El book or the libro? Insights from acceptability judgments into determiner/noun code-switches

Parafita Couto, M.C.; Stadthagen-González, H.

Research Note

The International Journal of Bilingualism

El book or the libro? Insights from acceptability judgments into determiner/noun code-switches

International Journal of Bilingualism
2019, Vol. 23(1) 349–360

DOI: 10.1177/1367006917728392
journals.sagepub.com/home/ijb


Maria Carmen Parafita Couto
Leiden University, The Netherlands

Hans Stadthagen-Gonzalez
University of Southern Mississippi, USA

Abstract

Objectives/research questions: We used two types of acceptability judgments to experimentally test the predictions of two theoretical models of code-switching regarding the surface realization of the determiner in nominal constructions: lexicalist approaches within the Minimalist Program (MP) versus the Bilingual NP Hypothesis within the Matrix Language Frame Model (MLF).

Methodology: Two separate groups of 40 early Spanish-English bilinguals evaluated the acceptability of sentences with code-switches between the determiner and the noun that reflected the predictions of the MP model, the MLF, of both or none. The first group rated them on a Likert scale, while the second group performed a two-alternative forced-choice acceptability task (2AFC).

Data and analysis: Ratings from the Likert ratings were subjected to an analysis of variance while results from the 2AFC were analyzed using Thurstone's Law of Comparative Judgment.

Conclusions: Both experiments yielded converging evidence supporting the predictions of the MLF. Results from the 2AFC provided a more detailed picture that suggests also a (smaller) contribution of the language of the determiner.

Originality: This is the first study to use acceptability judgments to directly contrast the predictions of these major theoretical models regarding switches between determiner and noun. An additional novelty is the use of the 2AFC and Thurstone's Law of Comparative Judgment, which yielded a more detailed picture than the more commonly used Likert-scale ratings.

Implications: Our results provide further support for Eppler, Luescher, and Deuchar's recent claim that we can only advance our understanding of grammaticality in code-switching if we combine the insights of the different frameworks rather than considering them in isolation.

Keywords

Determiners, Spanish, English, code-switching, comparative judgments

*Authors listed in alphabetical order; equal contribution from both authors.

Corresponding author:
Maria Carmen Parafita Couto, Leiden University Center for Linguistics, Witte Singel-complex, van Wijkplaats 3, 2311 BX Leiden, The Netherlands.
Email: m.parafita.couto@hum.leidenuniv.nl

Introduction

This paper focuses on the factors influencing the language of determiners in mixed nominal constructions. It has been reported that in the speech of Spanish/English bilinguals, Spanish determiners with English nouns (*el* book) are preferred over English determiners followed by Spanish nouns (*the* libro) (e.g. Liceras et al., 2008; Liceras, Spradlin, & Fernández-Fuertes, 2005; Moro Quintanilla, 2014; Valdés Kroff, 2016). These findings were argued to support a generativist view of feature spell-out where the language with the richest array of 'uninterpretable phi features' provides the surface realization of the functional category (Liceras et al., 2005). This means that since the Spanish determiner carries two uninterpretable features (gender and number), while English has no gender, the Spanish determiner will be preferred in mixed nominal constructions. Liceras et al. (2008) argue that the number of features is not as important as how the features are grammaticalized, that is, their degree of visibility. In other words, it is the more grammaticalized functional category (Spanish D containing the unvalued gen feature and the valued Φ feature) that will determine the direction of the switch, as it will be the one that will be spelled-out. In the context of English-Spanish switched Determiner Phrases (DPs), Spanish would be the 'dominant' language given that Spanish has gender. Also within generativism, Moro Quintanilla (2014) argues that the preference for Spanish determiners in mixed nominal constructions can be attributed to the presence of a gender feature on the Spanish determiner. Moro Quintanilla (2014) follows Chomsky in that 'we take deletion to be a "one fell swoop" operation, dealing with the phi-set as a unit. Its features cannot selectively delete: either all delete or none' (Chomsky, 2000: 124). In examples with an English determiner, the probe (the English D) lacks gender. Since it is entering into a construction with a noun from Spanish that has the additional gender phi-feature, it is incomplete and is unable to delete its uninterpretable features all at the same time. Consequently, the derivation does not converge. So Moro Quintanilla's analysis within the Minimalist approach predicts Spanish-English mixed nominals such as '*el* team' but not '*the *equipo*' ('team'). As is often the case in the code-switching literature, there is no agreement on the description of the facts. While there seems to be agreement that a Spanish determiner followed by an English noun is fine, the disagreement comes with an English determiner and a Spanish noun. Moro Quintanilla (2014), for example, argues that code-switching between an English determiner and a Spanish noun is less acceptable than the reverse, while Herring, Deuchar, Parafita Couto, & Moro Quintanilla (2010) and Blokzijl, Deuchar, & Parafita Couto (2017) doubt that this is always the case. Looking at Spanish-English naturalistic data from Miami, Herring et al. (2010) found that the language of the determiner tended to match the language of the morphosyntactic frame of the clause, which was usually Spanish (hence the higher frequency of Spanish determiners). Blokzijl et al. (2017) expanded on Herring et al.'s analysis and did an automatic analysis of the entire Bangor Miami corpus as well as a corpus of Nicaraguan Creole English and Spanish bilinguals. They found that in both corpora, the determiner matched the matrix language. In the Miami corpus, there was a higher tendency to have Spanish determiners in a Spanish matrix language clause. In the Nicaraguan corpus, on the other hand, determiners from Nicaraguan Creole English in a Nicaraguan Creole English matrix clause were more common. This observed match between the morphosyntactic frame and the determiner is precisely what Jake et al. (2002) postulated in their Bilingual NP Hypothesis within the Matrix Language Frame Model (MLF), stating that determiners in mixed nominal constructions should come from the matrix language of the clause. Under this view, Spanish determiners will surface in clauses with a Spanish matrix language and English determiners will surface in clauses with an English matrix language. However, neither Liceras et al. (2005, 2008) nor Moro Quintanilla (2014) took into consideration the morphosyntactic frame of the clause. It is also worth noting that, more recently, Myers-Scotton and Jake (2016) stated that since

