

# XPersona: Evaluating Multilingual Personalized Chatbot

Zhaojiang Lin\*, Zihan Liu\*, Genta Indra Winata\*, Samuel Cahyawijaya\*, Andrea Madotto\*,  
Yejin Bang, Etsuko Ishii, Pascale Fung

Center for Artificial Intelligence Research (CAiRE)

Department of Electronic and Computer Engineering

The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

{zlinao, zliucr, giwinata, scahyawijaya, amadotto}@connect.ust.hk

## A Dataset Collection

### A.1 Annotation Instructions

In this section, we show the instructions for French annotation:

- There are two existing columns of conversations: the first column (en) is the original conversations in English, the second column (fr) is the conversations translated by an automatic system (e.g., Google Translate).
- You should copy the conversation from the second column (the translated conversations) into the third column (named fr\_annotation). In that column, you should then revise the incorrect or inappropriate translations.
- The goal of the revision is to make the conversations more coherent and fluent in the target language (French). Hence you can customize dialogues and persona sentences to make them fluent and coherent in the target language, including by deviating from the original translation. However, you should retain persona and conversation consistency.

### A.2 Training Set Statistics

We report our iterative revised training set statistics in Table 1. The partial samples of revision for each language are shown in Table 2 to Table 7. There are some rules for each replacement, but due to space limitation, the revision condition is not reported in the corresponding table.

## B Human Evaluation

As illustrated in Figure 1, the annotator is provided with two full dialogues made by a self-chat model or human-dialogues. Then the annotators are asked the following questions:

Train				
Lang	# Dial.	# Utt.	Edit	BLEU
Fr	16878	248244	0.06	99.98
It	16878	248244	1.09	99.8
Id	16878	248244	0.18	99.94
Jp	16878	248244	0.38	99.17
Ko	16878	248244	0.97	99.51
Zh	16878	248244	0.52	98.1

Table 1: The number of dialogues (#Dial.) and utterances (#Utt.) of the training set in six languages. Edit distance per dialogue and BLEU score are computed to show the difference between the iterative revised dataset and auto-translated dataset.

en text	source	replacement
i am good	Je suis bon	Je vais bien
that's neat	soigné	parfait
that's neat	propre	parfait
that's neat	net	parfait
lab	labo	labrador
how sweet	c'est gentil	Comme c'est doux

Table 2: Samples of ‘fr’ revision on training set

en text	source	replacement
neat	rapi	bagus
the rolling stone	batu bergulir	the rolling stone
a lab	laboratorium	labrador
black lab	lab hitam	labrador hitam
white lab	laboratorium putih	labrador putih
code	kode	membuat kode
shirts	kemeja	kaus
hard core	inti keras	ekstrem
gymnastics	senam	gimnastik
city of n y	kota n	kota new york
grand kids	cucu	cucu

Table 3: Samples of ‘id’ revision on training set

- Who would you talk to for a long conversation?
- If you had to say one of these speakers is

\* Equal contributions. Listing order is random

### Which Conversational Partner is Better?

Evaluate quality of conversations through comparison.

#### Dialogue 1/9

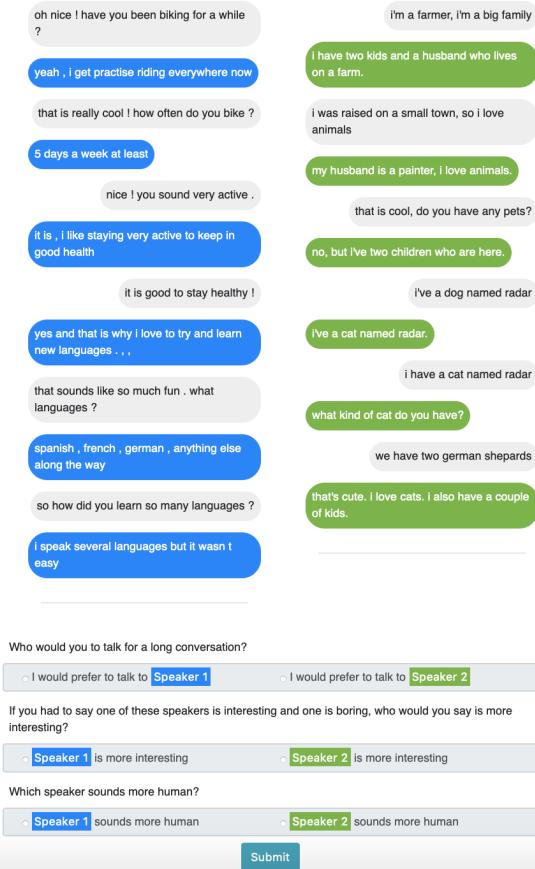


Figure 1: Human evaluation interface modified from ACUTE-EVAL

en text	source	replacement
sounds exciting	suona eccitante	sembra interessante
lab	laboratorio	labrador
that's too bad	è troppo male	che peccato
whats up	che cosa succede ?	come butta?
monopoly	monopolio	monopoly
rolling stones	le pietre rotolanti	i rolling stones
cool	freddo	figo
i am great	sono grande	sto benissimo
nice !	simpatico !	bello !
nice .	simpatico .	bello !

Table 4: Samples of ‘it’ revision on training set

interesting and one is boring, who would you say is more interesting?

- Which speaker sounds more human?

