Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues

Tanmay Parekh[1], Emily P Ahn[2], Yulia Tsvetkov[1], Alan W Black[1]

[1] Carnegie Mellon University
[2] University of Washington

Outline

- Introduction
- [C1] Bilingual Code-switching Dialogue System
- [C2] Data Collection
- [C3] Data Analysis
- Future Work

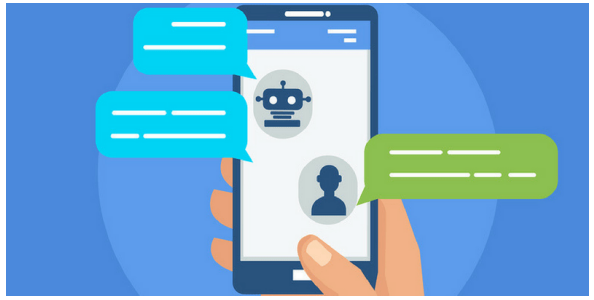
Introduction

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**

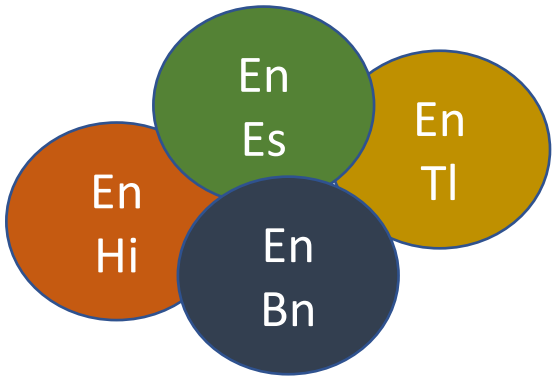


Introduction

Usage of two or more languages within the same utterance or across utterances

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

> 50% [1]



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

[I hope you will like my talk and presentation]

[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
- Dialogue Acts
- Linguistic Style
- Backchannels
- Posture
- Utterance Length

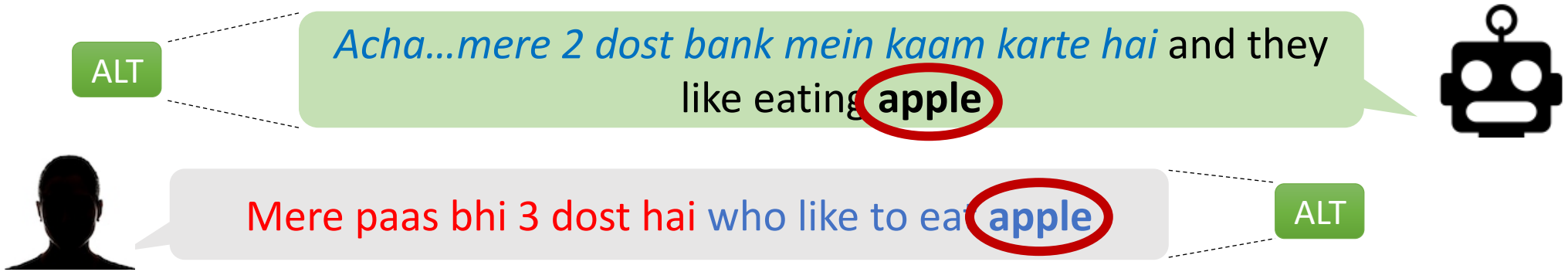
[2] H Giles, DM Taylor, and R Bourhis, 1973. Towards a theory of interpersonal accommodation through language: Some canadian data.

