Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues

Tanmay Parekh¹, Emily P Ahn², Yulia Tsvetkov¹, Alan W Black¹

¹ Carnegie Mellon University
² University of Washington
Outline

• Introduction

• [C1] Bilingual Code-switching Dialogue System

• [C2] Data Collection

• [C3] Data Analysis

• Future Work
Introduction

Understanding *Linguistic Accommodation* in Code-Switched Human-Machine Dialogues
Introduction

One of the participants in the dialogue is a machine

Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues
Usage of two or more languages within the same utterance or across utterances

Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues

I hope *aapko mera* talk and presentation *acha lagega*

[I hope you will like my talk and presentation]

> 50%[1]

[1] V Guadalupe. "Multilingualism."
**Introduction**

Dialogues between humans and machines wherein one or more participants are code-switching

Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues

*Kya tumhare paas koi dost hai* who like to eat mangoes?

*Nahi. Mere kisi friend ko aam pasand nahi*

*Acha...mere 2 dost bank mein kaam karte hai* and they like eating apple
Introduction

Phenomenon wherein people adjust their behaviors or speech styles to their conversational partners [2]

Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues

- Pause Length
- Head Nodding
- Backchannels
- Posture
- Dialogue Acts
- Linguistic Style
- Utterance Length

[2] H Giles, DM Taylor, and R Bourhis, 1973. Towards a theory of interpersonal accommodation through language: Some canadian data.
Introduction

Choice of language for specific words

Way of switching from one language to the other

Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues

Acha...mere 2 dost bank mein kaam karte hai and they like eating apple

Mere paas bhi 3 dost hai who like to eat apple
Motivation

Hey! How are you doing?

Hey! ¿Cómo estás?

Code-switching Chatbot

Make chatbots accommodate users

Hey! ¿Cómo estás?

I am doing good, ¿y tú?

Engagement

Naturalness
Contributions

• [C1] Development of a generalized bilingual dialogue system that can be easily adapted to a new language pair

• [C2] New dataset - CommonDost - comprising of 439 Hindi-English human-machine conversations
Contributions

• [C3a] Analysis of accommodation of language style and language choice in the code-switched dialogues

• [C3b] Exploratory analysis comparing code-switching patterns across Spanish-English and Hindi-English
Outline

• Introduction

• [C1] Bilingual Code-switching Dialogue System

• [C2] Data Collection

• [C3] Data Analysis

• Future Work
Goal-Oriented Collaborative Dialogue System

Find your Mutual Friend

Time Remaining: 2:10

[02/09/18 00:17:18] <You entered the room.>
[02/09/18 00:17:19] Partner: hi
[02/09/18 00:17:28] Partner: I have 1 university of illinois at springfield, 1 radford university.
[02/09/18 00:17:44] You: sup
[02/09/18 00:17:50] Partner: Do you have any friends who like outdoor?
[02/09/18 00:17:57] You: I have a couple at radford

Your friends

| # | School | Time Preference | Location Preference |
|---|--------|-----------------|---------------------|
| Select | University of the Ozarks | afternoon | outdoor |
| Select | Christian Brothers University | afternoon | outdoor |
| Select | Radford University | afternoon | indoor |
| Select | North Carolina Wesleyan College | morning | outdoor |

[3] H He, A Balakrishnan, M Eric, and P Liang. 2017. Learning symmetric collaborative dialogue agents with dynamic knowledge graph embeddings.
Bilingual Dialogue System

Bilingual Adaptations for Spanish-English \[4\]

Original Monolingual Dialogue System

Bilingual Adaptions

[4] E Ahn, C Jimenez, Y Tsvetkov, and A Black. 2020. What code-switching strategies are effective in dialogue systems?
Bilingual Dialogue System: Issues

1. Reliance on a robust machine translation system (which can handle code-switching)
Bilingual Dialogue System: Issues

Linguistic rule-based adaptations vary largely based on language pair similarity.
Bilingual Dialogue System: Issues

