Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues

Tanmay Parekh
Language Technologies Institute
Carnegie Mellon University
tparekh@cs.cmu.edu

Emily P. Ahn
Department of Linguistics
University of Washington
eahn@uw.edu

Yulia Tsvetkov
Language Technologies Institute
Carnegie Mellon University
ytsvetko@cs.cmu.edu

Alan W Black
Language Technologies Institute
Carnegie Mellon University
awb@cs.cmu.edu

Abstract

Code-switching is a ubiquitous phenomenon in multilingual communities. Natural language technologies that wish to communicate like humans must therefore adaptively incorporate code-switching techniques when they are deployed in multilingual settings. To this end, we propose a Hindi-English human–machine dialogue system that elicits code-switching conversations in a controlled setting. It uses different code-switching agent strategies to understand how users respond and accommodate to the agent's language choice. Through this system, we collect and release a new dataset COMMONDOST, comprising of 439 human–machine multilingual conversations. We adapt pre-defined metrics to discover linguistic accommodation from users to agents. Finally, we compare these dialogues with Spanish-English dialogues collected in a similar setting, and analyze the impact of linguistic and socio-cultural factors on code-switching patterns across the two language pairs.[1]

1 Introduction

When interlocutors share more than one language, they nearly inevitably engage in *code-switching (CS)*: shifting from one language to another (Sankoff and Poplack, 1981; Muysken, 2000; Auer, 2013). Since most people in the world today are multilingual (Grosjean and Li, 2013), CS is a ubiquitous phenomenon in multilingual communities. It goes beyond simple lexical borrowing to blending of languages at syntactic, grammatical and morphological levels (Sitaram et al., 2019). Code-switching has been studied in linguistics and sociolinguistics for decades (Poplack, 1980; Gumperz, 1982; Milroy et al., 1995; Auer, 2013; Gardner-Chloros and Weston, 2015) since it reveals various linguistic and socio-cultural behaviours (Heller, 1982). However, NLP studies of written CS are limited to social media texts, rather than natural conversation, and tend to focus on single sentences, rather than be contextualized in a dialogue (Rabinovich et al., 2019).

Figure 1: We present a bilingual dialogue system for human-machine conversations in Hindi-English (Red: Hindi, Blue: English). We discover that humans positively adopt the agent's code-switching style (ALT and INS) and the language choice for keywords (highlighted in bold).

Advances in dialogue research (Vinyals and Le, 2015; Zhang et al., 2020; Serban et al., 2016) have enabled conversational AI technologies for human–machine interactions, like Alexa and Siri. Although these technologies are pervasive, they still have limited abilities to accommodate to the user, and they do not account for the ubiquity of multilingual communication. Due to the lack of code-switching abilities in existing language technologies, there has been limited work in studying linguistic accommodation in written CS dialogues.

With the ultimate goal to enable adaptive code-switching dialogue agents, in this paper we study user accommodation, i.e., *entrainment* (Brennan

[1]The code and data is available at https://github.com/TanmayParekh/commonDost

Proceedings of the 24th Conference on Computational Natural Language Learning, pages 565–577

https://doi.org/10.18653/v1/P17

and Clark, 1996) in CS human–machine dialogues. Our exploratory analysis of user accommodation will facilitate better development of dialogue agents which can eventually accommodate to users in return. To this end, we adopt a collaborative dialogue framework of Ahn et al. (2020), which converses with Spanish–English (Spanglish) bilinguals. To facilitate a more general analysis, we extend this framework to Hindi–English (Hinglish), a language pair which is typologically distinct from Spanglish and is spoken by millions of people.

We begin by providing background on code-switching (§2) and linguistic accommodation (§3) We then introduce our generalized bilingual dialogue system (§4). In §5, we describe our experimental setup for Hinglish data collection and discuss the data statistics. We later provide our exploratory analysis of language accommodation and other socio-linguistic factors affecting the CS patterns in the user utterances (§6). A case-study comparing code-switching distributions across Hinglish and Spanglish is presented in §7. Finally, we discuss directions for future work in §8.

This paper's contributions include: (1) the development of a bilingual collaborative dialogue system easily generalizable to a new CS language pair, (2) a new dataset, COMMONDOST, comprising of 439 Hindi-English human–machine conversations, (3) adaptation of accommodation metrics and a corresponding analysis of accommodation of language style and choice in CS dialogues, and (4) an exploratory study of linguistic and socio-cultural factors on users' CS patterns across Spanglish and Hinglish.

2 Code-Switching Strategies

Given that CS is used in very nuanced ways, researchers have been studying *how people code-switch*, examining the switch-points of languages syntactically (Poplack, 1980; Solorio and Liu, 2008), prosodically (Fricke et al., 2016), lexically (Kootstra, 2012), pragmatically (Begum et al., 2016), and so forth. Many works have attempted to model code-switching text and speech from a statistical perspective (Garg et al., 2018a,b). Recent works and benchmarks such as Linguistic Code-switching Evaluation (LinCE) (Aguilar et al., 2020) and GLUECoS (Khanuja et al., 2020) have provided a unified platform to evaluate CS data for various NLP tasks across various language pairs. Our work is in line with these recent efforts to provide NLP capabilities to users with diverse linguistic backgrounds. We extend the human–machine CS dialogue system by Ahn et al. (2020) to a new language pair of Hindi-English.

