

To Ask LLMs about English Grammaticality, Prompt Them in a Different Language



Shabnam Behzad, Amir Zeldes, Nathan Schneider
Georgetown University

- Learners often become confused when constructing sentences and seek help from teachers, peers, or online resources for clarity. Many of these queries relate to grammar; in fact, the grammar tag is the most popular one on online question and answer forums like English Language Learners Stack Exchange.
- Previous work (Behzad et al., 2023) investigate the capabilities of large language models (LLMs) in responding to learner questions with a focus on questions asked in and about English.

- In this work, we investigate how multilingual LLMs perform at crosslingual metalinguistic question answering.
 - To ensure a controlled multilingual comparison, we extracted minimal pairs from annotated learner corpora in different L2s (sentence with learner error vs. corrected version), and inserted these into a forced-choice prompt template.
- Languages include [English](#), [German](#), [Korean](#), [Russian](#), and [Ukrainian](#).

is believed to grammar sentence, which one is correct?

Asked 6 years, 10 months ago Modified 6 years, 10 months ago Viewed 12k times

Which one is grammatically correct?
He is believed to be awarded the prize at the age of 17 when he was in London.
or
He is believed to have been awarded the prize at the age of 17 when he was in London.

2

grammar

Share Improve this question Follow

asked Dec 21, 2017 at 15:58
Mohamed Magdy
528 5 10 23

Template	Data Instance	Correct Answer
in eng about eng	Which sentence is more grammatical and native-like? 1) It helps me healthy . 2) <i>It helps me keep healthy .</i>	2
in eng about deu	Which sentence is more grammatical and native-like? 1) Fühlen sich etwas orientierungslos . 2) <i>Sie fühlen sich etwas orientierungslos .</i> (translation: You feel a bit disoriented.)	2
in kor about deu	어떤 문장이 더 문법적이고 자연스러운가요? 1) <i>Das nützt uns für unseren Beruf .</i> (translation: This is useful for our profession.) 2) Das nützt für unseren Beruf .	1

Examples instances from our evaluation benchmark for different languages. The correct sentence is shown in green. English translations are shown for clarity (they were not provided to the system).

Models:

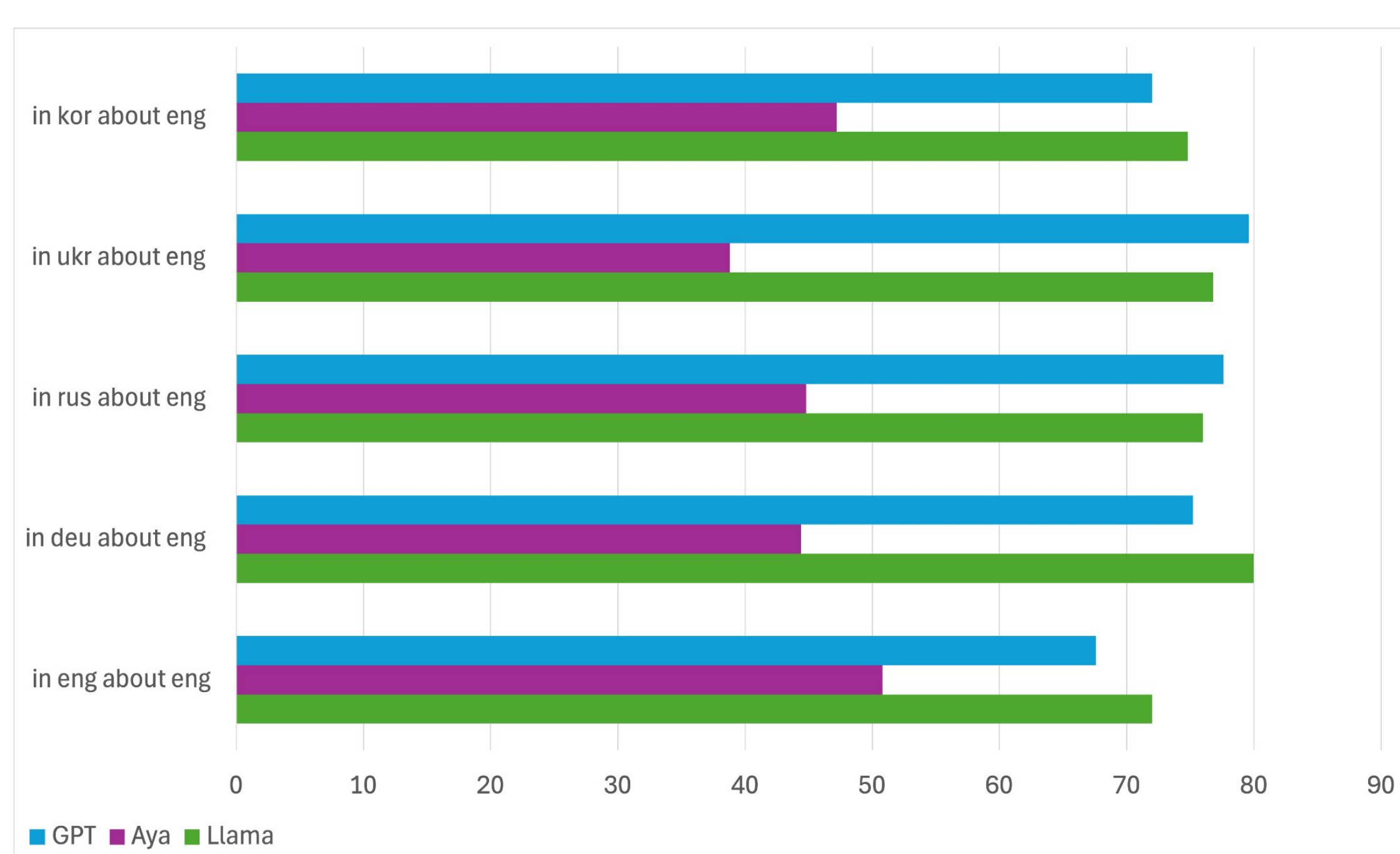
- ❖ GPT: gpt-4-0125-preview
- ❖ LLAMA: Meta-Llama-3-8B-Instruct
- ❖ Aya: Aya-101