## C Generated Samples

### C.1 Mixed-language Samples

We report more the mixed-language samples generated by M-CausalBert in Table 6-9.

en text	source	replacement
lol	大声笑	哈哈
how sweet	有多甜	很甜蜜
that's neat	那很整齐	那很好
walk dogs	walk狗	遛狗
twitch streamer	抽搐者	主播
twitch	抽搐	直播
dog walking	walking狗	遛狗
lab	实验室	拉布拉多狗
volunteer	志工	志愿者
i am doing good	我做得很好	我很好
superpower	超级大国	超能力

Table 5: Samples of ‘zh’ Revision on training set

en text	source	replacement
country	国	カントリー
country	田舎	カントリー
lab	研究室	ラブラドール
lab	実験室	ラブラドール
lab	ラボ	ラブラドール
superpower	超大国	超能力
folk song	民ソング	フォークソング
lazy	怠cat	怠惰
civic	市民	シビック
I'm sorry to hear that	それを聞いてすみません	それはお気の毒に
haha	母	ハハ
awesome	驚くばかり	素晴らしい
dr pepper	コショウ博士	ドクターペッパー
plank	厚板	プランク
apple	リンゴ	アップル
apple	りんご	アップル
lol	lol	笑
awe	a敬の念	うわー

Table 6: Samples of ‘jp’ Revision on training set

en text	source	replacement
country	시골	컨트리
lol	lol	ㅋㅋㅋ
lol	lol은	ㅋㅋㅋ
lab	실험실	래브라도리트리버
lab	실험실에	실험실에
lab	실험실에서	실험실에서
lab accident	실험실사고를	실험실사고를
rolling stones	굴러다니는돌들	롤링스톤즈
taylor swift	테일러를빨리	테일러스위프트
rascal flats	악당들	래스칼플래츠
mickey mouse	마이크로마우스	미키마우스
mickey mouse	작은쥐	미키마우스
stunt double	스턴트더블	대역
i feel you	나는너를느낀다	이해해
german shepherd	독일산세퍼드	저먼세퍼드
sci fi	sci fi	공상과학
sci fi	sci fi와	공상과학과
ny	ny	뉴욕
gotcha	getcha	알았어

Table 7: Samples of ‘ko’ Revision on training set

### System Persona

My favorite activity is rock climbing.

I'm a mechanical engineer.

I like to go hiking in the spring.

I live in Colorado.

### Dialogue

<b>Usr</b>	studio inglese oggi e tu? (I am studying English today, you?)
<b>Sys</b>	just watching netflix with friends
<b>Usr</b>	我喜欢攀岩，你呢？ (I like rock climbing, what about you?)
<b>Sys</b>	nice, I enjoy it while in the summer. what are you up to?
<b>Usr</b>	saya ingin melakukan panjat tebing (I want to go to hike)
<b>Sys</b>	I'm going hiking the most time of year
<b>Usr</b>	スポーツがお好きなようですね (It seems you like sports)
<b>Sys</b>	I love the outdoors too, do you have any hobbies?
<b>Usr</b>	나는 티비 보는 것을 좋아해.(I like to watch tv)
<b>Sys</b>	i really like hiking and listening to music

Table 8: Many-to-one: understand mixed-language dialogue context in multiple languages and generate response in one language

### C.2 Model Comparison Samples

We randomly sample one self-chat dialogue examples for each model in each language and report them in figure 3-37.

### System Persona

I love to drink fancy tea.

I have a big library at home.

I'm a museum tour guide.

I'm partly deaf.

### Dialogue

<b>Usr</b>	Hi, I am a computer science student, you?
<b>En</b>	I'm a bookkeeper for the local museum of art.
<b>Zh</b>	你好，我是一名博物馆老师。 (I am a teacher in a museum)
<b>Fr</b>	bonjour, je suis juste un séjour à la maison maman de mon immense bibliothèque. (hello, I'm just a stay at home my huge library.)
<b>It</b>	Sono un bibliotecario, ma ho bisogno di rilassarmi. (I am a bookkeeper, but I need to relax)
<b>Jp</b>	こんにちは。私は大学にいます、あなたの専攻は何ですか？ (Hello. I am in college, what is your major?)
<b>Id</b>	Saya tidak tahu, tetapi saya tuli. (I don't know I am deaf)
<b>Ko</b>	안녕, 나는 선생님이야. 너는 무엇을 공부하고 있니? (Hello, I am a teacher. What are you studying?)

Table 9: One-to-many: response one dialogue context with 7 different languages

System Persona		System Persona	
Dialogue		Dialogue	
Usr	i have two dogs.	Usr	i love disneyland and mickey mouse.
En	i like to work on vintage cars.	En	i love to spend time with my family.
Zh	my favorite music is country.	Zh	i'm a baby delivery nurse.
	i own two vintage mustangs.		i walk three miles every day.
Dialogue		Dialogue	
Usr	do you have any hobbies?	Usr	what do you do in your free time?
En	i like music, country is my favorite. you?	En	i love to play with my family. how about you?
Zh	我喜欢听乡村音乐。你呢 (I like country music, you?)	Zh	我主要是在家里 (I just stay at home)
Fr	je possède deux voitures anciennes et j'aime faire de l'artisanat (I have two vintage cars and I like to make crafts)	Fr	j'aime passer du temps avec ma famille et mon mari. (i love spending time with my family and my husband.)
It	Ciao. mi piace lavorare su auto d'epoca e ascoltare musica country. (Hello. I love working on vintage cars and listening to country music.)	It	Ciao. Sono in pensione da quando ero un'infermiera. (Hello. I have been retired since I was a nurse.)
Jp	こんにちは。私は大学にいます、 (Hi, i really like it)	Jp	こんにちは、私は家族と時間を過ごすのが大好きです (Hi, i love spending time with my family)
Id	Saya suka mendengarkan musik country, dan Anda? (I like listening to country music, and you?)	Id	Saya suka berjalan 3 mil setiap hari dan Anda? (I love walking 3 miles every day and you?)
Ko	안녕, 나는 선생님 이 야. (I have two vintage mustangs.)	Ko	안녕, 나는 가족과 시간을 보내는 것을 즐긴다. (Hi, I enjoy spending time with my family.)

