequent for one population than the other.

4 References

Balam, O., Couto, M. & Stadthagen-González, H. (2020). Bilingual verbs in three Spanish/English codeswitching communities. *International Journal of Bilingualism 24*(5-6). DOI: https://doi.org/10.1177/1367006920911449

Castaño-Nuñez, Y. (2019). DP internal codeswitching by Spanish and English bilinguals in the US. *Graduate Theses, Dissertations, and Problem Reports.* https://researchrepository.wvu.edu/etd/3891

Cruz, A. (2021). A syntactic approach to gender assignment in Spanish-English bilingual speech. *Glossa: A journal of general linguistics 6*(1). DOI: https://doi.org/10.16995/glossa.5878

Deuchar, M. *BilingBank Spanish-English Bangor Miami Corpus*. Retrieved from https://biling.talkbank.org/access/Bangor/Miami.html

Liceras, J.M. et al. (2008). Gender and gender agreement in bilingual native and non-native grammars: A view from child and adult functional-lexical mixings. *Lingua 118*, 827-851. DOI: http://dx.doi.org/10.1016/j.lingua.2007.05.006

Myers-Scotton, C. (2004). Precision tuning of the Matrix Language Frame (MLF) model of codeswitching. *Sociolinguistica 18*(1), 106-117. DOI: https://doi.org/10.1515/9783484604650.106

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching. *Linguistics 18*(7-8), 581-618. DOI: http://dx.doi.org/10.1515/ling.1980.18.7-8.581

Valdés Kroff, J. R. (2016). Mixed NPs in Spanish-English bilingual speech: Using a corpus-based approach to inform models of sentence processing. In Guzzardo Tamargo, R. E., Mazak, C. M., & Carmen Parafita Couto, M. (Eds.), *Issues in Spanish and Lusophone linguistics: Spanish-English codeswitching in the Caribbean and the US* (Vol. 11). John Benjamins Publishing Company. DOI: 10.1075/ihll.11