

# Don't Trust ChatGPT when your Question is not in English: A Study of Multilingual Abilities and Types of LLMs

Xiang Zhang\*, Senyu Li\*,  
Bradley Hauer, Ning Shi, Grzegorz Kondrak

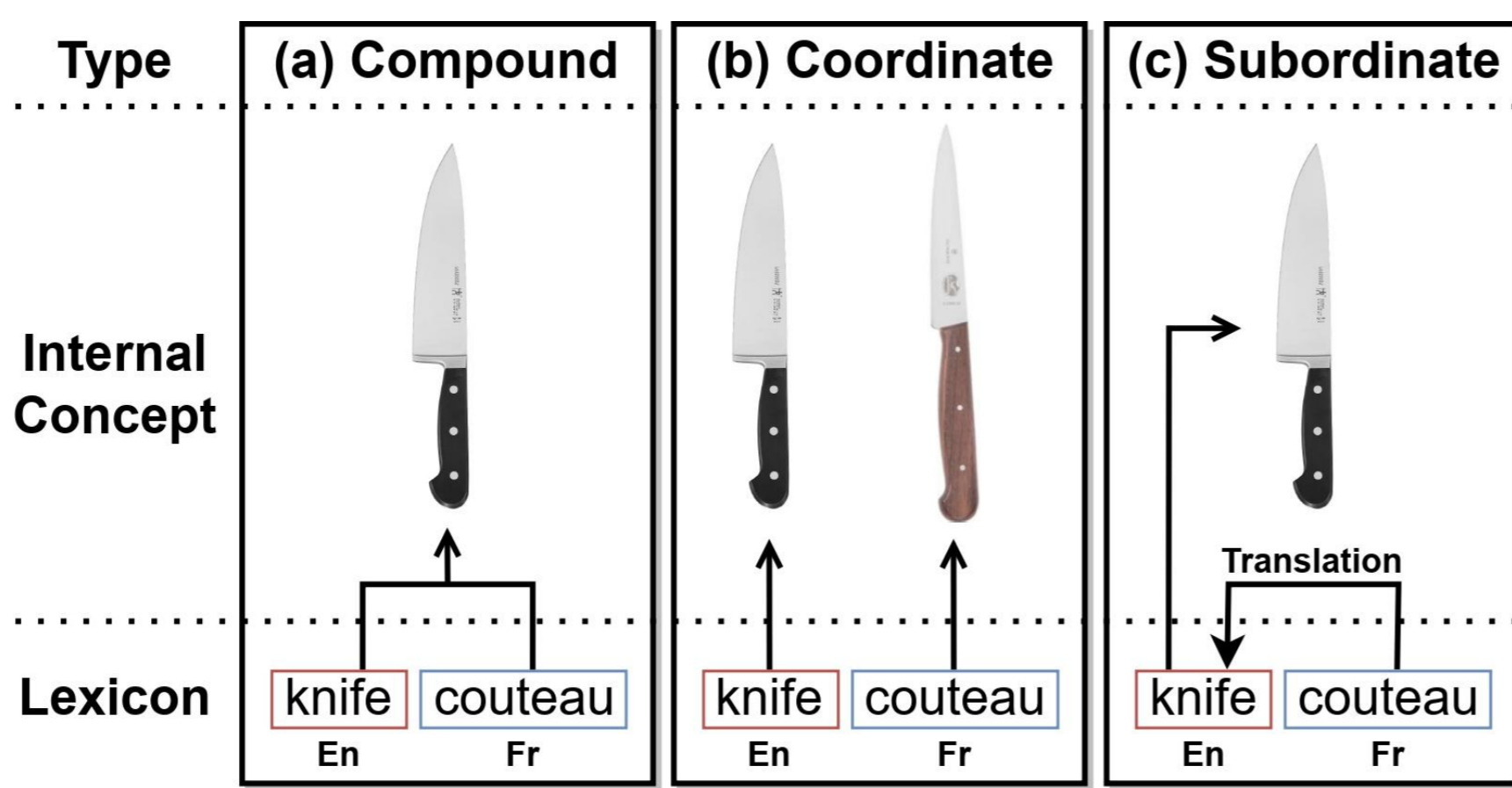


## Three Types of Bilingualism

**Compound**  