Spanish determiners carry grammatical gender, they are more salient at the level of lexical-conceptual structure to the speaker than English determiners and will usually be preferred, unless English is the matrix language. According to Myers-Scotton and Jake (2016), if Spanish is the matrix language in any code-switching corpus, then it is more likely that Spanish determiners will dominate, but they hypothesize that Spanish determiners may also dominate due to differences in how determiners are elected in the two languages. According to them, the production process that elects nominal DPs is completed at the level of lexical-conceptual structure. The phi-features in the Spanish determiner carry information about gender and number, which need to be available to structure the rest of the clause. This recent proposal by Myers-Scotton and Jake (2016) gets closer to the predictions of lexicalist/generativist approaches.

In this study, we extend the work of a series of studies that have been devoted to evaluating these contrasting hypotheses, that is, (i) the determiner should come from the matrix language (the Bilingual NP hypothesis, within the MLF) or (ii) the determiner should come from Spanish, given that it is the language with grammatized gender features (Minimalist Program (MP)).

Herring et al. (2010) tested the predictions of both accounts with spontaneous data from Spanish-English and Welsh-English bilinguals. Both Spanish and Welsh have gender, while English does not. Herring et al. tested Moro Quintanilla's generativist account (which predicted that Welsh and Spanish determiners should surface in mixed nominal constructions) against the MLF's Bilingual NP Hypothesis (which predicted that the determiner should match the matrix language of the clause). Moro Quintanilla's account was successful in explaining all their Welsh-English data and most of the Spanish-English data. However, when they examined the language of the verb in the clause containing the mixed nominal constructions as part of the process of testing the MLF, they observed that the success of the generativist account was due to the fact that the language of the verb was almost always Welsh or Spanish (which were also the languages with grammatical gender features). In the small number of clauses where the finite verb was in English, an English determiner was usually found, contrary to Moro Quintanilla's generativist predictions. It is possible then that if Spanish determiners are more frequent than English determiners in production data, it may just be because of the greater frequency of clauses with Spanish as a matrix language in the datasets that have been studied. Blokzijl et al. (submitted) looked at a larger dataset of Spanish-English bilinguals from Miami as well as production data from Nicaragua Creole English-Spanish bilinguals from the South Atlantic Autonomous Region of Nicaragua. They found that while in Miami, Spanish determiners followed by English nouns were more common, in Nicaragua, Nicaraguan Creole English determiners followed by Spanish nouns were more frequent and, in all cases, the determiner would match the matrix language of the clause. Their data supports the hypothesis that the matrix language is the influential factor that determines the language of the determiner in mixed nominal constructions. Further, the match between the matrix language and the language of the determiner seems to be unaffected by any grammaticized features in the language of the determiner, particularly in the Nicaraguan corpus where the language with no grammaticized gender features provide the determiner in mixed nominal constructions.

