Testing alternative theoretical accounts of code-switching: Insights from comparative judgments of adjective–noun order

Hans Stadthagen-González, M Carmen Parafita Couto, C Alejandro Párraga, and Markus F Damian

Abstract

Objectives:

Spanish and English contrast in adjective–noun word order: for example, brown dress (English) vs. vestido marrón (‘dress brown’, Spanish). According to the Matrix Language model (MLF) word order in code-switched sentences must be compatible with the word order of the matrix language, but working within the minimalist program (MP), Cantone and MacSwan arrived at the descriptive generalization that the position of the noun phrase relative to the adjective is determined by the adjective’s language. Our aim is to evaluate the predictions derived from these two models regarding adjective–noun order in Spanish–English code-switched sentences.

Methodology:

We contrasted the predictions from both models regarding the acceptability of code-switched sentences with different adjective–noun orders that were compatible with the MP, the MLF, both, or none. Acceptability was assessed in Experiment 1 with a 5-point Likert and in Experiment 2 with a 2-Alternative Forced Choice (2AFC) task.

Data and analysis:

Data from both experiments were subjected to linear mixed model analyses. Results from the 2AFC task were also analyzed using Thurstone’s law of comparative judgment.

Conclusions:

We found an additive effect in which both the language of the verb and the language of the adjective determine word order.

Originality:

Both experiments examine adjective–noun word order in English–Spanish code-switched sentences. Experiment 2 represents a novel application of Thurstone’s law of comparative judgements to the study of linguistic acceptability which yielded clearer results than Likert scales. We found convincing evidence that neither the MLF nor the MP can fully account for the acceptability of adjective–noun switches.

Implications:

We suggest that advances in our understanding of grammaticality in code-switching will be achieved by combining the insights of the two frameworks instead of considering them in isolation, or by espousing a probabilistic model of code-switching.

Testing alternative theoretical accounts of code-switching: Insights from comparative judgments

Code-switching is the phenomenon by which bilinguals go back and forth between their two languages in the same conversation (see Deuchar, 2012). It is generally accepted that switches between languages do not occur at random, but follow specific patterns and rules (Bullock & Toribio, 2009; Guzzardo Tamargo, Mazak & Parafita Couto, 2016, i.a.), and the typifying of those rules represents a rich and active field of research in the bilingualism literature. However, scholars do not agree on the best theoretical account of these regularities. Recently, an interest has emerged in evaluating the predictions of theoretical models to try to disentangle between theoretical predictions (see Gullberg & Parafita Couto, 2016; Herring, Deuchar, Parafita Couto & Moro Quintanilla, 2010; Parafita Couto, Deuchar & Fusser, 2015; Fairchild & Van Hell, 2015; vanden Wyngaerd, 2016, or Eppler, Luescher & Deuchar, 2016). These researchers have shown that different theories account for some aspects of the observed data, but there is no overarching theory that can explain all the code-switching patterns. In what follows we will experimentally examine the accuracy of predictions regarding adjective-noun order derived from an approach within the Minimalist Program (Cantone & MacSwan 2009) and the Matrix Language Framework (Myers-Scotton, 2002). Of particular interest are the so-called conflict sites (Poplack & Meechan, 1998), instances where the grammars of the two languages differ. For example, the default order for noun-adjective constructions in English is adjective-noun as in 'red book', whereas the default order in Spanish is the opposite (libro rojo –literally 'book red'). In Spanish, some adjectives may appear before or after the noun with a change in meaning. Post-nominal adjectives have a basic, attributive meaning (1a), while prenominal adjectives have a more restricted, intrinsic meaning (1b).

(1) a. un hombre grande b. un gran hombre

'a tall man' 'a great man'

However, in this paper, we will focus on post-nominal adjectives.

In the context of Spanish-English code-switching, we may consider whether libro red, red libro, rojo book, and book rojo are equally probable. It is unclear, how a conflict is resolved (i.e., what constrains word order) when a Spanish noun is combined with an English adjective, or vice versa. Some comparative analyses addressing the question of adjective-noun order in code-switched noun phrases have been done couched within the two theoretical models (MP and MLF) but with different outcomes (cf.; Parafita Couto, Deuchar & Fusser, 2015; vanden Wyngaerd, 2016).

Of the two models in question, Myers-Scotton's (2002) MLF differentiates the languages involved: one language is known as the matrix language (ML), the other as the embedded language (EL). In a code-switched clause, the ML is assumed to provide the morphosyntactic frame and allows predictions regarding acceptable constructions. On the other hand, generativist MacSwan (2005a, 2005b) argues that all instances of code-switching may be accounted for based on general mechanisms of grammar (for further details see MacSwan, 2005a, 2005b; Jake, Myers-Scotton & Gross, 2005).

The MLF predicts that (i) 'late outsider morphemes' such as finite verb morphology marking subject-verb agreement and (ii) word order within a clause that contains code-switching will be sourced from the ML. Previous studies of Welsh-English code-switching show that Welsh-English bilinguals tend to produce clauses with code-switching in which the language of the finite verb morphology matches clause word order (Davies, 2010; Davies & Deuchar, 2010; Deuchar & Davies, 2009; Deuchar, 2006; Parafita Couto, Deuchar & Fusser, 2015). This leads

us to predict that if the bound morphology of the finite verb is from language A, adjective and noun word order in a code-switched noun phrase should also be dictated by language A.