The sentence pairs were derived from benchmarks for grammatical error correction:

- ❖ English: W&I (Yannakoudakis et al., 2018)
- ❖ German: Falko (Reznicek et al., 2012; Boyd, 2018)
- ❖ Korean: Kor-Learner (Yoon et al., 2023)
- ❖ Russian: RULEC (Alsufieva et al., 2012; Rozovskaya and Roth, 2019)
- ❖ Ukrainian: UA-GEC (Syvokon et al., 2023)

Are LLMs consistent in performance when the same question is asked in different languages? No.

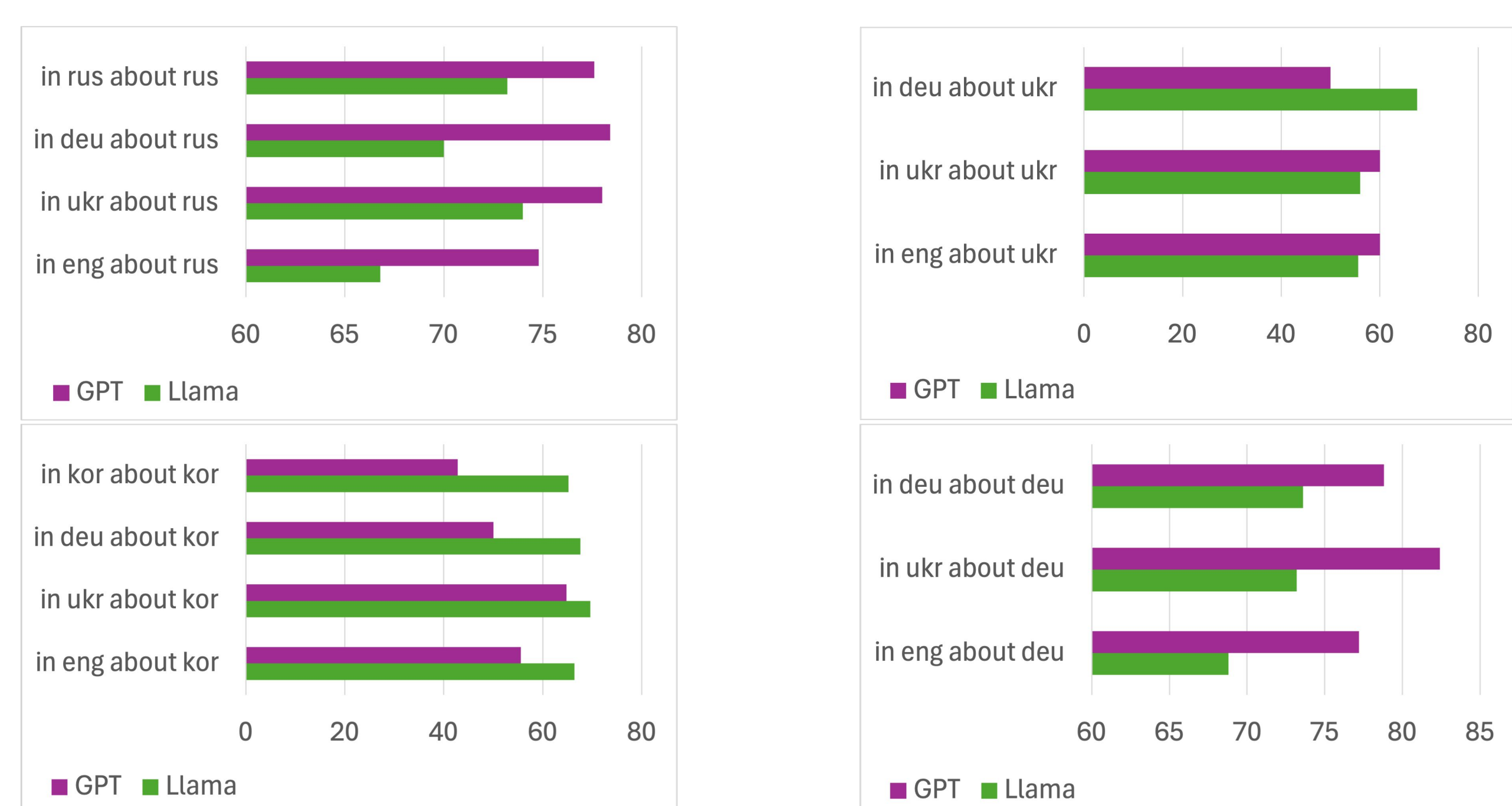
- We keep English sentence pairs constant, but translate the first part of the prompt ("Which sentence is more grammatical and native-like?") into different languages.
- Since the knowledge required to answer the question remains the same (English grammar) we expect language models to perform consistently among different languages.
- Both GPT and Llama perform better when question is asked in a language other than English.



Performance (accuracy) on 250 English sentence pairs, when questions are asked in different languages.

Do LLMs perform better when questions are asked in specific languages? Yes, and it's not English.

- Except for questions about Russian grammar, GPT seems to perform better when the question is asked in **Ukrainian** rather than in the same language as the sentence pair.
- With Llama, it appears that code-switching generally improves performance, particularly with questions about Ukrainian, which is considered a mid-resource language. There is an 11.6% performance boost when questions are asked in German rather than in Ukrainian.



Comparing model accuracy when the question is asked in the same language as sentence pairs, vs. when it is asked in another language and there is code switching.

Qualitative Analysis - Llama

Which sentence is more grammatical and native-like?