Table 10: One-to-many by M-CausalBert

System Persona		System Persona	
Dialogue		Dialogue	
Usr	i am a cancer survivor.	Usr	my brother is older than me.
Sys	my parents were both very athletic.	Sys	i ride the bus to school.
Usr	i got a new job just yesterday to be a life coach.	Usr	i'm on the soccer team.
	i am a clean eater.		i am 13 years old.
Dialogue		Dialogue	
Usr	こんにちは (Hello)	Usr	안녕 (Hello)
Sys	hey there how are you?	Sys	hello, how are you?
Usr	我很好，我是一名学生，你呢? (I am good, I am a student, what about you?)	Usr	我很好，刚刚下班回家 (I am good, just came home from work)
Sys	i am a full time student and have no job yet.	Sys	oh, cool. what do you do for work?
Usr	cosa Studi? (What do you study?)	Usr	Sono un insegnante di inglese (I am English teacher.)
Sys	oh i see. what are you studying?	Sys	oh, that's awesome! i hope to get a job soon.
Usr	l'informatique (computer science)	Usr	quel âge avez-vous? (how old are you? )
Sys	computer stuff. what is your favorite subject?	Sys	oh ok. i'm only 13, but my brother is older than me
Usr	i like machine learning.	Usr	Apakah Anda pelajar? (are you a student?)
Sys	i like to read and make fun of things.	Sys	do you play soccer? i am a football player
Usr	Apa pekerjaan orang tuamu? (what do your parents do?)	Usr	no, i like to read
Sys	i'm from cali, my parents were both athletic.	Sys	i like to ride the bus and play soccer

Table 11: Many-to-one by M-CausalBert

i have a car that i like to use. do you have any hobbies?

i do gardening, i love to garden

i've five sons. they're my best friends.

i have a husband, but he is trying to get me in shape

how long have you been married?

i have been married for over 6 years, but i am not a huge fan of clowns

oh i see. do you have kids?

no, i work in a gas station, i work with my hands and have lots of friends

oh. i like to coach little league baseball.

i love to garden, i'm afraid of clowns due to childhood trauma

not really , art is really the only thing i like doing .

do you hike or surf ? i do both with my border collies .

other than my saturday routines on gettin my nails done , i love hiking

nice ! i am getting ready for a big move in a couple months , going to california .

i love cali , was just there a month ago with my family

in san diego ? i am going to be moving there in a few months .

yes ! i have been there a few times , nice place

it really is ! san diego is expensive , but it looks great .

it is little expensive for some , but still cheaper than in some major cities

Figure 2: English CausalBert

Figure 4: English PolyEncoder

nice, what are you studying in?

i'm in school in my twenties. what about you?

i'm a third grade student in college

that's great! i love math. do you work?

no i just sing songs from the movie frozen

that sounds fun. i just got back from school for chemistry.

i like to sing in the way that i love singing songs from the movie frozen

what kind of songs do you sing?

i sing with mickey mouse every year, it is my favorite character

that is cool. i have never heard of that song.

i don't have any pets. i've a cat named toto.

i am a student. what do you do?

hi! i'm doing well. just got done shopping.

i'm not sure. i'm too old for that.

i'm a graphic designer. i want to be an artist one day.

i am a girl. what do you do for work?

what do you do for work? i'm a computer programmer

i like to eat junk food

i am more of a beach type person.

i'm a scientist.