Cantone and MacSwan (2009) in their analysis of noun-adjective order in Italian-German code-switching follow Cinque's (1994) proposal that a universal base underlies adjective position, with adjectives assumed to universally precede the noun. On this view, differences in word order between English and Spanish follow from overt movement of the noun in Spanish to a position to the left of the adjective, resulting in the contrasting surface word order outlined above, with adjectives preceding the noun in English and following it in Spanish. Cantone and MacSwan (2009) find that for Italian-German code-switching, "the language of the adjective determines the position of the NP relative to the adjective" (p. 268). We can test this descriptive generalization, which in the context of Spanish-English code-switching, would mean that English adjectives should precede Spanish nouns whereas Spanish adjectives should follow English nouns.

There have been attempts to evaluate these two approaches at different grammatical switching points, but they have been conducted in different language pairs and using different methodologies. Hence, it is perhaps not surprising that they have also yielded conflicting results. For example, Herring, et al. (2010) examined the predictions of both theories regarding determiner-noun switches from naturalistic Spanish-English and Welsh-English code-switching data. They do not find statistically significant differences between the accuracy of the predictions of the two theories. Fairchild and Van Hell (2015) experimentally examined (through a series of picture naming tasks) the ability of the MP and the MLF to explain determiner-noun switches in Spanish-English bilinguals, finding no support for either theory. Focusing specifically on adjective-noun order in code-switching, Parafita Couto et al. (2015) designed a study to evaluate

the predictions of these two models within Welsh-English mixed nominal constructions by using a multi-task approach comprising (1) naturalistic corpus data, (2) an elicitation task, and (3) an auditory judgment task. They found that Likert-style judgment tasks are not very useful in this community of code-switchers, due to the stigmatized nature of the phenomenon. However, the data from the naturalistic corpus and the elicitation task supported the relative superiority of the MLF model. Nevertheless, it was only a small proportion of their data that could distinguish between the two models, so no definite conclusion could be reached. On the other hand, vanden Wyngaerd (2016), basing her analysis on elicited data and judgment tasks, examined word order in French-Dutch mixed nominal constructions, finding support for Cantone and MacSwan's generalization.

As described above, the formulation of these theoretical models, particularly the MLF, has mainly relied on naturalistic corpora analysis. Multiple studies (Herring et al. 2010, Parafita Couto, Deuchar & Fusser, 2015; Parafita Couto & Gullberg, 2015; Eppler et al., 2016) have cited corpus data with no definitive conclusion with regards to word order within code-switching. This situation highlights a limitation of corpus data: despite its descriptive richness and ecological validity, it is not probative in nature. Other studies have used acceptability judgments (see Schütze, 1996), where informants provide a yes or no answer as to whether they accept or reject a given sentence as "correct" or "acceptable", or rate how well the sentence "sounds" on a given scale (e.g.: "On an ascending scale of 1 to 7, how acceptable do you find the following sentence?"). As pointed out by Cowart (1996), this technique is very similar to introspective judgments that are routinely studied in the field of psychophysics and signal detection theory to determine the limits of our sensory system (e.g.: how dim a light our eyes can sense). While acceptability judgments afford more control, and, potentially more probative value than corpus

data, they are vulnerable to different types of response biases that may obscure their results (see Parafita Couto et al. 2015 for a discussion of how the stigma associated with code-switching may affect grammaticality judgments).

From the signal detection theory literature (e.g.: Gescheider, 1997; Green & Swets, 1988) we can learn about some of the vulnerabilities of yes/no and Likert-style judgments. Yes/no judgments can suffer from criterion effects, where factors external to the variable under study may affect the willingness of the informant to report a "yes" or a "no" answer (for a more detailed discussion of detection thresholds, see Gescheider, 1997). Scaled responses present a different set of challenges, particularly regarding consistency of use of the scale. In order to position all items along the same scale, informants must not only "calibrate" their ratings along the entire range, but also keep a memory record of the ratings on earlier stimuli in order to avoid possible shifts in the internal rating scale when new items are presented (Parraga, 2015). Both yes/no and Likert-style acceptability judgments have attracted criticism because of lack inter-rater reliability (Labov, 1972, 1975; Ross, 1979; Stokes, 1974; Bader & Häussler, 2010) and stability within the same informants at different times or under different testing conditions (e.g: Carol, Bever & Pollack, 1981; Nagata, 1988; Snow & Meijer, 1977).

From very early on, the psychophysics literature (e.g.: Fechner, 1876) identified and addressed these problems concerning introspective judgments by using paired comparison scales, also known as the 2-Alternative Forced Choice (2AFC) method. This is still now preferred for quantifying detecting fine-grained differences between value judgments that can only be made based on subjective criteria (David, 1988; Parraga, 2015).

The Two-Alternative Forced Choice Task and the Law of Comparative Judgment.

In a 2AFC task, participants are presented with pairs of stimuli and must choose which item is "better" according to a specified criterion (without the possibility of a tie); in our case, this means deciding which of the sentences in a pair is more acceptable or "sounds better". Besides greatly reducing the response bias effects mentioned above (Green & Swets, 1966), comparative judgments are a more natural task for participants than rating scales. Nunnally (1976, p. 40) states that "People simply are not accustomed to making absolute judgments in daily life, since most judgments are inherently comparative [...] people are notoriously inaccurate when judging the absolute magnitude of stimuli." Paired comparisons also greatly reduce the demands on memory required for avoiding shifts in the internal rating scale (Parraga, 2015).

