# Code-switching aids the prediction of the unexpected

Aleksandra Tomic & Jorge Valdés Kroff

## Introduction

Code-switching (CS), or the use of several codes in the same conversation (Gardner-Chloros, 2009) is one of the hallmarks of bilingualism. Decades of sociolinguistic research on the rules governing the use of code-switching, its functions and meanings, found that code-switching serves a large spectrum of sociopragmatic purposes (Gumperz, 1982; Myers-Scotton, 1993). Nevertheless, experimental psycholinguistic research has so far mostly focused on the processing costs of code-switching (Litcofsky & Van Hell, 2017; Olson 2017), as well as potential attenuation of these costs in certain circumstances (Fricke et al., 2016; Guzzardo Tamargo et al., 2016; Valdés Kroff et al., 2017). While this line of research provides important insights into the cognitive control processes associated with switching languages, it does not take into account the sociopragmatic motivations for code-switching and potential benefits that code-switching could provide in processing subsequent segments. Despite the processing costs associated with it (at least in cued paradigms, e.g. Meuter & Allport, 1999), code-switched speech makes up a relatively large portion of bilingual discourse - around 20% (Beatty-Martinez & Dussias, 2017). This ubiquity of code-switching suggests that code-switching affords benefits to bilinguals which override the purported processing costs. The study reported here experimentally tested and confirmed one such processing benefit of CS, founded in sociopragmatic motivations: alerting to and aiding prediction of lower frequency, unexpected words.

## Background

Some of the most cited functions of code-switching are identity expression (Velasquez, 2010), situational marking, (re)negotiating social relations (Myers-Scotton, 1993), face-saving (Bentahila, 1983), discourse organization (Auer, 1988), emphasis (Gumperz, 1982), introducing indirect speech (Albirini, 2011). A more recently proposed function for code-switching is organizing discourse in terms of information content (Myslín & Levy, 2015). The authors analyzed bilingual corpora using statistical modeling, as well as the results of an offline cloze task to find that bilinguals code-switch from their more frequently used language to their less used, or more salient language at more information-rich, unpredictable words. The offline task, Shannon guessing game, showed that bilinguals expect harder-to-predict words to be the L1-L2 code-switch sites, suggesting that bilinguals use code-switches as a discourse marker for information content. Another sociolinguistically 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 content. 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.

This potential function of code-switching aligns it with other discourse-organizational markers, such as disfluencies. Disfluencies, or irregularities in fluent speech, such as “uh”, “um”, pauses, have been found to occur when referring to new vs. given information (Arnold et al., 2000; Barr, 2001). Experimental research showed that bilinguals make use of this distribution regularity in on-line language processing to help them predict 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 faster, shortly prior to or following the onset of the target word.

Given the parallels in the distribution of CS and disfluencies, as well as the role of disfluencies in on-line processing, we propose to test whether CS can serve a similar function, i.e. whether code-switchers are able to make use of this discourse marker during on-line processing to help them predict more unexpected, lower frequency words coming up.

## Research Design

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 independent variables are Language of the Instructions with 2 levels: Spanish (L1) and Code-switched (Spanish-English, L1-L2), and Frequency of the Fixated Image: Low and High. The study was thus geared towards Spanish-first bilinguals. Only the L1-L2 CS direction was 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 English monolingual conditions to more precisely determine the factors affecting frequency-level prediction in code-switching.

### Materials

### Instructions and audio recordings

In the instructions, 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 CS and Spanish monolingual carrier phrases and image names 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 picture names 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 onset of the code-switch was somewhat delayed (Mean = 912 ms) compared to the comparable point in Spanish-only instructions (Mean = 890) and the pre-CS-onset portion seemed to contain phonetic cues for the upcoming code-switch. This was not deemed a problem, since it mimics the way code-switches are produced in real-life 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 comprehenders to smoothly process the code-switch, focusing only on its informative value. The time frame from CS onset to target onset was longer in the Spanish only instructions (Mean = 955 ms) than in the CS condition (Mean = 832.33 ms), presumably due to an additional syllable in the Spanish equivalent for “drawing” and the additional syllable in the article for the female gender.

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.

### 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 with 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). All experimental pictures, picture names, and frequencies can be found in Appendix X.

