What Code-Switching Strategies are Effective in Dialogue Systems?

Emily Ahn[1*] **Cecilia Jimenez**[2] **Yulia Tsvetkov**[3] **Alan W Black**[3]
[1]University of Washington [2]University of Pittsburgh [3]Carnegie Mellon University
eahn@uw.edu, CCJ16@pitt.edu, {ytsvetko,awb}@cs.cmu.edu

Abstract

Since most people in the world today are multilingual (Grosjean and Li, 2013), code-switching is ubiquitous in spoken and written interactions. Paving the way for future adaptive, multilingual conversational agents, we incorporate linguistically-motivated strategies of code-switching into a rule-based goal-oriented dialogue system. We collect and release COMMONAMIGOS, a corpus of 587 human–computer text conversations between our dialogue system and human users in mixed Spanish and English. From this new corpus, we analyze the amount of elicited code-switching, preferred patterns of user code-switching, and the impact of user demographics on code-switching. Based on these exploratory findings, we give recommendations for future effective code-switching dialogue systems, highlighting user's language proficiency and gender as critical considerations.[1]

Figure 1: We build a bilingual goal-oriented agent that can converse in Spanish–English code-switching with human users. In controlled settings, we collect human–computer conversations that enable us to develop effective CS strategies for future dialogue systems.

1 Introduction

Humans seamlessly adjust their communication to their interlocutors (Gallois and Giles, 2015; Bell, 1984). We adapt our language, communication style, tone and gestures; when we share more than one language with our interlocutor, we inevitably resort to multilingual production or *code-switching*—shifting from one language to another within an utterance (Sankoff and Poplack, 1981).

We envision naturalistic conversational agents that communicate fluently and multilingually as humans do. However, existing dialogue systems are agnostic to the user, generating monolingual sentences which overfit to the language, domain, and style of their training data. To enable user-centric multilingual conversational agents, dialogue systems need to be extended to accommodate and converse with bilinguals, potentially using multiple languages in an utterance, as shown in Figure 1.

Before the rise of social media, code-switching (henceforth, CS) was primarily a spoken phenomenon, and it has been studied in spoken conversations (Lyu et al., 2010; Li and Fung, 2014; Deuchar et al., 2014). However, the spoken language domain is not directly comparable to the written one, and its spontaneous settings make it difficult to conduct controlled experiments to study accommodation in CS of one speaker to another. In controlled settings, CS has been extensively studied in psycholinguistics (Kootstra, 2012), but these are typically carefully designed experiments with few participants, which are hard to apply in large-scale data-driven scenarios like ours. In the written domain, which is the focus

*This work was done while the first author was a student at Carnegie Mellon University.

[1]This study was approved by the IRB. All code and collected data are available at https://github.com/emilyahn/commonamigos.

of our work, CS has been studied in broadcast texts such as social media (e.g. Reddit and Twitter) posts (Rabinovich et al., 2019; Aguilar et al., 2018) at the level of a single sentence and not contextualized in a dialogue.

Strikingly, little is known about human choices in written code-switching in conversations beyond the context of an individual utterance. In this paper, we introduce a novel framework which will allow us to fill this gap and study CS patterns contextualized in written conversations. Our focus languages are Spanish and English; these are often code-switched by people in Hispanic communities, who make up roughly 18% of the total US population (US Census Bureau, 2017).

We first introduce our bilingual goal-oriented dialogue system—an extension of a monolingual approach of He et al. (2017)—which controllably incorporates CS (§2). Then, we define our focus CS strategies, grounded theoretically and empirically (§3). In §4, we describe the experimental methodology and deployment of the dialogue system on crowdsourcing platforms. After collecting multilingual dialogues, we analyze patterns of CS along several axes such as the amount of CS, user accommodation (or *entrainment*) to dialogue systems that use different patterns of CS, and preferred CS patterns across user demographics (§5). Following the analysis, we provide additional background (§6) before concluding with areas for future work (§7).

Our three main contributions are (1) formulating a new task and framework of incorporating code-switching into a bilingual collaborative dialogue system. This framework has enabled us to apply and validate prior linguistic theories about CS. We show that it is useful to analyze CS along different strategies, as was suggested by Bullock et al. (2018), and we implement novel metrics to compute and generate these strategies. Our next contribution (2) is a publicly available corpus, CommonAmigos, of 587 code-switched Spanish–English human–computer text dialogues and surveys, useful for further development of multilingual dialogue systems and for explorations of sociolinguistic factors of accommodation in CS (cf. Danescu-Niculescu-Mizil et al., 2011). Finally, (3) our exploratory analyses of CS patterns in this corpus serve as a crucial first step to enable naturalistic bilingual dialogue systems in the future.

2 Bilingual Collaborative Human–Computer Dialogue System

Our ultimate goal is to study human preferences in written code-switching, and to integrate this knowledge into bilingual, adaptive dialogue systems. To gain insights into human CS patterns and to enable such systems, however, we first need to collect examples of multilingual human–computer dialogues, a resource that does not yet exist.

To collect human–computer dialogues in a controlled manner, we (1) modify an existing goal-oriented dialogue framework to code-switch; (2) create multiple instances of code-switching dialogue systems, where each instance follows one pre-defined strategy of CS as described in §3; and (3) analyze collected dialogues and study how people communicate differently with dialogue agents following a particular strategy.