Gustav Fechner (1876), the pioneer of psychophysics, proposed that the proportion of times an items is chosen over another in a series of pairwise comparisons provides a measure of the distance between the two items in some pleasantness continuum. For example, if item A is chosen over item B half the time, both objects are equally pleasant. Correspondingly, if object C is consistently chosen over item A, then by the same measure, object C is likely to be the most pleasant item of the three. Thurstone (1927), with his law of comparative judgment, generalized this concept into a measurement model which converts simple pairwise comparisons between stimuli into one-dimensional quality scores (for further explanation see Bock and Jones, 1968 and Torgerson, 1958). Thurstone proposed that the proportion of times a stimulus is judged greater than another is related to the number of psychological scale units separating the two sensations. In this method, which is the standard scaling method in many disciplines today, pairwise judgments are performed many times by a comparatively large number of subjects, resulting in statistically robust results. The outcome of the analysis is a ranking of preference for

stimuli along an interval scale that reflects the relative distance between conditions. Moreover, the results from this analysis can be tested for significance using standard statistical methods such as standard errors and ANOVA. Despite the multiple benefits associated with the 2AFC method attested by its widespread use in psychophysics and other areas of research, its use for linguistic acceptability judgments has been conspicuously absent so far.

In the present study we used both a traditional Likert scale task and a novel application of the 2AFC task and Thurstone's (1927) Law of Comparative Judgments to test the predictions on adjective-noun order derived from the the MLF and the MP. We directly contrasted sentences that reflected (or not) the structures derived from such predictions. This contrast generated four conditions: sentences that followed the CS pattern predicted by either the MP (but not the MLF), or the MLF (but not the MP), both, or neither. Table 1 presents all possible combinations of noun- and adjective order and language (for each direction of switch), as well as the predictions from each model as to whether each combination should be acceptable to bilinguals or not. We asked participants to engage in pairwise comparisons of acceptability in all possible combinations of those conditions and the results were analyzed by applying Thurstone's law of comparative judgment case V to generate an interval scale based on comparisons of pairs of code-switched sentences that indicates a ranking of acceptability but also the relative distance between conditions.

[Table 1 about here]

If code-switching is indeed constrained by the grammatical properties proposed by one model or the other, the presence of such property should predict the acceptability of a given sentence. For example, if the matrix language determines word order within the noun phrase (as

proposed by the MLF model), sentences 1, 3, 5, and 7 should show higher acceptability than the others; if, on the other hand, word order is defined by the language of the adjective itself (as according to Cantone and MacSwan´s generalization), sentences 2, 4, 6, and 8 would have higher scores. Furthermore, if code-switching acceptability is entirely predicted by one of the models, there should be no difference between the preferred construction (whichever one may be) and those sentences "acceptable" according to both models (exemplars 1 and 5); conversely, the least favored construction according to that model should not show an advantage over sentences rejected by both models (exemplars 4 and 8).

Experiment 1

Participants

A total of 40 early English/Spanish bilinguals took part in this experiment. Most were born in the United States (N=34) or moved there before the age of 5 (N=6; mean age when they migrated =3:6 years). There were 24 men and 16 women and their mean age was 30:3 years (S.D.: 9.6; Range: 18-56). All stated that they spoke the Mexican variety of Spanish. The highest level of education for 22 participants was high school, 17 had attended at least some college, and 1 attended graduate school.

Participants were recruited through Amazon Mechanical Turk, an online crowdsourcing website that has been shown to be a good source of participants for collection of acceptability judgments (Gibson, Piantadosi & Fedorenko, 2011). Participants were paid a small fee for completing the study and only workers with an acceptance rate of 90% or above and at least 100

tasks completed were allowed to take part in the study (following the guidelines proposed by Peer, Vosgerau & Acquisti, 2014).

Language Proficiency and Dominance. In order to take part in the acceptability judgment task, each participant completed an English and a Spanish test to confirm their proficiency in both languages. These tests were an adaptation from the Online Placement Tests used by Oxford University's Language Center (Oxford University Language Center, n.d.). The tests were modified to reflect Latin American (rather than Iberian) verb conjugations and vocabulary (e.g.: "ustedes" instead of "vosotros" for the second person plural pronoun), and geographical landmarks to reflect U.S. or Latin American locations (e.g.: New York instead of London). Only participants that attained a score of 34 (out of 50) or more were allowed to continue with the study. This range of scores is classified as "Higher proficiency" by the Oxford website.

Participant's average scores were about the same in both languages (English: M =44.9; S.D. =2.8; Spanish: M =43.9; S.D. =3.9; $t(39) = 2.02$, $p = .2$). All participants stated that they were able to speak Spanish before age four and English by the time they entered elementary school. They declared to be slightly more proficient in English than Spanish, with ratings of 4.00 (S.D.: 0.00) vs. 3.80 (S.D.: 0.46) respectively, in a scale of 1 to 4, where 4 indicates that they are "Confident in extended conversations".