In order to better understand the style and usage of languages in a code-switched utterance, we cluster and characterize these utterances by a set of predefined CS strategies. Previous works have mainly identified two commonly used code-switching (CS) strategies: *Insertional* and *Alternational*, and these strategy distinctions are important in implementations of CS technology (Muysken, 2000; Bullock et al., 2018).

Insertional CS strategy involves one language to be the matrix language (MatL) with the other serving as the embedded language (EmbL). Words/phrases from EmbL are inserted in the sentence while maintaining the grammar and structure of MatL (Myers-Scotton, 1993). On the other hand, Alternational CS strategy involves alternating between separate independent clauses of the languages, switching from one MatL to another.

In our work, we focus on the Hindi-English language pair. We experiment with 4 CS strategies - (1) EN$\xrightarrow{ins}$HI (inserting English phrases into Hindi MatL), (2) HI$\xrightarrow{ins}$EN (inserting Hindi phrases into English MatL), (3) HI$\xrightarrow{alt}$EN (alternating from Hindi MatL to English MatL), and (4) EN$\xrightarrow{alt}$HI (alternating from English MatL to Hindi MatL).

CS is also observed more often in informal and casual settings than formal ones (Sitaram et al., 2019). We test this hypothesis by inducing informality in the agent's strategies. Although recent works (Madaan et al., 2020) have introduced neural methods to induce informality, we deploy a simple way to moderate formality by adding discourse markers (e.g. "so", "well") at the beginning and ending of sentences. These markers are independent of context and syntax (Schiffrin, 1988), and are often associated with informality (Jucker, 2002). Thus, we define four more agent strategies by infusing informality (+ *Informality*) in each of the previously described 4 CS strategies.

3 Measuring Accommodation in Dialogue

Communication Accommodation Theory posits that people adjust their behaviors or speech styles to their conversational partners' (Giles et al., 1973). Linguistic accommodation has proven to reduce

interpersonal distance (Camilleri, 1996) and is correlated with dialogue success and engagement (Nenkova et al., 2008). Although well-studied in the monolingual dialogues (Brennan and Clark, 1996; Niederhoffer and Pennebaker, 2002), it is relatively new in the CS setting. Soto et al. (2018) found rate of code-switching to be accommodated in human–human Spanish-English dialogues. Choice of language when code-switching can also be adapted in dialogues (Bawa et al., 2018). Fricke et al. (2016) further discover that part-of-speech of a CS utterance may impact the following language choice. Our work adds to this field by studying accommodation of language choice for lexical classes. In terms of quantifying accommodation, we adapt a metric from Mizukami et al. (2016) to measure accommodation (we refer it to as *global* accommodation).

Global accommodation extends the score proposed in Nenkova et al. (2008) by aggregating a speaker's word usage across an entire dialogue and biases it relatively with other non-partners in the corpus. For two partners a and b, we denote $E_{a,b} = -\sum_{w \in V} |Pr_a(w) - Pr_b(w)|$ for a given word class V (where $Pr(w)$ is the empirical probability of word w). Denoting the set of non-partners for the speaker a by $\mathcal{N}_a$, we define *ratio* as

$$ratio(E_{(a,b)}, E_{(a,np)}) = \begin{cases} 1 & E_{(a,b)} > E_{(a,np)} \\ 0.5 & E_{(a,b)} = E_{(a,np)} \\ 0 & E_{(a,b)} < E_{(a,np)} \end{cases}$$

for all non-partners $np \in \mathcal{N}_a$. The *global* score for the speaker a is the average of *ratio* over all the non-partners. The final *global* score for the dataset is the average of the scores over all the speakers in the dataset. In context of human–machine conversations, we choose the set of non-partners for an agent to be the set of humans that did not interact with this agent. Since this metric is defined primarily for lexical accommodation, we redefine different styles as a lexical class to adapt it for measuring stylistic accommodation.

Danescu-Niculescu-Mizil et al. (2011) presented another interesting metric which measures accommodation *locally* across turns within a single dialogue. For two partners a and b, we can formulate this metric as

$$local_{(a,b)}(C) = Pr\left(T_b^C | T_a^C\right) - Pr\left(T_b^C\right)$$

where T_a and T_b denote the messages of a and b respectively. Here, T_b is the reply to T_a. T_b^C (and T_a^C) denote the prevalence of style C in T_b (and T_a). In essence, it attempts to measure an increase/decrese in the usage of a style C by b grounded on the usage of C by a. In our setting of human–machine conversations, since the agent's strategy is fixed, it's not as interesting to use this metric for our analysis. Thus, we focus our analysis only using *global* accommodation metric.

4 Bilingual Dialogue System

Our bilingual human–machine dialogue system mainly serves two important purposes: (1) collection of CS data and (2) experimentation of new agent strategies. Previous work (Ramanarayanan and Suendermann-Oeft, 2017) developed a rule-based CS dialogue system restricted to a fixed set of prompts. Ahn et al. (2020) proposed a more flexible bilingual system for English-Spanish as an extension of a monolingual goal-oriented collaborative dialogue framework (He et al., 2017), originally designed for the MUTUALFRIENDS task. This task provides the two conversational partners A and B individually with a knowledge base (KB) of friends, out of which there is exactly one friend common in both KBs. Each friend in the KB has several attributes such as hobby, location of work, etc. The goal of the task is to collaboratively find this mutual friend by text conversations between the two partners–which can be human or machine.