We begin by modifying an existing goal-oriented collaborative dialogue framework (He et al., 2017). The framework implements a scenario of discussing mutual friends given a knowledge base, private to each interlocutor. Each of the interlocutors has a list of friends with attributes such as hobby and major. Only one friend is the same across both lists, and the goal is to find that mutual friend via collaborative discussion over text chat.

We extend this framework to a bilingual Spanish–English goal-oriented collaborative dialogue. In our bilingual interface, users see the private table of friends and attributes in both Spanish and English.

To code-switch in language generation, we add modifications (visualized in green in Figure 2) to the original monolingual generation (in blue). The rule-based agent generates English strings, which are passed to an Automatic Machine Translation (MT) system[2] in order to receive the Spanish translations. With parallel English and Spanish utterances, we define rules and templates to output a bilingual utterance following one of the CS strategies described in §3 for the full duration of the chat (see examples in Table 1).

To process text from the users, utterances are first passed to the MT whose target language is English. The monolingual dialogue system receives English strings and parses utterances into basic entities, and this informs the next turn from the dia-

[2] We use Google Translate API, a state-of-the-art MT that produced reliable translations.

Strategy		Example Sentence	Miami	Twitter
Monolingual	*EN*	Do you have any friend who studies linguistics?	–	–
	SP	*¿Tienes algún amigo que estudie lingüística?*	–	–
Insertional	$SP \xrightarrow{ins} EN$	Do you have any *amigo* who studies *lingüística*?	9.0%	5.5%
	$EN \xrightarrow{ins} SP$	*¿Tienes algún* friend *que estudie* linguistics?	25.7%	30.1%
Alternational	$EN \xrightarrow{alt} SP$	Do you have any friend *que estudie lingüística*?	12.2%	12.0%
	$SP \xrightarrow{alt} EN$	*Tienes algún amigo* that studies linguistics?	15.7%	10.5%
Informal	$+\ EN \xrightarrow{ins} SP$	hey *tienes algún* friend *que estudie* linguistics?	–	–
	$+\ SP \xrightarrow{alt} EN$	*pues tienes algún amigo* that studies linguistics?	–	–
Neither	–	*pero* she is the case manager for those patients	37.5%	41.9%

Table 1: We show transformations of the same example sentence (references first given monolingually) in each CS strategy, as would be generated by our dialogue system. The example for Neither is from the Miami corpus and is not an utterance we generate. Note that the Informal setting can be added to either Insertional or Alternational strategies, so 2 of the possible 4 informal settings are given in this set. We also verify that our two main strategies have a presence in existing corpora (Miami and Twitter).

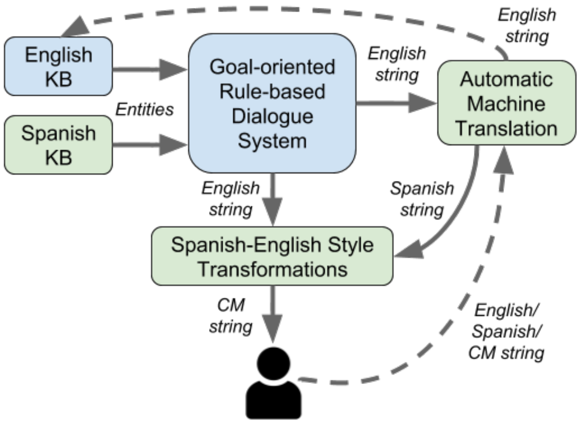

Figure 2: We add bilingual adaptations (in green) to the existing monolingual rule-based generation (in blue). The main dialogue system generates code-switched text via MT and a set of linguistically-informed code-switching rules. It receives the user's (code-switched) text after it was translated into English.

logue agent.

3 Code-Switching Strategies

We explore a variety of code-switching strategies and integrate these in our bilingual dialogue systems; each system follows one pre-defined strategy throughout the whole conversation. In this section, we describe the strategies we use, and how we operationalize them to detect and generate varied CS utterances in our dialogue system. We also verify the prevalence of these strategies in Spanish–English corpora in related domains: the Miami corpus of transcribed spontaneous speech (Deuchar et al., 2014), and a Twitter corpus (Molina et al., 2016). Examples of an utterance in each strategy along with the distribution of these strategies in both Twitter and Miami corpora are given in Table 1.

We follow Muysken's (2000) approach. The first strategy from Muysken (2000) is **Insertional code-switching**, which follows the Myers-Scotton framework of a Matrix Language (MatL) and an Embedded Language (EmbL). The structure and grammar of the MatL is maintained while inserting the EmbL (often single words or phrases) in certain spots (Myers-Scotton, 1993). According to Joshi (1982), closed class items such as determiners, quantifiers, etc., would remain in the MatL. This has also been shown to be more commonly used when the speakers are not equally proficient in both languages (Deuchar et al., 2007).

We experiment with two conditions: (1) retaining the grammar of English while inserting Spanish nouns ($SP \xrightarrow{ins} EN$), and (2) using Spanish grammar while inserting English nouns ($EN \xrightarrow{ins} SP$).

Next, we experiment with **Alternational code-switching**, when the two languages remain more separate and alternate after clauses. Switch-points adhere to constituent boundaries (Sankoff and Poplack, 1981) and can separate topics or sentences (Ardila, 2005). This has been shown to be more prevalent among fluent or highly proficient bilinguals as a form of more stable bilingualism (Deuchar et al., 2007).