Materials

Critical trials: We compiled 5 translation-equivalent "base sentences" in Spanish and English (please see the Appendix for a list of base sentences). We created two sets of code-switched sentences, one in which the first switch went from English to Spanish (E.g.: I like the MARRÓN

dress; I like the dress MARRÓN) and another one in which it went from Spanish to English (E.g.: ME GUSTA EL VESTIDO brown; ME GUSTA EL brown VESTIDO). Each base sentence was modified into code-switched forms following four switching patterns going from Spanish to English and four going from English to Spanish as follows:

MP-/MLF+: Sentences that follow the predictions of the Matrix Language Frame but not the Minimalist Program (E.g.: I like the dress MARRÓN; ME GUSTA EL brown VESTIDO)

MP+/MLF-: Sentences that follow the predictions of the Minimalist Program but not the Matrix Language Frame (E.g.: I like the MARRÓN dress; ME GUSTA EL VESTIDO brown)

MP-/MLF-: Sentences that do not follow the predicted pattern from either theory (e.g.: I like the VESTIDO brown; ME GUSTA EL MARRÓN dress)

MP+/MLF+: Sentences that follow the predictions of both models (E.g.: I like the brown VESTIDO; ME GUSTA EL dress MARRÓN)

In Spanish, all nouns were masculine or invariable (e.g.: vestido – 'dress', baño – 'bath', cuaderno – 'notebook', cantante – 'singer', abogado – 'attorney') and all adjectives were invariable with respect to gender (marron – 'brown', caliente – 'hot', verde – 'green', nicaragüense – 'Nicaraguan', independiente – 'independent'). Proper names in the sentences were chosen so that they were commonplace in both Spanish and English (e.g.: Max, Claudia). We avoided using nouns and adjectives whose onset would elicit changes in the preceding

indefinite article depending on their position in some conditions (that is, avoid changes such as "Hugo is a singer estadounidense" vs. "Hugo is an estadounidense singer").

Filler trials: In each session we included 72 non-critical trials where the focus of the CS was not the adjective but the determiner or the adverb. Some of the results for those trials will be reported elsewhere. By including these filler trials plus the quality control trials described below, critical trials made up only about a third of all pairs seen by participants. This was done to make it harder for raters to engage in strategic choices for their response (Cowart, 1996).

Quality control trials: There were 8 quality-control trials that consisted of sentences with inter-sentential code-switches. Half of the sentences had an uncontroversial error that could be easily detected if the sentences were read carefully (e.g.: LA PASÉ MUY BIEN, the music *were excellent). These errors were equally distributed among the following factors: first vs. second half of the sentence, English vs. Spanish portion, and type of error (verb tense, number agreement, gender agreement, and word order). If a participant failed more than 2 of these trials, they were removed from the sample and substituted with a new participant.

Procedure

The survey was administered online using Qualtrics and testing occurred across 3 separate sessions: one in which participants completed the language tests and the background questionnaire, and two counterbalanced sessions in which participants were presented with NPs in which the switch went from English to Spanish and from Spanish to English, respectively. Test sessions were about a week apart from each other. Participants were given the choice of

reading the instructions in English or Spanish; all but 4 participants chose to complete the questionnaire in English. Participants were informed that they would see a series of sentences and that they were to indicate on a 5-point scale how "permitted" a sentence was according to the way they would speak to- or hear from another bilingual person. In the scale, a score of 1 stood for "never permitted" while 5 stood for "always permitted". Participants were then presented with the 76 code-switched sentences as described above. Each sentence was presented one at a time and the order of presentation was individually randomized for each participant. Participants had to rate each item before progressing to the next one and could not go back to previous sentences.

Results

Table 2 shows the mean acceptability ratings for each condition (on a scale from 1 to 5). For sentences where the switch went from English to Spanish, Mauchly's test indicated that the assumption of sphericity had been violated, $\chi 2$ (5) = 17.91, p= .003, therefore degrees of freedom were corrected using the Huynh-Feldt estimates of sphericity (ε = .880). The results show that there was a significant effect of CS pattern, F(2.40, 93.57) = 23.52; p < .001. Post-hoc pairwise comparisons revealed that there was no significant difference between CS patterns MP+/MLF- and MP-/MLF+ (p = 1.0), while all other differences between CS patterns were significant (all p values ≤ .02). For sentences where the switch went from Spanish to English, Mauchly's test also indicated that the assumption of sphericity had been violated, $\chi 2$ (5) = 17.72, p= .003, therefore degrees of freedom were corrected using the Huynh-Feldt estimates of sphericity (ε = .877). The results show that there was a significant effect of CS pattern, F(2.63, 102.57) = 11.05; p < .001. Post-hoc pairwise comparisons revealed that the only significant differences were between CS

pattern MP+/MLF+ and each of the other CS patterns (all p values ≤ .01), while all other contrasts were not significant (all p values ≥ .11).

[Table 2 about here]

In summary, MP+/MLF+ switches are the most preferred in both directions, while MP-/MLF- switches are the least preferred in the Spanish to English direction (with a non-significant trend in the same direction for English to Spanish). There were no significant differences between MP+/MLF- and MP-/MLF+ sentences. The results from Experiment 1 show evidence of a hierarchy of preference for certain CS patterns, particularly in the English to Spanish direction, but do not differentiate between the two models of CS under evaluation. The average acceptability scores were all close to each other, slightly on the positive side of the middle of the scale.