Introduction

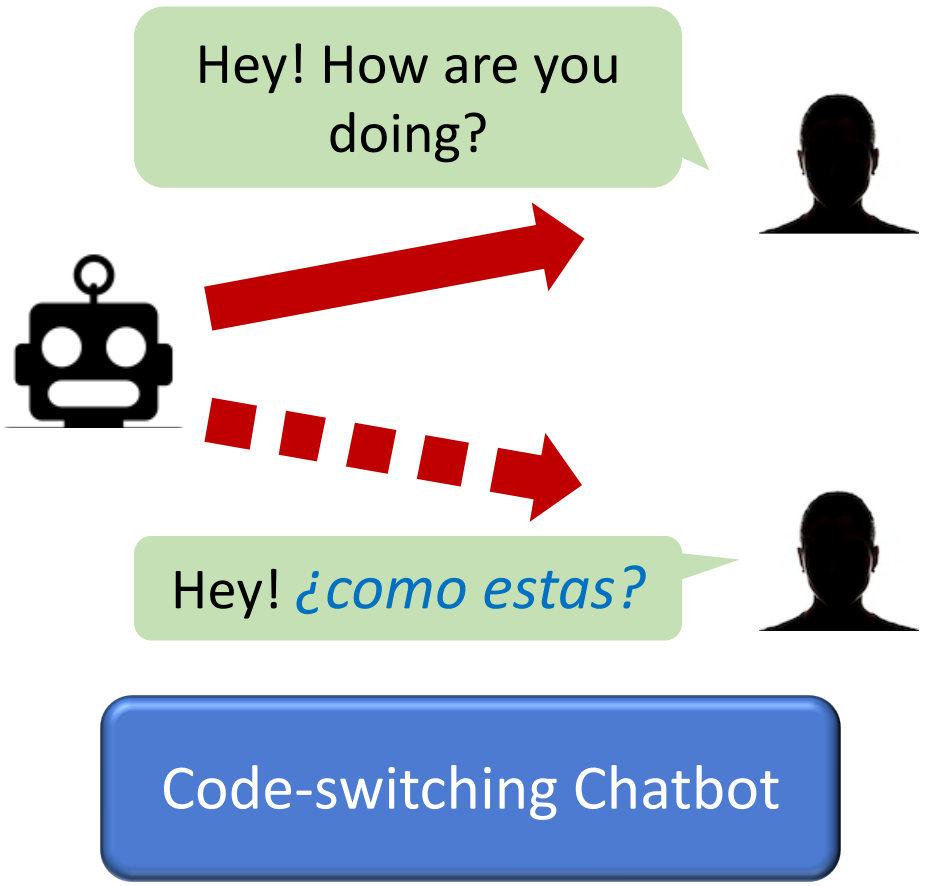
Choice of language for specific words

Way of switching from one language to the other

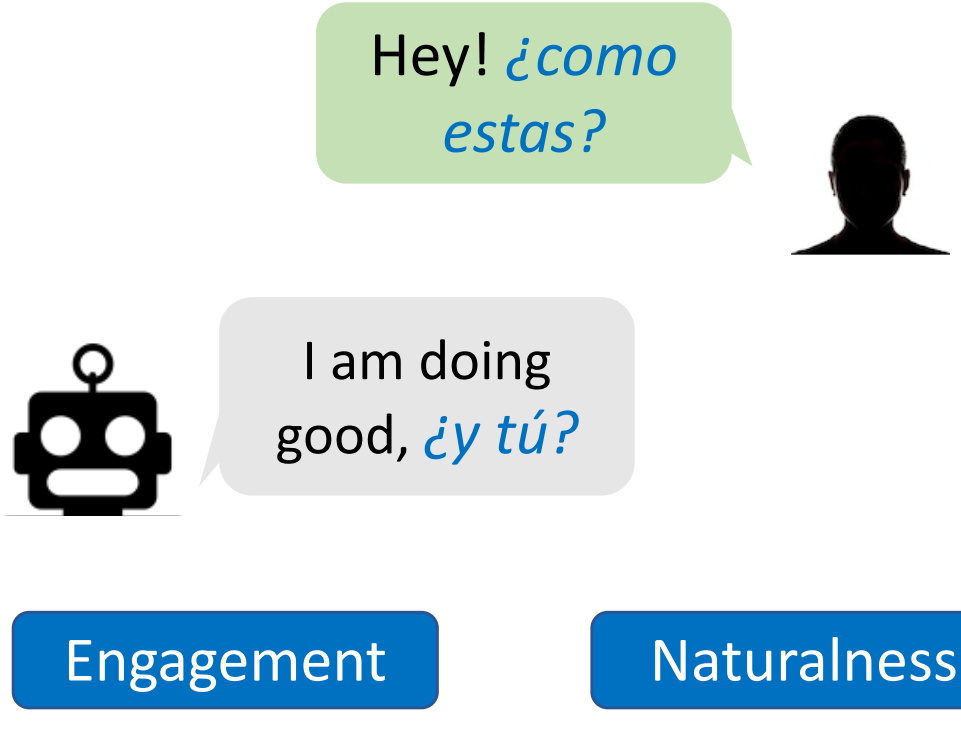
Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues



Motivation



Make chatbots accommodate users



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

Monolingual Dialogue System

Goal-Oriented Collaborative Dialogue System[3]

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

Enter your message here

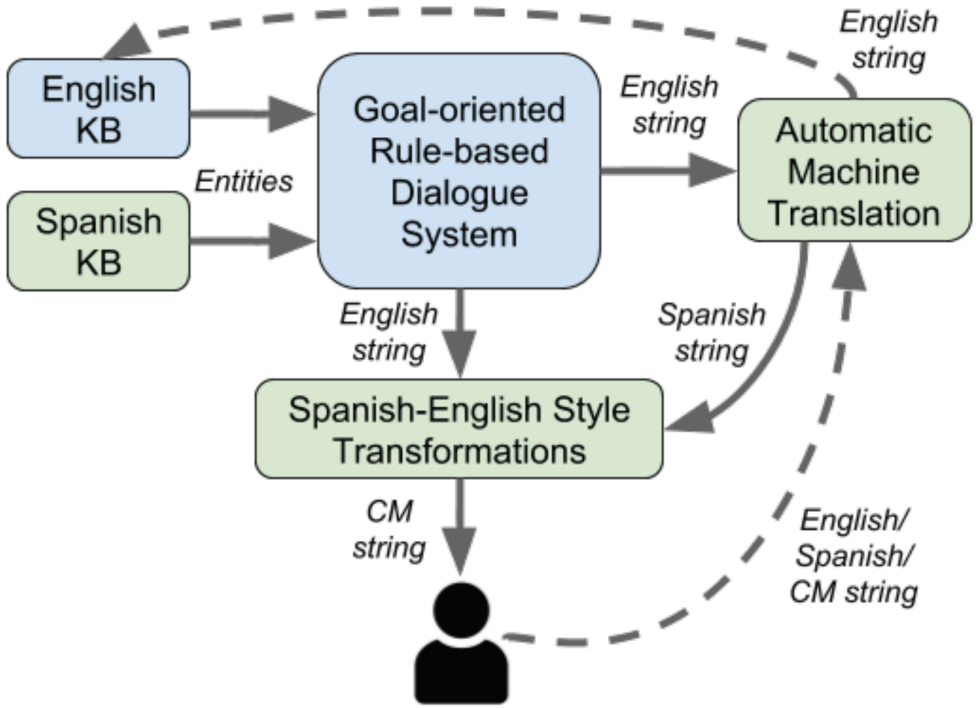
Your friends

# ▲▼	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]

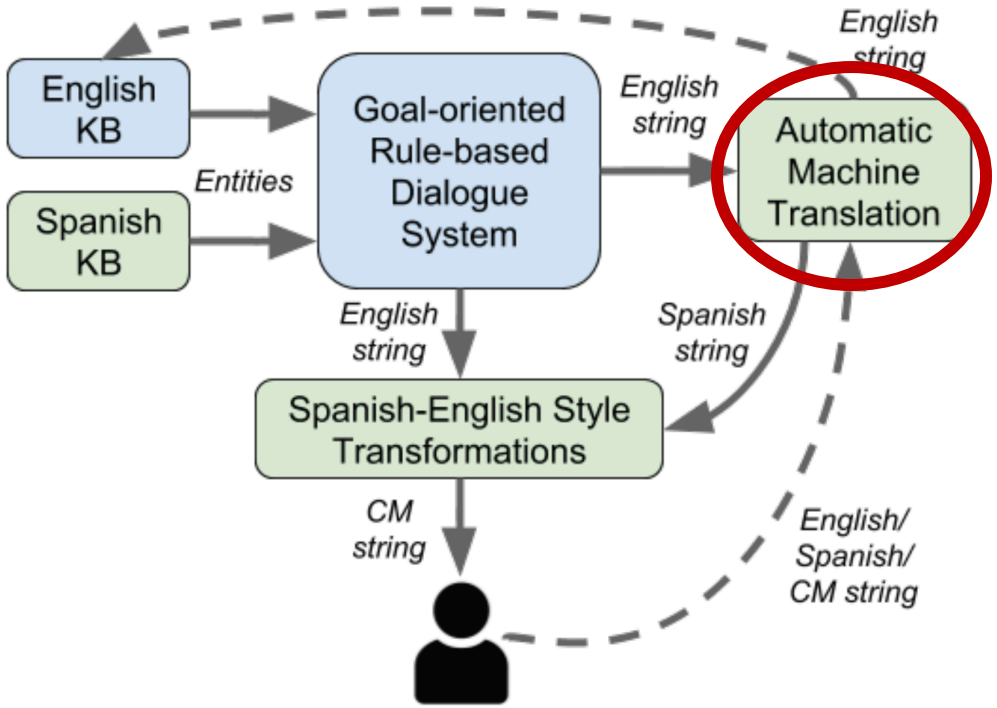


Original Monolingual Dialogue System

Bilingual Adaptions

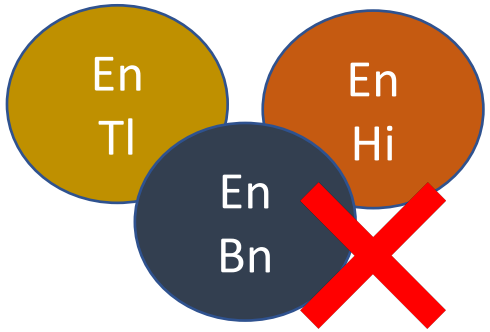
[4] E Ahn, C Jimenez, Y Tsvetkov, and A Black. 2020. What code-switching strategies are effective in dialogue systems?