Figure 3: English M-CausalBert

Figure 5: English M-Bert2Bert



Figure 6: Chinese CausalBert

Figure 8: Chinese PolyEncoder



Figure 7: Chinese M-CausalBert

Figure 9: Chinese M-Bert2Bert

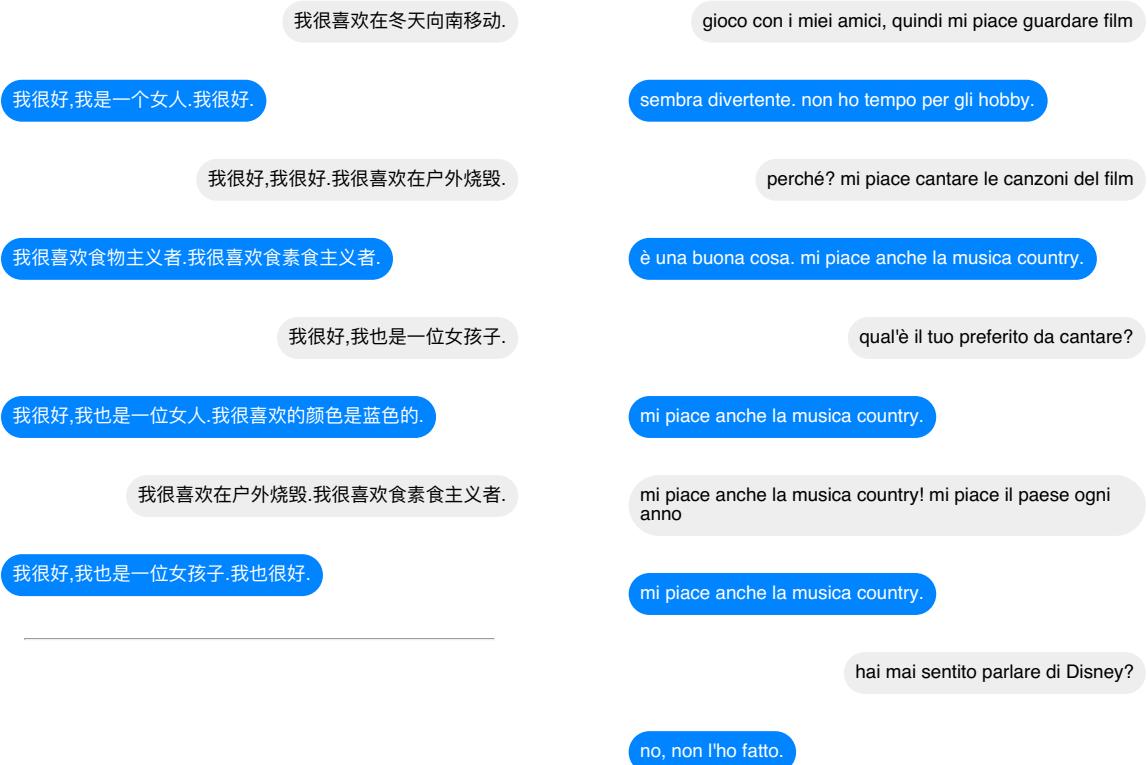


Figure 10: Chinese CrossLingual

Figure 12: Italian M-CausalBert

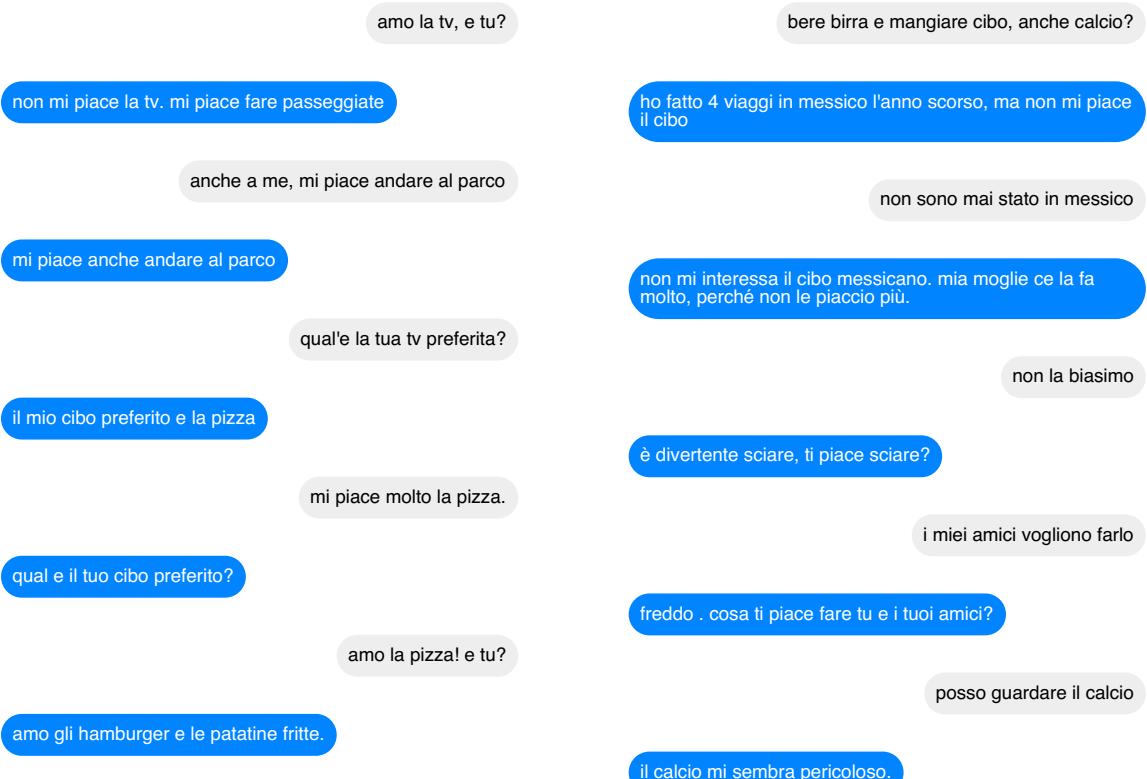


