Investigating code-switching and disfluencies in bilingual dialogue

Fahima Ayub Khan and **Bill Noble**

Department of Linguistics, Philosophy and Theory of Science
University of Gothenburg
`fahima.ayub.khan@gu.se`; `bill.noble@gu.se`

Abstract

This paper investigates the relationship between disfluency and code-switching in bilingual dialogue. We examine a corpus of 41 bilingual (Spanish-English) conversations and test the hypothesis that code-switching can be a response to negative evidence of grounding in the form of disfluencies. We find that there is a statistically significant relation between disfluencies and code-switching. Particularly, disfluencies have a positive effect on within turn code-switching.

1 Introduction

This paper investigates the communicative function of code-switching (switching between languages) in bilingual dialogue, as it relates to disfluencies, such as filled pauses and self-repair.

Disfluency is not just an interruption to the normal flow of conversation. On the contrary, disfluencies have a crucial role to play in facilitating communicative alignment and coordinating interaction (Hlavac, 2011). It is well-established that repair is crucial for aligning speakers and establishing common ground in dialogue (Healey et al., 2018, 2013, 2011). Furthermore, psycholinguistic studies have noted that disfluencies facilitate referential disambiguation, leading to greater interactive efficiency (Bailey and Ferreira, 2007). In bilingual dialogue, code-switching is an additional resource that speakers can draw on to facilitate conversational alignment (Wei and Milroy, 1995; Cromdal and Aronsson, 2000). Furthermore, previous work has demonstrated that code-switches tend to occur in turns with disfluencies (such as "um", "er", etc.,) and in clarification requests (Beatty-Martínez et al., 2020; Kootstra et al., 2020).

However, the mechanisms of interaction between disfluencies and code-switching are not well-studied. Is code-switching a response to communicative problems indicated by disfluency? If so, is it a response to disfluencies produced by other speakers, or also by oneself? More generally, why is it that disfluencies appear with greater frequency in the vicinity of code-switches? In this paper, we test the hypothesis that code-switching is a repair mechanism in bilingual dialogue—i.e., that code-switching can be a response to negative evidence of grounding, such as disfluency or negative feedback.

2 Method

2.1 Data

We investigated this hypothesis using the Bangor Miami corpus (Deuchar, 2010) which is a set of 56 spontaneous conversations between Spanish-English bilinguals living in Miami (USA). Of these we excluded 15 conversations that included only one of the two participant's turns. The final dataset contains 41 dialogues, 40 841 turns, and 254 739 tokens.

The conversations are transcribed in the CHAT format (MacWhinney, 2022), and include token-level language annotation. This makes it possible to pinpoint code-switches both within and between turns.[1] For the purpose of this study, we define intersentential switching as using both languages between turns. Instances where participants code-switch in the same turn has been coded as intrasentential switching in our analysis. In order to determine switch between turns, we coded turns whose language tag was different from the language tag corresponding to the last token of the previous turn. The data contains 4971 turns (12.2%) that switch from the language of the previous turn (i.e., between turn switches) and 2215 turns (5.4%) with within-turn switching. In total, 6539 turns (16%) are code-switched.

The CHAT transcription format also includes

---

[1]Some tokens, such as proper names, are coded as belonging to both languages. For the purposes of detecting code-switches, we consider these tokens to have the language tag of the previous token in the turn (or the final token in the previous turn, if it is the first token).

fine-grained annotations for disfluencies. In total, 6694 turns out of 40 841 (16.4%) contained disfluencies. Under the category of disfluencies, we included repetition, repairs, alteration, filled and unfilled pauses.

2.2 Statistical models

The analysis was done using the lme4-package (Bates et al., 2015) in R Studio (Team, 2021). The data was fitted with mixed-effects logistic regression models to predict the interaction between code-switching and disfluencies.

For the first part of the analysis, we built a baseline model with code-switching as the dependent measure. We incrementally added 'turns with disfluencies' variable as the predictor along with 'previous turns with disfluencies' variable as an additional predictor. The speakers and the dialogue ID were included as random effects. In order to check the model fit, the models were compared using a log-likelihood ratio chi-square test. Based on the results of the chi-square distribution, predictors were added or removed in the best-fit model.

For a detailed analysis on the interaction between code-switching and disfluencies, we built models to test the effect of disfluencies on code-switching within and between turns.

3 Results

The model we built for the first analysis yielded a positive and significant effect of disfluencies on code-switching ($\beta = 0.38$, 95% CI [0.30, 0.45], p < .001). The best fitting model predicts that speakers code-switch within and between turns after encountering disfluencies in their turn and in the preceding turn. The model's total explanatory power is substantial (conditional R2 = 0.30) and the part related to the fixed effects alone (marginal R2) is of 4.13e-03. The results indicate that speakers code-switched in turns where disfluencies occurred.