The modifications made by Ahn et al. (2020) for extending this monolingual system to support bilingual Spanish-English dialogues were mainly in three components: (1) Bilingual Readability: Supporting instructions and KB available to the users in Spanish as well as English, (2) Bilingual Response Generation: Procuring parallel Spanish sentences using a Machine Translation (MT) system and applying rule-based transformations for generating code-switched Spanglish, (3) Bilingual Response Understanding: Translating code-switched Spanish-English to monolingual English (using a MT system) and passing it to the pre-existing response understanding system for English.

Ahn et al. (2020)'s modified Spanish-English dialogue system cannot be directly applied across other language pairs due to three key reasons: (1) The dialogue system relies on a robust CS MT system[2] which is more readily available for resource-rich languages like Spanish and English. Such

[2]Translation from code-switched Spanish-English to monolingual English.

systems might not be accessible for languages like Tagalog and Swahili. (2) The linguistic rule-based adaptations for generation are simple in the case of Spanish-English as they are typologically closer. On the contrary, linguistically diverse pairs like Telugu-English might need further adaptations due to differences in word order and morphology. (3) Spanish and English are written using the same script. Many other language pairs within which CS is pervasive, like Hindi-English, are written in different scripts, and are typically romanized in the CS setting. Lack of normalization and robust transliteration models pose challenges to multiple system components for such pairs.

In our work, we build a more generalized dialogue system to tackle the challenges stated above. One highlight of this modified system is its simplicity, which helps in adapting to new language pairs easily. We briefly discuss these challenges and our enhancements to various components for our Hindi-English dialogue system below.

Language Bias in KB Due to social and cultural priors, certain domains and topics in the KB might not be equally represented in both languages. In order to avoid biasing the language usage in the dialogue and promote code-switching, it is necessary to carefully choose equilingual domains. In the case of Hinglish, we replace the domain of *college majors*, which is highly anglicized with respect to Hindi, with *favourite fruit* which is more equally represented in both languages.

Handling gender-markings Third person pronouns and verb forms in Hindi are usually gender-marked (eg. *karta/karti* [he/she does], *uska/uski* [his/her]). Since the Spanglish KB does not provide any information about the gender of friends, we consequently notice the dialogues using this system to be gender-skewed. In the COMMONAMIGOS Spanglish data (Ahn et al., 2020), the ratio of masculine to feminine word usage was 3.9; whereas for Hinglish[3], this gender-ratio is 27.7. We mitigate this by simply adding a new "gender" attribute to the KB and correspondingly, notice a drastic drop of the gender-ratio to 3.4 for Hinglish.

Dialogue Generation The Spanglish dialogue system utilizes a MT system[4] to generate parallel Spanish-English sentences and leverages rule-based transformations (specific to Spanish) to generate code-switched sentences. For language-pairs written in a non-native script (e.g. Hinglish written in English), there is a need of an additional transliteration model alongside a MT model for script-conversion. This agglomeration of the models leads to a cascade of errors that results in a poor overall translation. We circumvent this issue by building a simple phrase-based translation system. Despite its simplicity, the translation performance of the system is qualitatively better owing to the closed domain nature of the task.

Furthermore, the rule-based transformations need appropriate modifications to accommodate the new language pair. For Hinglish, we synthesize additional transformations to handle differences in word order and verb conjugations.

Natural Language Understanding (NLU) The Spanglish dialogue system relies on a robust MT system for converting CS user utterances to English and then exploits an English NLU component for entity extraction. Procuring such MT systems[5] for other language-pairs is not feasible. This issue is amplified for languages written in non-native script (Hinglish) due to lack of normalization in user sentences. We overcome this challenge by building a simple dictionary-based NLU component which can directly understand and extract entities from CS Hinglish text. Although it cannot handle complex inputs, this simple model still outperforms the translation-based NLU pipeline.

5 Data

We use the modified bilingual dialogue framework (§4) to collect romanized Hindi-English CS data for human–machine dialogues. Here, we first describe this data collection process and later discuss statistics for the collected data.

5.1 Data Collection

The majority of our data (80%) was collected by crowdsourcing our task on the Amazon Mechanical Turk[6] (AMT) platform, while the other 20% of the data was collected via participation from local Indian communities. A pre-requisite audio-based question-answering test is used to ensure the proficiency of the participants to chat in Hinglish. We limit three attempts per participant to boost diversity of the data.

[3]Tested on a set of 65 pilot dialogues.

[4]Google Translate API in the original implementation.

[5]Phrase-based translation systems perform poorly because user utterances are open-domain.

[6]https://www.mturk.com.

A:	hey do you have any friends working at the zoo *ya dost hai jise sona pasand hai* [or friends who like sleeping] ?
H:	*mere paas 2 dost hai jo* zoo *mei kaam karte hai aur unko* photography *ya* drawing *pasand hai* respectively [I have 2 friends who work in the zoo and they like photography and drawing respectively]
A:	*toh* [so] i have some female friends *jinhe aam khana pasand hai* [who like eating mango]
H:	*mere paas ek* female friend *hai jisko aam khana pasand hai aur usko* dancing *pasand hai* [I have 1 female friend who likes eating mango and likes dancing]

Table 1: Excerpt from a dialogue from our COMMONDOST dataset. We highlight the Hindi content in *italics* along with its English translation in []. H: human and A: agent.