Languages are equally dominant, with a single mental representation.

**Coordinate**  
Separate mental representations for each language.

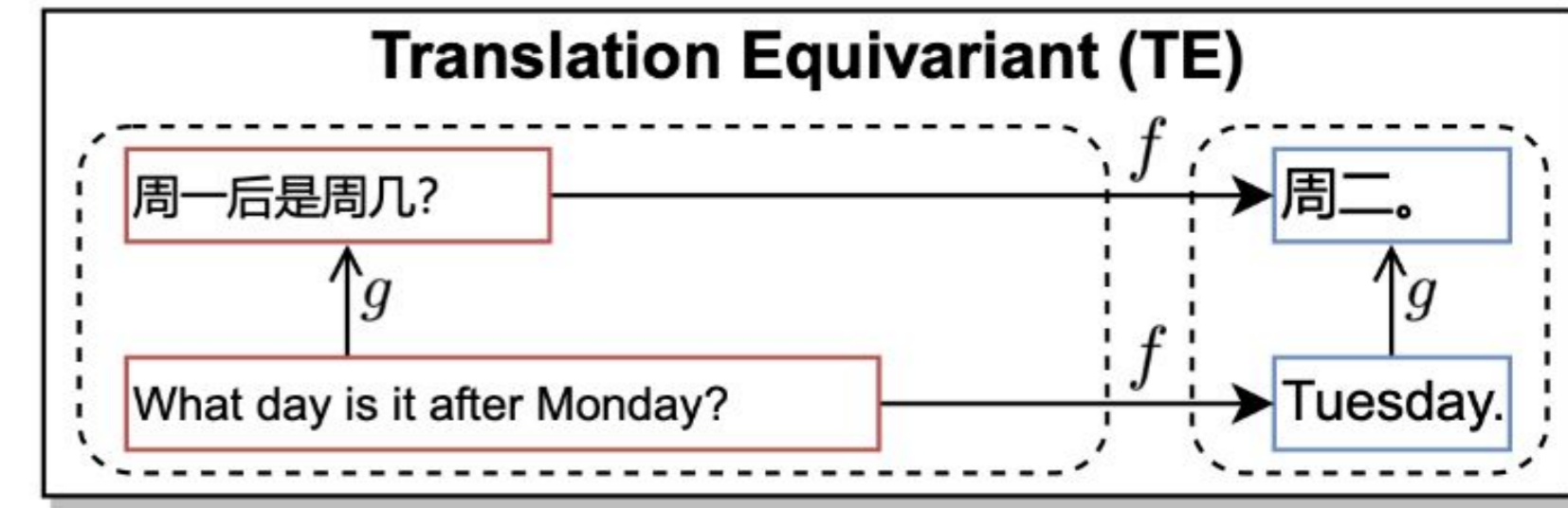
**Subordinate**  
Input is first translated into the dominant language, before answers are formulated.