Experimental images were chosen such that there is the smallest possible discrepancy in the frequency of the Spanish and English picture name. Descriptive statistics for English and Spanish words are presented in Table X.

|  |  |  |
| --- | --- | --- |
|  | English | Spanish |
| Overall Mean (SD) | 3.145 (1.87) | 3.245 (1.908) |
| High Freq Mean (SD) | 4.892 (.713) | 5.028 (.772) |
| Low Freq Mean (SD) | 1.398 (.542) | 1.48 (.552) |

Table X: Means and standard deviations for all experimental words in English and Spanish, and split by the frequency group.

Average absolute difference between the English and Spanish object name counterparts was 0.444 frequency counts (SD = 0.371, range: 0 – 1.61). A two-tailed paired t-test showed no significant difference between English and Spanish picture name counterparts, ensuring no extraneous factors would affect the results of the study (t[63] = -1.527, p = 0.131, mean of differences = -0.109).

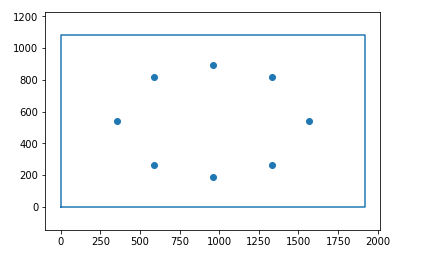
Pictures were paired such that there is the largest possible frequency gap between them. Average difference between the high and low frequency members of a pair was 3.535894 frequency counts for CS trials, and 3.595042 counts for Spanish trials. Paired t-tests for both English and Spanish trials show 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). The experimental two-picture panels were also matched for the gender of the Spanish picture name/translation equivalent to prevent participants from using gender of the article as a cue. 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 appropriate pictures from the database, 12 pictures representing English-Spanish cognates were 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.

Four 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 trials per condition in a list and ensured that a participant sees a given experimental item in only one of the 4 conditions. The order of experimental and filler trials was pseudorandomized using nested lists to ensure that there was never more than 3 experimental items in a row. There were 64 filler trials in each list, twice as many as there were experimental trials. The filler images were drawn from the same database. They were mostly neutral and similar in terms of frequency. The fillers were the same across lists, but their order and image position was randomized in presentation.

The experiment was programed in Experiment Builder (SR Research, 2011). Images were presented two at a time in the panels, on white background. The pictures could appear anywhere on the circular pattern shown in Graph X. The position of the images was randomized prior to feeding the stimuli list into Experiment Builder, by using a prepared Python script. To make sure that there is no overlap between the looks to the target vs. distractor items, the items were never adjacent.



Graph X: Possible positions for the images in the panel. Two images were not allowed to appear next to each other.

### Participants

Thirty Spanish-English bilingual participants (4 male), self-reported code-switchers, age range 18-32 (M = 20.83, SD = 3.53), were recruited on the UF campus. They were compensated for their participation in course credit or cash. All participants reported having begun learning both English and Spanish before puberty (Spanish age of acquisition [AoA] Mean = 0.67, SD = 2.35, Max = 12; English AoA Mean = 3.67, SD = 2.48, Max = 10).

Participants completed Language History Questionnaire (LHQ), 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 **[not sure whether to include this if we’re not reporting the results here].** Proficiency results are summarized in Table X. 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. These ratings would have been used in case there was a large incorrect trial percentage, or a large discrepancy between the frequency counts of experimental words taken from CELEX compared to participants own exposure to and use of the words.

Twenty-eight participants reported code-switching in the LHQ, whereas 2 participants responded with “Not sure”. All participants filled out additional questions on the frequency of use and exposure to code-switching in speaking and writing. The mean response to the question of the frequency of code-switching use was 4.1, SD = 0.845, with 1 standing for “Never” and 5 for “Always”. The mean response to the question on the exposure to code-switching in spoken conversation was 3.867, SD = 0.973, with the same response coding.

|  |  |  |  |
| --- | --- | --- | --- |
|  | English: M (SD) | Spanish: M (SD) | Difference: Paired 2-tailed T-test |
| LHQ – Speaking | 9.73 (.64) | 8.27 (1.55) | \*\*\* |
| LHQ – Listening | 9.73 (.58) | 9.17 (1.15) | \* |
| LHQ – Writing | 9.53 (1.2) | 7.67 (1.73) | \*\*\* |
| LHQ – Reading | 9.67 (.84) | 8.07 (1.6) | \*\*\* |
| MELICET | DELE | 43.73 (3.36) | 30.97 (6.97) | \*\*\* |
| Semantic fluency |  |  |  |

Table X: Proficiency profile for participants (n = 30). LHQ values represent self-reported ratings of proficiency from the Language History Questionnaire.