In order to draw direct comparisons between the collected Hinglish and Spanglish data, we closely follow the task setup as in Ahn et al. (2020). The instructions for the task are provided in Hinglish. We further use a post-task survey to gather sociolinguistic information about the participants. More details about the data collection process are described in the Appendix.

5.2 Data Processing

Owing to the lack of normalization of romanized Hindi, data processing and analysis is a non-trivial task. Further, due to paucity of CS data, there are fewer commercial systems available. To circumvent this issue, we develop simple custom tools and describe them below.

Language ID (LID) Tagger It is an important component to identify language usage (Hindi/English) by users in the dialogue data. We build a dictionary-based tagger using pre-populated lists of most common English words. We mark the remaining words with the Hindi LID[7]. This simple model achieves an accuracy of 94.5% on an unseen set of 84 human-annotated sentences. The tagged data is further corrected by human annotators.

CS Strategy Classifier We classify the user utterances into one of 7 strategies - 4 CS (see §2), 2 *monolingual* (Hindi and English), or *Neither*. We develop a simple rule-based system utilizing the LIDs for detection of these strategies. Although this system uses simple heuristics, it achieves an F1 score of 0.85 on an unseen set of 84 CS sentences. Two independent human linguists achieve an average F1 score of 0.85 on this set, thus validating the performance of the classifier.

5.3 Data Statistics

We collect a total of 439 human–machine conversations (we provide an example dialogue in Table 1) across a pool of 164 unique people, wherein close to 85% participants attempted the task more than once. The distributions of the data collected via AMT and the local community are nearly the same except for age[8]. The self-reported survey further reveals that among unique users, 72% were male, 91% have a college degree, and 90% originate from the Indian subcontinent. Nearly 72% of users speak an additional regional Indian language other than Hindi or English.

	Hinglish	**Spanglish**
# Dialogues	439	587
# User Utterances	4,361	4,617
# User Tokens	29,117	28,452
% Task Success	59%	64%
Avg dialogue length	9.93	7.9
Avg utterance length	6.68	6.2
EN vocab size	539	571
HI/SP vocab size	1,280	846
% EN utterances	19%	16%
% HI/SP utterances	34%	44%
% CS utterances	47%	39%
% CS dialogues	92%	70%

Table 2: Data statistics for the Hinglish COMMONDOST dataset and its comparison with the Spanglish COMMONAMIGOS dataset. EN: English, HI: Hindi and SP: Spanish.

We present the general statistics of our COMMONDOST data and compare it with the Spanglish COMMONAMIGOS dialogue dataset (Ahn et al., 2020) in Table 2. Although our absolute task success rate is not very high, we procure good quality code-switched dialogue data due to the agent's engagement. Notably, we observe longer chats (12.44 utterances per dialogue) for unsuccessful dialogues compared to successful ones (8.17). We

[7]Ambiguous words are handled with separate rules.

[8]The data collected from the local Indian community is skewed towards a younger age group.

Agent Strategy	# Dial	Task Success	Avg Utts	Avg Tok/Utt	% EN Utt	% HI Utt	% CS Utt
HI$\xrightarrow{alt}$EN	39	64%	10.87	6.79	**14%**	**41%**	45%
+ Informal	42	60%	9.88	**7.51**	**13%**	33%	**54%**
EN$\xrightarrow{alt}$HI	39	54%	9.36	**5.86**	**23%**	31%	46%
+ Informal	42	48%	10.38	6.95	**12%**	34%	**54%**
EN$\xrightarrow{ins}$HI	41	**76%**	9.56	6.39	20%	**27%**	**53%**
+ Informal	35	57%	8.46	**8.2**	**5%**	**25%**	**69%**
HI$\xrightarrow{ins}$EN	41	63%	8.66	**6.05**	**28%**	**24%**	47%
+ Informal	41	**73%**	11.12	6.39	19%	37%	44%
HI mono	40	**45%**	9.53	6.86	**9%**	**55%**	**36%**
EN mono	39	51%	10.41	6.08	**57%**	**19%**	**24%**
random	40	55%	10.88	6.63	**12%**	**41%**	46%

Table 3: General statistics of the COMMONDOST user dialogues filtered by agent strategy. We highlight the statistically significant (with $p < 0.05$) in **bold**. # Dial: Number of dialogues, Avg Utts: Average number of utterances per dialogue, Avg Tok/Utt: Average number of tokens per utterance, % EN Utt: Percentage of English utterances, % HI Utt: Percentage of Hindi utterances, % CS Utt: Percentage of code-switched utterances.

also observe that utterances in the Hinglish data are generally longer than that of their Spanglish counterparts. In terms of vocabulary sizes, we observe that COMMONAMIGOS data has a smaller Spanish vocabulary size when compared to the Hindi vocabulary size in the COMMONDOST data. Overall, there is more CS in Hinglish data compared to Spanglish data.

6 Analysis of Hinglish Conversations

We study the impact of each of the agent strategies (4 CS strategies and their informal counterparts) on the user dialogues using various dialogue- and language-oriented dimensions, as shown in Table 3. We also introduce monolingual agent strategies (*HI mono* and *EN mono*) and a random CS strategy[9] as baselines for our analysis. We procure roughly 40 dialogues for each agent strategy for a principled comparison across these metrics.

6.1 Code-switching and Task Success

Our data substantiates the prevalence of CS in the language pair of Hindi-English. Although no explicit instructions were provided to exhibit code-switching, 92% of the dialogues and 47% of the user utterances are code-switched (Table 2). Furthermore, even when the agent converses completely in Hindi or English, we observe CS in nearly 30% of the user utterances (Table 3). Thus clearly, our dialogue system elicits code-switching.

