Code-switching hallmark of bilingualism

Mostly focus on switch costs and processes involved

Nevertheless, cs is ubiquitous and has various socio-pragmatic motivations. Therefore, it is very likely that it affords processing benefits overriding switch costs

The study reported here tested the role of CS in prediction in terms frequency.

One place where CS is frequently found is on or prior to a less frequent, more unexpected portions of speech. Myslin and Levy (2015) showed that this is the case by modeling bilingual corpora and conducting an offline sentence completion task. Bilinguals

This function of CS is similar to the processing benefits of disfluencies

Stimuli:

Pics from IPNP

Difs in freq for stimuli pairs

Start with the same letter?, Gender matching

4 lists

Speaker, PR variety

Presentation order, appear on screen when? 200 ms before the audio onset

Length of CS portion in CS and Sp sentences

Cutting and intensity scaling?

No reaction times due to error in coding

Participants:

Age 18-32 (M = 20.83, SD = 3.53)

Gender (26 female, 4 male)

Age of acquisition of Spanish and English

All participants reported having begun learning both English and Spanish before puberty (Spanish AoA Mean = 0.67, SD = 2.35, Max = 12; English AoA Mean = 3.67, SD = 2.48, Max = 10).

Kroff

**Proficiency – dominance**

**Graphical dominance but continuous variable**

**Not necessary AX-CPT for this**

**Debra Titone**

**Bins? Not typical**

**autocorrelated**

**Observed just plot observed**

**lm**

## **Introduction and Background**

Code-switching (CS), or the use of several codes in the same conversation (Gardner-Chloros, 2009), is fairly frequent in bilingual discourse (Beatty-Martinez & Dussias, 2017). Nevertheless, studies have found that there is a cost associated with its processing (Altarriba et al., 1996; Meuter & Allport, 1999), albeit reduced with the increase in experiments’ ecological validity (Blanco-Elorrieta & Pylkkänen, 2017). This seeming contradiction between the ubiquity of code-switching and the associated processing costs has not found a clear explanation in the current psycholinguistic research on the online processing of CS.

Nevertheless, offline psycholinguistic studies and sociolinguistic observations offer converging socio-pragmatic and information-organizational motivations for CS. These motivations might outweigh any purported costs of processing for bilinguals and solidify the use of CS as a discourse marker. Myslín and Levy (2015) showed that code-switching to the more marked, i.e. less used, language can be used to signal a more information-rich (i.e., unexpected) portion of speech, both in their corpus analysis and the offline experiment, where bilinguals assumed a less frequent word followed a first- to second-language (L1-L2) CS. Another attested motivation for switching to the more marked language is speaking about emotional, predominantly taboo topics (Bentahila, 1983). CS to the more salient code could thus alert to increased emotionality as well, which is another type of information. These observations suggest that code-switching could offer online-processing benefits by alerting to and thus aiding the prediction of highly informative portions of speech.

Disfluencies, or irregularities in fluent speech, such as “uh”, “um”, pauses, etc, seem to aid prediction in a similar way. Disfluencies aid the prediction of unexpected, new (Arnold, Fagnano, & Tanenhaus, 2003; Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Arnold, Kam, Hudson, & Tanenhaus, 2007) or low-frequency words (Bosker, Quené, Sanders, & de Jong, 2014). In these visual world paradigm studies (eye-tracking while looking at images and listening to instructions), disfluent instructions to select an image caused participants to start looking at the unexpected item earlier. Given the parallels between CS and disfluencies, we propose to test whether CS can serve a similar function in online-processing, i.e. whether it can alert bilingual comprehenders to a more informative, lower frequency word coming up.

## **Research Design**

In this study, a behavioral proxy to the eye-tracking visual world paradigm was used, with the reaction time of button presses as the dependent variable. The participants were presented with two-picture panels, one representing a low-frequency word, and the other a high-frequency word. The participants listened to the instructions on which object to click on. The instructions were either in monolingual Spanish, or took the form of Spanish-English code-switched sentences. The code-switch preceded the name of the object by 3 words, one content and two function ones, to avoid any immediate effects of the code-switch processing affecting the results. The code-switch was placed within a noun phrase, after an article and at the noun, which was deemed an acceptable place for a code-switch by 3 bilingual speakers.

The independent variables are Language of the Instructions with 2 levels: Spanish (L2) and Spanish-English (L1-L2 code-switch), and Frequency of the Target image: Low and High. The study was thus geared towards Spanish-dominant or Spanish-first bilinguals. Only Spanish-English code-switched instructions were chosen to simplify the design and since L1-L2 is the attested switch direction used to signal more informative portions of speech to come (Myslín & Levy, 2015). Further studies will include English-Spanish and monolingual conditions to more precisely determine the factors affecting frequency-level prediction in code-switching.