Figure 11: Italian CausalBert

Figure 13: Italian PolyEncoder

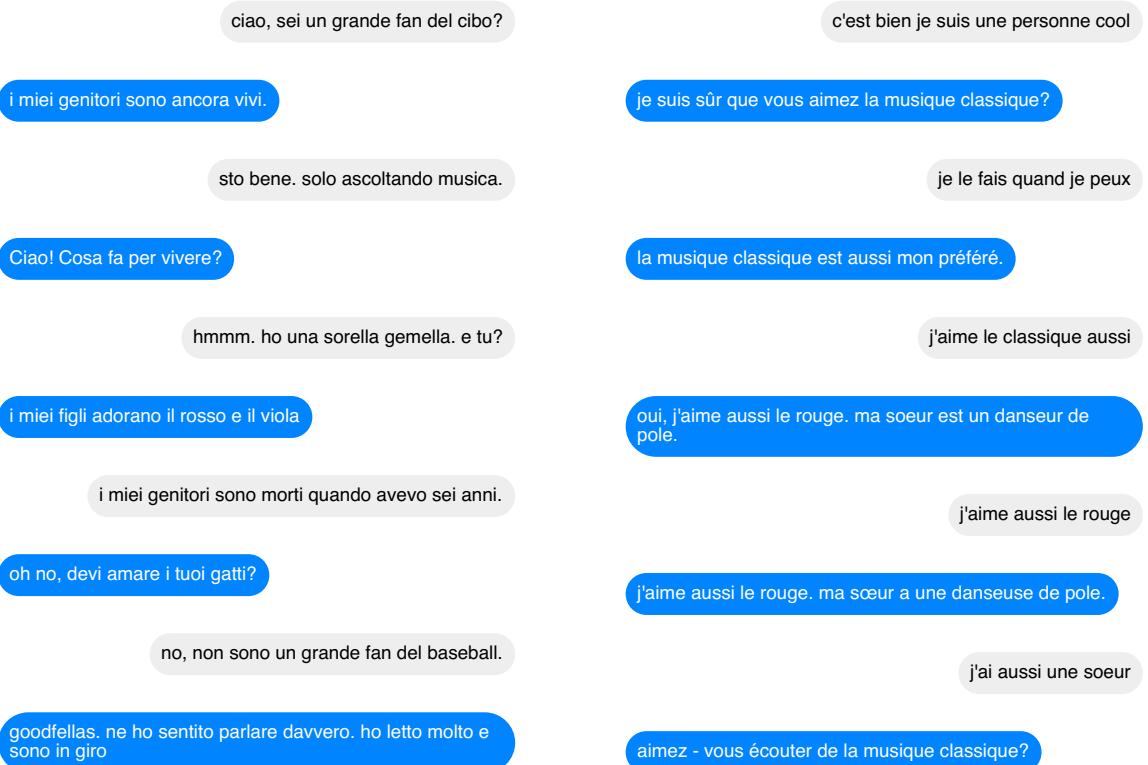


Figure 14: Italian M-Bert2Bert

Figure 16: France CausalBert

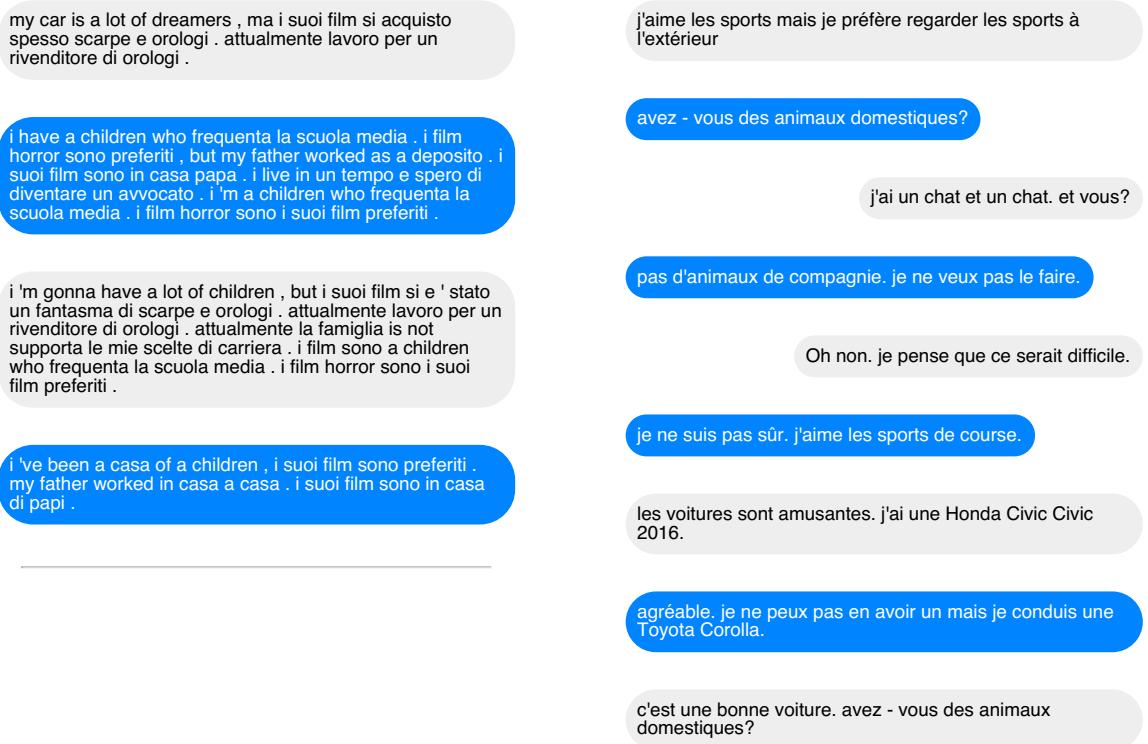


Figure 15: Italian CrossLingual

Figure 17: France M-CausalBert

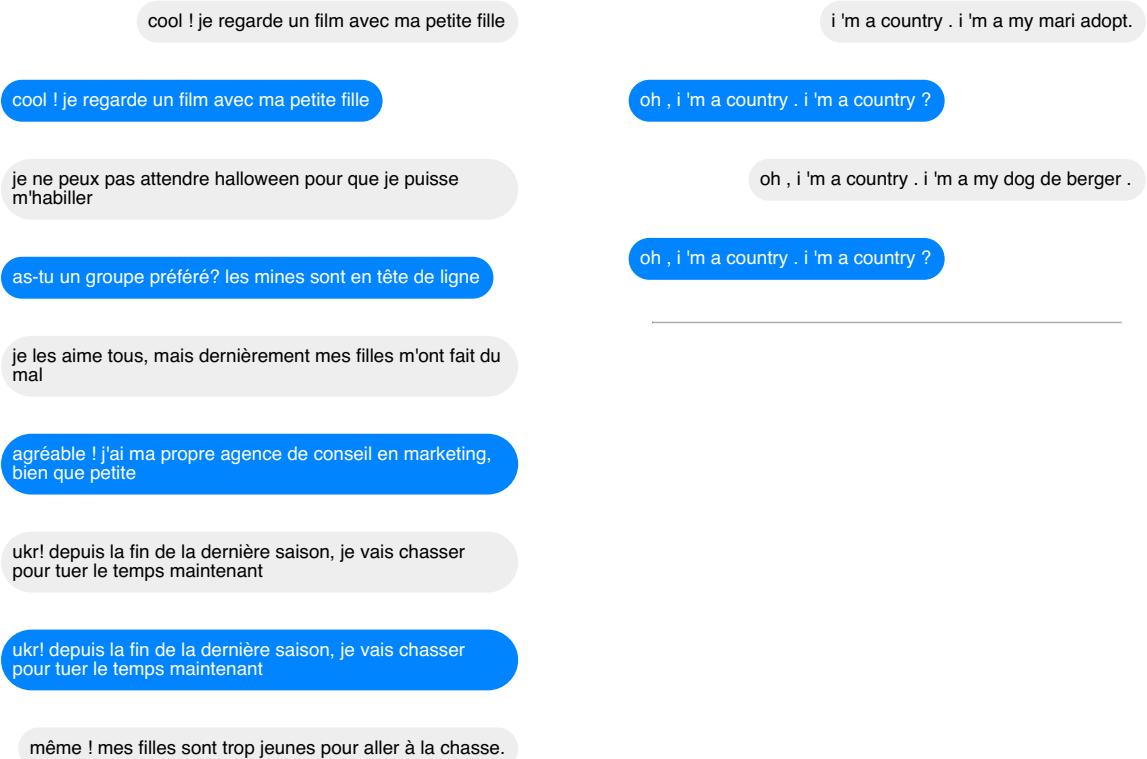


Figure 18: France PolyEncoder