Bilingual Dialogue System: Issues

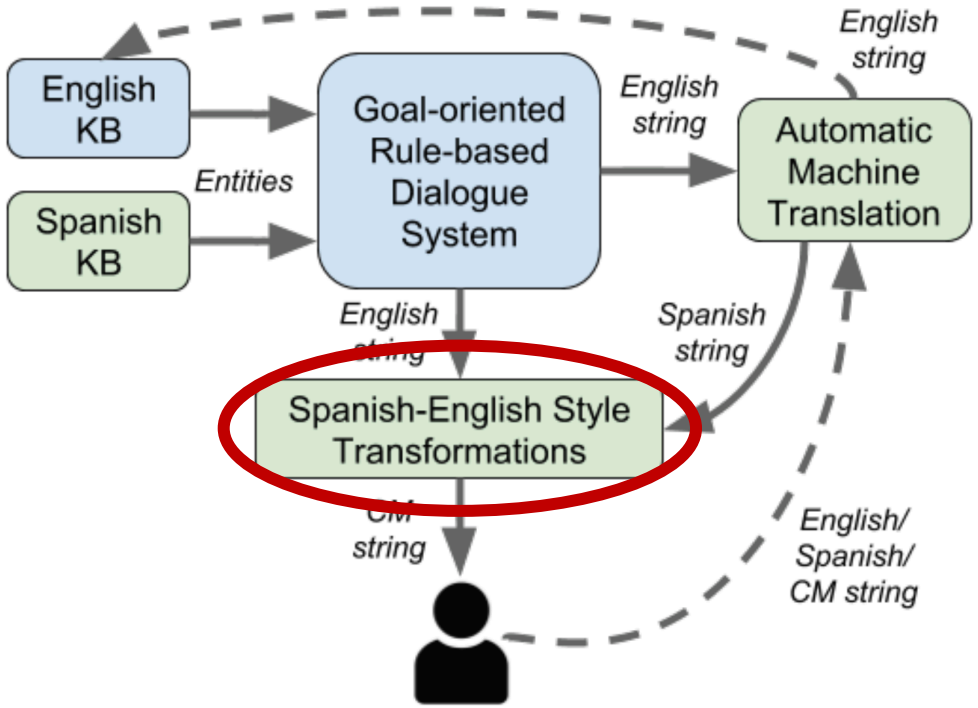


1

Reliance on a robust machine translation system (which can handle code-switching)