Lack of normalization for writing code-switched text

- Code-switching

| Latin Alphabet | Devanagiri Alphabet |
|----------------|---------------------|
| En             | Hi                  |
| Es             | Hi                  |
| Hi             | Hi                  |

- Devanagiri Alphabet
  - Hi
  - Ma

- Latin Alphabet
  - Hi

Examples:
- एक (one)
  - ek
  - ik
  - eik
  - eek
  - ikk

[one]
Generalized Bilingual Dialogue System

Language Bias in KB

1
Generalized Bilingual Dialogue System

Handling Gender Markings

Pronouns
uska/uski [his/her]

Verbs
karta/karti [he/she does]

Gender Ratio (M:F)

Favorite Fruit
Profession
Hobby
Time of Work
Gender

En
Hi
Generalized Bilingual Dialogue System

Phrase-based MT

Transformations

Transformations

Transformations

Dialogue Generation

En → Es

En → Hi

English Text

MT

Spanish Text

Spanish Text

Hindi Text

Hinglish Text

Romanized Hindi Text

Hinglish Text

3

MT

English Text

Spanish Text

Hindi Text
Generalized Bilingual Dialogue System

Spanglish Text → MT → English Text → NLU → Extracted Information

Hinglish Text → Rule-based NLU → Extracted Information

Natural Language Understanding
Outline

• Introduction

• [C1] Bilingual Code-switching Dialogue System
  • [C2] Data Collection
  • [C3] Data Analysis

• Future Work
Data Collection

International Location
Screening Test (Audio Based)
Instructions in Hinglish

Conversational Task (Timer: 8 mins)
Post-Task Survey
CommonDost
Common dost koun hai?

Tum yaha ek aur online partner se baat karo. Tumhare aur us online user mein sир ek common dost hai. Tumhe apne har dost ke baare mein kuch jaankari hai (jaise shauk, favourite phal, etc.). Iss jaankari se tumhe apna common dost dhoonthena hai!

Instructions

- Please natural sentences ka upyog kii jiyte jitna ho sake.
  - Sahi: mere 3 dost bank mein kaam karte hai
  - Galat: 3 bank
- Seedha doston ki jaankari (kaam ki jagah ya samay, etc) mat likhiye. Pura sentence likhiye.
- Right side mein ek table mein tumhe apne doston ki jaankari milegi
- Tumhare partner ke paas bhi aisa hi ek table hai. Niche chat box mein partner se baat karke tumhe uske doston ki jaankari milegi. Tumhe us jaankaari ka upyog karke common dost dhoonthena hai
- Jab tumhe common dost mil jaaye, toh tum Select button dabake us dost ko chun sakte ho. Agar tumne aur tumhare partner ne same dost ko chuna toh tum is task mein safal ho jaoge
- Agar samay khatam ho bhi jaata hai par tumme achi koshish kii, tab bhi tumhe paise milenge.
- Kripya dhyaan se chune. Yadi tumne kisi galat dost ko chuna, toh tumhe aple 10 second tak ko e aur dost ko chunne ka mauka nahi milega. Uske baad tumne phir se partner se baat karke dusra dost chunna hoga

Samay / Time: 7:51

[02/06/20 11:04:15] <You entered the room.>
[02/06/20 11:04:16] Partner: namaste
Data Collection

Demographics

• 164 unique participants
• 72% Male
• 90% originate from Indian subcontinent
• 91% have a college degree

|                           | 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%       |
Data Collection

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]

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]
Outline

• Introduction

• [C1] Bilingual Code-switching Dialogue System

• [C2] Data Collection

• [C3] Data Analysis

• Future Work
Kya tumhare paas koi dost hai jise seb pasand hai?

Kya tumhare paas koi friend hai jise apple pasand hai?

[Do you have any friend who likes apple?]