In Experiment 2 we repeated the acceptability judgments for the same code-switched sentences with a different group of participants, but this time using a two-alternative forced choice presentation of the stimulus and using Thurstone's Law of Comparative Judgment for the analysis of the results to find out if this alternative method provides a clearer picture of this phenomenon. Given the suitability of this method for measuring subtle differences in introspective judgments, we expected to obtain a more fine-grained picture of the contrasts under study.

Experiment 2

A total of 40 early English/Spanish bilinguals living in the United States took part in this experiment, none of which participated in Experiment 1. Thirty-five participants were born in the United States, while the rest immigrated before age 5 (mean age when they migrated =3:7 years). There were 22 men and 18 women and their mean age was 29:5 years (S.D.: 8.8; Range: 18-57). All stated that they spoke the Mexican variety of Spanish. The highest level of education for 21 participants was high school, 17 had attended at least some college, and 2 attended graduate school. Participants were also recruited through Amazon Mechanical Turk and completed the same questionnaire and language tests as those in Experiment 1.

Language Proficiency and Dominance. Participant's average scores in the language proficiency tests were slightly but significantly better for English (M =45.2; S.D. =2.8) than for Spanish (M =43.7; S.D. =3.8); $t(78) = 2.01$, $p < .05$. All participants stated that they were able to speak Spanish before age four and English by the time they entered elementary school. They declared to be slightly more proficient in English than Spanish, with ratings of 3.98 (S.D.: 0.16) vs. 3.68 (S.D.: 0.47) respectively.

Materials

Critical trials: Participants were presented with the same critical sentences as Experiment 1, but they were compared pairwise among themselves to yield 30 judgments for each direction of the switch (English to Spanish and Spanish to English). Sentences were contrasted only within their own directionality set in two separate sessions, that is, code-switched sentences going from English to Spanish were only contrasted with other sentences also going from English to Spanish and so on. This was done in order to control for possible preferences in the directionality of the

switch (Parafita Couto et al. 2015) and insulate the effect of noun-adjective order, which is the focus of the present study. This was also a pragmatic decision because the combinatorial outcome of comparing all possible pairs of the resulting 8 variations for each sentence made the experiment unfeasible (it would have generated 28 combinations per sentence instead of 6).

Monolingual trials: We presented a separate group of 40 early Spanish/English bilinguals with monolingual versions (either English or Spanish) of the base sentences used in this experiment. English and Spanish sentences were presented separately and the order of presentation of language blocks was counterbalanced across participants. The monolingual sentences were presented in pairs in which the adjective was pre- or post-nominal and asked them to choose which one was more acceptable. In Spanish, 99% of responses indicated a preference for a post-nominal adjective, while in English 98.5% of responses indicated a preference for pre-nominal adjectives.

Filler trials: In each session we included 72 non-critical trials where the focus of contrast between choices in a pair was not the adjective but the determiner or the adverb so that the critical trials made up only about a third of all pairs seen by participants. Some of the results for the filler trials will be reported elsewhere.

Quality control trials: There were 12 quality-control trials that consisted of pairs of sentences with inter-sentential code-switches. One of the members of each pair had an uncontroversial error that could be easily detected if the sentences were read carefully. If a participant failed

more than 2 of these trials, they were removed from the sample and substituted with a new participant.

Procedure

The survey was administered online using Qualtrics and testing occurred across 3 separate sessions: one in which participants completed the language tests and the background questionnaire, and two in which they completed half of the 2AFC trials. Participants were presented with NPs in which the switch went from English to Spanish in one session and from Spanish to English in the other, counterbalancing across participants. At the beginning of each session participants were given the choice of reading the instructions in English or Spanish; all but 5 participants chose to complete the questionnaire in English. The instructions informed participants that they would see a series of sentence pairs, and asked them to pick the one closer to the way they would speak to another bilingual person. They were asked to make a choice even if both sentences sounded "right" or both sounded "wrong". Then participants were presented with the 114 pairs of code-switched sentences described above. The pairs of sentences were presented one at a time and the order of presentation of the pairs (critical trials, fillers, and quality control trials), as well as the order of each sentence within each pair, was individually randomized for each participant. Participants had to make a choice for each trial before progressing to the next one and could not go back to previous sentences.

Results

Participants' responses were analyzed using Thurstone's (1927) Law of Comparative Judgment, Case V, which analyzes participants' pairwise comparison of the stimuli to generate a ranking of

preference among conditions as well as a measure for relative comparison between them. The results of Thurstone's analysis must be interpreted within the context of signal detection theory (c.f. Cowart, 1996): every measurement includes a noise component that usually follows a normal distribution. If we repeatedly measure the same construct, we will end up with a normal distribution centered on the likely "ideal" value of the magnitude, with a lesser probability of obtaining large errors of judgment than small errors in either direction. Thurstone's "measure" provides the center or mean of such a normal distribution, which Thurstone called "discriminal dispersions", for each condition. These measures can be interpreted values on an interval scale that represents a psychological continuum (in our case, the acceptability of the sentences). The unit of measurement along that scale is defined as the standard deviation of the distribution. (Brown & Peterson, 2009).