Also making use of naturalistic production data, but focusing on a different language pair (German-English), Eppler et al. (2017) conducted a comparison of the predictions of these two frameworks as well as Word Grammar (Hudson, 2010). They found that all three models had a high level of accuracy for their data, although they made predictions about different aspects. While Word Grammar and the MP deal mainly with agreement between the features of the determiner and the noun, the MLF deals mainly with agreement in language between the determiner and the finite verb. Related to this, they claim that the 'observation due to the MLF of the co-occurrence relation between the source language of a determiner and of the verb in the same clause will need to be

accounted for in any fully accurate theory' (p. 23). They argue that a theory that can handle all aspects is necessary and conclude that in order to achieve an optimal theory, insights from all frameworks should be incorporated (p. 25). They also raise the issue of whether it is sufficient to limit ourselves to naturally occurring data, due to the inherent limitations of these data (no negative evidence, for example).

Moving from naturalistic to experimental methods, Fairchild and Van Hell (2017) is, to our knowledge, the only study that has experimentally examined, through a series of picture naming tasks, the ability of the two models to explain determiner-noun switches in Spanish-English bilinguals. Their results did not align with either theory, but they note that this could be due to the fact that their focus was on externally induced switching rather than naturalistic or spontaneous switching. They also point out that language dominance may have played a role, given that the bilinguals they tested were English-dominant. Hence, more controlled studies are called for to be able to provide objective evaluation of the predictions from these two mainstream theoretical accounts of code-switching. In this study we set out to compare Spanish/English mixed nominal constructions and control for the language of the morphosyntactic frame.

In the present study we used both traditional Likert scale ratings and a novel application of the two-alternative forced-choice task (2AFC) and Thurstone's (1927) Law of Comparative Judgments to test the predictions on determiners in code-switching derived from the different theoretical views (MP versus MLF) that have been under scrutiny in previous work (Blokzijl et al., 2017; Eppler et al., 2017; Fairchild & Van Hell, 2017; Herring et al., 2010).

Experiment 1: Acceptability judgments using Likert scales

Method

Participants. A total of 40 early Spanish/English bilinguals took part in each of the two experiments. All stated that they spoke the Mexican variety of Spanish, that they used both English and Spanish on a daily basis, and that they were able to speak Spanish before age four and English by the time they entered elementary school. Table 1 shows participant characteristics for each experiment. None of the participants took part in both experiments.

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).

Each participant's proficiency in English and Spanish was confirmed through English and Spanish tests prior to the main experiment. These tests were adapted from the Online Placement Tests used by Oxford University's Language Centre (Oxford University Language Centre, 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 US 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. Participants on both experiments had scores that were slightly better for English. In addition, participants were asked to rate their own language ability (c.f. Gathercole, 2007; Oscarson, 1989) on a scale of 1–4, where 4 indicates 'Confident in extended conversations', and, on average, they indicated to be slightly more confident in English than Spanish (see Table 1).

Table 1. Participant characteristics for Experiments 1 and 2.

	Experiment 1	Experiment 2
Number of female/male participants	18/22	18/22
Mean age in years (SD; range)	30;6 (8.5; 19–57)	29;5 (8.8; 18–56)
Number of participants born in the USA	36	35
Mean age of immigration to USA in years (for those not born in USA)	4:0	3:7
English exam score (out of 50)	44.7 (SD: 2.8)	45.2 (SD: 2.8)
Spanish exam score (out of 50)	42.8 (SD: 4.3)	43.7 (SD: 3.8)
Self-assessment of English proficiency[a]	3.98 (SD: 0.2)	3.98 (SD: 0.2)
Self-assessment of Spanish proficiency[a]	3.60 (SD: 0.6)	3.68 (SD: 0.5)

[a]On a scale of 1–4, where 4 indicates 'Confident in extended conversations'.

Materials

Critical sentences. We constructed 12 base sentences that included a subject, a verb, a determiner and a noun. Each base sentence was modified into code-switched forms according to the following patterns:

$$\text{MLF-/MP+: } V_{En} + D_{Sp} + OBJ_{En} - \text{ Edgar wanted } \textit{estos shoes}$$

$$\text{MLF+/MP+: } V_{Sp} + D_{Sp} + OBJ_{En} - \textit{Edgar quería estos} \text{ shoes}$$

$$\text{MLF+/MP-: } V_{En} + D_{En} + OBJ_{Sp} - \text{ Edgar wanted these } \textit{zapatos}$$

$$\text{MLF-/MP-: } V_{Sp} + D_{En} + OBJ_{Sp} - \textit{Edgar quería} \text{ these } \textit{zapatos}$$

This generated 48 code-switched critical sentences that were evaluated by participants using the 2AFC and the Likert scale techniques. These sentences were split into two sets, one for each testing session.