- 1- In spite of this , summer is a very good time to enjoy **the little things** life gives us and have fun with them .
- 2- In spite of this , summer is a very good time to enjoy **little things** life gives us and have fun with them .

The difference between the two sentences is the presence of **the indefinite article "a" before the noun "little things"**.

In sentence 1, the sentence is a bit awkward because "a very good time" is a singular noun phrase, and it's not clear why we need to modify it with "little". The phrase "little things" is a plural noun phrase, and it's not clear what "a" is doing there.

In sentence 2, the sentence is more natural and idiomatic because **we're using "little things" as a plural noun phrase, and we're not trying to modify it with "a"**. We're simply saying that summer is a good time to enjoy the small pleasures in life.

In general, when we use an indefinite article like "a" or "an", we're usually referring to a single, specific thing. When we're referring to a plural noun phrase, we usually don't use an indefinite article. Instead, we might use a determiner like "these", "those", or "the".

So, **sentence 2 is more grammatical and native-like** because it's using the indefinite article correctly and using the plural noun phrase "little things" without trying to modify it with "a".

Takeaways

- There is still room for improvement in this task across different languages including English.
- Models are too sensitive to the prompt and the language of the prompt;
 - We see a clear trend favoring asking in a different language, possibly because they sensitize the model to a meta-linguistic environment.
- The results vary widely across models—without a clear correlation with parameter count, language, or error type—suggesting that more attention needs to be paid to metalinguistic Q&A in order to obtain satisfactory and consistent performance.