Table 3 shows the rank order and Thurstone's measure for each condition. The measure values are relative to the pattern with the lowest acceptability (which is by convention set to 0). The 95% confidence interval for this set of data was 0.071. A within-participants ANOVA revealed a significant effect of sentence type both for sentences in which the switch went from English to Spanish ($F(3,796)=526.22$, $p < .001$) as well as when it went from Spanish to English ($F(3,796) =184.48$, $p < .001$). Post-Hoc Tukey HSD tests indicated that all contrasts were significant (all p values < .01).

[Table 3 about here]

The results of Experiment 2 show a clear hierarchy of preference for the CS patterns tested.

the predictions of the MP over MLF. It is important to point out, however, that MP-/MLF+ condition is preferred over the MP-/MLF- condition, and that, similar to the results of Experiment 1, the preferred condition was that in which both constraints were satisfied.

Discussion

In this study, we investigated the contrasting predictions derived from the MLF and the MP (specifically, the generalization by Cantone & MacSwan, 2009) regarding the mechanisms underpinning code-switching between the noun and the adjective. In two critical conditions, the predictions for adjective positioning made by the MLF and the MP differed. The MLF predicts a violation when the adjective position is incompatible with the word order of the sentence's Matrix Language, while the MP predicts a violation when the adjective position is disallowed by the language of the adjective. While Experiment 1 (with traditional Likert scales) didn't yield a clear differentiation between the predictive power of each model, Experiment 2 indicated an advantage for the MP. However, the difference between the MP+/MLF- and MP+/MLF+ conditions points towards an additive effect in which both the language of the verb and the language of the adjective are used to determine word order in the NP. If decisions were based only on the language of the adjective (as predicted by the MP), there would be no difference between MP+/MLF- and MP+/MLF+ because the language of the verb would have nothing to add above and beyond what can be explained by the language of the adjective in regards to the acceptability of a given sentence. On the other hand, the fact that there is a difference between MP-/MLF+ and MP-/MLF- confirms that the language of the verb contributes to the acceptability decisions of informants.

Our results constitute strong support for Eppler et al's (2016, p. 1) suggestion that "advances in our understanding of grammaticality in code-switching will be achieved by combining the insights of different theoretical models instead of considering them in isolation". Our findings could be interpreted as evidence for a potential parallel between the MLF and the MP, as already pointed out by Radford, Kupisch, Köppe, and Azzaro (2007), who argued that it is possible to equate the MLF's notion of morphosyntactic frame with the MP's notion of Phase (a derivational unit in minimalist syntax). The rationale behind this possibility is that "the head of a phase is responsible (via a form of selection) for "handing over" functional features to subordinate items within the phase" (Radford et al, 2007, p. 245). Hence, our data could be interpreted as support for the suggestion that the MLF may dominate the whole Complementizer Phrase (CP) phase, thus making both theoretical models compatible and complementary.

Additionally, our results seem to provide evidence against the more general proposal within Minimalism that the DP (Determiner Phrase) is a separate phase (Svenonius, 2004, Hiraiwa, 2005), since information outside the DP needs to be taken into account when building a nominal construction.

In addition to understanding the mechanisms involved in solving conflict-sites in the grammar of bilinguals, the use of the 2AFC task provided robust and unprecedented insights into the validity of theories of code-switching. As discussed above (and more extensively in the psychophysics literature), the 2AFC task represents an improvement over other methods of measuring introspective judgments and provides clear, interpretable data that can be used for directly contrasting the predictions of linguistic theoretical models. This approach opens the way for systematic testing of predictions from linguistic theory in general, especially in cases where corpus data appear insufficient to test all possible combinations of theoretical predictions.

REFERENCES

Bader, M., & Häussler, J. (2010). Toward a model of grammaticality judgments. Journal of Linguistics, 46, 273-330.

Bock, R. D., & Jones, L.V. (1968). The measurement and prediction of judgment and choice. San Francisco, CA: Holden-Day.

Brown, T.C., Peterson, G.L. (2009). An enquiry into the method of paired comparison: reliability, scaling, and Thurstone's Law of Comparative Judgment. Fort Collins, CO: U.S. Department of Agriculture, Forest Service.

Bullock, B.E. & Toribio, A. J. (Eds.) (2009). The Cambridge Handbook of Linguistic Code-switching. Cambridge: Cambridge University Press.

Cantone, K. F., & MacSwan, J. (2009). The syntax of DP-internal codeswitching. In L. Isurin, D. Winford & K. de Bot (eds.) Multidisciplinary Approaches to Codeswitching (pp. 243-78). Amsterdam: John Benjamins Publishing.

Carol, J.M., Bever, T.G., & Pollack, C.R. (1981). The non-uniqueness of linguistic intuitions. Language, 57, 368-383.

Cinque, G. (1994). On the evidence for partial N-movement in the Romance DP. In G. Cinque, J. Koster, J-Y Pollock, L. Rizzi & R. Zanuttini (Eds.), Paths towards Universal Grammar. Studies in honor of Richard S. Kayne (pp. 85-110). Washington, DC: Georgetown University Press.

Cowart, W. (1996). Experimental Syntax: Applying Objective Methods to Sentence Judgments. California: Sage Publications Inc.

David, H.A. (1988). The method of paired comparisons (2nd ed.). New York, NY: Oxford

Davies, P. (2010). Identifying word-order convergence in the speech of Welsh-English bilinguals. PhD dissertation, Bangor University.