### Procedure

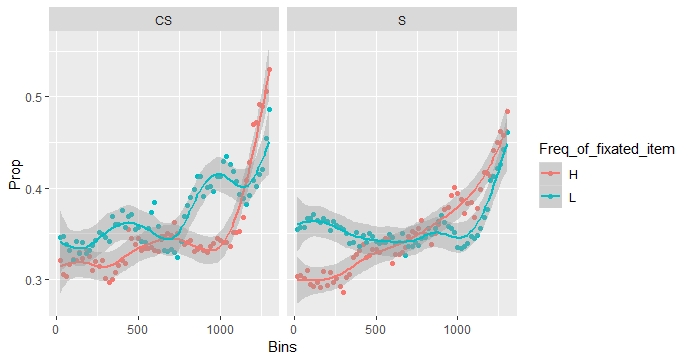
The study took place at the Bilingual Sentence Processing Lab, Department of Spanish and Portuguese Studies, University of Florida. Participants were greeted and given instructions in the form of a code-switched written and spoken passage to listen to audio instructions and click on the correct image. Eight practice items followed the eye-tracker calibration process. The eye-movements were recorded using SR Research Eyelink 1000 Plus desk-mounted eye-tracker. Audio instructions playback was delayed by 200ms after the presentation of the panel. The participants were instructed to use a mouse to click on the correct image.

## Data Analysis

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 (Graph X). Planning and launching an eye-movement has been shown to take around 200 ms (Allopenna, Magnuson, & Tanenhaus, 1998; Travis, 1936), so the 200 ms post-target-word region still reflects predictive processes which had taken place prior to the presentation of the target.

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 despite relative low frequency of certain items.

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 out of 160 ms total that each participant spent looking towards a particular image. After calculating the proportions, blinks and in-saccade eye-states were removed from the data **[percentage of removed time per participant……………..].** The independent variables were Language Context (Spanish, Code-switched), Frequency of the Fixated Image (FreqFix; High, Low), and Dominance (continuous). The values for the independent variable Frequency of the Fixated Image were assigned to each fixated image regardless of their target/distractor status. Dominance was operationalized as the ratio between the DELE and MELICET scores, with higher ratio meaning more Spanish dominant. Due to an error in procedure, the Dominance data was not available for 1 participant. Their Dominance score was filled with the Dominance average. Dominance and Proportion of Looks have been standardized by z-scoring.

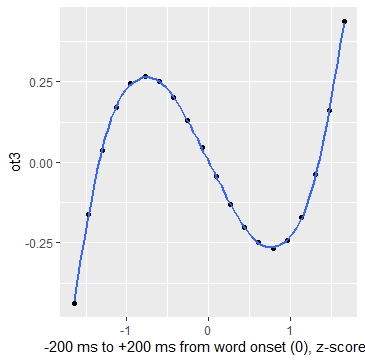
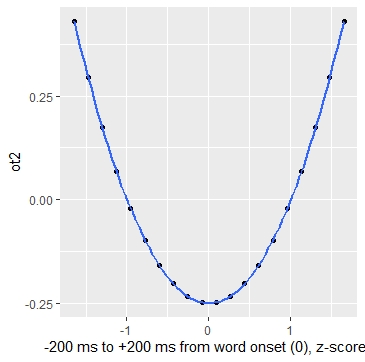
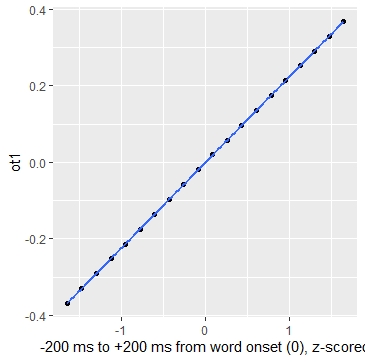


Graph X: Y-axis represents the Proportion of looks to High and Low frequency images out of 160 ms total per bin, for Code-switched and Spanish language context. X-axis represents the time course of -800 ms (approximate CS onset) to +500 ms from the target word onset (red line). The blue rectangle represents the target time period used in the analysis.