In a goal-oriented framework like ours, task success is an essential metric to assess our agent (Column 3 in Table 3). We observe that task success is significantly better when the agent uses a CS strategy (62%) compared to agent's monolingual strategies (48%). Furthermore, when users were asked to rate the agent for how non-native the agent seemed (1-5, 5 is most non-native), agents using CS strategies were rated 2.62, which is much lower than 2.92 and 3.11 when agents used monolingual and random strategies respectively. Thus, CS aids in better communication and engagement between the agent and the user as suggested in Camilleri (1996), which eventually translates to better success rate.

6.2 Informality improves Dialogue Quality

Infusion of informality in the agent's CS strategies has two major observable effects on the user dialogue. First, we observe an increased user utterance length (column 5 in Table 3), which is in contrast to the finding in Ahn et al. (2020). We attribute this to the users being less curt[10] as they find the informal agent is relatively more friendly. The us-

[9] We randomly switch between languages at a phrase-level.

[10] 1-2 word user utterances reduce by 7% when agent uses informal strategies.

age of discourse markers per dialogue by the users increases from 1.87 to 2.44 when conversing with an informal agent compared to a formal one. Second, we witness a higher CS and reduced English usage in the user utterances (column 6 and 8 in Table 3), similar to the finding in Spanglish (Ahn et al., 2020). Finally, when users were asked to rate the agent for how human-like it seemed, (1-5, 5 is most human-like), informal agents were rated 3.99, which is higher than 3.54 for an agent without informality. We conclude that informality helps the agent be perceived as friendlier. It induces longer and more code-switched user responses, improving the quality of the conversation.

6.3 Linguistic Accommodation in Dialogue

We focus our analysis of accommodation on the choice and style of language usage in the CS setting. We utilize the *global* accommodation metric (§3) to quantify our learnings.

Stylistic Dimension	Global score
Lexical Items (KB)	0.790
- English	0.648
- Hindi	0.700
CS Strategies	0.665

Table 4: Reporting the *global* accommodation metric for the word class of lexical items (KB) (row 1). We further report the accommodation score for the choice of language for these items (row 2 and 3). Finally we report the score for accommodation of CS strategies (row 4). Divergence is indicated by 0 while 1 indicates convergence, and 0.5 is no accommodation.

6.3.1 Language choice for lexical items

Lexical accommodation of a word class is a common and well-observed phenomenon in monolingual dialogues. In the CS setting, we study an additional dimension of language choice for the word class. For example, *if the agent uses the English word for mentioning fruits in its utterance, will the user also use the English word for referring fruits in their utterance*? We focus this analysis on the word class of all the lexical items in the knowledge base (in Hindi and English). First, we evaluate the overall language-independent score for the word class and then the language-dependent scores highlighting the accommodation of language choice for referring to the word class (first three rows in Table 4).

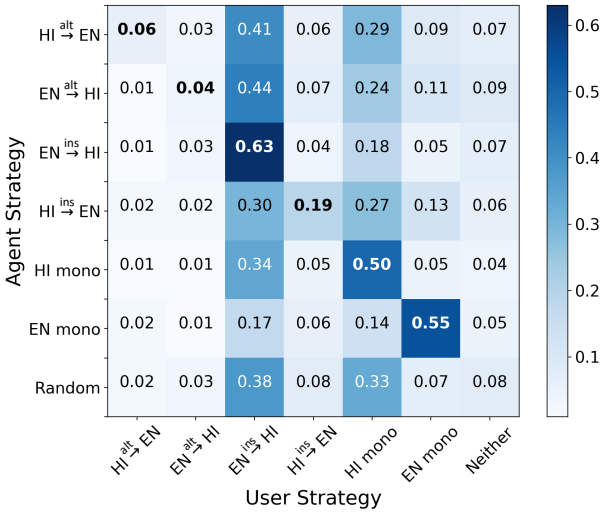

Figure 2: Probability distributions of the usage of CS strategies in user utterances (across the columns) for each agent strategy (across the rows). We highlight the statistically significant ($p < 0.05$) numbers in **bold**.

We notice a high accommodation score for the overall word class of lexical items, which is expected due to the nature of the task. We further witness a positive accommodation score for the usage of both languages (Hindi and English) while referring to this word class. This result is more motivating given there is an inherent user bias[11] to use Hindi words (60%) compared to English words (40%) while referring to lexical items in the dialogue. Thus, we conclude that the agent's language choice for lexical items positively influences the users' language choice too.

6.3.2 Accommodation of CS Strategy

In Table 3, we observe higher CS in user utterances when the agent is code-switching (51%) compared to when the agent is conversing monolingually (30%). This elicits the global accommodation of the phenomenon of code-switching by the users. Here, we focus on studying the accommodation of the style of CS between the user and the agent.

Using the rule-based CS strategy classifier (§5.2), we cluster the user utterances into one of the seven strategies - 4 CS, 2 monolingual, or neither. In Figure 2, we present a confusion matrix to study the impact of each agent's CS strategy[12] on the usage of the users' strategy. Each row in the matrix represents the normalized distribution[13] of the user

[11] Calculated by comparing usage on the *random* baseline.
[12] We do not maintain the distinction of informal strategies.
[13] We exclude sentences with length < 3 as they do not follow any particular CS strategy.