Determiners in Spanish and English differ in that Spanish obligatorily encodes for grammatical gender on some determiners (masculine versus feminine), whereas English only sometimes encodes for number on determiners (e.g. 'those', where the D explicitly encodes number). It has been reported that bilingualism type may influence gender assignment strategies (i.e. analogical criterion, masculine default or phonological clues, cf. Liceras et al. 2008; Parafita Couto, Munarriz, Epelde, Deuchar, & Oyharcabal, 2015 for an overview). For bilinguals similar to those included in this study (early Spanish/English bilinguals in the USA), there is strong evidence that masculine serves as the default gender for English nouns in code-switched noun phrases (Valdés Kroff, 2016), irrespective of the gender of the Spanish translation of the noun. Valdés Kroff (2016) reports a strong preference for the use of masculine gender in the production of mixed nominal constructions in the Spanish-English bilingual speech (from the Bangor Miami corpus; 296 out of 304 phrases with Spanish determiner and English noun had a masculine determiner; Table 3 in Valdés Kroff (2016)). He shows that the usage of masculine Spanish articles with English nouns is not constrained by gender assignment in Spanish. This apparent neutralization of gender assignment in bilingual speech is in contrast to how Spanish gender is assigned in unilingual Spanish, where gender concord between a noun and a determiner is obligatory. On the other hand, mixed nominal constructions with feminine marked determiners do not follow the same usage pattern. As reported by Valdés Kroff, although the number of feminine tokens was quite small (3% of the entire corpus), these infrequent mixed nominal constructions were categorically restricted to English nouns with feminine Spanish translation

equivalents, for example, *la* cookie, Sp. la.fem galleta.fem. This immensely asymmetric distribution favors a masculine default gender assignment strategy in bilingual speech production. According to Valdés Kroff (2016), a planned code-switched utterance in these Spanish-English bilinguals involves the adoption of a default gender strategy and, hence, forms that are marked with masculine determiners. It follows, according to him, that feminine-marked mixed noun phrases are not planned code-switched utterances. Instead, they are exceptional switches that occur on-the-fly in speech planning. Valdés Kroff suggests that these tokens may be the exception to regular production patterns in code-switching.[1] Beatty-Martínez & Dussias (2017) followed this up and tested whether bilingual speakers are sensitive to this production asymmetry observed by Valdés Kroff (2016) in ways that constraint the comprehension system. They compared the processing of gender congruent and incongruent mixed nominal constructions in sentential contexts using Event Related Potentials. Their results indicate that the comprehension system becomes optimally attuned to the input and capitalize on the importance of assessing code-switching experience in bilingualism research. In a similar fashion, Valdés Kroff, Dussias, Gerfen, Perrotti, and Bajo (2017) also used code-switching as a tool to illustrate how language experience modulates comprehension. They employed the visual world paradigm to examine the extent to which gender-marked Spanish determiners facilitate upcoming target nouns in a group of Spanish-English bilingual code-switchers. Across their experiments, bilinguals revealed an asymmetric gender effect in processing, reflecting the asymmetric use of gender in the production of code-switched speech.

Given that (i) our study concerned the choice of language of determiner (English or Spanish), rather than the strategies that bilinguals use for gender assignment when the determiner comes from Spanish and the noun from English, (ii) the fact that feminine determiners are so infrequent in mixed nominal constructions in regular production (Valdés Kroff, 2016) and (iii) there is evidence that the comprehension system is attuned to the input (Valdés Kroff et al. 2017; Beatty-Martínez & Dussias, 2017), we decided to only include English nouns with a masculine translation equivalent in Spanish (e.g. juego – game, tornillos – screws). This was done in order to avoid confounds that may be due to different gender assignment strategies in code-switching. Hence, all our Spanish determiners had to be masculine (independently of a default gender assignment strategy or a translation equivalence strategy), matching also the patterns found in production data. If we had included nouns with a feminine translation equivalence (e.g. table '*mesa*'), it is possible that a bilingual participant would react negatively to 'la.fem *mesa*', simply because they use the masculine default strategy reported by Valdés Kroff (2016).

Proper names in the sentences were chosen so that they were commonplace in both Spanish and English (e.g. Edgar, Anabel). Each of the two sets contained one sentence with one of the following determiners: este – this, ese – that, estos – these, esos – those, el – the (singular), los – the (plural).

Filler sentences. We also created 96 non-critical sentences (half for each session) where the focus of contrast between sentences was not the determiner but the adjective or the adverb. 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 sentences. We included eight quality control sentences with inter-sentential code-switches. Each sentence 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 versus second half of the sentence; English versus Spanish portion; and type of error (verb tense, number agreement, gender agreement and word order). If a participant failed more than two of these trials, they were removed from the sample and substituted with a new participant.