### Growth Curve Analysis

We initially performed a Growth Curve Analysis model, with the time variable transformed with orthogonal polynomials of 1st, 2nd, and 3rd order: OT1 – linear term, OT2 – quadratic term, OT3 – cubic term (Graph X). This analysis would have allowed us to potentially take into account the curvature of the time series data and increasing trends in the dependent variable. Nevertheless, the OT2 and OT3 terms as fixed effects did not improve the model significantly[[1]](#footnote-1), as confirmed by comparing the models with and without OT2 and OT3 terms as fixed effects, using the anova function (Fox & Weisberg, 2011). Therefore, we only report the more parsimonious model without these terms as fixed effects.[[2]](#footnote-2)

a) b) c)



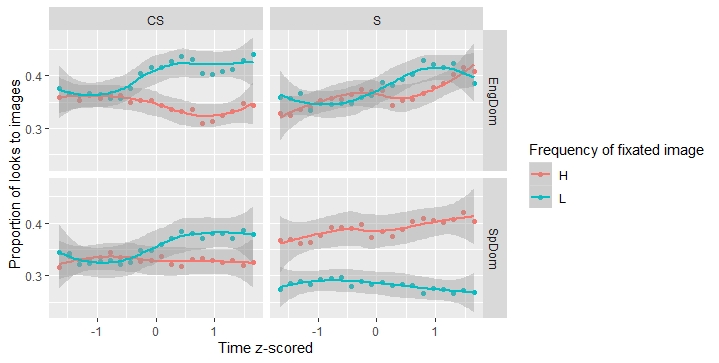
Graph X: Graphical representations of the Time variable transformed using orthogonal polynomials of 1st, 2nd, and 3rd order: a) OT1, linear term, b) quadratic term, c) cubic term.

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 using lme4 package in R (Bates et al., 2011). The model included Language (contrast coded as Spanish -0.5, Code-switched +0.5), FreqFix (contrast coded as High -0.5, Low +0.5), Dominance (continuous), and linear orthogonal polynomial term, OT1, (continuous), as well as their interactions, as predictors. Participant Language, FreqFix, and orthogonal polynomial terms OT1, OT2, and OT3 intercepts were included in the model as random effects.[[3]](#footnote-3) Results of the analysis are summarized in the Table X.

|  |  |  |  |
| --- | --- | --- | --- |
| Fixed effects: | Estimate | St. Error | t-value |
| (Intercept) | -0.02832 | 0.08300 | -0.341 |
| Language | 0.01940 | 0.05761 | 0.337 |
| Dominance | **-0.15906** | **0.08345** | **-1.906** |
| FreqFix | -0.03351 | 0.12170 | -0.275 |
| ot1 | **0.28034** | **0.09074** | **3.090** |
| Language:Dominance | -0.05961 | 0.05781 | -1.031 |
| Language:FreqFix | **0.49200** | **0.04740** | **10.380** |
| Dominance:FreqFix | -0.14483 | 0.12224 | -1.185 |
| Language:ot1 | -0.18068 | 0.10538 | -1.714 |
| Dominance:ot1 | -0.09330 | 0.09103 | -1.025 |
| FreqFix:ot1 | **0.30105** | **0.10554** | **2.852** |
| Language:Dominance:FreqFix | **0.28593** | **0.04722** | **6.055** |
| Language:Dominance:ot1 | 0.12296 | 0.10548 | 1.166 |
| Language:FreqFix:ot1 | **0.89128** | **0.21102** | **4.224** |
| Dominance:FreqFix:ot1 | 0.01930 | 0.10559 | 0.183 |
| Language:Dominance:FreqFix:ot1 | **0.82272** | **0.21112** | **3.897** |