Sound recordings and preprocessing steps

Instructions were recorded with the help of a balanced Spanish-English speaker of Puerto Rican descent. Two carrier verbs were chosen for the carrier phrases: eligir - ‘choose’, and encontrar - ‘find’, to increase variety and approach ecological validity. The carrier phrases were as follows:

‘Encuentra el dibujo de un/una/Ø \_\_\_\_\_.’

‘Elige el dibujo de un/una/Ø \_\_\_\_\_.’

‘Encuentra el drawing of a/an/Ø \_\_\_\_\_.’

‘Elige el drawing of a/an/Ø \_\_\_\_\_.’

The names of the pictures were recorded in isolation, with the speaker being given the instructions to imagine the noun coming at the end of a sentence. The carrier phrases were recorded in combination with a mock noun, and subsequently cut, to ensure that the intonation and article pronunciation were as natural as possible. The carrier phrases containing code-switches were somewhat longer in the initial Spanish portion and contained cues for the upcoming code-switch. This was not deemed a problem, since it mimics the way code-switches are uttered in real-world conversations and it could potentially alter the comprehenders to the code-switch and make the code-switch processing “easier”, i.e. less of a surprise. This would ultimately allow the comprehenders to smoothly process the code-switch, focusing only on its informative value.

All carrier phrases were scaled to the average intensity of 70 decibels, and all nouns were scaled to the average intensity of 66 decibels, using sound manipulation software Praat (Broersma & Weenink, 2018), to ensure volume uniformity among carrier phrases and noun, and a natural volume decline at the end of a sentence. The final versions of carriers and nouns were concatenated without a pause, again for the sake of greater ecological validity as there was no pause in the full carrier phrase + noun recordings.