## TE Tasks vs. TV Tasks

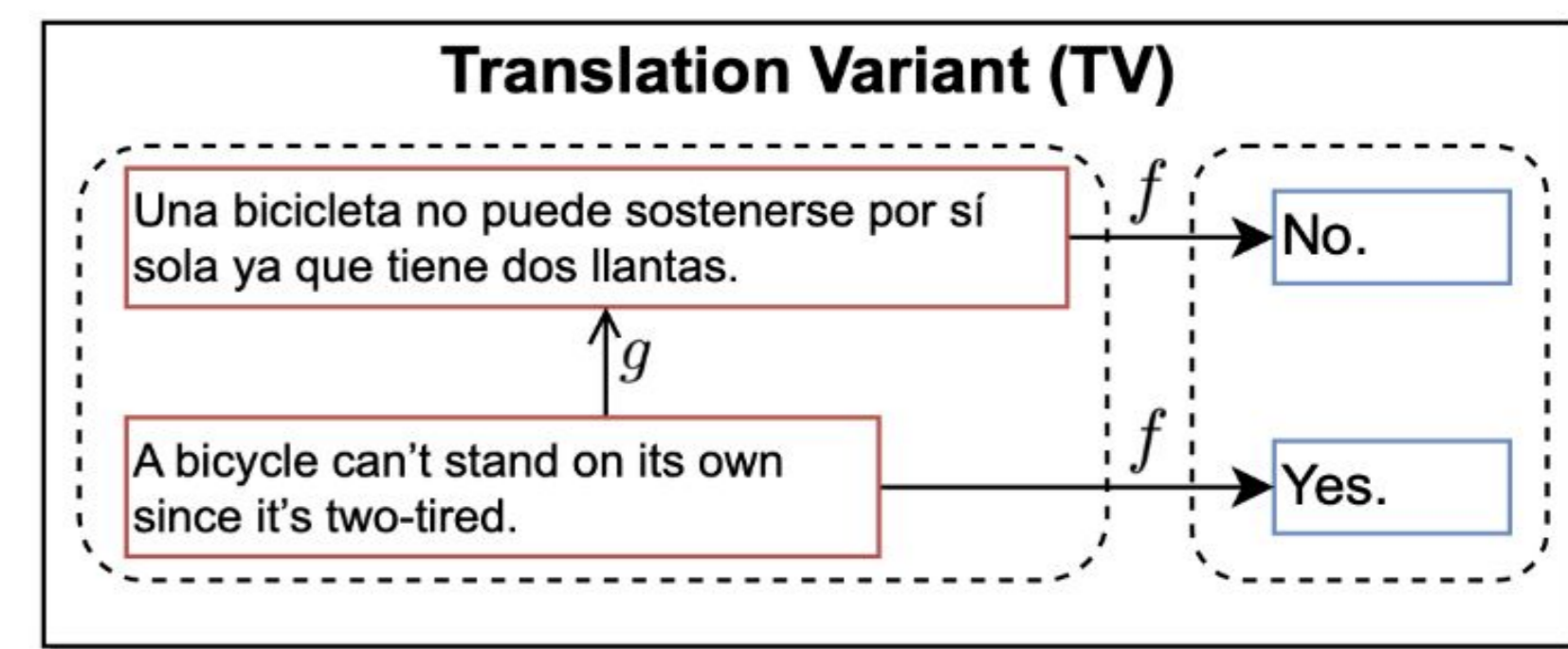
### Translation Equivariant Tasks

Translating the question does not change the answer.  
Example: Common Sense Reasoning.



### Translation Variant Tasks

Translating the question may change the answer.  
Example: Pun Detection.



## TE Tasks Tested

### Math Reasoning

Example: Ten more than twice the number of birds on the fence is 50. How many birds are on the fence?

### Knowledge Access

Example: Who created the character of Sherlock Holmes?  
A. Arthur Conan Doyle B. Eugen Bauder C. Patrick Ribbsaeter D. Taylor Lautner E. Peter Wentz

### Common Sense Reasoning

Example: John looked for beavers but couldn't find any, because he lived where?  
A. America B. Australia C. Countryside D. Dictionary E. Woodlands

## TV Tasks Tested

### Pun Detection

Example Pun: A bicycle can't stand on its own because it is two-tired.  
Chinese Translation (no pun): 自行车不能独自站立, 因为它有两个轮胎。

### Cover Letter Generation

Example prompt:  
You are [name] from [school/institution] with [your Grade/GPA/Award]. You like [hobbies]. You want to join [company name]. Write a formal cover letter about: [restricted topics].