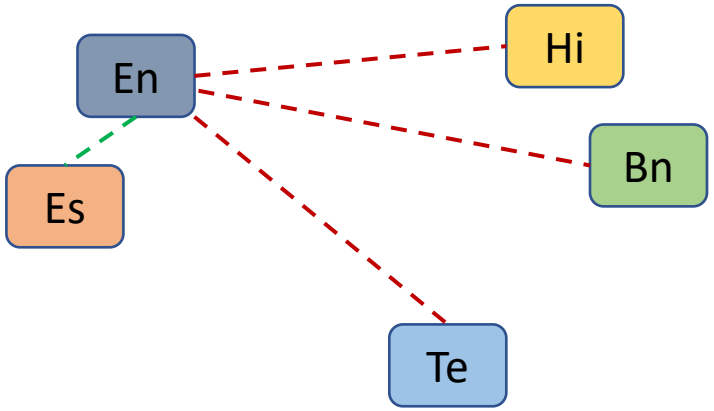


Bilingual Dialogue System: Issues

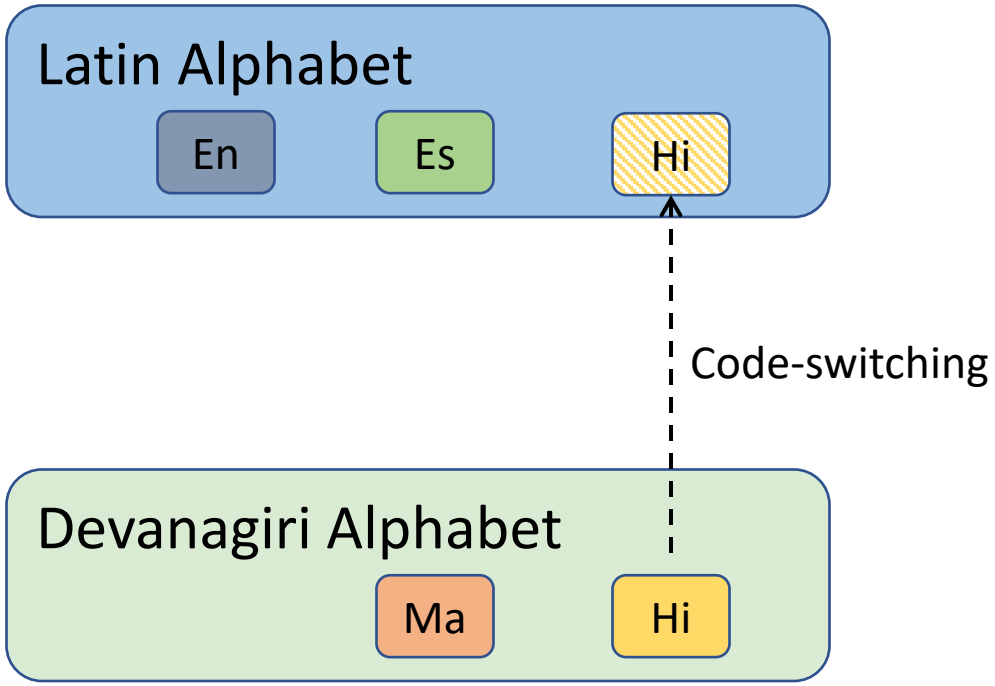


2

Linguistic rule-based adaptations vary largely based on language pair similarity

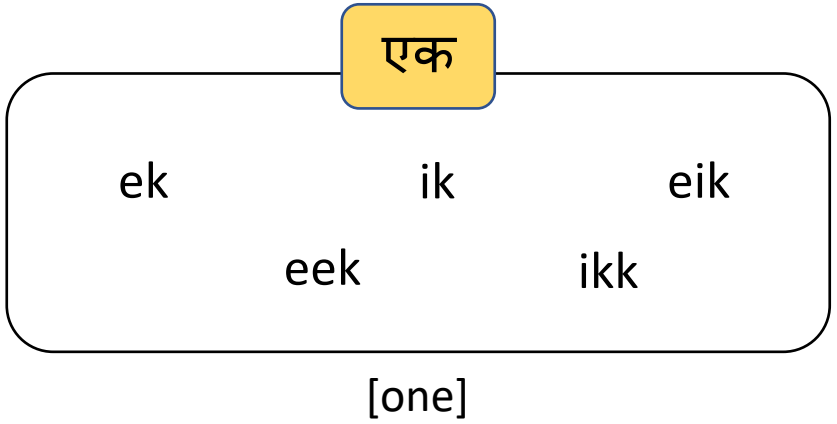


Bilingual Dialogue System: Issues

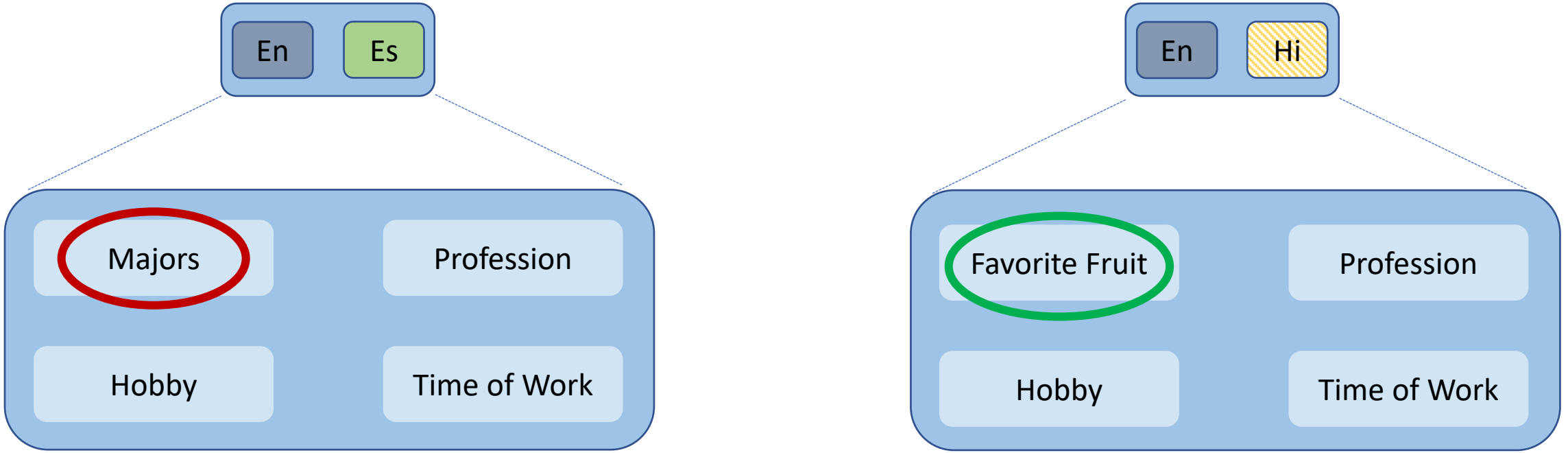


3

Lack of normalization for writing code-switched text

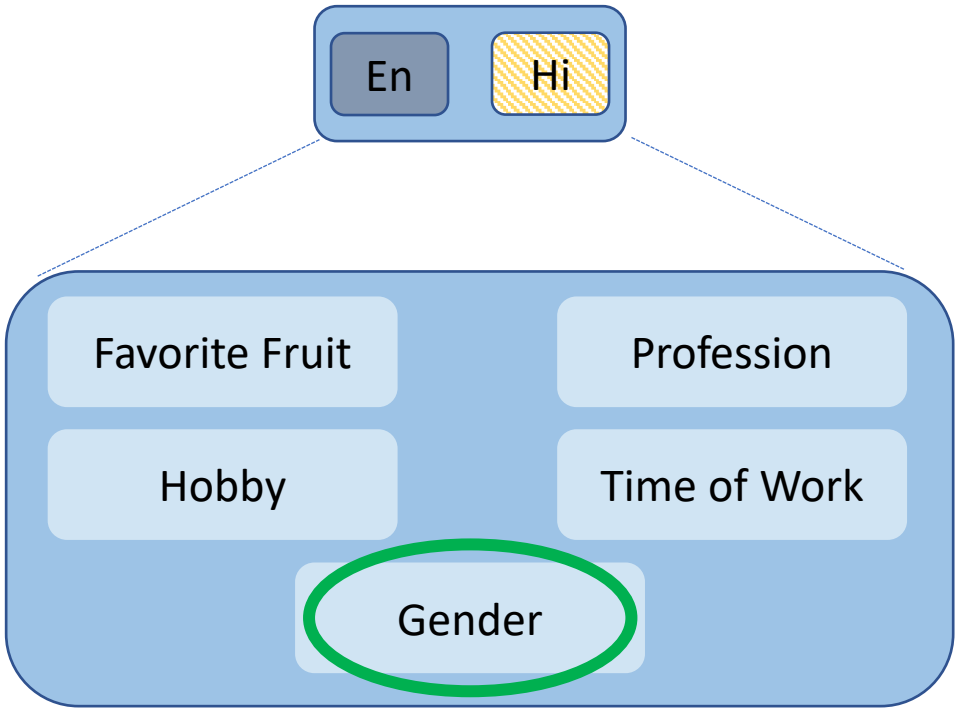


Generalized Bilingual Dialogue System



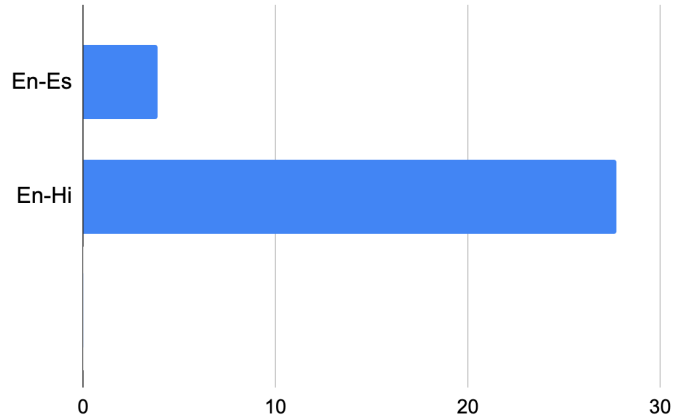
1 Language Bias in KB

Generalized Bilingual Dialogue System



Pronouns — *uska/uski* [his/her]