Table 2. Results for Experiments 1 and 2.

Condition	Structure	Example	Thurstone's measure	Likert rating (SD)
MLF+/MP+	V_{Sp} + D_{Sp} + OBJ_{En}	*Edgar quería estos* shoes	1.64	2.96 (0.67)
MLF+/MP-	V_{En} + D_{En} + OBJ_{Sp}	Edgar wanted these *zapatos*	1.57	3.03 (0.58)
MLF-/MP+	V_{En} + D_{Sp} + OBJ_{En}	Edgar wanted *estos* shoes	0.44	2.56 (0.61)
MLF-/MP-	V_{Sp} + D_{En} + OBJ_{Sp}	*Edgar quería* these *zapatos*	0	2.39 (0.63)

Procedure. The survey was administered online using Qualtrics and testing occurred across three separate sessions: one in which participants completed the language tests and the background questionnaire; and two in which they rated half of the sentences described in the *Materials* section. These last two sessions happened about a week apart. The order of presentation of each block of sentences was counterbalanced across participants. Participants were given the choice of reading the instructions in English or Spanish. The instructions informed participants that they would see a series of sentences and that they were to indicate on a five-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 80 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 make a choice for each item before progressing to the next one and could not go back to previous sentences.

Results. A preliminary data analysis revealed that there was no significant interaction between sentence block and acceptability per condition ($F(3,117) = .758$, $p = .52$), so the analyses presented herewith were performed collapsing data across testing sessions.

Table 2 presents the mean and standard deviation for the ratings for each type of sentence on a five-point Likert scale. A within-participants analysis of variance (ANOVA) revealed a significant effect of CS pattern ($F(3,160) = 10.0$, $p < .001$). Post-hoc Tukey honest significant difference (HSD) tests indicated that there was no significant difference in ratings between conditions MLF+/MP+ (e.g. *Edgar quería estos* shoes) and MLF+/MP- (e.g. Edgar wanted these *zapatos*) ($p = 0.61$) or between conditions MLF-/MP+ (e.g. Edgar wanted *estos* shoes) and MLF-/MP- (e.g. *Edgar quería* these *zapatos*) ($p = 0.9$). All other differences were significant (all p values $<.05$). Participants showed preference for the two conditions in which CS patterns follow the MLF predictions, irrespective of whether they correspond with the MP predictions.

Experiment 2: Acceptability judgments using two-alternative forced-choice tasks

In order to corroborate the results obtained in Experiment 1, we collected acceptability ratings using the 2AFC and analyzed the data using Thurstone's (1927) Law of Comparative Judgments. This technique is considered the gold standard for collecting and interpreting subjective introspective data in most areas of behavioral research (e.g. Cattelan, 2012; Montag, 2006; Párraga, 2015), but has been hitherto largely absent from use in linguistic acceptability judgments (for more details, see Stadthagen-Gonzalez, López, Parafita Couto, & Párraga, 2018). In 2AFCs participants are presented with pairs of stimuli and must choose which item is more acceptable, with pairwise comparisons covering all possible contrasts between conditions. Forced-choice tasks have been shown to have higher statistical power than Likert-style ratings (e.g. Gigerenzer, Krauss, & Vitouch, 2004; Gigerenzer & Richter, 1990; Sprouse & Almeida, 2011), while Likert ratings

Table 3. Example of paired comparisons for Experiment 2.

Sentence 1	Sentence 2
Edgar quería estos shoes	Edgar wanted these zapatos
Edgar wanted estos shoes	Edgar wanted these zapatos
Edgar wanted estos shoes	Edgar quería these zapatos
Edgar quería estos shoes	Edgar wanted these zapatos
Edgar quería estos shoes	Edgar quería these zapatos
Edgar wanted these zapatos	Edgar quería these zapatos

sometimes suffer from a 'rush to the middle', where participants use only a subset of available values on the scale and avoid the extremes.

The method of choice for analyzing 2AFC data is derived from Thurstone's Law of Comparative Judgment (Bock & Jones, 1968; Cattelan, 2012; Engen, 1971; Jones & Thissen, 2007; Párraga, 2015; Reber, 1995), which places the results of multiple pairwise comparisons along a single interval scale that represents a one-dimensional quality (in our case, 'acceptability'). The units of this scale are not pre-determined but are given by the standard deviation of the distribution of responses for a particular set of data, thus providing an unrestricted scale with potentially infinite granularity (Sprouse & Almeida, 2011). This combination of high statistical power and granularity allows us to investigate subtler differences than those available from Likert-scale results (Stadthagen-Gonzalez et al., 2018).