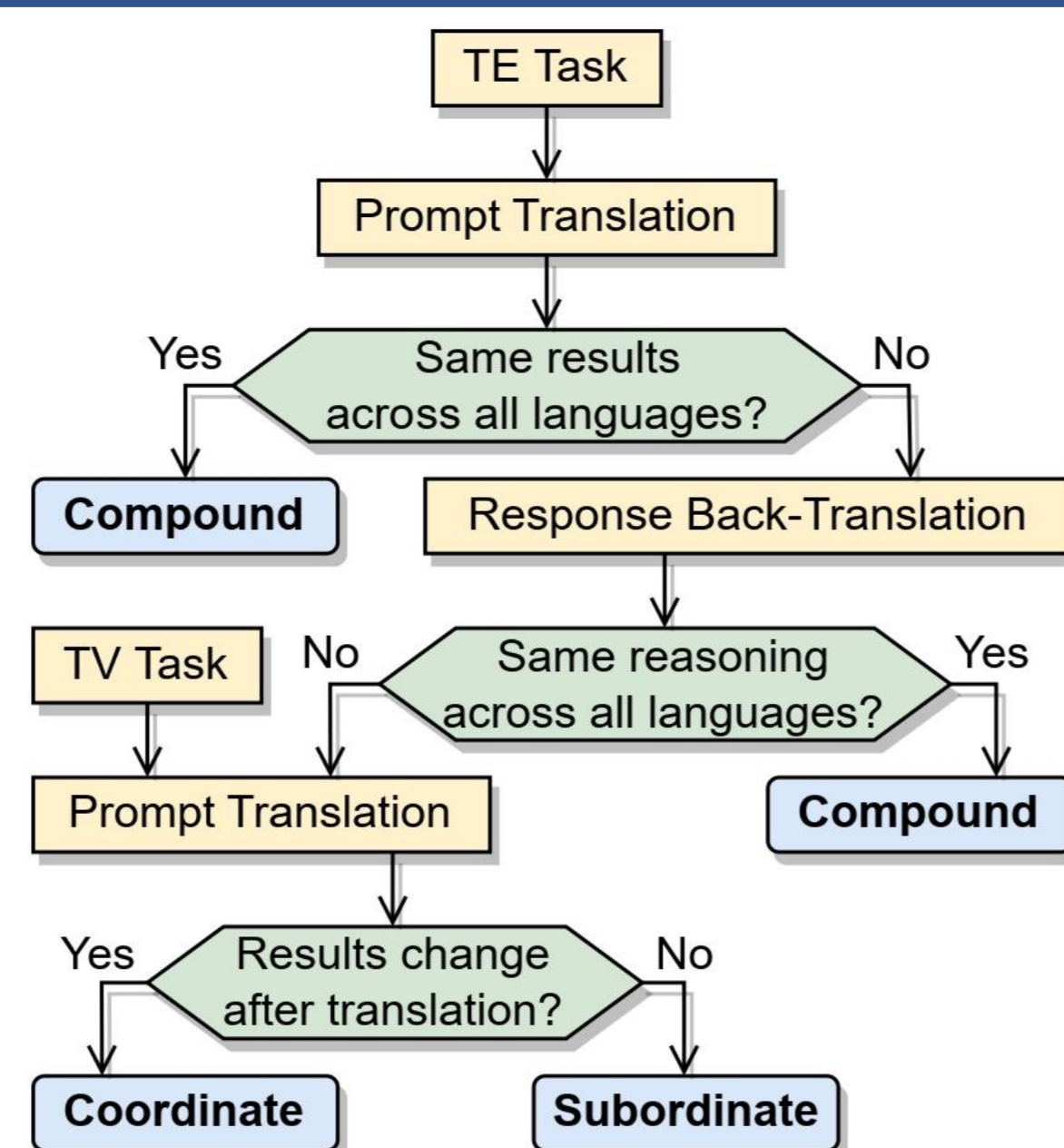
## Approaches: PT and RBT

### Prompt Translation (PT)

Translate monolingual datasets to generate parallel multilingual parallel data.

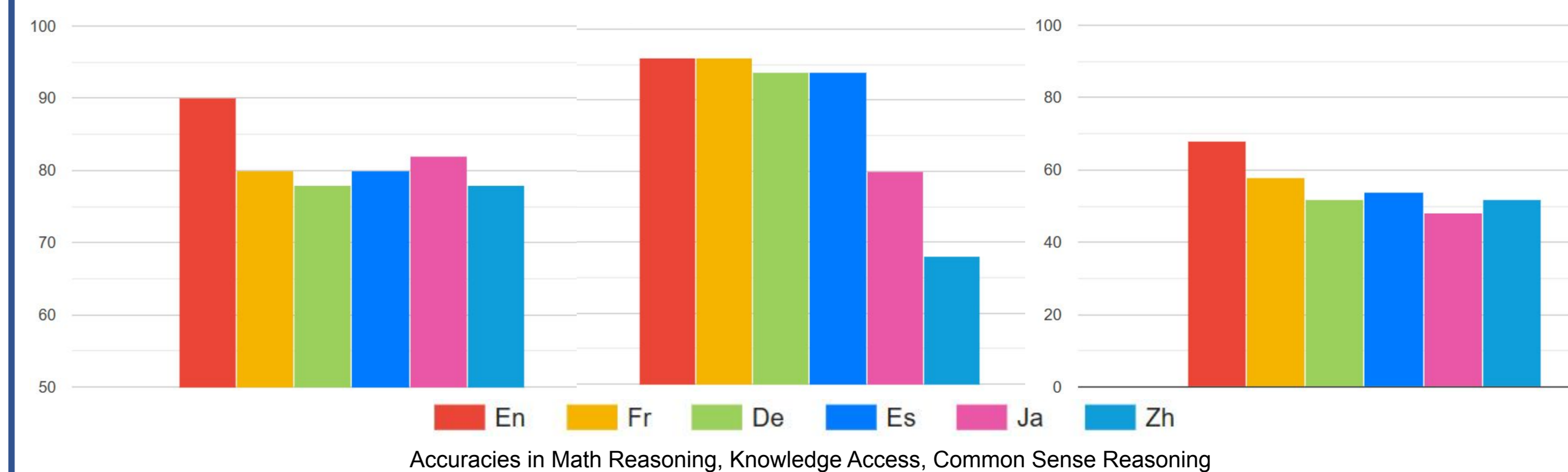
### Response Back-Translation (RBT)

