INVESTIGATING CODE-SWITCHING AND DISFLUENCIES IN BILINGUAL DIALOGUE

Fahima Ayub Khan and **Bill Noble**

Centre for Linguistic Theory and Studies in Probability (CLASP)
Department of Philosophy, Linguistics and Theory of Science

UNIVERSITY OF GOTHENBURG

disfluencies are a ubiquitous feature of dialogue, and have a role to play in conversational grounding.

Forward looking — the moment of interruption is followed not by an alteration, but just by a completion of the utterance

Backward looking — the moment of interruption is followed by an alteration that refers back to an already uttered reparandum. [Ginzburg et al., 2014]

FORWARD LOOKING (FLD)	(.)	**unfilled pause —** a pause in speech within an utterance
	um er	**filled pause —** filler words that hold the floor before the same utterance continues
	[/]	**repetition —** repetition of the same content one or more times
BACKWARD LOOKING (BLD)	[//]	**retracing —** repetition of the same basic phrase, changes the syntax but mantains the same idea
	[///]	**reformulation —** a full and complete restating of the message without any specific corrections

[MacWhinney, 2022]

code-switching is the use of two or more languages in the same utterance and/or dialogue.

between utterances — the current utterance begins in a different language from the language in which the previous utterance ended

within utterance — the current utterance uses multiple languages

three observations:

1. Disfluency plays a role in facilitating communicative alignment and coordinating interaction. [Hlavac, 2011]
2. Code-switching also facilitates conversational alignment. [Wei & Milroy 1995]
3. Code-switches tend to co-occur with disfluency. [Beatty-Martínez et al., 2020]

research question: Is code-switching a response to negative evidence of grounding?

the dataset

is extracted from the Bangor Miami (Spanish-English) corpus contains 41 dialogues, 40 841 turns, and 254 739 tokens. [Deuchar, 2010]

utterances with **code-switching:**

- 12.2% — switch between utterances
- 5.4% — switch within utterance
- 16.0% — total utterances with switches

utterances with **disfluencies:**

- 14.2% — forward-looking disfluencies
- 4.5% — backward-looking disfluencies
- 16.8% — total utterances with disfluencies

MAD: I went
MAD: and I got **[///]** era hoy chickpeas **um [//]**
today it was
sopa de garbanzos
chickpea soup.
SAM: yeah
SAM: oh ok
SAM: yeah **[/]** yeah I know yeah la sopa de garbanzos
chickpea soup.
MAD: tú sabes er
you know.
MAD: uhuh
SAM: yeah
MAD: sopa de garbanzos
chickpea soup.
MAD: so I had that
MAD: but it was hot as heck so I had to wait
MAD: and I ordered café con leche **(..)** with no sugar
a latté
MAD: entonces **(.)** I kept putting off taking aspirins
then
MAD: because I was like
MAD: you know I m gonna **[/]** I m gonna see
SAM: yeah I ll do it
SAM: and que le tuvieron **[/]** que le tu vieron s spa
they had to, they had to

exploratory analysis

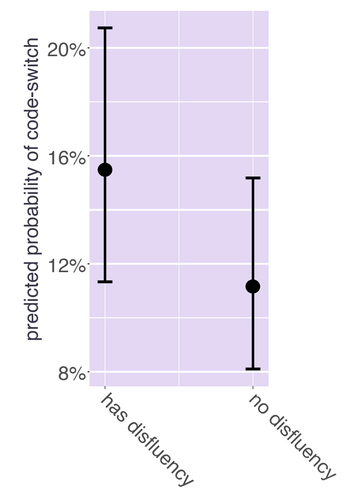
The data was fitted with mixed-effects logistic regression models to predict the effect of code-switching and disfluencies on each other in dialogue.

Speaker and dialogue IDs were included as random effects. Models were compared using a log-likelihood ratio chi-square test to test model fit.

M1

What is the effect of disfluency on whether or not speakers code-switch?

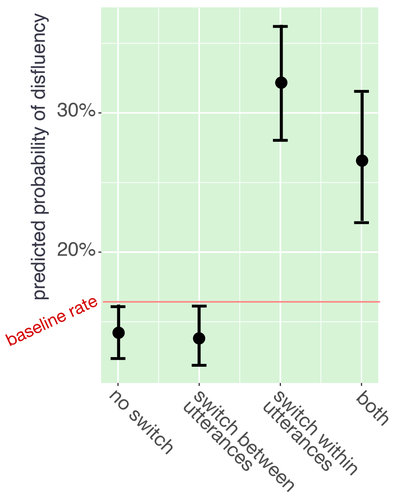
predictor: disfluencies (T/F)
dependant variable: code-switching (T/F)
explanatory power: substantial (conditional R2 = 0.30)
conclusion: Speakers code-switched more in turns where disfluencies occured.
beta = 0.38, 95% CI [0.30, 0.45], p < .001; Std. beta = 0.38, 95% CI [0.30, 0.45]



M2

What is the effect of each switch type on disfluencies in each turn?

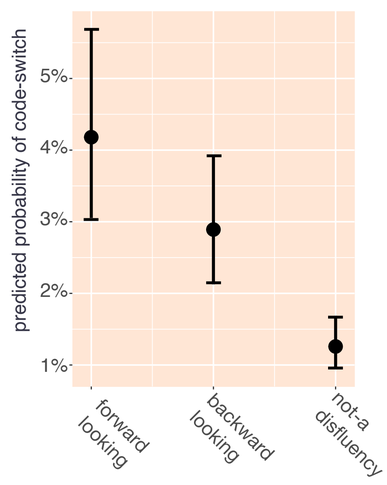
predictor: code-switching type
dependant variable: disfluencies (T/F)
explanatory power: weak (conditional R2 = 0.11)
conclusion: The effect of within-turn switching on disfluencies is statistically significant and positive.
beta = 0.27, 95% CI [0.06, 0.47], p = 0.013; Std. beta = 0.27, 95% CI [0.06, 0.47]



M3

What is the effect of disfluency type on code-switching in the next token in the same turn?

predictor: disfluency type
dependant variable: code-switch (T/F)
explanatory power: substantial (conditional R2 = 0.28)
conclusion: Speakers are more likely to code-switch in the next token after encountering BLD in the previous token in the same turn.
beta = -3.24 (95% CI [-3.61, -2.86], p < .001)



references

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

Deuchar, Margaret. "BilingBank Spanish-English Bangor Miami Corpus." TalkBank, 2010.

Ginzburg, Jonathan, Raquel Fernández, and David Schlangen. "Disfluencies as Intra-Utterance Dialogue Moves." Semantics and Pragmatics 7, no. 0 (June 10, 2014): 9–64.

Hlavac, Jim. "Hesitation and Monitoring Phenomena in Bilingual Speech: A Consequence of Code-Switching or a Strategy to Facilitate Its Incorporation?" *Journal of Pragmatics* 43, no. 15 (December 2011): 3793–3806.

MacWhinney, Brian. "The CHILDES Project: Tools for Analyzing Talk. 3rd Edition.," May 8, 2022.

Wei, Li, and Lesley Milroy. "Conversational Code-Switching in a Chinese Community in Britain: A Sequential Analysis." Journal of Pragmatics 23, no. 3 (March 1995): 281–99.

CLASP centre for linguistic theory and studies in probability