[5] Pieter Muysken. 2000 Bilingual speech: a typology of code-mixing
Data Analysis: Background

Kya tumhare paas koi dost hai jise seb pasand hai?

Switch from one language to another at a syntactic point

Kya tumhare paas koi dost hai who likes apple?

[Do you have any friend who likes apple?]
Kya tumhare paas koi dost hai jise seb pasand hai?

So kya tumhare paas koi dost hai jise seb pasand hai?

[Do you have any friend who likes apple?]

[6] Deborah Schiffrin. 1988. Discourse markers
Linguistic Accommodation

Global Accommodation

Measuring speaker’s word usage and biasing it with all non-partners

\[
E_{a,b} = - \sum_{w \in V} |Pr_a(w) - Pr_b(w)|
\]

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

\[
global = \frac{1}{|S|} \sum_{s \in S} \frac{1}{N_s} \sum_{np \in N_s} \text{ratio}(E_{s,p(s)}, E_{s,np})
\]

[7] Masahiro Mizukami, Koichiro Yoshino, Graham Neubig, David Traum, and Satoshi Nakamura. 2016. Analyzing the effect of entrainment on dialogue acts.
Data Analysis

We didn’t provide any explicit instructions to users to code-switch

30% CS utterances even if agent is monolingual

47% CS user utterances

92% CS dialogues

Our dataset elicits code-switching
Data Analysis

Task Success

- CS Agent: 62%
- Mono Agent: 48%

How non-native is your partner?

- CS Agent: 2.62
- Mono Agent: 2.92
- Random Agent: 3.11

B: Code-switching leads to better engagement
Data Analysis

Informality improves dialogue quality

Formal Agent
Informal Agent

Avg user utterance length

|                | Formal Agent | Informal Agent |
|----------------|--------------|----------------|
| Avg user utterance length | 6.27         | 7.23           |

# Avg Discourse Markers

|                | Formal Agent | Informal Agent |
|----------------|--------------|----------------|
| # Avg Discourse Markers | 1.87         | 2.44           |

How human-like was your partner?

|                | Formal Agent | Informal Agent |
|----------------|--------------|----------------|
| How human-like was your partner? | 3.54         | 3.99           |

C
Informality improves dialogue quality
Data Analysis

Accommodation of lexical items

All

If the agent uses any word for mentioning fruits in its utterance, will the user also use a word that refers to fruits in their utterance?

English

If the agent uses the English word for mentioning fruits in its utterance, will the user also use the English word that refers to fruits in their utterance?

Global Accommodation Score

All

0.79

English

0.648

Hindi

0.7

Agent’s language choice positively influences users’ language choice
Data Analysis

Confusion matrix for usage of user CS strategies compared to agent’s CS strategies

User uses a given strategy the most when the agent uses the same strategy

Agent’s CS strategy positively influences users’ CS strategy
Data Analysis

We ask users their other language of proficiency in the post-task survey

South Indian Speakers
- Ins Hi->En
- Alt En->Hi

North Indian Speakers
- Ins En->Hi
- Alt Hi->En

Higher Influence of English/Hindi in South/North Indian languages

Language Proficiency influences usage of CS strategies
Data Analysis

Equivalence Constraint: CS at point where syntactic rules are not violated

Word Order

Higher Alt in En-Es v/s En-Hi

Comparison of Spanglish and Hinglish

[8] Shana Poplack. 1980. Sometimes i’ll start a sentence in english y termino en espaniol.
Outline

• Introduction

• [C1] Bilingual Code-switching Dialogue System

• [C2] Data Collection

• [C3] Data Analysis

• Future Work
Future Work

• Adaptation of the dialogue system to collect dialogue data to more CS languages

• Comparison of our findings about CS patterns and accommodation with human-human conversations

• Encourage further research for how linguistic and socio-political factors affect users’ CS distribution across language pairs
Team

Tanmay Parekh

Emily P Ahn

Yulia Tsvetkov

Alan W Black
Thank You

[dhanyawaad]