Method

Participants. Participants were a total of 40 early English/Spanish bilinguals with very similar characteristics of those in Experiment 1. Table 1 shows participant characteristics for Experiment 2. None of the participants took part in both experiments.

Materials. The materials used in Experiment 2 were identical to those in Experiment 1, but presented in pairs, as described in the *Procedure* section.

Procedure. The general procedure for Experiment 2 was similar to that in Experiment 1, but this time sentences were presented in pairs, contrasting each version of a given base sentence with all its other variations (Table 3 provides an example of such contrasts). This yielded a total of 120 trials (including fillers and quality control sentences) for each of the two sessions. 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'.[2] Then participants were presented with the 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, 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

A preliminary data analysis revealed that there was no significant interaction between sentence block and acceptability per condition ($F(3,123) = 1.415$, $p = .24$), so analyses presented herewith were performed collapsing data across testing sessions.

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. 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). Stadthagen-Gonzalez, et al. (2018) provide further details on how to perform this type of analysis.

Table 2 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 0.04. A within-participants ANOVA (from summary data) revealed a significant effect of sentence type ($F(3,1916) = 322.8$, $p < .001$). Post-hoc Tukey HSD tests indicated that all contrasts were significant (all p values <.001) except for the difference between conditions MLF+/MP+ and MLF+/MP-, which was not significant ($p = 0.7$).

Overall, the pattern of results from both methods indicates that the most preferred patterns are those that the follow the predictions of the MLF, independently of whether the matrix language is English or Spanish. This means that English determiners are accepted in switched nominal constructions at a similar rate as Spanish determiners (contra Liceras et al.'s 2008 and Moro Quintanilla's 2014 predictions). Experiment 2 shows more granularity in the results in that it yielded a significant difference between the MLF-/MP+ and the MLF-/MP- patterns, suggesting the possibility of gradience, where if there is a mismatch between the determiner and the matrix language, then the Spanish determiner is preferred. In this regard, our results do not categorically rule out an influence of the determiner's language; they indicate that, although the matrix language seems to carry the most weight regarding judgments of acceptability, the language of the determiner can also play a (lesser) role.

Conclusion and discussion

We set out to test the acceptability of sentences built to present different combinations of determiner and nouns, within a set syntactic frame. Twelve base sentences (Subject + Verb + Determiner + Noun) were modified into code-switched forms according to the patterns shown in Table 1. The results were similar whether using Likert-scale ratings or the 2AFC and Thurstone's Law of Comparative Judgment: patterns MLF+/MP+ (e.g. *Edgar quería estos* shoes) and MLF+/MP- (e.g. Edgar wanted these *zapatos*) exhibit the highest acceptability with no significant difference between them (all ps > .6). Acceptability for patterns MLF-/MP+ (e.g. Edgar wanted *estos* shoes) and MLF-/MP- (e.g. *Edgar quería* these *zapatos*) was significantly lower (all ps < .05 in both experiments). For the Likert scale results there was no difference between patterns MLF-/MP+ and MLF-/MP- ($p = .9$), but that difference was significant ($p < .001$) for 2AFC data, which typically yields greater discriminability between conditions (Stadthagen-Gonzalez, et al., 2018). The overall pattern of results offers converging evidence in support of the MLF's predictions. If decisions were based only on the language of the determiner (as predicted by the MP), there would be no difference between the MLF+/MP+ and the MLF-/MP+ conditions because the language of the verb would not affect the acceptability of a sentence above and beyond what can be explained by the language of the determiner. This is precisely what Eppler et al., 2017, Herring et al. (2010), Blokzijl et al., (2017) and Parafita Couto and Gullberg (2017) found in their naturalistic data in different communities (German/English, Spanish/English, Nicaraguan Creole/Spanish, Papiamento/Dutch and Welsh/English). Our study confirms this previous observation due to the MLF of the co-occurrence relation between the source language of a determiner and of the verb in the same clause. Hence, our experimental results reinforce Eppler et al.'s (2017) claim, based on naturalistic data, that we can only advance our understanding of grammaticality in code-switching if we combine the insights of the different frameworks rather than considering them in isolation. As argued by

