ELQA: A Corpus of Metalinguistic Questions and Answers about English

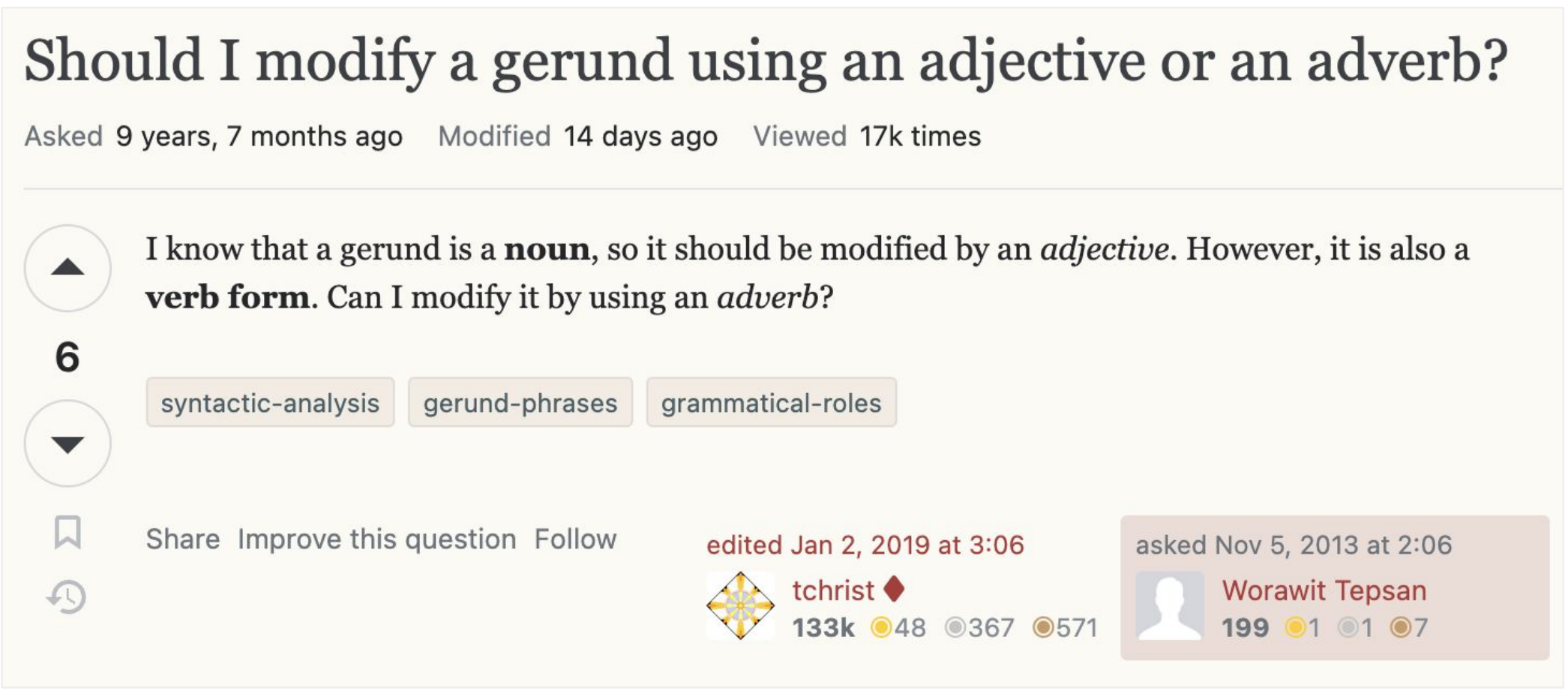
Shabnam Behzad[1], Keisuke Sakaguchi[2], Nathan Schneider[1], Amir Zeldes[1]

[1]Georgetown University, [2]Tohoku University

Motivation

A long-form question answering benchmark that can be used for:

- Educational NLP applications: Can we build models to answer questions posed by English language learners?
- Investigating the metalinguistic capabilities of LLMs



Should I modify a gerund using an adjective or an adverb?