We again experiment with two conditions, either beginning in English for a phrase and then switching to Spanish ($EN \xrightarrow{alt} SP$), or beginning in Spanish and then switching to English ($SP \xrightarrow{alt} EN$).

Since people may code-switch more often in informal, casual settings or when there is higher rapport, we experiment with the above four CS strategies with our agent speaking either informally or formally. We modulate formality by adding discourse markers. Discourse markers are known to be actively used by speakers in improving the flow of dialogue, and they remain relatively independent of syntax or semantics (Schiffrin, 1988). Within CS speech, these markers can be adopted as an easy form of lexical borrowing by bilinguals of varying proficiency. In particular, Spanish markers within English speech can be used to signify a less formal tone or to reveal Latino social identity (Torres, 2011). Therefore we define our agent's informal setting (*+Informal*) to have discourse markers added to either Insertional CS or Alternational CS utterances.

3.1 Detecting Insertional and Alternational Code-Switching

The two strategies can be manually detected by linguists, but there has not been a direct attempt to automatically label CS utterances as Insertional or Alternational.[3] We therefore introduce a novel method to computationally classify CS utterances into $EN \xrightarrow{alt} SP$, $SP \xrightarrow{alt} EN$, $SP \xrightarrow{ins} EN$, $EN \xrightarrow{ins} SP$, or *Neither*.[4]

An utterance is Alternational when it switches from Lang_A to Lang_B under 2 conditions: (1) there is a contiguous span of 2+ words in Lang_A followed by a contiguous span of 2+ words in Lang_B, and (2) there is at least 1 finite (i.e. conjugated) verb form or auxiliary word in each language.[5]

If the utterance is not first classified as Alternational, it is next tested for Insertional. We define Insertional CS to occur under 3 conditions: (1) the MatL has at least 1 function word or finite verb, (2) the EmbL has at least one content word (either a noun or an adjective), and (3) the MatL has more tokens than the EmbL. This metric ensures maintaining the grammar of the MatL with insertions of the EmbL.

We test our implementation of this metric on a gold set of 150 CS utterances (50 each from Miami, Twitter, and COMMONAMIGOS datasets) annotated for strategy jointly by two linguists proficient in both Spanish and English. A third linguist achieves a Cohen's κ of 0.75 (substantial agreement) or an F1 of 0.8 against the adjudicated gold set. Our implementation receives an F1 of 0.76 on the same gold set.

To verify the coverage of these types of CS, we analyze their prevalence in the Miami and Twitter corpora, with distributions given in Table 1. We observe that the most commonly used strategy is Insertional CS, specifically $EN \xrightarrow{ins} SP$, which mirrors findings from a Spanish–English corpus of blogs from Montes-Alcalá (2007).

4 Data Collection

In order to examine effects of different CS strategies with human bilingual speakers, we modify an existing dialogue system (§2) and deploy it to chat with online crowdworkers.

4.1 Crowdsourcing

We release this task on two crowdsourcing platforms: Amazon Mechanical Turk and Figure Eight.[6] In order to target Spanish–English bilinguals, we limit workers to be in the US,[7] and then include several ungraded Spanish proficiency test questions.[8]

Additionally, the introduction and instructions to the task are purely written in Spanish to prime the user in both languages, given that English is usually the default language for tasks released in the US. For each chat, there are always 10 friends with 3 attributes each (randomly selected with varying complexity). Users have up to 8 minutes to complete the task. Besides the 8 CS conditions, we have 2 more monolingual conditions (Spanish and English), as well as a Random CS condition where a switch point could occur with 50% chance at every smallest word unit.

[3] Bullock et al. (2018) gathered metrics to identify those two strategies across an entire corpus but not across a single utterance.

[4] This method has been refined after several iterations of discussions with linguists and examining the implementation's coverage over annotations.

[5] Detecting verbs and auxiliaries was made possible by generating English and Spanish POS tags from Spacy, available at https://spacy.io/.

[6] https://www.mturk.com; https://www.figure-eight.com.

[7] Other countries were not included in order to limit the variance of cultural factors for Spanish–English CS.

[8] 92% of all users scored 67%+ accuracy on 3 questions.

# Dialogues	587
% Extrinsic task success	64%
Avg # user utterances	7.9
Avg # tokens / utterance	6.2
EN vocab size	571
SP vocab size	846
% EN utterances	16%
% SP utterances	44%
% CS utterances	39%
% dialogues w/ CS	70%

Table 2: COMMONAMIGOS, our bilingual corpus of crowdsourced chats, has a strong presence of CS.

4.2 Collected Dialogues

We report general statistics of our collected dialogues in Table 2.

A total of 737 dialogues are collected, but 587 remain for analysis after removing chats with missing text or surveys from users. From the pool of 587 valid chats, there are 296 unique workers because some did more than one task. The self-reported survey reveals that the mean age of the workers is 31, 60% of them are male, and the most frequently reported countries of origin are USA, Venezuela, and Mexico.

Examples of conversations gathered with crowdsourced bilinguals are given in Table 3. An interesting observation is that the user chooses to emulate the strategy instead of echoing that lexical item in the $SP\xrightarrow{alt}EN$ Alternational condition. Even when the agent uses the Spanish word *contabilidad*, the user says the equivalent meaning in English, which is *accounting*. Similarly, when the $SP\xrightarrow{alt}EN$ agent discusses *dancing*, the user replies with the Spanish equivalent, *bailar*, thus prioritizing strategy over lexicon.