CS strategies for the given agent strategy.

The diagonal elements in the matrix represent the percentage when the user adopts the same strategy as the agent. In any column of the matrix, we observe that these elements are the highest (statistically significant with $p < 0.05$). This implies that the user's usage of any given CS strategy increases significantly when the agent is using the same strategy. Using the *global* accommodation metric to quantify this phenomenon (last row of Table 4), we observe a high positive accommodation score. Based on these observations, we conclude that users synchronize their style of language use with the agent in a CS setting.

6.4 Language Proficiency influences Hinglish CS

We cluster and analyze user dialogues based on the self-reported additional languages of proficiency in the post-task survey. Diving deeper into the usage of CS strategies filtered by their language of proficiency, we find a general peculiarity amongst speakers proficient in South Indian languages (Malayalam, Telugu, Tamil and Kannada). These speakers (specifically Telugu) have a higher usage of HI$\xrightarrow{ins}$EN and EN$\xrightarrow{alt}$HI and a relatively lower usage of EN$\xrightarrow{ins}$HI CS strategies in their utterances. These strategies indicate that such speakers are using English as their MatL, or at least starting with it. Alternatively, we find a higher usage of EN$\xrightarrow{ins}$HI and HI$\xrightarrow{alt}$EN strategies for speakers proficient in North Indian languages (Gujarati, Marathi, Punjabi, Odia and Bengali). These two strategies indicate that such speakers are adopting Hindi as their MatL, or at least in the beginning (which is opposite as observed for South Indian speakers). This phenomenon can be attributed to a higher influence of English in the South Indian languages and correspondingly, Hindi in the North Indian languages as suggested in Baldridge (2002).

Overall, we believe that other languages of proficiency, as a proxy for geographical region and cultural factors, largely impact the dialogue and CS patterns, and are understudied in general. Studying such patterns in human–human dialogues (eg. when a North Indian speaker converses with a South Indian speaker in a CS setting) would reveal further various socio-cultural factors influencing code-switching.

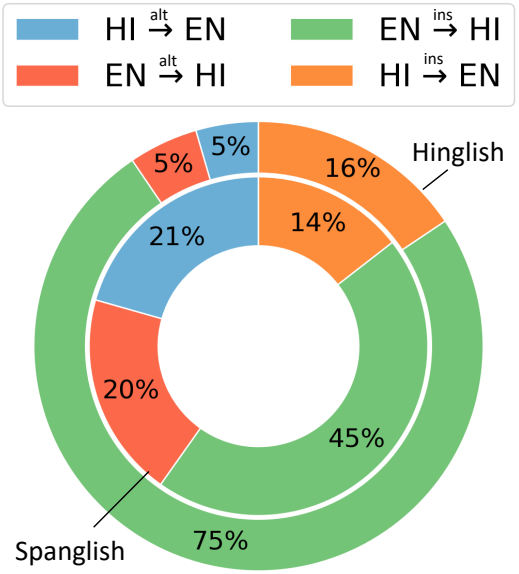

Figure 3: Comparing the distributions of users' usage of CS strategies in Hinglish and Spanglish. Alternational CS is relatively less frequent in Hinglish. Inner circle: Spanglish and Outer circle: Hinglish

7 Comparison of Spanglish and Hinglish

To gain better insights into how linguistic and sociocultural factors influence code-switching patterns, we compare the distributions of the users' usage of various CS strategies in Hinglish and Spanglish in Figure 3. We observe that EN$\xrightarrow{ins}$HI and EN$\xrightarrow{ins}$SP are the most dominant CS strategies in Hinglish and Spanglish respectively. On the other hand, we notice a large difference in the usage of Alternational CS strategies in the language pairs. For Spanglish, it accounts to roughly 40% while it is merely 10% for Hinglish.

As attributed by the Equivalence Constraint, CS points tend to occur only if a syntactic rule is not violated in either of the two languages being mixed (Poplack, 1980). Given this requirement, a pair of languages that have differing word order could have more constraints on where switches can occur. We hypothesize that Alternational CS may not work within a verb clause in Hindi as it is a verb-final (SOV) language while English is verb-medial (SVO). Spanish is verb-medial like English, and their word order similarity may facilitate the use of Alternational CS.

Beyond structural differences, sociolinguistic factors may affect CS strategies of speakers. Backus (1998) describes a gradient of strategy usage across generations of immigrants. Earlier gen-

erations of immigrants would progress from simple to complex insertions, and later generations would alternate the two languages, eventually using reverse insertion. As the Spanglish dataset includes later generations of immigrants to the US, 90% of Hinglish speakers are 1st generation. This would highlight Hinglish speakers' affinity towards insertion into the Hindi matrix language.

Additionally, the status of English in the US (for Spanglish) and English in India (for Hinglish) is different. As found in §6.4, the status of English can vary within regions of India itself, and this can lead to varying uses of CS strategy. Attitudes towards language use have been shown to affect code choice in bilingual speakers (Redinger, 2010). It is likely that attitudes towards CS is not the same in the Spanglish and Hinglish populations, which can provide further variability in the speakers' language choice.

8 Conclusion and Future Work

In our work, we proposed a generalized bilingual dialogue system and procured human–machine dialogue data (COMMONDOST) for the language pair of Hindi-English using this system. Adaptation of this dialogue system for newer CS languages could promote collection of more bilingual dialogue data.