Asked 9 years, 7 months ago Modified 14 days ago Viewed 17k times

6

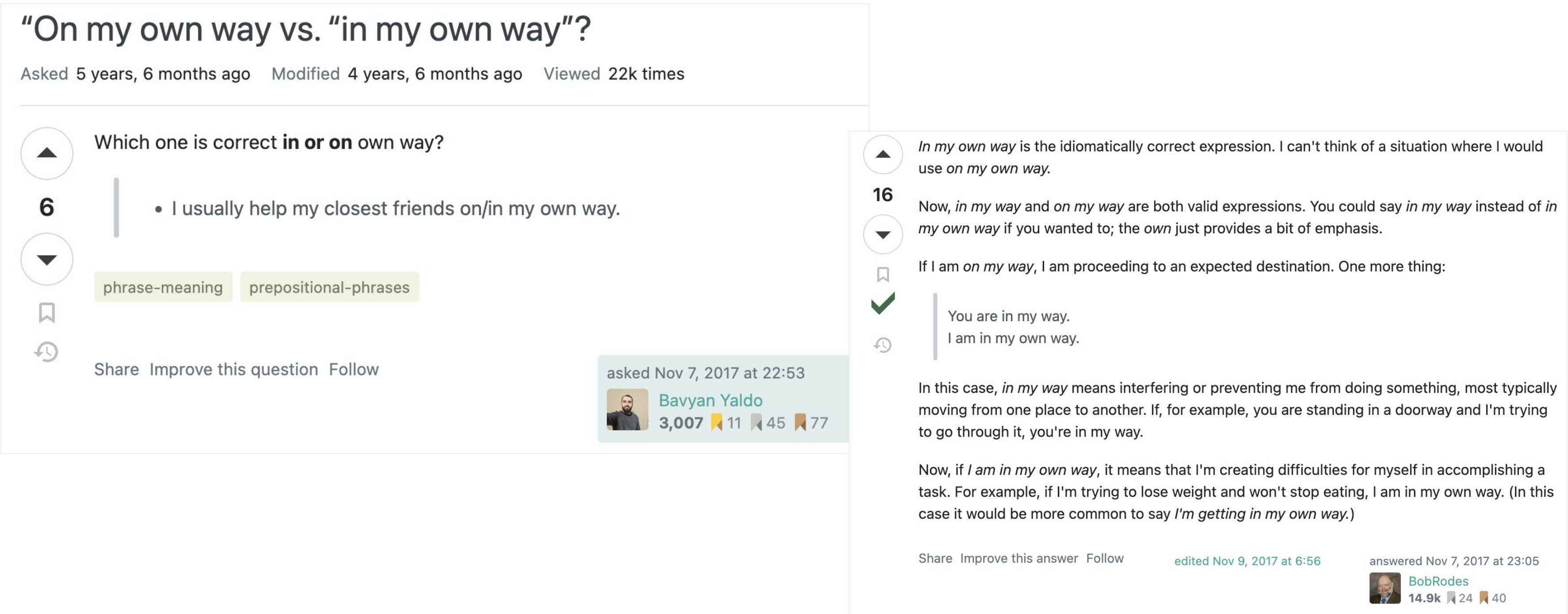
I know that a gerund is a **noun**, so it should be modified by an *adjective*. However, it is also a **verb form**. Can I modify it by using an *adverb*?

syntactic-analysis gerund-phrases grammatical-roles

Share Improve this question Follow

edited Jan 2, 2019 at 3:06
tchrist ♦
133k ●48 ●367 ●571

asked Nov 5, 2013 at 2:06
Worawit Tepsan
199 ●1 ●1 ●7



"On my own way vs. "in my own way"?

Asked 5 years, 6 months ago Modified 4 years, 6 months ago Viewed 22k times

6

Which one is correct **in or on** own way?

> - I usually help my closest friends on/in my own way.

phrase-meaning prepositional-phrases

Share Improve this question Follow

asked Nov 7, 2017 at 22:53
Bavyan Yaldo
3,007 11 45 77

16 ✓

In my own way is the idiomatically correct expression. I can't think of a situation where I would use *on my own way.*

Now, *in my way* and *on my way* are both valid expressions. You could say *in my way* instead of *in my own way* if you wanted to; the *own* just provides a bit of emphasis.

If I am *on my way*, I am proceeding to an expected destination. One more thing:

> You are in my way.
> I am in my own way.

In this case, *in my way* means interfering or preventing me from doing something, most typically moving from one place to another. If, for example, you are standing in a doorway and I'm trying to go through it, you're in my way.

Now, if *I am in my own way*, it means that I'm creating difficulties for myself in accomplishing a task. For example, if I'm trying to lose weight and won't stop eating, I am in my own way. (In this case it would be more common to say *I'm getting in my own way.*)

Share Improve this answer Follow

edited Nov 9, 2017 at 6:56

answered Nov 7, 2017 at 23:05
BobRodes
14.9k 24 40

Data

ELQA: community-sourced dataset for English Language Question Answering

- Consists of more than 70k questions and answers on **numerous topics about English language such as grammar, meaning, fluency, and etymology.**
- Questions and answers are gathered from two Stack Exchange sites: *English Language & Usage (ENG) and English Language Learners (ELL)*.