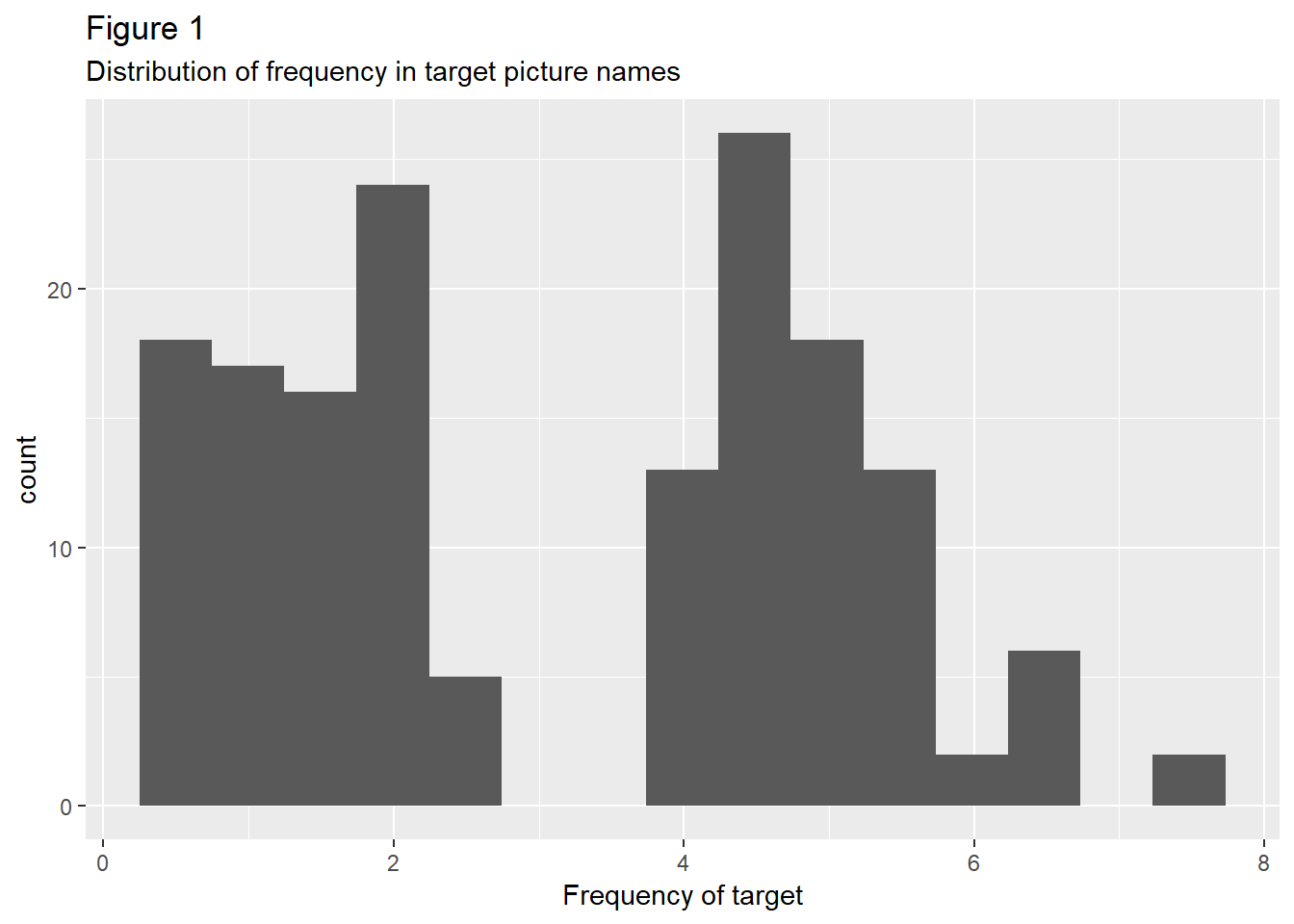
## **Materials**

### Picture panels

Sixty two images, 32 representing high-frequency words and 32 representing low frequency words were extracted from the International Picture Naming Project database (Szekely et al., 2004, 2005), to form 32 experimental panels - pairs of images. The frequency measures used in this database are log natural transformed frequency counts, taken from the CELEX Lexical database (Baayen, Piepenbrock, & Gulikers, 1995). In the English counterpart low frequency group, the frequency ranged between 0.693 and 2.303 frequency counts, and in the Spanish counterpart low frequency group, it ranged between 0.602 and 2.303 frequency counts. In the English counterpart high frequency group, the frequency ranged between 3.989 and 7.396 frequency counts, and in the Spanish counterpart high frequency group, it ranged between 4.007 and 7.584 frequency counts. Pictures were paired such that there is the largest possible frequency gap between them.

Average difference between the high and low frequency members of a pair was 3.535894

frequency counts for CS trials, and 3.595042 counts for Spanish trials. Paired t-tests for both English and Spanish trials show that there is a significant frequency difference between the high and low frequency members of the experimental pairs (Spanish counterparts: t(33) = 21.778, p-value < 0.0001; English counterparts: t(33) = 33.993, p-value < 0.0001). Figure 1 below is a histogram representing the distribution of the target picture name frequencies in both lists. There is a marked gap between the frequencies of the two groups, ensuring that the research design includes the desired manipulation.



Only pictures whose names had similar frequency in both Spanish and English were chosen for the study. The frequency for English words ranges from 0.69 to 6.08 log natural transformed frequency counts, whereas the Spanish word frequency ranges from 0.6 to 7.58 counts. Average absolute difference between the English and Spanish object name counterparts was 0.444 frequency counts (range: 0 – 1.61). A paired t-test showed no significant difference between English and Spanish picture name counterparts, ensuring that no extraneous factors affect the results of the study (t[63] = -1.5286, p-value = 0.1314, mean of differences -0.1090411).

In addition to being matched for the frequency difference between low and high frequency groups, as well as the frequency similarity between Spanish and English counterparts, the experimental two-picture panels were also matched for the gender of the Spanish picture name/translation equivalent. Out of 32 experimental trials, there were 15 where both pictures had a feminine gender Spanish name.

Due to difficulties in attaining a sufficient number of pictures from the database, 12 pictures representing English-Spanish cognates had to be included in the experimental trials. These pictures were paired with each other to control for any possible cognate effects, resulting in 6 cognate experimental trials.

Two experimental lists were created, with one experimental item pair appearing in one of the 4 conditions/versions in one of the lists:

Spanish, Low Frequency target;

CS Low Frequency Target;

Spanish High Frequency target;

CS High Frequency target.

This process resulted in 8 items per condition in a list, and ensured that a participant sees a given experimental pair in only one of the 4 conditions.

Participants

Thirty Spanish-English bilingual participants who self-report code-switching were recruited on the UF campus. They were compensated for their participation in course credit or cash.