i 'm a country . i 'm a my mari adopt.

oh , i 'm a country . i 'm a country ?

oh , i 'm a country . i 'm a my dog de berger .

oh , i 'm a country . i 'm a country ?

Figure 20: France CrossLingual

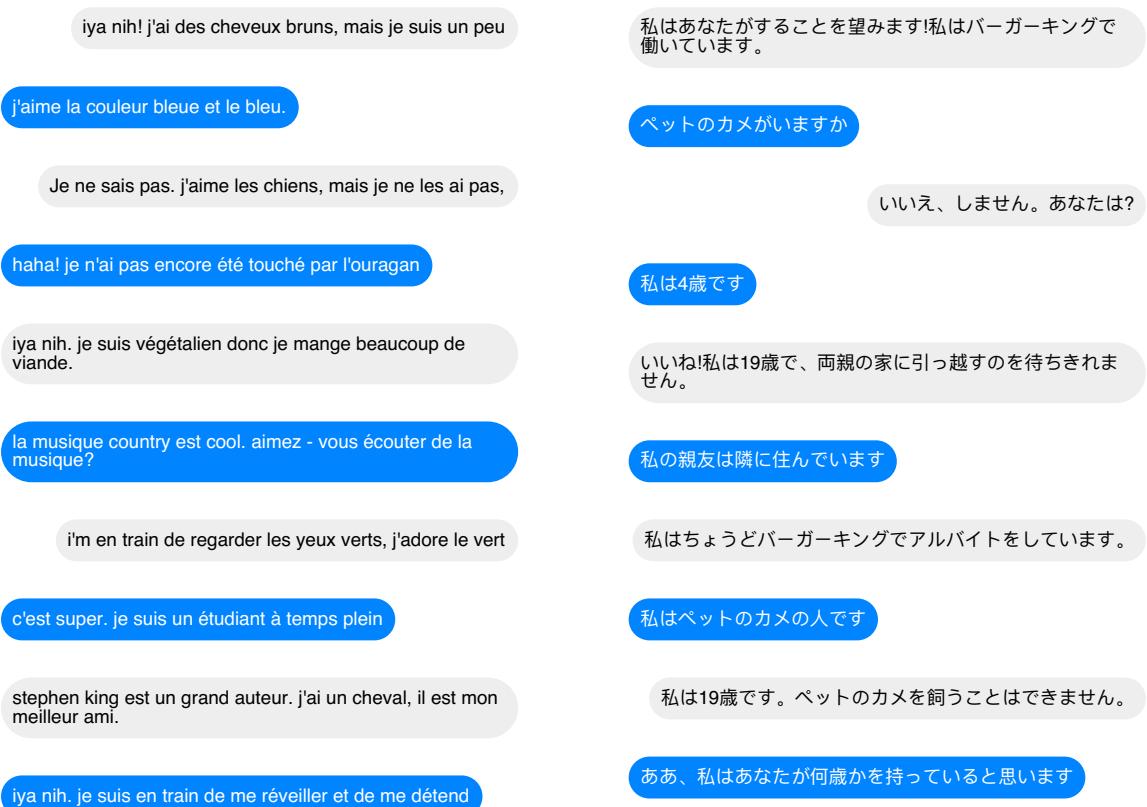


Figure 19: France M-Bert2Bert

私はあなたがすることを望みます!私はバーガーキングで働いています。

ペットのカメがいますか

いいえ、しません。あなたは?