Eppler et al. (2017), we may be able to provide a better explanation for our data if we abandon a strict version of lexicalism and adopt constructionist approaches instead. Although providing a precise theoretical formulation of our results within the MP is outside the scope of this paper, we could at least speculate that one implication of our results would be that the DP cannot be a separate phase (contra Hiraiwa, 2005; Svenonius, 2004), since we have evidence that information outside the DP needs to be taken into account when deriving a mixed nominal construction. If determiners are not phase heads, an interesting theoretical question arises, already posited by Radford, Kupisch, Köppe, and Azzaro's (2007), as to whether we can posit some sort of parallelism between the MLF and the MP if we equate the MLF notion of frame with the Minimalist notion of Phase. According to the MLF, a mixed utterance has a morphosyntactic frame (from the matrix language) and morphemes from an embedded language can be inserted into this frame. Properties of the matrix frame will determine the nature of the embedded morphemes that can be inserted within it. In a similar fashion, within Minimalism, Radford et al. (2007) argued that 'the head of a phase is responsible (via a form of selection) for "handing over" functional features to subordinate items within the phase' (p. 245). Further research on different switching points could elucidate whether the potential parallels between the notion of *frame* within the MLF and the notion of *phase* within the MP can indeed be equated.

It is also worth noting that the theoretical approaches examined here do not address the asymmetric distributions of switches observed in the production data. An emergent approach that accepts asymmetric differences in usage is highly likely (Valdés Kroff, 2016). To this end, and in order to further explore the comprehension–production link, it would be informative to run a similar comprehension study in the South Atlantic Autonomous Region of Nicaragua, to see whether we could replicate our results in a community with a different switching direction in the production data (as reported by Blokzijl et al., 2017).

From a methodological point of view, the application of the 2AFC and Thurstone's Law of Comparative Judgment to this type of issue shows promise in providing a more detailed, granular, picture of acceptability judgment data. From a theoretical point of view, our results point to a theory where insights from both the MP and the MLF frameworks are combined: while features are important (as in the lexicalist/generative view), we should consider abandoning a strict version of lexicalism and adopting constructionist approaches (cf. Eppler et al., 2017).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Notice that this masculine default strategy is not operative for all language contact situations. Different bilingual communities may use different strategies for gender assignment. For example, Parafita Couto et al. (2015) reported a feminine default strategy for Basque-Spanish bilinguals; Duran Eppler (2010), looking at gender assignment in German-English, reported that phonological factors were frequently overridden by morphological and semantic factors; Poplack, Pousada, & Sankoff (1982), looking at Spanish-English in Puerto Rico and French-English in Montreal, found that the role of phonological and semantic factors was different in both bilingual communities and argued that the factors governing gender assignment are language specific; Treffers-Daller (1994) observed that the gender of the noun in the donor language is important in her French-Dutch data.
2. The probability of two different stimuli having exactly the same value on the judgment scale is considered negligible, and thus no 'tie' is allowed when making the pairwise judgment. Also, if indeed there

was no difference, the probability of a particular judge to pick one option over the other would be 50%, so any differences caused by the forced choice on a particular data point would even out across several judges and instances of the comparison (David, 1988).

References

Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language*, *95*, 173–189.

Blokzijl, J., Deuchar, M., & Parafita Couto, M. (2017). Determiner asymmetry in mixed nominal constructions: The role of grammatical factors in data from Miami and Nicaragua. *Languages*, *2*, 20.

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.

Cattelan, M. (2012). Models for paired comparison data: A review with emphasis on dependent data, *Statistical Science*, *27*, 412–433.

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*. Cambridge, MA: The MIT Press.

Cowart, W. (1996). *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage.

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

Duran Eppler, E. (2010). *Emigranto: The syntax of German/English code-switching*. Vienna: Braumueller.

Engen, T. (1971). Psychophysics, Vol. II: Scaling methods. In J. Kling & L. Riggs (Eds.), *Woodworth and Schlosberg's experimental psychology* (pp. 89–91). New York, NY: Holt, Rinehart, & Winston.

Eppler, E. D., Luescher, A., & Deuchar, M. (2017). Evaluating the predictions of three syntactic frameworks for mixed determiner–noun constructions. *Corpus Linguistics and Linguistic Theory*, *13*(1), 27–63. DOI: 10.1515/cllt-2015-0006

Fairchild, S., & Van Hell, J. (2017). Determiner-noun code-switching in Spanish heritage speakers. *Bilingualism: Language and Cognition*, *20*(1), 150–161. DOI: 10.1017/S1366728915000619

Gathercole, V. (Ed.). (2007). *Language transmission in bilingual families in Wales*. Cardiff: Welsh Language Board.

Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using mechanical turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass*, *5*, 509–524.

Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual: What you always wanted to know about significance testing but were afraid to ask. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences*. Thousand Oaks, CA: Sage.

Gigerenzer, G., & Richter, H. (1990). Context effects and their interaction with development: Area judgments. *Cognitive Development*, *5*, 235–264.

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, MA.