Participants completed Language History Questionnaire (………………), semantic fluency tests in Spanish and English, adapted standardized grammar tests: Michigan English Language Institute College English Test (MELICET) and Diplomas of Spanish as a Foreign Language (DELE), and the cognitive control task AX-CPT. The tests were administered either prior to or following the main reading experiment. The order of the main experiment and tests was counterbalanced, as well as the language order. Upon the completion of the main experiment, the participants rated the words which appeared in the experiment on the frequency of use and amount of exposure. !!!!!!!!!!!!!!1 do this analysis! With their own continuous ratings

Twenty-eight participants reported code-switching in LHQ, whereas 2 participants responded with “Not sure”. All participants filled out additional questions on the frequency of use and exposure to code-switching.

Two different proficiency measures were collected in addition to self-reports: standardized grammar tests and semantic fluency tests. Proficiency scores were transformed into dominance by taking the ratio between the Spanish- and English-language standardized tests DELE and MELICET, and the ration between Data on semantic fluency was also collected and

Data Analysis

Models

The target time period for eye-movement analysis was 200 ms before the onset of the target and 200 ms after the onset of the target. Planning and executing an eye-movement has been shown to take around 200 ms, so the 200 ms post-target-word region still reflects processes which occurred prior to the presentation of the target.

(13 out of 960)

We removed the trials in which participants clicked on the incorrect image prior to analyzing the eye-tracking data. This resulted in 1.35% trial loss, evenly distributed among participants (max. 2 trials per participant), suggesting that participants were performing at ceiling.

The Time variable in the eye-tracking data was binned into 20 ms bins. Each participant saw 8 items in the same condition, amounting to 160 ms total time which one participant could spend looking at the set of items in a particular condition within one time-bin. The dependent variable in the analyses was the proportion of milliseconds within a time-bin that each participant spent looking towards the High vs. Low Frequency items, regardless of their target/distractor status, in the Spanish and the Code-switched language context. The proportion of time spent looking at a set of images in a particular condition was normalized by z-scoring. ………………. (………After calculating the proportions, blinks and in-saccade eye-states were removed from the data. )

The participant variables included in the model were Dominance (DELE/MELICET), CS Use, and Exposure to CS. All continuous predictors have been normalized by z-scoring. Bins/Time were excluded from the analysis due to autocorrelation………………..

A step-wise regression was performed by comparing models which individually included one of the three participant variables, using the anova function ().

Z-scored proportions of time spent looking at a set of items in a particular condition in each bin was fit to a linear mixed-effects model (……………………..) with the independent variables Language Context (Spanish vs. Code-switched), Frequency of the Fixated Item (FreqFix; High vs. Low), and Language Dominance (continuous), as well as their interactions as predictors. Participant intercept and Language Context and FreqFix slopes were included in the model as random effects. None of the main effects was found to significantly affect the dependent variable. Nevertheless, the Language Context X FreqFix interaction was significant, beta………….. = -0.481, t = -2.026, such that participants were looking at the low frequency item in the Spanish condition less than in the Code-switched condition (Table X……). Graph X ……………. shows the data predicted by the model (line) vs. observed data (points).

**Otho poly …………..**

**We report the results of a model with the time variable transformed into orthogonal polynomials from 1st to 3rd order. ……………. Nadji negde gde se pominje**

**The time transformed using orthogonal polynomials were included in the model along Language Context, FreqFix, and Dominance, as well as their interactions with each variable. The interactions among the three transformed time variables were not included in the model. All continuous variables have been standardized as described above. Participant Language Context, and FreqFix intercepts, as well as intercepts for each transformed time variable, were included in the model as random effects.**

**Significant effects of the analysis are summarized in the Table X …………………….**

**Only Language main effect was found to be significant, …………….., such that there were more looks overall in the Spanish condition compared to CS. Importantly , several interactions were found to be significant. The Language x FreqFix interaction proved significant, beta = -0.487, t(…………) = -10.499, such that Low frequency items were fixated less in the Spanish condition compared to CS condition. Also, the interaction of Language, Dominance, and FreqFix was significant, b = -0.276, t() = -5.967, such that** **the participants with higher relative Spanish dominance looked at the Low frequency items less in the Spanish condition compared to CS condition. The interaction of Language, FreqFix, and the 1st order polynomial transformed time variable was also found to be significant, b = -0.885, t() = -4.184 such that the looks towards the Low frequency item decreased over time in the Spanish condition compared to CS condition. The interaction of Language, Dominance, FreqFix, and 1st order polynomial transformed time variable was significant, b = , t = -4.023, such that the participants with higher relative Spanish dominance looked less over time at the Low frequency item compared to the CS condition. Graph X ………… shows the model fit, with modeled data represented as lines and observed as points.**

**Items as thingies random effects**

**Both as random effects?**

Reml was true here, but what if it should be false?

I’ve done it once like a pro, I’ll do it again.