5 Analysis

We examine the subtleties of how users code-switched under different conditions, and share our main findings below. The questions we now explore are how much do the users code-switch, how do they do it, and how do agent strategies factor into response style?

5.1 Our bilingual dialogue system elicits code-switching

Our first encouraging finding is that a high majority of dialogues contain CS from the user (Table 2), although the users were not explicitly required to code-switch. This implies that CS is a prevalent communication style and that conversational agents could benefit from supporting multilinguality.

We first analyze the amount or presence of CS from the users. Guzmán et al. (2017) defined several metrics based on quantifying token counts and span lengths of continuous monolingual tokens. The Multilingual-index (**M-idx**) reflects how balanced the tokens are in each language, where 0 is fully monolingual and 1 is an equal number of tokens per language. The Integration-index (**I-idx**) is the probability of switching languages between any two tokens, where 0 is fully monolingual and 1 is a perfectly interleaved corpus, with a switch at every word.[9] Higher values of both indices imply a higher quantity of CS.

Table 4 shows that $SP\xrightarrow{ins}EN$ +*Informal* and Alternational conditions result in higher M-indices than average. Most notably, the $EN\xrightarrow{ins}SP$ condition results in the lowest M-idx and I-idx from users. We reason that this is due to receiving more monolingual Spanish text from users than in any other condition, a potential result of having the crowdworkers primed to be in Spanish mode. Conversely, the $SP\xrightarrow{ins}EN$ conditions maintain markedly high CS indices from users. $SP\xrightarrow{ins}EN$, the agent with the highest number of English tokens, could have encouraged users to balance their Spanish tokens with more English. We advise future CS systems to be aware of their target audience's assumptions of the agent's default language.

The added formality setting has a number of effects on the two main strategies. Across all 4 Insertional and Alternational conditions, +*Informal* reduces the average number of tokens in a user's reply (seen in Table 4), which could be a result of users being more casual with the dialogue system. M-idx increases for both Insertional strategies while sharply decreasing for both Alternational strategies. I-idx slightly increases for all strategies except $SP\xrightarrow{ins}EN$. We can recommend that if the goal of a future CS dialogue system is to be efficient in number of turns, the $SP\xrightarrow{ins}EN$ strategy is useful, but if the goal is to keep the user engaged and chat for longer, $SP\xrightarrow{ins}EN$ +*Informal*

[9]To calculate I-idx in a given dialogue, all utterances by one party are concatenated in order, so switch-points can occur across utterance boundaries.

$EN \xrightarrow{alt} SP$	$EN \xrightarrow{ins} SP$
A: I have 2 friends *que estudiaron la contabilidad* [that studied accounting] H: *yo tambien* [me too]. one that studies accounting *trabaja en el concesionario de coches y el otro en la oficina* [works at the car dealership and the other in the office] A: Do you have any friend who likes dancing *o amigos a los que les guste hornear* [or friends who like to bake]? H: *nadie le gusta bailar* [no one likes to dance]. one likes baking–*el/ella estudia fisica* [he/she studies physics]	A: *¿Tiene* [Do you have] friends *que trabajen en el* [who work at the] theater *o un* [or a] friend *que trabaje en la* [that works at the] jewelry store ? H: *si. la del* [yes. the one from] jewelry store *le gusta dormir* [likes to sleep] A: *tengo* [I have] 1 friend *que le gusta* [who likes] acting, 1 friend *que trabaja en el* [who works at the] zoo H: *la del teatro le gusta* [the one from the theater likes] photography
$SP \xrightarrow{alt} EN$ +Informal	**$SP \xrightarrow{ins} EN$ +Informal**
A: *tengo un amigo* [I have a friend] who studied english.. *y tú* [and you]? H: *no tengo... solo tengo un amigo que estudio* [I don't have... I only have a friend that studied] linguistics A: hey *tengo dos amigos* [I have two friends] who like sewing H: *yo tengo un amigo que le gusta* [I have a friend that likes] sewing!	A: do you have any *amigos* [friends] who studied *derecho* [law] ? H: no i don't H: *tienes un amigo a quien le gusta cocinar* [do you have a friend who likes to cook]? A: nah i have no *amigo* [friend] who likes *cocinar* [to cook]..

Table 3: These examples from our corpus of human (H) interactions with rule-based CS agents (A) show a diversity of CS strategies, given the static agent strategy in **bold**.

or $SP \xrightarrow{alt} EN$ *+Informal* could yield more turns. We encourage CS dialogue systems to consider implementing casual styles of speech in CS, as our simple additions of discourse markers produced patterned changes in token length and amount of CS.

5.2 Agent strategy can affect user strategy

We see the presence of entrainment between agent strategy (condition) and user strategy. In the matrix in Figure 3, perfect entrainment (where all the users' CS utterances use the same fixed agent strategy) would be shown with a normalized value of 1.0 along the diagonal. We compare values across CS conditions (without examining *+Informal* for now) to the random baseline, which ideally reveals the natural unconditioned distribution of user strategy.[10] Because the values on the diagonal are significantly greater than in the random condition ($p < .05$), we conclude that the agent's strategy had influence on the user's code-switching.

[10]Reassuringly, the percentages in this random condition are similar to the distribution of the Miami and Twitter corpora from Table 1.