Davies, P., & Deuchar, M. (2010). Using the Matrix Language Frame model to measure the extent of word order convergence in Welsh-English bilingual speech. In A. Breitbarth, C. Lucas, S. Watts & D. Willis (eds.) Continuity and change in grammar (pp. 77–96). Philadelphia: John Benjamins.

Deuchar, M. (2012). Code Switching. In Chapelle, C.A. (ed.) The Encyclopedia of Applied Linguistics (pp. 664-675). Chichester, UK: Wiley-Blackwell.

Deuchar, M. (2006). Welsh-English code-switching and the Matrix Language frame model. Lingua, 116, 1986-2011.

Deuchar, M. and Davies, P. (2009). Code-switching and the future of Welsh. International Journal of the Sociology of Language, 195, 15–38.

Eppler, E.D., Luescher, A. & Deuchar, M. (2016) Evaluating the predictions of three syntactic frameworks for mixed determiner–noun constructions. Corpus Linguistics and Linguistic Theory, Advance online publication. DOI: 10.1515/cllt-2015-0006

Fairchild, S., & Van Hell, J. (2015). Determiner-noun code-switching in Spanish heritage speakers. Bilingualism: Language and Cognition. Advance online publication. DOI: 10.1017/S1366728915000619.

Fechner, G. T. 1876. Vorschule der Aesthetik, Vol. 2. Leipzig: Breitkopf & Härtel.

Gescheider, G. (1997). Psychophysics: the fundamentals (3rd ed.). New Jersey: Lawrence Erlbaum Associates.

Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using Mechanical Turk to Obtain and Analyze English Acceptability Judgments. Language and Linguistics Compass, 5, 509-524.

Green, D. M., & Swets, J. A. (1988). Signal Detection Theory and Psychophysics. Los Altos, CA: Penninsula.

Guzzardo Tamargo, R.E., Mazak, C., & Parafita Couto, M.C. (Eds.) (2016) Spanish-English CodeSwitching in the Caribbean and the U.S. Amsterdam: John Benjamins.

Gullberg, M. & Parafita Couto, M.C. (2016) An integrated perspective on code-mixing patterns beyond doubling? Bilingualism: Language and Cognition. Advance online publication. DOI: 10.1017/S1366728916000080

Herring , J., Deuchar, M., Parafita Couto, M.C., & Moro Quintanilla, M. (2010). "When I went to Canada, I saw the madre": evaluating two theories' predictions about codeswitching between determiners and nouns using Spanish-English and Welsh-English bilingual corpora, International journal of bilingual education and bilingualism, 13, 553-573.

Hiraiwa, K. (2005) Dimensions of Symmetry in Syntax: Agreement and Clausal Architecture. PhD dissertation, Massachusetts Institute of Technology.

Jake, J. L., Myers-Scotton, C. M., & Gross, S. (2005). A response to MacSwan (2005): Keeping the Matrix Language. Bilingualism: Language and Cognition, 8, 271–276.

Labov, W. (1972). Some principles of linguistic methodology. Language in Society, 1, 97-120.

Labov, W. (1975). What is a linguistic fact? Lisse: Peter de Ridder.

MacSwan, J. (2009). Generative approaches to codeswitching. In A. J. Toribio & B. E. Bullock (eds.), Cambridge Handbook of Linguistic Codeswitching (pp. 309-335). Cambridge, UK: Cambridge University Press.

MacSwan, J. (2005a) Codeswitching and generative grammar: A critique of the MLF model and some remarks on "modified minimalism." Bilingualism: Language and Congition, 8,1-22.

MacSwan, J. (2005b) Comments on Jake, Myers-Scotton and Gross's response: There is no "matrix language." Bilingualism: Language and Cognition, 8, 277-284.

MacSwan, J. (2000) The architecture of the bilingual language faculty: Evidence from codeswitching. Bilingualism: Language and Cognition, 3, 37–54.

Myers-Scotton, C. (2002). Contact Linguistics: Bilingual Encounters and Grammatical Outcomes. Oxford, UK: Oxford University Press.

Nagata, H. (1988). The relativity of linguistic intuition: The effect of repetition on grammaticality judgments. Journal of Psycholinguistic Research, 171, 1-17.

Nunnally, J.C. (1976). Psychometric theory. New York: McGraw Hill.

Oxford University Language Centre (n.d.). "Placement Tests." lang.ox.ac.uk. http://www.lang.ox.ac.uk/tests/index.html (accessed July 1st, 2015)

Parafita Couto, M. C., Deuchar, M., & Fusser, M. (2015). How do Welsh-English bilinguals deal with conflict? Adjective-noun order resolution. In G. Stell, & K. Yakpo (eds.) Code-Switching Between Structural and Sociolinguistic Perspectives (pp. 65-84). Berlin: De Gruyter.

Parafita Couto, M.C., & Gullberg, M. (2015). Code-switching within the noun phrase. Evidence from three corpora. Paper presented at the 10th International Symposium on Bilingualism. Rutgers University, New Jersey 20-24 May, 2015.

Parraga, C.A. (2015). Perceptual Psychophysics. In G. Cristobal, M. Keil & L. Perrinet (eds.) Biologically-Inspired Computer Vision: Fundamentals and Applications (pp. 81-108). New York, NY: Wiley.