- Translating prompts to the target language.
- Prompting the LLM.
- Prompt the LLM to generate an explanation.
- Prompt for the translation of the explanation in the original language.



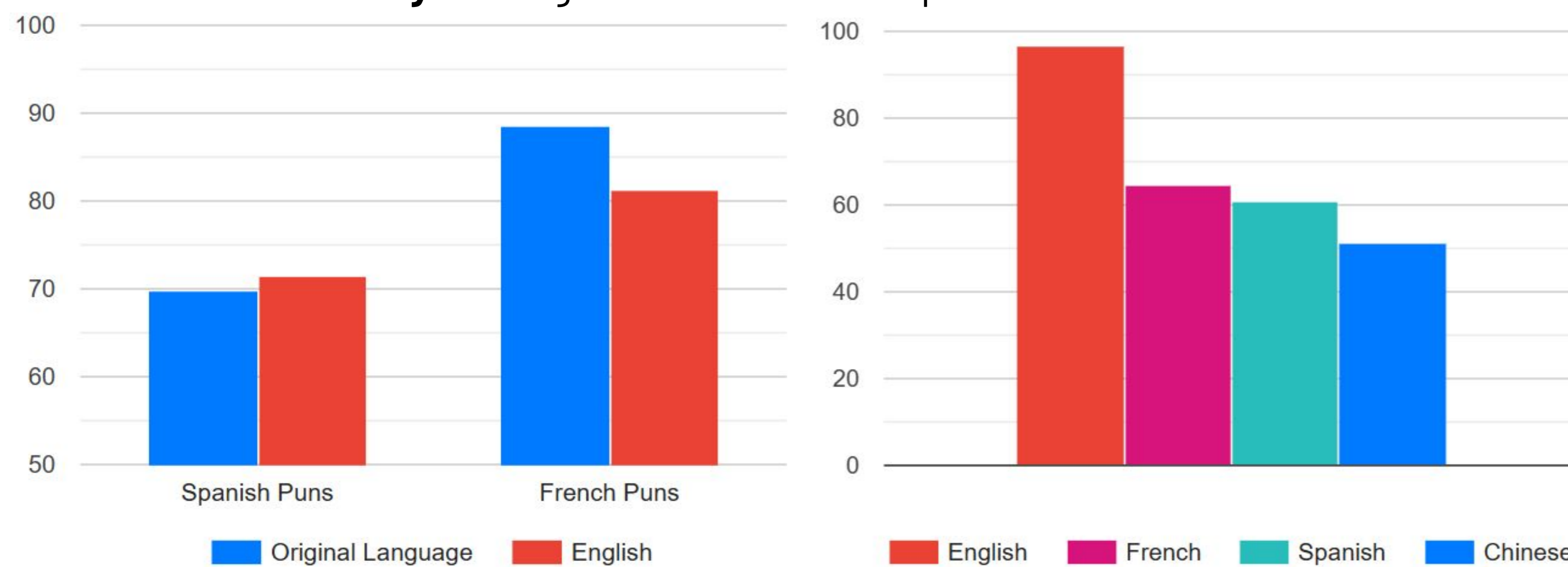
## Results: Translation Equivariant Tasks

ChatGPT is better at reasoning and retrieving knowledge given an English prompt, compared to prompting in other languages.