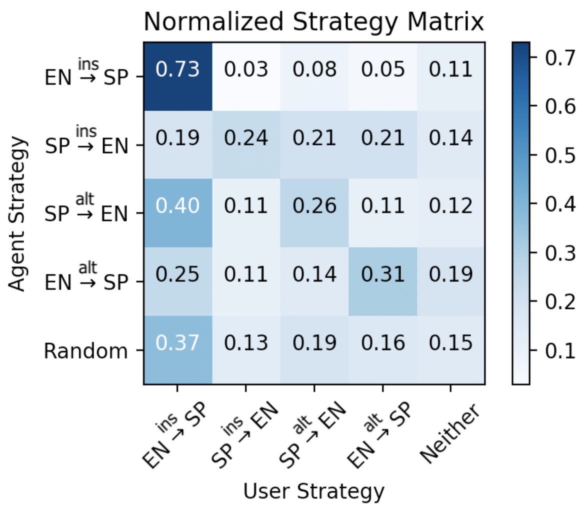

Figure 3: We find entrainment in our data. Given each agent strategy condition (per row), we display the normalized distribution of which strategies the users used (only accounting for utterances that are code-switched). Darker colors along the major diagonal indicate complete entrainment, and the random agent strategy at the bottom is shown for comparison.

For conditions where English is the main (or

Agent	# Dial	% Success	Avg Utts	Avg Tok	% CS Dial	% CS Utts	M-idx	I-idx
Average	53.4	64	7.9	6.2	70	39	0.74	0.23
Std Dev	(7.8)	(11)	(0.9)	(0.4)	(8)	(8)	(0.20)	(0.04)
$EN \xrightarrow{ins} SP$	**70**	**47**	8.4	6.3	74	42	**0.51**	0.23
+Informal	**44**	**77**	7.4	**5.7**	**80**	44	0.57	0.26
$SP \xrightarrow{ins} EN$	58	62	7.2	**6.9**	74	**52**	0.93	0.26
+ Informal	**44**	64	8.6	6.0	75	37	**0.99**	0.26
$SP \xrightarrow{alt} EN$	54	74	7.5	6.4	76	39	0.88	0.24
+Informal	56	**45**	**9.7**	6.1	75	40	0.71	0.26
$EN \xrightarrow{alt} SP$	55	**76**	7.9	6.3	71	40	0.91	0.23
+Informal	47	64	7.7	6.1	72	37	0.70	0.23
Mono SP	46	72	7.2	6.1	**57**	**26**	**0.37**	**0.16**
Mono EN	54	69	**6.4**	6.5	**54**	**25**	0.74	**0.16**
Random	59	64	8.2	**5.3**	66	39	0.86	0.22

Table 4: These general statistics show dialogue quantity, length, and extrinsic success of users, as well as user quantity of CS under different agent strategies. Values further than 1 standard deviation away from the mean are in **bold**.

starting) MatL, $EN \xrightarrow{ins} SP$ occurs less often, while other English-based CS strategies are used more often. There is also more sensitivity to the specific English strategy because more utterances are classified as $SP \xrightarrow{ins} EN$ in $SP \xrightarrow{ins} EN$ conditions and $EN \xrightarrow{alt} SP$ in $EN \xrightarrow{alt} SP$ conditions. Overall, $EN \xrightarrow{ins} SP$ is the most popular strategy used—it is most common in the $EN \xrightarrow{ins} SP$ condition, but it still keeps a strong presence in other conditions. We recommend $EN \xrightarrow{ins} SP$ to be a good default strategy in future CS agents, as that also follows the prevalent styles in the Miami and Twitter corpora (§3.1).

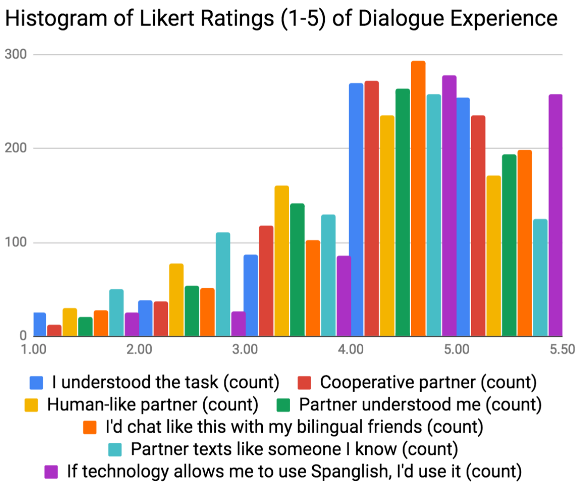

Figure 4: As an aggregate, users have generally positive experiences with our CS agent. They would rate their agreement with statements given in the legend, where 1 = Strongly Disagree, and 5 = Strongly Agree.

5.3 Users succeed in their dialogues

We define two types of success in the dialogues: (1) Extrinsic success (the binary task of finding the mutual friend in 8 minutes), and (2) User experience (self-reported measures on an agreement scale of 1-5, e.g. *"I understood the task perfectly"*, or *"My task partner texts like someone I know"*).

From Table 4, all Alternational and monolingual conditions achieve consistently high rates of extrinsic task success. This could reveal that longer spans of monolingual tokens aid in users comprehending the task, so we recommend CS systems to adhere to Alternational strategies if they desire specific goals to be achieved. As for user experience, Figure 4 displays users generally agreeing with statements such as *"I'd chat like this with my bilingual friends"*. Full exploration of variables affecting these ratings can be done with our COMMONAMIGOS corpus. Regarding entrainment, we do not find significant correlations with any type of success metric.