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. Behavior Research Methods, 46, 1023-1031.

Poplack, S., & Meechan, M. (1998). How languages fit together in codemixing. International Journal of Bilingualism, 2, 127-38.

Radford, A., Kupisch, T., Köppe, R., & G. Azzaro, G. (2007). Concord, Convergence and Crash-Avoidance in bilingual children. Bilingualism: Language and Cognition, 10, 239-256.

Ross, J.R. (1979). Where's English? In C.J. Fillmore, D. Kemper & W.S. Wang (eds.) Individual differences in language ability and language behavior (pp. 127-163). New York: Academic Press.

Schütze, C.T. (1996). The empirical base of linguistics: Grammaticality judgments and linguistic methodology. Chicago, Il: University of Chicago Press.

Snow, C., & Meijer, G. (1977). On the secondary nature of syntactic intuitions. In S. Greenbaum (ed.) Acceptability in language (pp. 163-177). The Hague, the Netherlands: Mouton.

Stokes, W. (1974). All of the work on quantifier-negation isn't convincing. In M.W. La Galy, R.A. Fox & A. Bruck (eds.) Papers from the tenth regional meeting, Chicago Linguistic Society (pp. 692-700). Chicago: Chicago Linguistic Society.

Svenonius, P. (2004). On the edge. In D. Adger, C. de Cat & G. Tsoulas (eds.) Peripheries: Syntactic edges and their effects (pp. 261-287). Dordrecht: Kluwer.

Thurstone, L. (1927). A Law of Comparative Judgment. Psychological Review, 34, 273-286.

Torgerson, W.S. (1958). Theory and methods of scaling. New York, NY: John Wiley & Sons.

vanden Wyngaerd, E.V. (2016) The adjective in Dutch-French codeswitching: Word order and agreement. International Journal of Bilingualism. Advance online publication. DOI: 10.1177/1367006916632302

Table 1. Prediction of acceptability for code-switched constructions according to the MP and MLF models.

	English→Spanish			Spanish→English	
	Exemplar	Predicted Acceptability		Exemplar	Predicted Acceptability
(1)	I like the marrón dress	MP-/MLF+	(5)	Me gusta el vestido brown	MP-/MLF+
(2)	I like the dress marrón	MP+/MLF-	(6)	Me gusta el brown vestido	MP+/MLF-
(3)	I like the vestido brown	MP-/MLF-	(7)	Me gusta el marrón dress	MP-/MLF-
(4)	I like the brown vestido	MP+/MLF+	(8)	Me gusta el dress marrón	MP+/MLF+

MP+ =Acceptable according to the Minimalist Program; MP- =Not acceptable according to the Minimalist Program; MLF+ =Acceptable according to the Matrix Language Frame model; MLF- =Not acceptable according to the Matrix Language Frame model.

Table 2. Mean acceptability rating for each condition

Condition	English→Spanish	Spanish→English
MP+/MLF+	3.71 (S.D. 0.93)	3.71 (S.D. 0.86)
MP+/MLF-	3.08 (S.D. 0.90)	3.25 (S.D. 0.96)
MP-/MLF+	3.24 (S.D. 0.94)	3.12 (S.D. 1.04)
MP-/MLF-	2.60 (S.D. 0.88)	2.89 (S.D. 1.04)

Table 3. Ranking and measure for each condition

		Thurstone's Measure	
Rank	Condition	English→Spanish	Spanish→English
1	MP+/MLF+	2.26	1.32
2	MP+/MLF-	1.29	0.82
3	MP-/MLF+	0.90	0.50
4	MP-/MLF-	0	0

APPENDIX

Monolingual base sentences

English	Spanish
I like the brown dress	Me gusta el vestido marrón
I need a hot bath	Necesito un baño caliente
Samuel wrote it all in his green notebook	Samuel lo escribió todo en su cuaderno verde
Hugo is a Nicaraguan singer	Hugo es un cantante Nicaragüense
They had to hire an independent attorney	Tuvieron que contratar un abogado independiente

Code-Switched Critical Sentences

MP-/MLF+

English → Spanish	Spanish → English
I like the marrón dress	Me gusta el vestido brown
I need a caliente bath	Necesito un baño hot
Samuel wrote it all in his verde notebook	Samuel lo escribió todo en su cuaderno green
Hugo is a Nicaragüense singer	Hugo es un cantante Nicaraguan
They had to hire an independiente attorney	Tuvieron que contratar un abogado independent

MP+/MLF-

MP-/MLF-

English → Spanish	Spanish → English
I like the vestido brown	Me gusta el marrón dress
I need a baño hot	Necesito un caliente bath
Samuel wrote it all in his cuaderno green	Samuel lo escribió todo en su verde notebook
Hugo is a cantante Nicaraguan	Hugo es un Nicaragüense singer
They had to hire an abogado independent	Tuvieron que contratar un independiente attorney

MP+/MP+

English → Spanish	Spanish → English
I like the brown vestido	Me gusta el dress marrón
I need a hot baño	Necesito un bath caliente
Samuel wrote it all in his green cuaderno	Samuel lo escribió todo en su notebook verde
Hugo is a Nicaraguan cantante	Hugo es un singer Nicaragüense
They had to hire an independent abogado	Tuvieron que contratar un attorney independiente