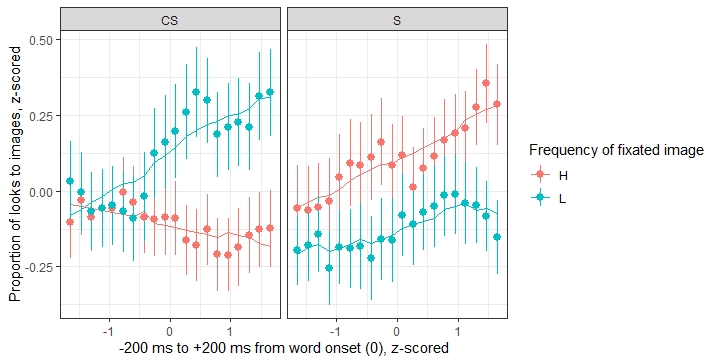
Table X: Coefficients, standard errors, and t-values for the GCA model reported here. Significant values are bolded. Relevant interactions are underlined.

OT1 main effect was found to be significant, β = 0.28, SE = 0.091,t = 3.090, such that overall looks increased over time, as well as the Dominance main effect, β = -0.159, SE = 0.084, t = -1.906, such that more Spanish-dominant participants had overall fewer looks to images. Importantly, several interactions were found to be significant. The Language x FreqFix interaction proved significant, β = 0.492, SE = 0.047, t = 10.380, such that Low frequency items were fixated more in the CS condition compared to Spanish condition. Also, the interaction of Language, Dominance, and FreqFix was significant, β = 0.285, SE = 0.047, t = 6.055, such that the participants with higher relative Spanish dominance looked at the Low frequency items more in the CS condition compared to Spanish condition (Graph X).



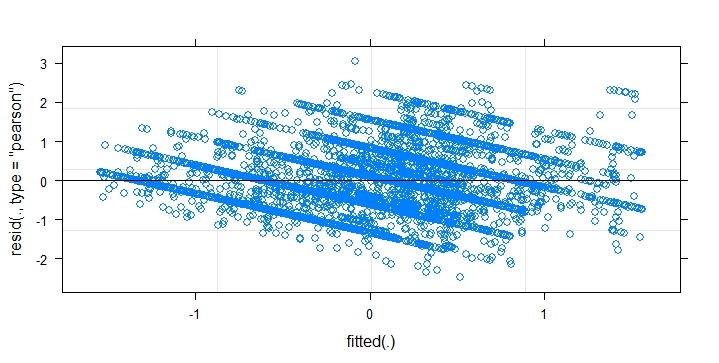
Graph X: Looks to High and Low Frequency items in CS and Spanish condition split by Dominance. Dominance was transformed into a categorical variable for graphing purposes by assigning the participants with a Dominance score lower/higher than the average Dominance score the value of Low/High Spanish Dominance.

The interaction of Language, FreqFix, and the 1st order orthogonal polynomial term was also found to be significant, β = 0.891, SE = 0.211, t = 4.224 such that the looks towards the Low frequency item increased over time in the CS condition compared to Spanish condition. The interaction of Language, Dominance, FreqFix, and 1st order polynomial term was significant as well, β = 0.822, SE = 0.211, t = 3.897, such that the participants with higher relative Spanish dominance looked more over time at the Low frequency item in the CS condition compared to the Spanish condition. Graph X shows the model fit.

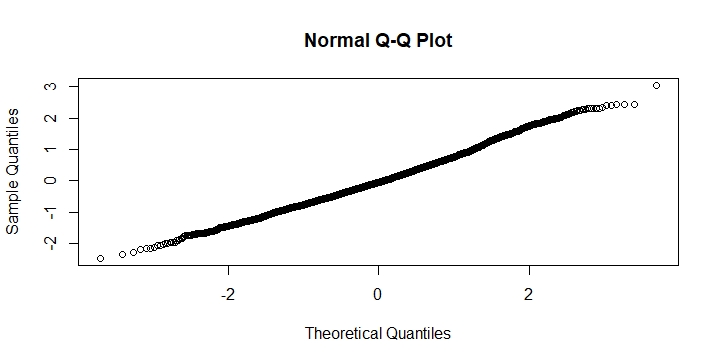


Graph X: Reported GCA model fit, with modeled data represented as lines and observed data as points.

Graph X shows the residuals from the model. Residuals seem to be normally distributed and do not show traces of autocorrelation. This is confirmed by the normal distribution (straight line) in the Quantile-Quantile plot (Q-Q plot, Graph X).



Graph X: Residuals from the reported GCA model.