Analysis of the COMMONDOST conversations revealed how users positively adopt and accommodate the agent's style of using language in a CS utterance. We also studied how informality and cultural factors independently affect the users' CS patterns. This proves that our findings are extendable across two CS pairs of Hinglish and Spanglish (Ahn et al., 2020). Similar analysis can be done for new language pairs (such as Arabic-English) and datasets from different domains. Another area of potential research would be to compare our findings of the CS patterns and accommodation with human–human CS conversations.

Finally, we discussed how linguistic and sociopolitical factors affect the distribution of users' CS patterns across the language pairs of Hinglish and Spanglish. Despite their dissimilarities, the similarities across these language pairs is encouraging, as it open avenues to learn about how code-switching functions cross–linguistically. We pave the path for future research on comparisons of multiple CS language pairs.

Acknowledgments

The authors are grateful to the anonymous reviewers for their invaluable feedback. This material is based upon work supported by the National Science Foundation under Grant No. IIS2007960.

References

Gustavo Aguilar, Sudipta Kar, and Thamar Solorio. 2020. Lince: A centralized benchmark for linguistic code-switching evaluation. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 1803–1813.

Emily Ahn, Cecilia Jimenez, Yulia Tsvetkov, and Alan Black. 2020. What code-switching strategies are effective in dialogue systems? *Proceedings of the Society for Computation in Linguistics*, 3(1):308–318.

Peter Auer. 2013. *Code-switching in conversation: Language, interaction and identity*. Routledge.

Albert Marie Backus. 1998. *Two in one: Bilingual speech of Turkish immigrants in The Netherlands.* Tilburg University Press.

Jason Baldridge. 2002. Linguistic and social characteristics of indian english. *Language in India*, 2(4).

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.

Susan E Brennan and Herbert H Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6):1482.

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

Antoinette Camilleri. 1996. Language values and identities: Code switching in secondary classrooms in malta. *Linguistics and education*, 8(1):85–103.

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.

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.

Penelope Gardner-Chloros and Daniel Weston. 2015. Code-switching and multilingualism in literature. *Language and Literature*, 24(3):182–193.

Saurabh Garg, Tanmay Parekh, and Preethi Jyothi. 2018a. Code-switched language models using dual RNNs and same-source pretraining. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3078–3083, Brussels, Belgium. Association for Computational Linguistics.

Saurabh Garg, Tanmay Parekh, and Preethi Jyothi. 2018b. Dual language models for code switched speech recognition. In *Proceedings of the 19th Annual Conference of the International Speech Communication Association (INTERSPEECH)*.

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.

John J Gumperz. 1982. *Discourse strategies*, volume 1. Cambridge University Press.

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 Association for Computational Linguistics (ACL)*.

Monica Heller. 1982. Negotiations of language choice in montreal. *Language and social identity*, pages 108–118.

Andreas H Jucker. 2002. *Discourse markers in Early Modern English*. Routledge.

Simran Khanuja, Sandipan Dandapat, Anirudh Srinivasan, Sunayana Sitaram, and Monojit Choudhury. 2020. GLUECoS: An evaluation benchmark for code-switched NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3575–3585, Online. Association for Computational Linguistics.

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

Aman Madaan, Amrith Setlur, Tanmay Parekh, Barnabas Poczos, Graham Neubig, Yiming Yang, Ruslan Salakhutdinov, Alan W Black, and Shrimai Prabhumoye. 2020. Politeness transfer: A tag and generate approach. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1869–1881, Online. Association for Computational Linguistics.

James Milroy et al. 1995. *One speaker, two languages: Cross-disciplinary perspectives on code-switching*. Cambridge University Press.

Masahiro Mizukami, Koichiro Yoshino, Graham Neubig, David Traum, and Satoshi Nakamura. 2016. Analyzing the effect of entrainment on dialogue acts. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 310–318.

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 46th Annual Meeting of the Association for Computational Linguistics ACL-08: HLT, Short Papers*, pages 169–172.

Kate G Niederhoffer and James W Pennebaker. 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, 21(4):337–360.

Shana Poplack. 1980. Sometimes i'll start a sentence in english y termino en espafiol. *Linguistics*, 18:581–618.

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

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.

Daniel Redinger. 2010. *Language attitudes and code-switching behaviour in a multilingual educational context: the case of Luxembourg*. Ph.D. thesis, University of York.

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.

Iulian V Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. 2016. Building end-to-end dialogue systems using generative hierarchical neural network models. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*.

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.

Oriol Vinyals and Quoc V. Le. 2015. A neural conversational model. In *Proceedings of the 31st International Conference on Machine Learning*.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2020. DIALOGPT : Large-scale generative pre-training for conversational response generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 270–278, Online. Association for Computational Linguistics.

A Task Illustration

We present the user interface for the MTurk task in Figure 4. The users are shown the task-related information and instructions at all points in time for quick reference. We present the knowledge-base (KB) of the friends in a tabular format on the right. The information in the table has both the English and Hindi translations of the entities and items in the KB. At the bottom of the page, we have the main chat interface. It comprises of a running timer (maximum of 8 mins), a chat history window showing the user and agent's responses and finally a text input box for writing responses. The user can chat with the agent using the chat window or guess the mutual friend using *Select* button corresponding to that friend in the table on the right.

B Example Dialogues

In addition to Table 1, we present additional samples of dialogues from the collected COMMON-DOST dataset (Table 5, 6, 7). Table 5 elicits a case of style accommodation by the user to the agent's CS strategy. Initially the user conversed in monolingual English, but later used a HI$\xrightarrow{alt}$EN CS strategy following the agent.