5.4 User demographics affect CS

Beyond analysis of the aggregate data, we find strong effects of the following user attributes.

Language Proficiency Our findings support the hypothesis from Deuchar et al. (2007) in that more proficient bilinguals (balanced in both languages) use Alternational strategies more often than asymmetrical bilinguals. We examine this by binning the groups into three categories from the self-reported language ability metric: highly proficient in both English and Spanish, dominant English only, and dominant Spanish only.[11] Compared to the aggregate report of user CS, dominant English speakers use $SP \xrightarrow{ins} EN$ more heavily, while dominant Spanish speakers use $EN \xrightarrow{ins} SP$ more heavily. Alternational CS occurs in those two groups but is more present in the balanced bilingual group.

For the dominant English speakers, a higher M-idx correlates with better agreement on statements such as *"My task partner was very cooperative"*. When these users entrain more to the agent's CS strategy, the number of turns in the dialogue also increases. Also, even though their extrinsic task success is low in the monolingual Spanish condition, almost all CS conditions boosted task success. Together, these findings show that the dialogue experience overall improves for less-balanced bilinguals when the agent uses CS instead of their weaker monolingual language. This supports a line of pedagogy that advocates incorporation of CS in second language instruction (cf. Moore, 2002).

Gender Reported gender[12] yields strong correlations in user CS strategy. When females chat with higher M-idx and I-idx values, they agree more with the statement *"I am very likely to chat like I did in this task when messaging with my bilingual friends"*. Under informal conditions, females also have longer dialogues, a higher percentage of CS utterances, and a higher percentage of dialogues containing any CS—all of which prove to be an opposite effect for males. These findings reflect that females may code-switch more naturally and will respond better to more informal CS dialogue systems.

6 Related Work

We provide a brief overview of previous works in the domains of CS and dialogue.

Most closely related to ours is the work of Ramanarayanan and Suendermann-Oeft (2017) who introduced a chatbot that spoke from a fixed set of Spanish–English and Hindi–English machine prompts to encourage human bilinguals to code-switch back to the agent. Our work takes this interaction further and does not assume a restricted set of sentences. Rather, we control one side of the spontaneous dialogue based on different CS strategies in order to learn human preferences when code-switching.

Sitaram et al. (2019) have surveyed attempts to integrate CS into NLP and Speech processing domains. These domains include Part-of-Speech tagging (Solorio and Liu, 2008; Soto and Hirschberg, 2018), Language Identification (Ramanarayanan and Pugh, 2018; Rijhwani et al., 2017), Named Entity Recognition (Aguilar et al., 2018), Language Modeling (Chandu et al., 2018b), Automatic Speech Recognition (ASR) (Yilmaz et al., 2018), and Speech Synthesis (Rallabandi and Black, 2017). There also has been a push to generate CS datasets synthetically to improve CS language modeling (Pratapa et al., 2018), or manually crowdsource CS utterances towards CS Question–Answering and dialogue systems (Chandu et al., 2018a; Banerjee et al., 2018).

Various other research has centered around understanding when and why people code-switch. Linguistically-driven methods have found that cognates and acoustic cues allow for more fluid switching between the languages (Kootstra et al., 2012; Fricke et al., 2016).

When pertaining to a dialogue setting, CS has been found to fulfill different goals of speakers (Begum et al., 2016). Solorio and Liu (2008) discussed how sociopragmatic factors, such as the topic being discussed and the rapport between the speakers, could influence the style of CS. Additionally, choosing to use one language over another can be a pragmatic way to mark sentiment, as Rudra et al. (2016) found in Hindi–English Twitter data. These findings support our aim of understanding CS in nuanced contexts of dialogue.

In dialogue generally, entrainment between conversational partners has been shown to improve task success and perceived naturalness (Reitter and Moore, 2014; Nenkova et al., 2008). In bilingual settings, accommodation has been recorded since Giles et al. (1973), where French–English speakers would choose their language according to their audience. More recently in entrainment of CS, Soto et al. (2018) showed a con-

[11] This is the strongest among various weak signals indicating language proficiency, namely the Spanish proficiency quiz, reported age of acquisition for each language, country of origin, and frequency of language use.

[12] "Other" gender constitutes 1% of users and is set aside for this analysis.

vergence in the quantity of CS between speakers over the course of long conversations in the Miami data. Fricke and Kootstra (2016) also found that the presence of CS can affect the utterance following it. Our work is the first to identify entrainment of diverse CS strategies beyond language choice in Bawa et al. (2018).

7 Conclusion

Through our novel Spanish–English dialogue framework, we generate code-switching utterances to which bilingual users also respond in various forms of code-switching. We find that users sometimes adapt to the agent's code-switching, but their choice of CS strategy primarily depends on their bilingual language proficiency. Adding discourse markers to make the agent less formal also affects patterns of user CS among female participants. Finally, extrinsic task success is not significantly affected by CS strategy, though users indicated positive dialogue experiences.

There are numerous follow-up directions that can be taken with our framework and with the novel COMMONAMIGOS corpus. For example, analyses can be done on the types of switch points, investigating attributes such as simplicity or frequency of the word that is switched, the nature of it being a cognate (Soto et al., 2018), or even the cognitive accessibility of switch words from users' mental lexicons.