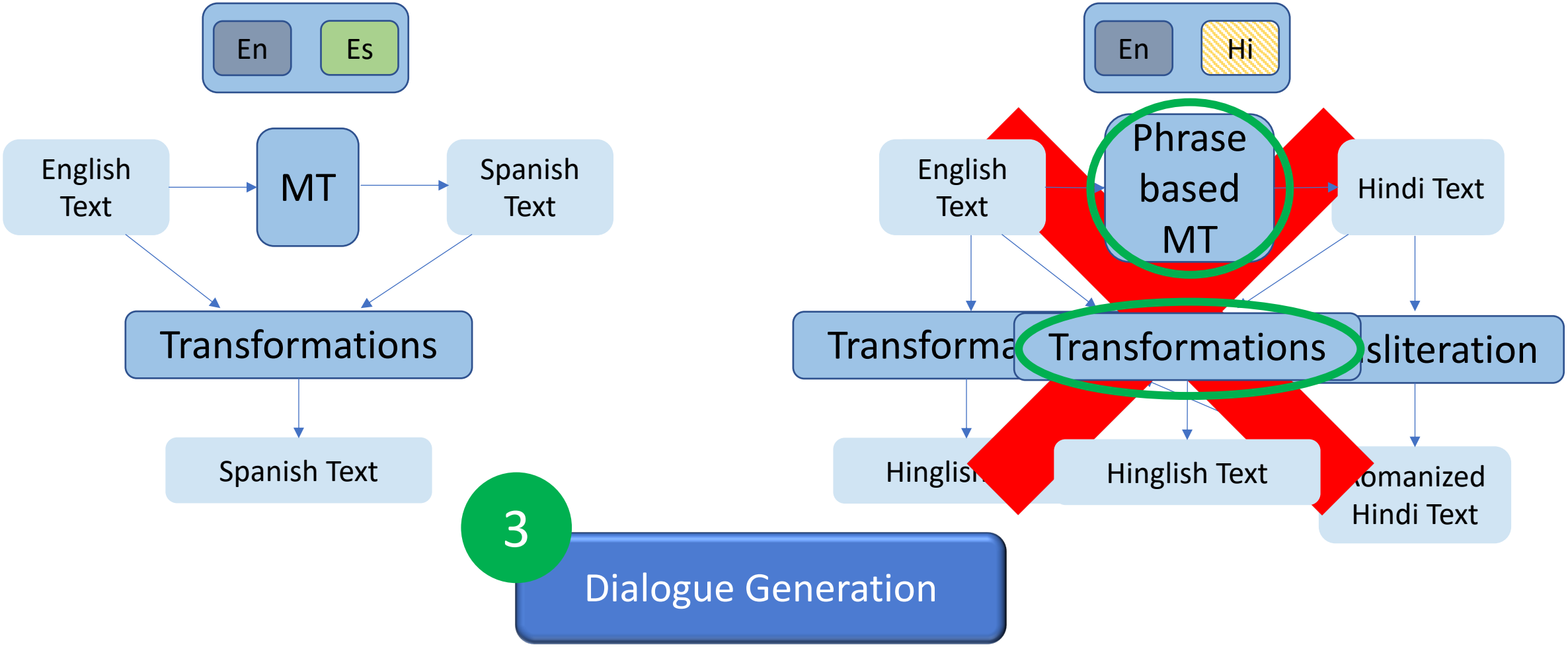
Verbs — *karta/karti* [he/she does]



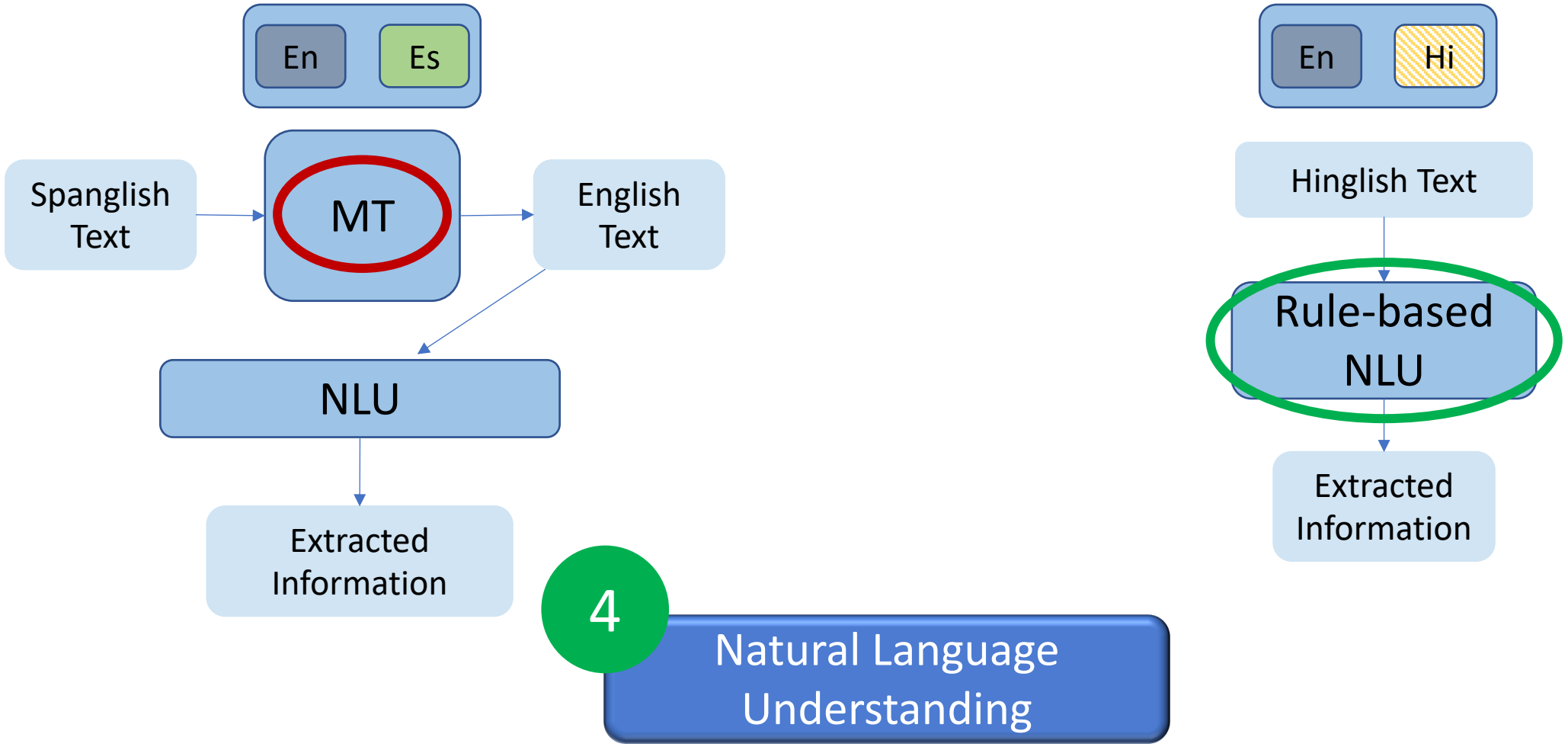
Gender Ratio (M:F)

2 Handling Gender Markings

Generalized Bilingual Dialogue System



Generalized Bilingual Dialogue System



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

Data Collection

Common dost koun hai?