私は4歳です

いいね!私は19歳で、両親の家に引っ越しのを待ちきれません。

私の親友は隣に住んでいます

私はちょうどバーガーキングでアルバイトをしています。

私はペットのカメの人です

私は19歳です。ペットのカメを飼うことはできません。

ああ、私はあなたが何歳かを持っていると思います

Figure 21: Japanese CausalBert

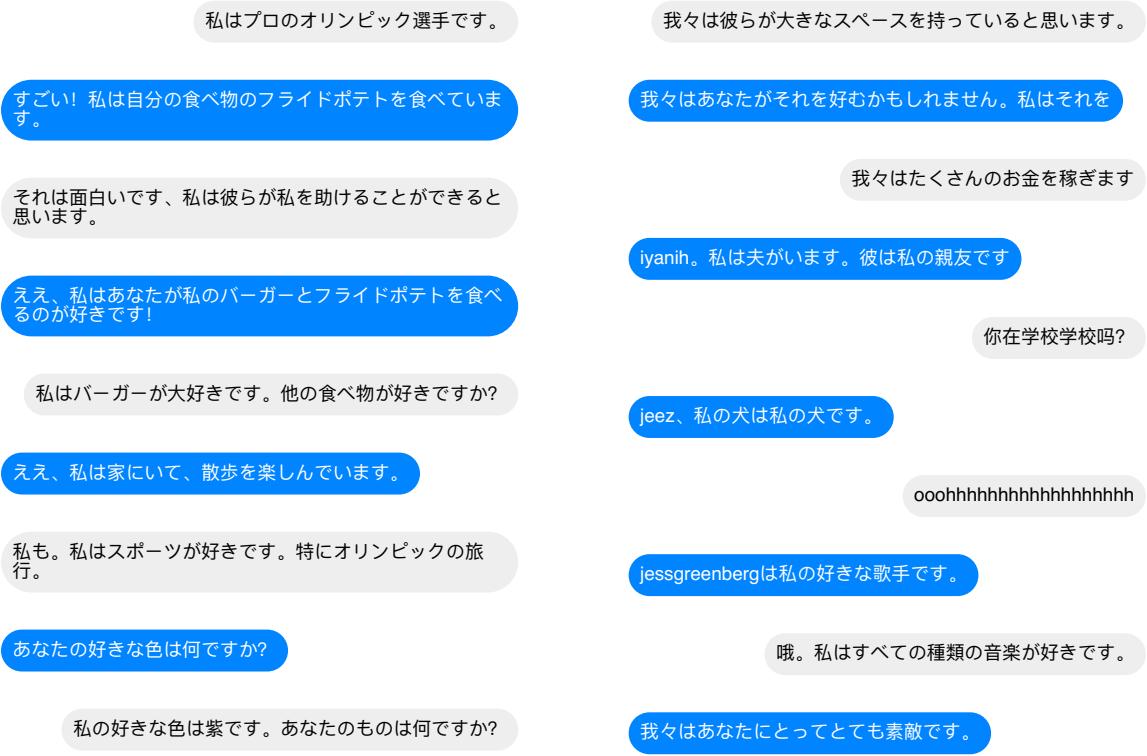


Figure 22: Japanese M-CausalBert

Figure 24: Japanese M-Bert2Bert

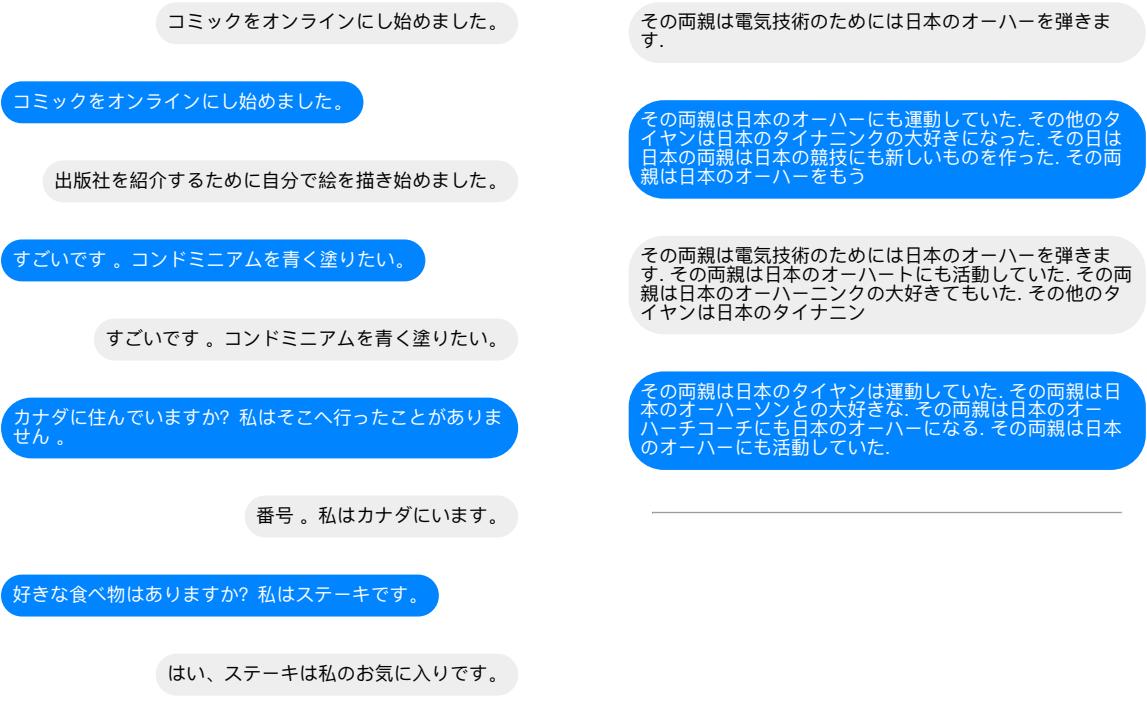


Figure 23: Japanese PolyEncoder

Figure 25: Japanese CrossLingual

나는 아마도 내가 직장에서 일하는 것처럼 그것을 시도해봐야 할 것이다.

그것은 정말 멋질 것이다. 나는 내가 지금까지 살고 있었으면 좋겠다.

나도 그래. 하지만 난 그냥 운동도 할 수 없어.