We acknowledge that COMMONAMIGOS reflects a specific population of users that would not represent all Spanish–English speakers across the world, and the crowdworker population may also be skewed in ways we cannot identify. Future work should consider other groups of Spanish–English speakers, as well as other language pairs such as Hindi–English or Tagalog–English, in order to learn how these varieties may be linguistically or functionally comparative to our findings.

The implications of our current work, which reveal which CS strategies are more entrainable than others, could help CS agents adapt to users and to better parse and predict user utterances with a more informed CS language model.[13] Future agents should incorporate different CS strategies dynamically within a single conversation that entrain to the user. In order to move beyond a rule-based agent, in future work we can leverage neural language generation systems (e.g., Park and Tsvetkov, 2019) trained on CS data. From here, we can usher in an era of bilingual dialogue systems that brings human–computer interactions to a more personalized space.

[13] This approach is similar to a method where ASR systems that lexically entrain users can lower ASR error rates (Levitan, 2013).

8 Acknowledgments

We acknowledge helpful input from the anonymous reviewers. We also thank Gayatri Bhat, Cindy Blanco, Anjalie Field, Melinda Fricke, Shirley Anugrah Hayati, He He, Sachin Kumar, Chan Young Park, Anat Prior, Sai Krishna Rallabandi, Shruti Rijhwani, Shuly Wintner, and Yiheng Zhou for fruitful discussions. Finally, we sincerely thank our annotator, Joshua Baumgarten. This work was supported by NSF grant IIS-1812327 and NSF GRFP grant DGE1745016.

References

Gustavo Aguilar, Fahad AlGhamdi, Victor Soto, Mona Diab, Julia Hirschberg, and Thamar Solorio. 2018. Named entity recognition on code-switched data: Overview of the CALCS 2018 shared task. In *Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching*, pages 138–147.

Alfredo Ardila. 2005. Spanglish: An Anglicized Spanish Dialect. *Hispanic Journal of Behavioral Sciences*, 27(1):60–81.

Suman Banerjee, Nikita Moghe, Siddhartha Arora, and Mitesh M Khapra. 2018. A dataset for building code-mixed goal oriented conversation systems. *arXiv preprint arXiv:1806.05997*.

Anshul Bawa, Monojit Choudhury, and Kalika Bali. 2018. Accommodation of conversational code-choice. In *Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching*, pages 82–91.

Rafiya Begum, Kalika Bali, Monojit Choudhury, Koustav Rudra, and Niloy Ganguly. 2016. Functions of code-switching in tweets: An annotation scheme and some initial experiments. *LREC. i*, pages 1644–1650.

Allan Bell. 1984. Language style as audience design. *Language in society*, 13(2):145–204.

Barbara E Bullock, Gualberto Guzmán, Jacqueline Serigos, and Almeida Jacqueline Toribio. 2018. Should code-switching models be asymmetric? *Proc. Interspeech 2018*, pages 2534–2538.

Khyathi Chandu, Ekaterina Loginova, Vishal Gupta, Josef van Genabith, Günter Neuman, Manoj Chinnakotla, Eric Nyberg, and Alan W Black. 2018a. Code-mixed question answering challenge: Crowdsourcing data and techniques. In *Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching*, pages 29–38.

Khyathi Chandu, Thomas Manzini, Sumeet Singh, and Alan W Black. 2018b. Language informed modeling of code-switched text. In *Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching*, pages 92–97.

Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. 2011. Mark my words!: linguistic style accommodation in social media. In *Proceedings of the 20th international conference on World wide web*, pages 745–754. ACM.

Margaret Deuchar, Peredur Davies, Jon Herring, M Carmen Parafita Couto, and Diana Carter. 2014. Building bilingual corpora. *Advances in the Study of Bilingualism*, pages 93–111.

Margaret Deuchar, Pieter Muysken, and Sung Lan Wang. 2007. Structured variation in codeswitching: Towards an empirically based typology of bilingual speech patterns. *International Journal of Bilingual Education and Bilingualism*, 10(3):298–340.

Melinda Fricke and Gerrit Jan Kootstra. 2016. Primed codeswitching in spontaneous bilingual dialogue. *Journal of Memory and Language*, 91:181–201.

Melinda Fricke, Judith F Kroll, and Paola E Dussias. 2016. Phonetic variation in bilingual speech: A lens for studying the production-comprehension link. *Journal of Memory and Language*, 89:110–137.

Cindy Gallois and Howard Giles. 2015. Communication accommodation theory. *The international encyclopedia of language and social interaction*, pages 1–18.

Howard Giles, Donald M Taylor, and Richard Bourhis. 1973. Towards a theory of interpersonal accommodation through language: Some canadian data. *Language in society*, 2(2):177–192.

François Grosjean and Ping Li. 2013. *The Psycholinguistics of Bilingualism*. Wiley-Blackwell.

Gualberto Guzmán, Joseph Ricard, Jacqueline Serigos, Barbara E Bullock, and Almeida Jacqueline Toribio. 2017. Metrics for modeling code-switching across corpora. In *Proc. Interspeech 2017*.

He He, Anusha Balakrishnan, Mihail Eric, and Percy Liang. 2017. Learning symmetric collaborative dialogue agents with dynamic knowledge graph embeddings. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1766–1776.

Aravind K Joshi. 1982. Processing of sentences with intra-sentential code-switching. *Proceedings of the 9th conference on Computational Linguistics*, 1:145–150.

Gerrit Jan Kootstra. 2012. *Code-switching in monologue and dialogue: Activation and alignment in bilingual language production*. [Sl: sn].