Tum yaha ek aur online partner se baat karoge. Tumhare aur us online user mein sirf 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 dhoondhna hai!

Instructions

- Please **natural sentences** ka upyog kijiye 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 dhoondhna 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 iss task mein safal ho jaoge
- Agar samay khatam ho bhi jaata hai par tumne achi koshish ki, tab bhi **tumhe paise milenge.**
- **Kripiya dhyaan se chune**. Yadi tumne kisi galat dost ko chuna, toh tumhe agle 10 second tak koi aur dost ko chunne ka mauka nahi milega. Uske baad tumhe 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

Enter your message here

Tumhare dost / Your friends

#	kaam ki jagah work location	kaam ka samay work time	favourite phal favourite fruit
Select	machhaleeghar the aquarium	raat night	imli tamarind
Select	havaee adda the airport	raat night	seb apple
Select	machhaleeghar the aquarium	subah morning	santra orange
Select	machhaleeghar the aquarium	subah morning	tarbuj watermelon
Select	machhaleeghar the aquarium	raat night	aadoo peach
Select	machhaleeghar the aquarium	dopahar afternoon	imli tamarind
Select	machhaleeghar the aquarium	raat night	santra orange
Select	chidiyaaghar the zoo	raat night	seb apple
Select	havaee adda the airport	subah morning	tarbuj watermelon
Select	daak ghar the post office	subah morning	tarbuj watermelon

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]

Examples

Outline

- Introduction
- [C1] Bilingual Code-switching Dialogue System
- [C2] Data Collection
- [C3] Data Analysis
- Future Work

Data Analysis: Background

Code-Switching Strategies

Insertional CS [5]

Replace content words in matrix language with embedded language

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

Code-Switching Strategies

Alternational CS [5]

Switch from one language to another at a syntactic point

Kya tumhare paas koi dost hai jise seb pasand hai?

Kya tumhare paas koi dost *hai* who likes apple?

[Do you have any friend who likes apple?]

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

Data Analysis: Background

Code-Switching Strategies

Informality in CS [6]

Addition of discourse markers

Kya tumhare paas koi dost hai jise seb pasand hai?

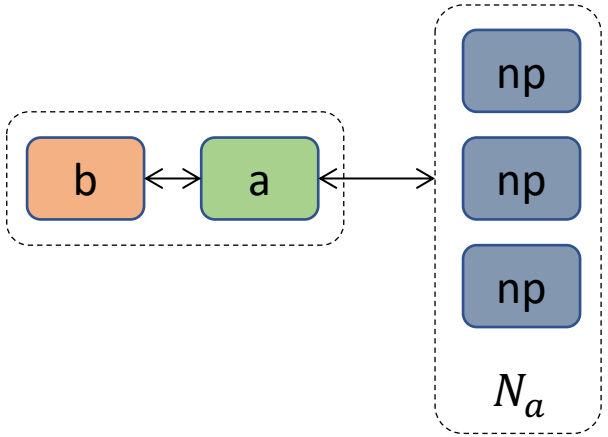
So *kya tumhare paas koi dost hai jise seb pasand hai?*

[Do you have any friend who likes apple?]

[6] Deborah Schiffrin. 1988. Discourse markers

Data Analysis: Background

Linguistic Accommodation



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

Global Accommodation [7]