H:	dont you have any friends who works in zoo in and like photography
A:	*mere paas ek dost hai* [I have one friend] working at the factory
H:	*kya tere pass* [Do you have] any friend works in the office
A:	*mere paas do dost hai* [I have two friends] working at the office

Table 5: Excerpt from a dialogue from our COMMON-DOST dataset. We highlight the Hindi content in *italics* along with its English translation in []. H: human and A: agent.

In Table 6, we observe that the user doesn't change its strategy with respect to the agent's CS strategy and continues conversing in monolingual Hindi. Further, this example also highlights the challenge of normalization of romanized Hindi as many words like *dosth* and *dho* are spelt differently owing to the language proficiency of the user.

A few dialogues in our dataset are extremely short as well if the user is able to guess the common friend early enough. One example of such a dialogue is presented in Table 7. The "SELECT" notation denotes the selection of the mutual friend. The conversation comprised of just a single user utterance (which was a greeting). Such dialogues are usually very sparse as guessing the common friend without any information is a low probability event.

A:	i have two *dost* [friends] working at the *machalighar* [aquarium].
H:	*haaa* [yes]
H:	*eek dosth hai* [there is one friend]
A:	i have one *dost* [friend] who likes *sona* [sleeping], one *dost* [friend] who likes *machali pakadna* [fishing] . . *aur tere* [and yours] ?
H:	*machali pakadna dho hai* [two for fishing]

Table 6: Excerpt from a dialogue from our COMMON-DOST dataset. We highlight the Hindi content in *italics* along with its English translation in []. H: human and A: agent.

A:	*suprabhat* [Good morning]
H:	*namaste* [Hello]
A:	oh do you have any *dost* [friend] who like eating *santra* [orange] ?
H:	SELECT {"fruits": "orange", "hobbies": "traveling", "loc": "the office", "gender": "female"}
A:	SELECT {"fruits": "orange", "hobbies": "traveling", "loc": "the office", "gender": "female"}

Table 7: Excerpt from a dialogue from our COMMON-DOST dataset. We highlight the Hindi content in *italics* along with its English translation in []. H: human and A: agent.

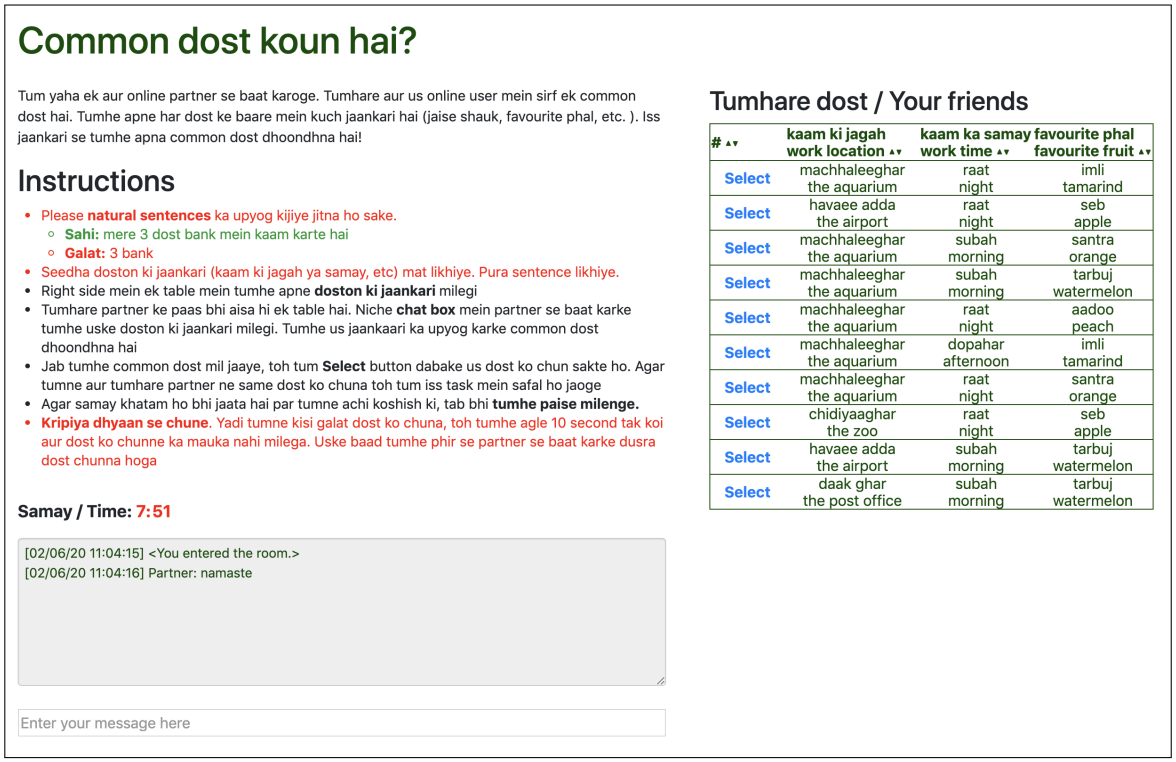

Figure 4: Illustration of the chat screen shown/used by the MTurk users for attempting the task - *Common dost koun hai?*[Who's the mutual friend?]. We collected the CommonDost using this setup.