Hudson, R. (2010). *An introduction to Word Grammar*. Cambridge, UK: Cambridge University Press.

Jake, J. L., Myers-Scotton, C., & Gross, S. (2002). Making a minimalist to codeswitching work: Adding the Matrix Language. *Bilingualism, Language and Cognition*, *5*, 69–91.

Jones, L. V., & Thissen, D. A. (2007). A history and overview of psychometrics. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, volume 26: Psychometrics* (pp. 1–27). New York, NY: Elsevier.

Liceras, J. M., Fernández Fuertes, R., Perales, S., Pérez-Tattam, R., & Spradlin, K. T. (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.

Liceras, J. M., Spradlin, K. T., & Fernández-Fuertes, R. (2005). Bilingual early functional-lexical mixing and the activation of formal features. *International Journal of Bilingualism*, *9*(2), 227–251.

Montag, E. D. (2006). Empirical formula for creating error bars for the method of paired comparisons. *Journal of Electronic Imaging*, *15*, 222–230.

Moro Quintanilla, M. (2014). The semantic interpretation and syntactic distribution of determiner phrases in Spanish/English codeswitching. In J. MacSwan (Ed.), *Grammatical theory and bilingual codeswitching*. Cambridge, MA: MIT Press.

Myers-Scotton, C., & Jake, J. (2016). Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. *International Journal of Bilingualism*, *21*(3), 340–366.

Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing, 6*, l–13.

Oxford University Language Centre. (n.d.). *Placement tests*. Retrieved July 1, 2015 from www.lang.ox.ac.uk/tests/index.html

Parafita Couto, M. C., & Gullberg, M. (2017). Code-switching within the Noun Phrase. Evidence from three corpora. *International Journal of Bilingualism*. Advance online publication. doi: 10.1177/1367006917729543

Parafita Couto, M. C., Munarriz, A., Epelde, I., Deuchar, M., & Oyharcabal, B. (2015). Gender conflict resolution in Spanish–Basque mixed DPs. *Bilingualism: Language and Cognition*, *18*, 304–323.

Párraga, 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., Pousada, A., & Sankoff, D. (1982). Competing influences on gender assignment: Variable process, stable outcome. *Lingua*, *57*, 1–28.

Radford, A., Kupisch, T., Köppe, R., & Azzaro, G. (2007). Concord, convergence and crash- avoidance in bilingual children. *Bilingualism: Language and Cognition*, *10*, 239–256.

Reber, A. (1995). *The Penguin dictionary of psychology*. New York, NY: Penguin Books.

Sprouse, J., & Almeida, D. (2011). Power in acceptability judgment experiments and the reliability of data in syntax (Unpublished manuscript). University of California, Irvine & Michigan State University.

Stadthagen-Gonzalez, H., López, L., Parafita Couto, M. C., & Párraga, C. A. (2018). Using two-alternative forced choice tasks and Thurstone's law of comparative judgment for acceptability judgments in code switching. *Linguistic Approaches to Bilingualism*, *8*(1), 67–97.

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, The Netherlands: Kluwer.

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273–286.

Treffers-Daller, J. (1994). *Mixing two languages: French-Dutch contact in a comparative perspective*. Berlin, Germany: Mouton de Gruyter.

Valdés Kroff, J. R. (2016). Mixed NPs in Spanish-English bilingual speech: Using a corpus based approach to inform models of sentence processing. In R. E. Guzzardo Tamargo, C. M. Mazak, & M. C. Parafita Couto (Eds.), *Spanish-English code-switching in the Caribbean and the US* (pp. 281–300). Amsterdam, The Netherlands: John Benjamins.

Valdés Kroff, J. R., Dussias, P. E., Gerfen, C., Perrotti, L., & Bajo, M. T. (2017). Experience with code-switching modulates the use of grammatical gender during sentence processing. *Linguistic Approaches to Bilingualism*, *7*(2), 163–198. doi:10.1075/lab.15010.val

Author biographies

Maria Carmen Parafita Couto is a university Lecturer at the Leiden University Center for Linguistics (Netherlands). She obtained her PhD in Linguistics at the University of Kansas and subsequently worked at Kenyon College (Ohio, USA), the ESRC Centre for Research on Bilingualism at Bangor University (UK), and the University of Strathclyde (UK). She studies the impact of bilingualism and language contact on language structure, mostly the syntax of codeswitching.

Hans Stadthagen-Gonzalez is currently an Assistant Professor of Psychology at The University of Southern Mississippi. He previously worked at the Bilingualism Research Centre at Bangor University (UK). He studies different aspects of the interaction between languages in bilinguals, including syntax, semantics, and pragmatics.