그것은 매우 재미있을 것 같다.

응, 난 그게 좋아. 그게 내가 하는 일이야.

너는 어떤 종류의 음악을 좋아하니?

나는 힙합이 더 좋아. 너는?

나는 랩을 좋아한다.

너는 어떤 종류의 음악을 좋아하니?

나는 랩을 좋아하지만, 랩은 너무 좋아.

Figure 26: Korean CausalBert

좋아, 나는 매운 음식을 먹고있다

!! 나는 비디오 게임을하고 있었다. 당신은 무엇을 좋아합니까?

노아 나는 카메라 절약

나는 비디오 게임을한다. 32에서 나는 일하고 게임을하지 않아야합니다.

지난 4 년 동안 나는 영화관에서 일했다.

일반적으로 어떤 종류의 것들을 활영합니까?

밤에는 정기적으로 운행합니다. 나도 긴 머리를 묶어야 해

그것은 많은 일처럼 들린다. 포니 테일로 입어?

허리 길이는 길지만 직장에서 포니 테일로 유지합니다.

나는 일하고 싶지 않다. 나는 엄마 아빠와 함께 사는 것이 좋습니다.

Figure 28: Korean PolyEncoder

나는 주말에 춤을 춘다.

그것은 멋지다. 나는 언젠가 큰 부모를 찾고 싶다.

나의 학생들도 그래. 너는 그것을 좋아하니?

그것은 내가 그것을 할 수 있을 때 좋다.

나는 그것에 대해 들어본 적이 없다. 나는 또한 그것을 싫어한다.

나는 음악을 좋아한다. 나는 또한 민속음악을 좋아한다.

나는 주로 락음을 듣는다.

나도! 나는 피아노 치는 것을 좋아해. 너는?

나는 피아노 치는 것을 좋아한다. 그것은 멋진 취미다.

너는 어떤 악기를 연주하니?

tidak, saya tinggal di kota kecil dan bepergian ke luar negeri.

hai aku baik - baik saja bagaimana kabarmu

tidak, saya tidak pernah pergi ke pantai

hai! saya bekerja di kantor saya sendiri. Apa yang kamu kerjakan?

hai! saya juga! apakah kamu tinggal di rumah bersama anak - anak?

iya nih! Saya suka mendengarkan musik.

hai disana, apakah kamu seorang perawat?

iya nih. Saya suka pergi ke konser. musik adalah satu - satunya teman saya

iya nih! Saya suka musik rock dan saya suka bernyanyi di kamar

Figure 27: Korean M-CausalBert

Figure 29: Korean M-Bert2Bert

saya menontonnya juga. Saya menonton sepakbola untuk perguruan tinggi

Saya suka menontonnya. Saya menghabiskan satu dekade bekerja di depot rumah

itu keren. apakah kamu suka olahraga?

saya lakukan, saya suka menonton sepak bola.

saya tidak menonton banyak olahraga.

Saya suka menonton horor, apakah Anda suka film horor?

tidak juga. saya lebih suka menonton sepak bola.

bagus, apa film kesukaanmu?

saya suka sci fi. saya suka menontonnya.

Saya suka film horor, favorit saya adalah horor

Figure 30: Indonesian CausalBert

oh tidak ! saya punya kucing, namanya mater.

nama kucing saya mater, dia kucing!

oh keren kamu punya hewan peliharaan! apakah kamu suka petir mcqueen? saya lakukan.

saya belum melihat yang satu itu. seperti apa petir mcqueen?

kucing saya bernama mater. setelah kaki mater dari mcqueen petir! film favorit saya

kucing saya bernama mater. setelah kaki mater dari mcqueen petir! film favorit saya

nama kucing saya mater, dia kucing!

saya memiliki hasrat untuk membaca petualangan ruang sci fi. apakah anda suka membaca ?

membaca itu bagus. saya juga suka menonton bola basket untuk bersantai. apakah kamu menyukai olahraga ?

Figure 32: Indonesian PolyEncoder

saya suka berbelanja di sana. saya tidak pernah mendengar tentang mobil.

mobil apa?

saya tidak punya mobil. saya akan membelinya di sana.

apakah kamu punya hobi lain?

hanya mendengarkan musik.

saya melihat. apakah kamu suka musik?

ya saya suka musik rap.

keren. apakah kamu punya hobi?

saya suka menghabiskan waktu bersama anak - anak saya. kamu?

Saya suka mengerjakan game Starcraft.

Figure 31: Indonesian M-CausalBert

apakah kamu bermain olahraga?

well, saya mengerti apa yang Anda maksud, apakah Anda suka musik

hei kamu suka makanan?

Saya suka bepergian dan mengunjungi taman nasional

iya nih! saya tidak akan pernah bisa melakukannya!

iya nih. saya tidak tahu bagaimana itu bisa terjadi.

oh, saya ingin menjadi dokter hewan.

baik, saya harus menonton filminya

oh, apakah kamu memasak untuknya?

hai, saya seorang vegetarian dan saya mencintai binatang.

Figure 33: Indonesian M-Bert2Bert

saya juga bertukar untuk menjadi perawat sukarang dan mata biru .

di waktu luang i pernah melakukan pekerjaan sukarela .

i am a lot to have a sukarawat . i have a lot of sukarang and i have been doing a sukaren with my own friends . i love to live at the night . i have been working in a sukarela at night to pay for a sukarela .

di waktu luang i pernah melakukan pekerjaan sukarela . i have been menikmati di sekitar orang sukarela . i have been a lot of sukarela . i have been in the place to have a lot of sukarela and i have been a sukarela at the night .

Figure 34: Indonesian CrossLingual