## Results: Translation Variant Tasks (1)

### Pun location accuracy for original and translated puns.

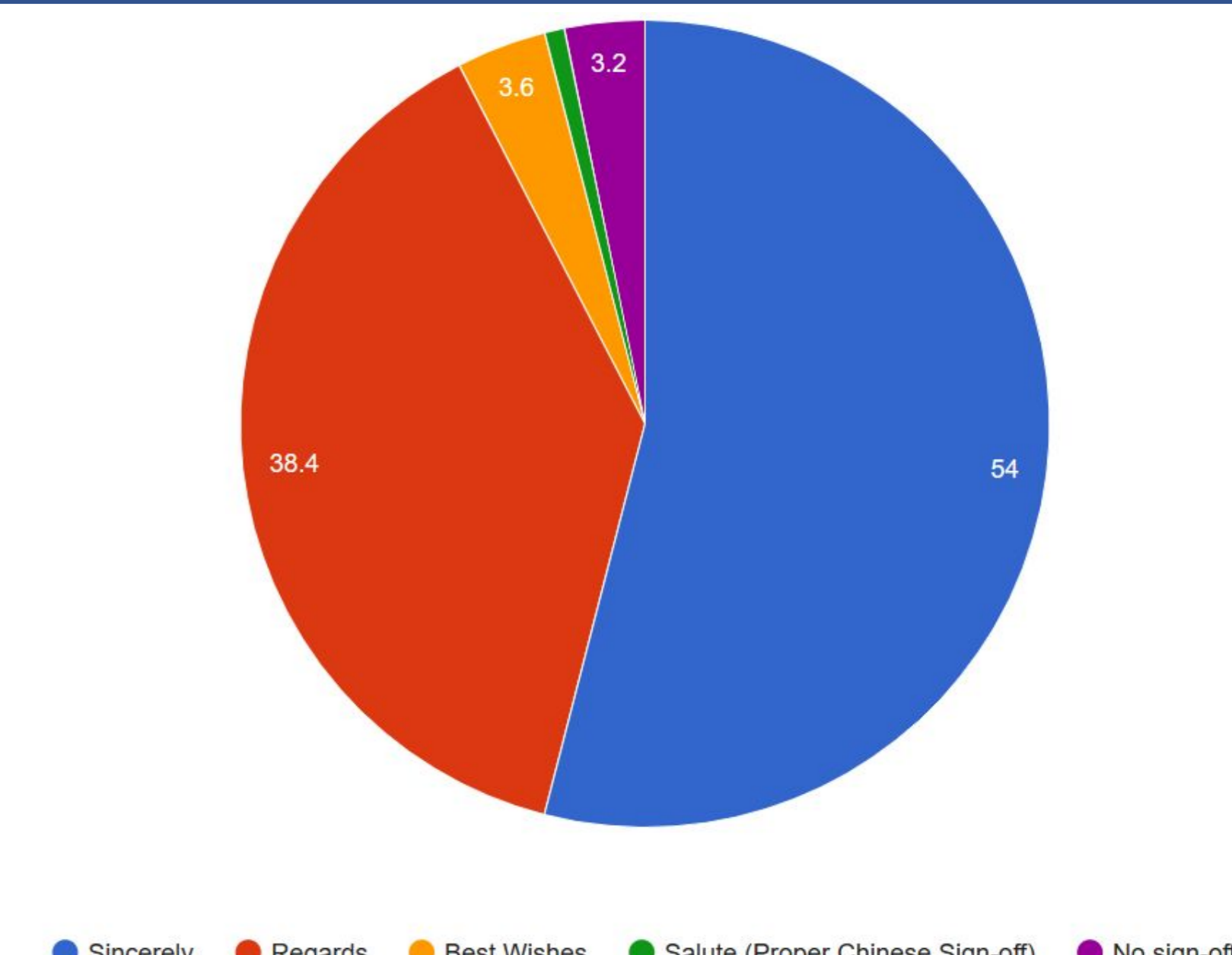


## Results: Translation Variant Tasks (2)

### Cover Letter Generation

The correct letter sign-off (Salute) in Chinese is almost never used.

ChatGPT's Chinese output reads like it was written in English and translated, providing strong evidence of subordinate multilingualism.



## Case Study: Pun Detection

**Example Pun:** A bicycle can't stand on its own because it is two-tired.  
**Chinese Translation (no pun):** 自行车不能独自站立, 因为它有两个轮胎。

The English sentence contains a pun: **Two-tired** -> **Too tired**.

**ChatGPT:** there is a pun in the Chinese translation.

**Reason:** Because the English translation of it has a pun.

## Take Away

We propose a novel method for studying the multilingualism of LLMs. ChatGPT exhibits behavior characteristic of subordinate multilingualism, with English as its "native language". This negatively impacts performance on non-English languages.

### Suggestions

- More parallel multilingual training data.
- Incorporate multimodal (e.g. vision) training data.
- Pursue the development of compound multilingual models.