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

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

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

[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%

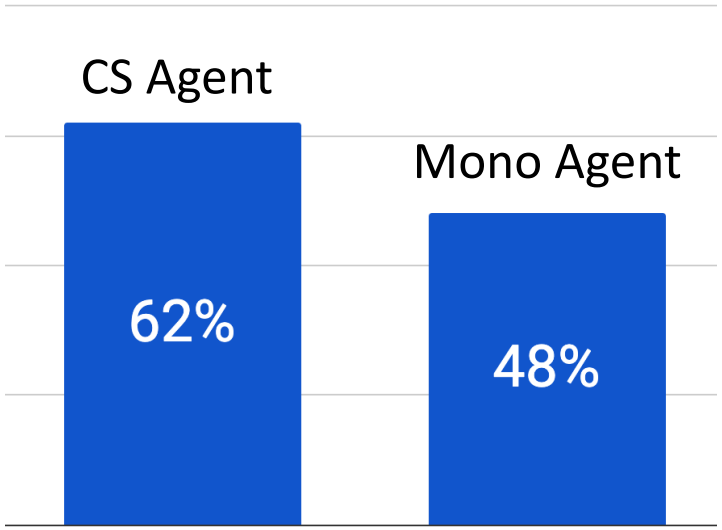
CS user utterances

92%

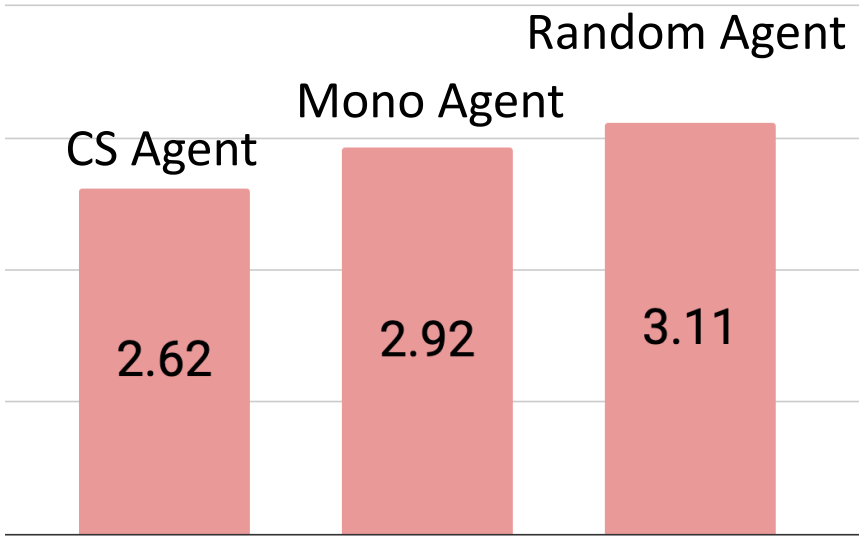
CS dialogues

A Our dataset elicits code-switching

Data Analysis



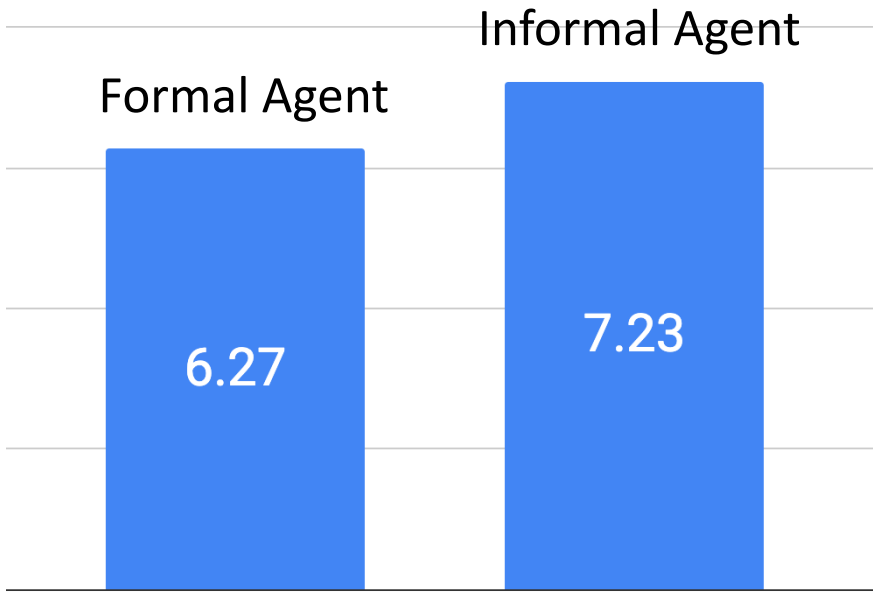
Task Success



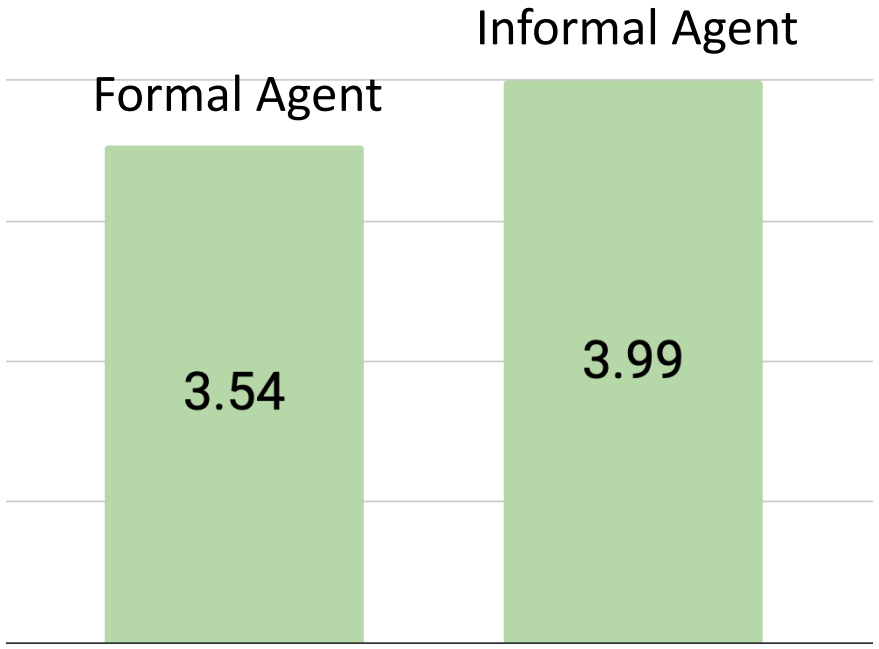
How non-native is your partner?

B Code-switching leads to better engagement

Data Analysis



Avg user utterance length



How human-like was your partner?

Avg Discourse Markers

1.87 → 2.44

C

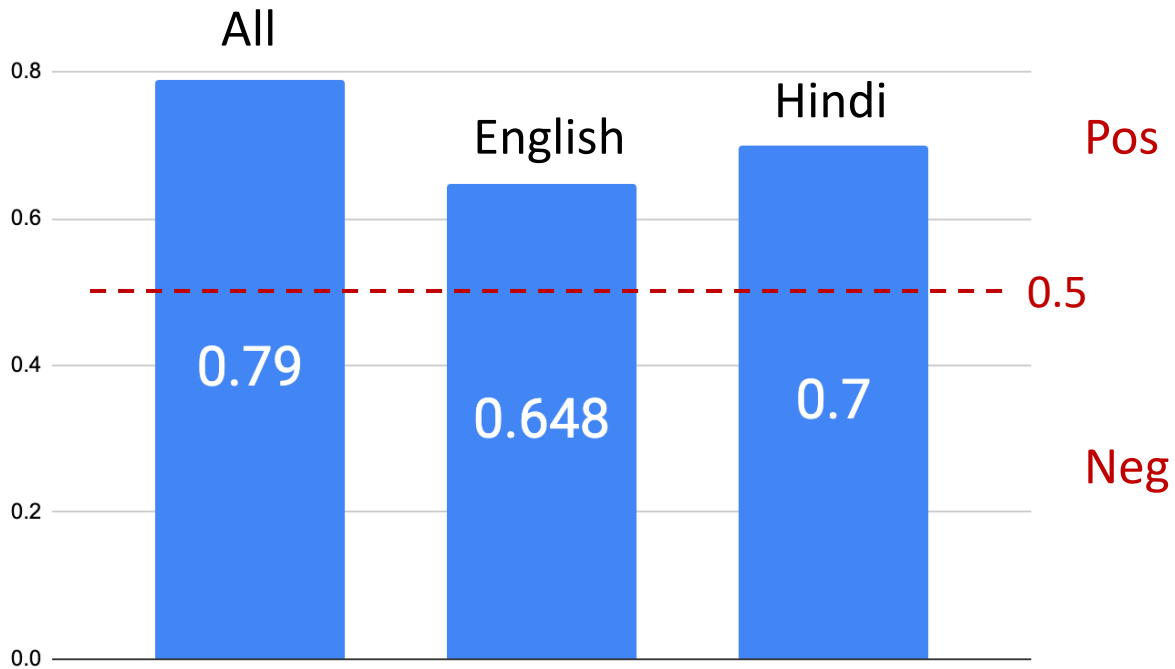
Informality improves dialogue quality

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?