The second set of models investigating the type of code-switching that is predicted by disfluencies yielded a significant effect on within turn switching ($\beta = 0.14$, 95% CI [0.02, 0.25], p = 0.022). While disfluencies in general have a significant effect on code-switching within a given turn, disfluencies in the previous turn also have a statistically significant effect on within turn switching. The models testing the effect of disfluencies on between turn switching did not yield a significant effect.

Predicted probabilities of code-switched turns
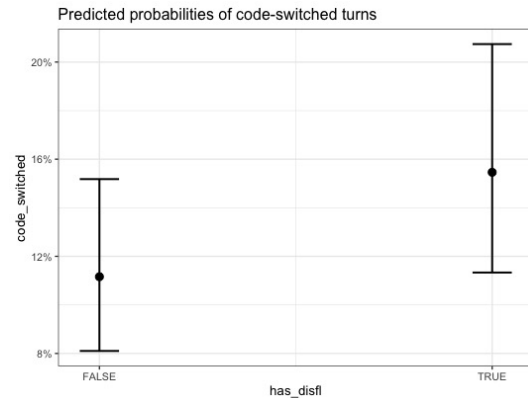
4 Discussion

This study is a point of departure for investigating code-switching as an interactive resource to facilitate grounding in bilingual dialogue. We analysed an available bilingual dialogue corpus (Deuchar, 2010) in order to gain some preliminary insights on the effect of disfluencies on code-switching in dialogue. It has to be noted that these results are specific to the context of Spanish-English bilinguals who are fluent in both languages. The results from the analysis indicate a strong relation between disfluencies and code-switching in bilingual dialogue. The results are similar to the findings from previous studies (Hlavac, 2011) on the frequency of pauses and repairs occurring alongside code-switching.

We are extending our analysis to further investigate the effect of each type of disfluency (repairs, pauses, etc.,) on each type of code-switching (within-turn and between-turn). Alternatively, code-switching could also trigger disfluencies since code-switching is cognitively demanding (Green and Abutalebi, 2013). The additional models we built to test this revealed that disfluencies can be predicted by code-switching. To what extent code-switching and disfluencies in bilingual dialogue affect each other can only be investigated further by testing this effect within a controlled experimental setting.

5 Conclusion

The results of this study have confirmed that there is a clear relationship between code-switching and disfluency in spontaneous bilingual interaction. We intend to investigate this further in dialogue-based experiments where we can closely control the effect of disfluencies and the interactive context. In our future work, we will additionally examine the effect of code-switching in interaction across language pairs.

References

Karl GD Bailey and Fernanda Ferreira. 2007. The processing of filled pause disfluencies in the visual world. In *Eye movements*, pages 487–502. Elsevier.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Anne L. Beatty-Martínez, Christian A. Navarro-Torres, and Paola E. Dussias. 2020. Codeswitching: A Bilingual Toolkit for Opportunistic Speech Planning. *Frontiers in Psychology*, 11.

Jakob Cromdal and Karin Aronsson. 2000. Footing in bilingual play. *Journal of Sociolinguistics*, 4(3):435–457.

Margaret Deuchar. 2010. BilingBank Spanish-English Bangor Miami Corpus.

David W Green and Jubin Abutalebi. 2013. Language control in bilinguals: The adaptive control hypothesis. *Journal of cognitive psychology*, 25(5):515–530.

Patrick Healey, Mary Lavelle, Christine Howes, Stuart Battersby, and Rosemarie McCabe. 2013. How listeners respond to speaker's troubles. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 35.

Patrick GT Healey, Arash Eshghi, Christine Howes, and Matthew Purver. 2011. Making a contribution: Processing clarification requests in dialogue. In *Proceedings of the 21st Annual Meeting of the Society for Text and Discourse*, pages 11–13. Citeseer.

Patrick GT Healey, Gregory J Mills, Arash Eshghi, and Christine Howes. 2018. Running repairs: Coordinating meaning in dialogue. *Topics in cognitive science*, 10(2):367–388.

Jim Hlavac. 2011. Hesitation and monitoring phenomena in bilingual speech: A consequence of code-switching or a strategy to facilitate its incorporation? *Journal of Pragmatics*, 43(15):3793–3806.

Gerrit Jan Kootstra, Ton Dijkstra, and Janet G Van Hell. 2020. Interactive alignment and lexical triggering of code-switching in bilingual dialogue. *Frontiers in psychology*, page 1747.

Brian MacWhinney. 2022. The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. *Applied Psycholinguistics*.

RStudio Team. 2021. Rstudio: integrated development for r. rstudio, pbc, boston, ma. 2020.

Li Wei and Lesley Milroy. 1995. Conversational code-switching in a chinese community in britain: A sequential analysis. *Journal of Pragmatics*, 23(3):281–299.