Gerrit Jan Kootstra, Janet G Van Hell, and Ton Dijkstra. 2012. Priming of code-switches in sentences: The role of lexical repetition, cognates, and language proficiency. *Bilingualism: Language and Cognition*, 15(4):797–819.

Rivka Levitan. 2013. Entrainment in spoken dialogue systems: Adopting, predicting and influencing user behavior. In *Proceedings of the 2013 NAACL HLT Student Research Workshop*, pages 84–90.

Ying Li and Pascale Fung. 2014. Language Modeling with Functional Head Constraint for Code Switching Speech Recognition. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 907–916.

Dau-Cheng Lyu, Tien Ping Tan, Engsiong Chng, and Haizhou Li. 2010. SEAME: a mandarin-english code-switching speech corpus in south-east asia. In *INTERSPEECH 2010*, pages 1986–1989.

Giovanni Molina, Nicolas Rey-Villamizar, Thamar Solorio, Fahad Al-Ghamdi, Mahmoud Gohneim, Mona Diab, and Abdelati Hawwari. 2016. Overview for the Second Shared Task on Language Identification in Code-Switched Data. *Proceedings of the Second Workshop on Computational Approaches to Code Switching*, pages 40–49.

Cecilia Montes-Alcalá. 2007. Blogging in Two Languages : Code-Switching in Bilingual Blogs. *Selected Proceedings of the Third Workshop on Spanish Sociolinguistics*, pages 162–170.

Danièle Moore. 2002. Code-switching and learning in the classroom. *International journal of bilingual education and bilingualism*, 5(5):279–293.

Pieter Muysken. 2000. Bilingual speech: a typology of code-mixing.

Carol Myers-Scotton. 1993. Common and uncommon ground: Social and structural factors in codeswitching. *Language in society*, 22(4):475–503.

Ani Nenkova, Agustin Gravano, and Julia Hirschberg. 2008. High frequency word entrainment in spoken dialogue. In *Proceedings of the 46th annual meeting of the association for computational linguistics on human language technologies: Short papers*, pages 169–172. Association for Computational Linguistics.

Chan Young Park and Yulia Tsvetkov. 2019. Learning to generate word- and phrase-embeddings for efficient phrase-based neural machine translation. In

Proceedings of the 3rd Workshop on Neural Machine Translation and Generation.

Adithya Pratapa, Gayatri Bhat, Monojit Choudhury, Sunayana Sitaram, Sandipan Dandapat, and Kalika Bali. 2018. Language modeling for code-mixing: The role of linguistic theory based synthetic data. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1543–1553.

Ella Rabinovich, Masih Sultani, and Suzanne Stevenson. 2019. Codeswitch-reddit: Exploration of written multilingual discourse in online discussion forums. In *Proc. of EMNLP*.

Saikrishna Rallabandi and Alan W Black. 2017. On Building Mixed Lingual Speech Synthesis Systems. In *Interspeech 2017*, pages 52–56.

Vikram Ramanarayanan and Robert Pugh. 2018. Automatic token and turn level language identification for code-switched text dialog: An analysis across language pairs and corpora. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 80–88.

Vikram Ramanarayanan and David Suendermann-Oeft. 2017. Jee haan, i'd like both, por favor: Elicitation of a code-switched corpus of hindi–english and spanish–english human–machine dialog. In *Proc. Interspeech 2017*, pages 47–51.

David Reitter and Johanna D Moore. 2014. Alignment and task success in spoken dialogue. *Journal of Memory and Language*, 76:29–46.

Shruti Rijhwani, Royal Sequiera, Monojit Choudhury, Kalika Bali, and Chandra Sekhar Maddila. 2017. Estimating Code-Switching on Twitter with a Novel Generalized Word-Level Language Detection Technique. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 1971–1982.

Koustav Rudra, Shruti Rijhwani, Rafiya Begum, Kalika Bali, Monojit Choudhury, and Niloy Ganguly. 2016. Understanding language preference for expression of opinion and sentiment: What do hindi-english speakers do on twitter? In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1131–1141.

David Sankoff and Shana Poplack. 1981. A formal grammar for code-switching. *Research on Language & Social Interaction*, 14(1):3–45.

Deborah Schiffrin. 1988. *Discourse markers*. 5. Cambridge University Press.

Sunayana Sitaram, Khyathi Raghavi Chandu, Sai Krishna Rallabandi, and Alan W Black. 2019. A survey of code-switched speech and language processing. *arXiv preprint arXiv:1904.00784*.

Thamar Solorio and Yang Liu. 2008. Learning to predict code-switching points. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, (October):973–981.

Victor Soto, Nishmar Cestero, and Julia Hirschberg. 2018. The role of cognate words, POS tags, and entrainment in code-switching. *Proc. Interspeech 2018*, pages 1938–1942.

Victor Soto and Julia Hirschberg. 2018. Joint part-of-speech and language ID tagging for code-switched data. In *Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching*, pages 1–10.

Lourdes Torres. 2011. Spanish in the United States: Bilingual Discourse Markers. In *The Handbook of Hispanic Sociolinguistics*.

US Census Bureau. 2017. B03003 - hispanic or latino origin. American Community Survey 1-Year Estimates.

E Yilmaz, A Biswas, E Van der Westhuizen, F De Wet, and T Niesler. 2018. Building a unified code-switching ASR system for south african languages. In *Proc. Interspeech 2018*.