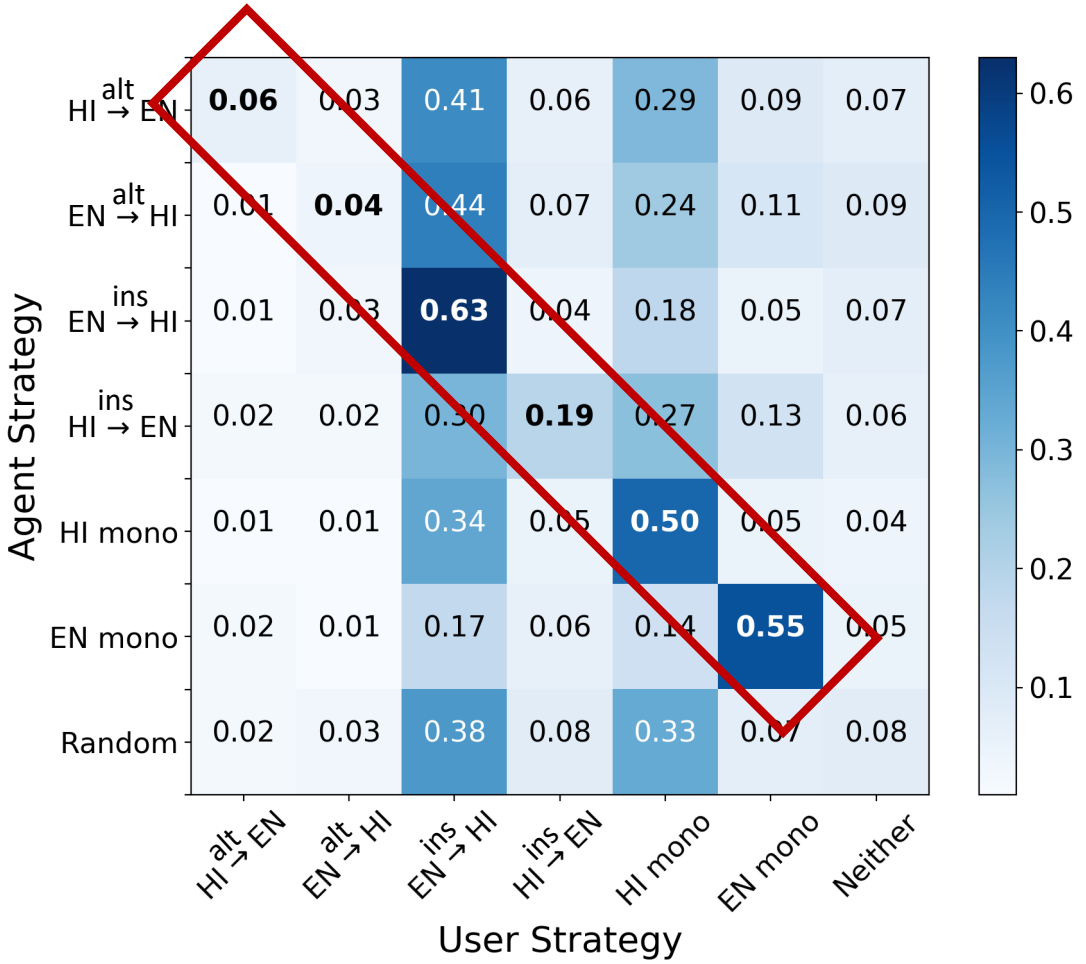
Global Accommodation Score

D 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



E

Agent's CS strategy positively influences users' CS strategy

Data Analysis

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

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

South Indian Speakers

- Ins Hi->En
- Alt En->Hi

North Indian Speakers

- Ins En->Hi
- Alt Hi->En

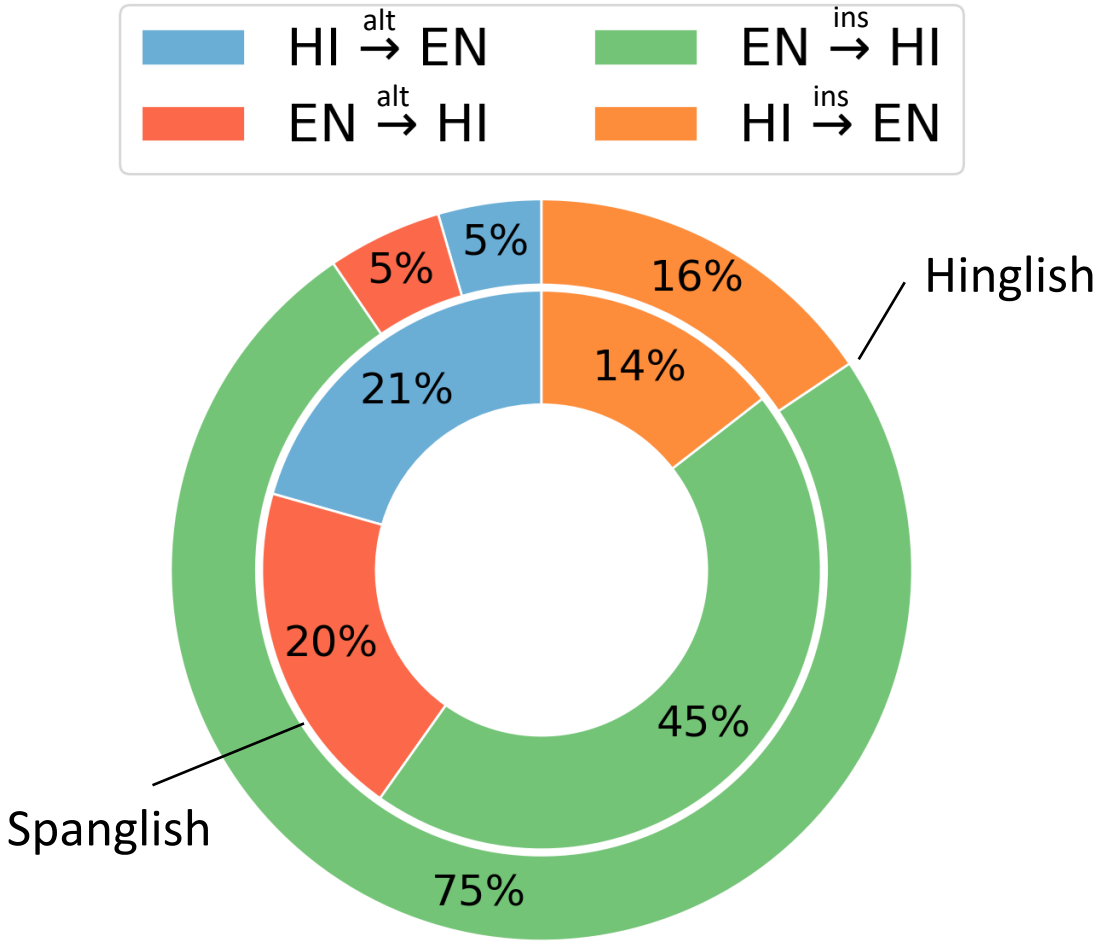
F Language Proficiency influences usage of CS strategies

Data Analysis

Equivalence Constraint: CS at point where syntactic rules are not violated [8]

Word Order

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



G Comparison of Spanglish and Hinglish

[8] Shana Poplack. 1980. Sometimes i'll start a sentence in english y termino en espafiol.

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]