Graph X: Quantile-Quantile plot with theoretical quantiles from a normal distribution plotted on the x-axis, and sample quantiles on the y-axis. The diagonal line suggests that the sample residuals from the reported GCA model come from a normal distribution.

## Discussion

The results from the Growth Curve Analysis point to a global increase of looks to the low frequency item in the CS condition in the period of 200 ms prior to the onset of the target word and 200 ms following the onset of the target word. The chosen interval following the target onset still reflects the processing prior to the onset, since it takes about 200 ms to program and launch a saccade (Allopenna, Magnuson, & Tanenhaus, 1998; Travis, 1936). This result provides support to our hypothesis that code-switching from L1 to L2 plays a role in on-line language processing: it changes bilinguals’ expectations in terms of the information load, i.e. frequency, of the following items. This result is especially significant given the Frequency Bias, i.e. a general tendency to look at higher frequency, more familiar items (e.g. Dahan, Magnuson, & Tanenhaus, 2001). Moreover, the results for the interaction of Language, FreqFix, and OT1 further support our predictions, since there is a continuous increasing trend for the looks to the lower frequency items in the CS context compared to the Spanish context as the target onset approaches.

The significant interactions Language x Dominance x FreqFix and Language x Dominance x FreqFix x 1st order polynomial term, are more privy to the cues provided by the L1-L2 code-switch or the absence of it, or can simply make use of it in real time more efficiently than relatively lower Spanish proficiency bilinguals. The Looks-by-Dominance graph (Graph X) suggests that the bilinguals with a higher proficiency in Spanish exhibit the Frequency bias in Spanish (e.g. Dahan, Magnuson, & Tanenhaus, 2001), i.e. the tendency to look at images with higher frequency. This makes the flip of the trend in the CS condition, i.e. increased looking at the low frequency items, all the more extreme.

Future studies will include the English-Spanish code-switch direction condition and the main effect of CS Direction and its interaction with Dominance in the analysis. This will allow us to determine whether it is lower dominance or higher salience of the language being switched into that are driving the predictive process.

## Conclusion

The primary goal of the study was to account for the discrepancy between the processing costs and the ubiquity of code-switching, as well as to bridge the gap between the sociolinguistic and experimental psycholinguistic research on code-switching. The psycholinguistic focus on code-switching costs may be neglecting the fact that code-switching does not occur randomly and can serve various socio-pragmatic functions. It is crucial for any sound psycholinguistic theory of code-switching to account for sociopragmatic functions associated with it (Myers-Scotton, 2006), such as the one experimentally probed here: discourse organization in terms of information distribution.

The results of the study corroborate the findings of the off-line cloze study and corpus modeling of the information distribution within bilingual discourse (Myslín & Levy, 2015). The results also confirm our hypothesis that code-switching provides experimentally detectable processing benefits as a cue to bilinguals to anticipate unexpected information, much like disfluencies in studies on monolinguals (Arnold, Fagnano, & Tanenhaus, 2003; Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Arnold, Kam, Hudson, & Tanenhaus, 2007). The fact that the CS distribution patterns in the discourse affected the ease of processing in comprehension is also in line with the models linking production and comprehension, such as Production-Distribution-Comprehension model (PDC; MacDonald, 2013). These models posit that production pressures shape the language distribution. Subsequently, emerging distribution patterns affect the ease of language comprehension.

In this study, we began with a simple operationalization of unexpectancy or salience as lexical frequency. However, this function of L1-L2 code-switching could extend to other non-salient/salient information contrasts. Future studies could thus probe the role of code-switching in on-line processing of given vs. new and emotionally neutral vs. taboo information (Tomic & Valdés Kroff, in prep.). We hope that this and similar studies will further open the scientific conversation on the roles of code-switching in language processing and continue to bring psycholinguistic research closer into alignment with sociolinguistic approaches to code-switching.

1. Only the FreqFix:OT3 interaction was found to be significant, β = -0.265, SE = 0.105, t = -2.518. [↑](#footnote-ref-1)
2. We report the results of the basic linear mixed-effects model without the orthogonal polynomials in the Appendix X. [↑](#footnote-ref-2)
3. The model without OT2 and OT3 Participant intercepts failed to converge. [↑](#footnote-ref-3)