ELQA-large	**ELL**	**ENG**
Total # of Qs	23,520	47,532
Total # of As	49,345	152,315
Avg. Q length	92.41	102.41
Avg. A length	158.25	137.90
Max. A score	392	581
Min. A score	−13	−28
Avg. A score	4.85	5.15
Total # of available tags	513	951
ELQA-small	**ELL**	**ENG**
Total # of Qs	6,477	14,234
Total # of As	18,389	62,744
Avg. Q length	84.21	89.25
Avg. A length	156.29	118.66
Max. A score	392	581
Min. A score	−13	−13
Avg. A score	6.63	6.73
Total # of available tags	437	823

Table 1: ELQA statistics on Qs (questions) and As (answers).

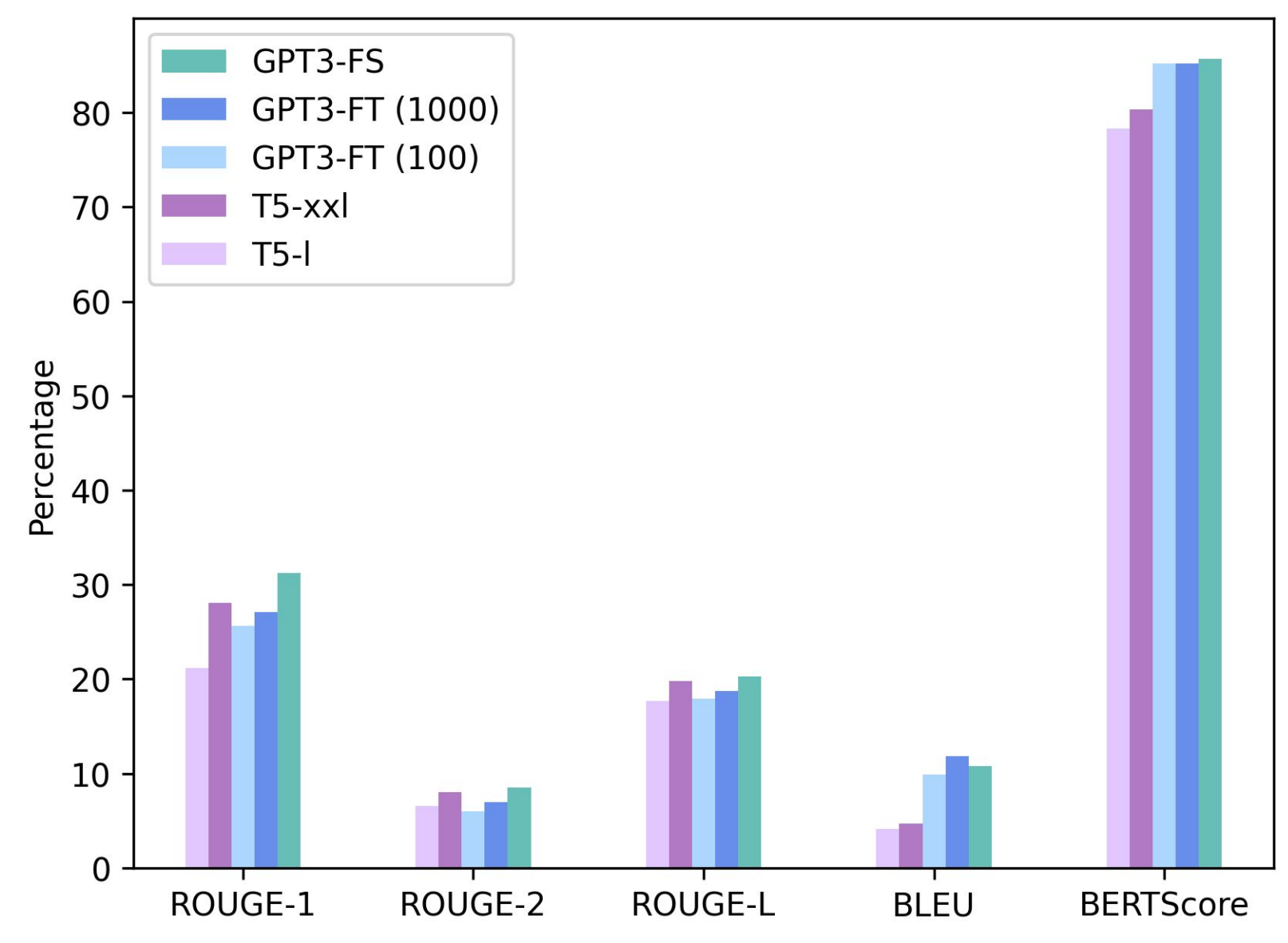
Salient types of questions in ELQA based on manual inspection:

- Fluency (≈38% of questions)
- Form to Meaning (Interpretation, ≈19% of questions)
- Meaning to Form (Encoding, ≈20% of questions)
- Grammatical Analysis (≈11% of questions)
- Other (Any type of question not listed above; including questions about pronunciation, etymology, etc.; ≈10% of questions)

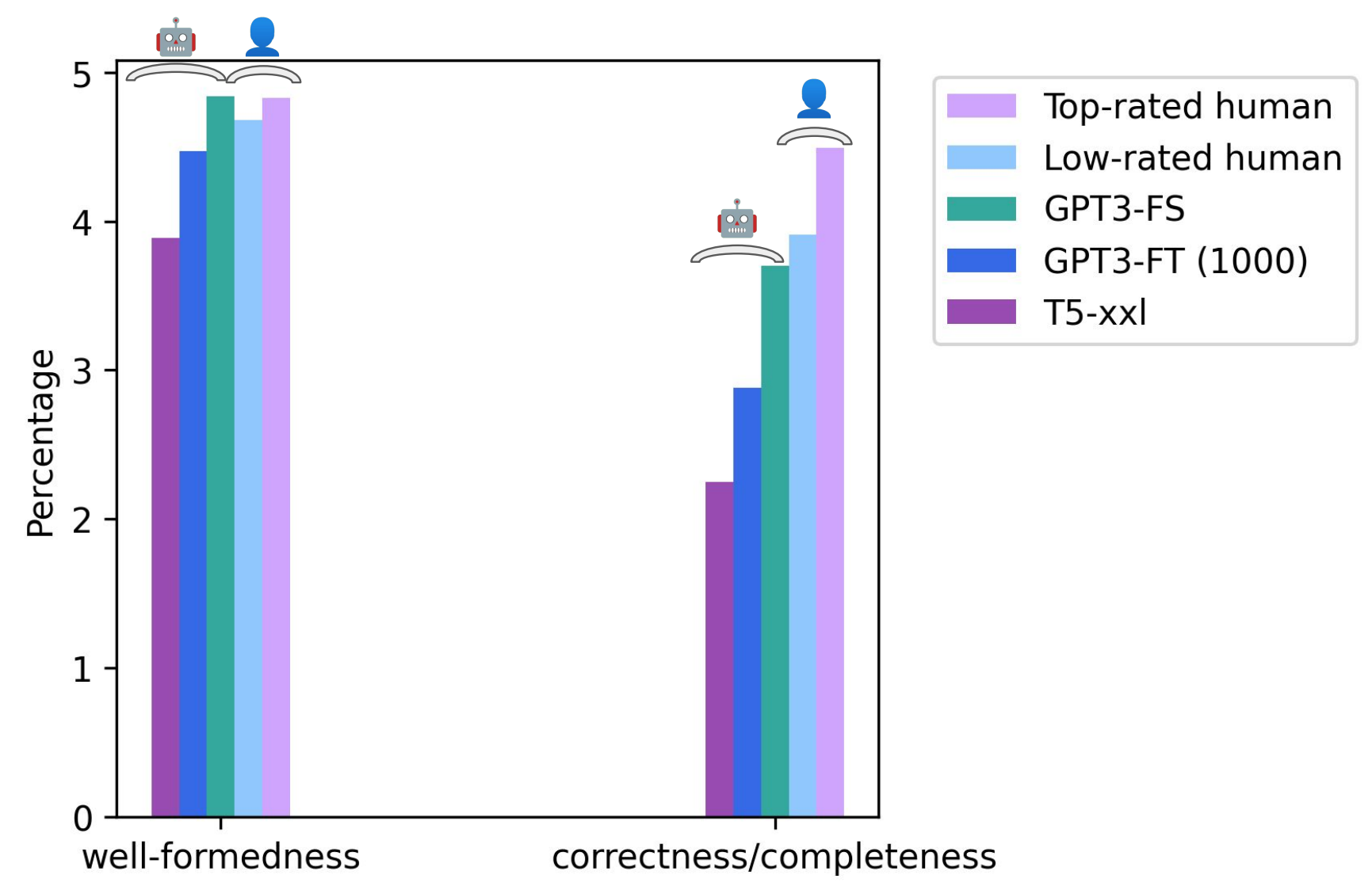
Scan to access our paper and data!

Experiments

We experimented on closed-book question answering with T5-xxl and T5-l plus GPT3 fine-tuned (FT) on 100 and 1000 samples and a few-shot (FS) setting in which we included 5 examples at inference time.



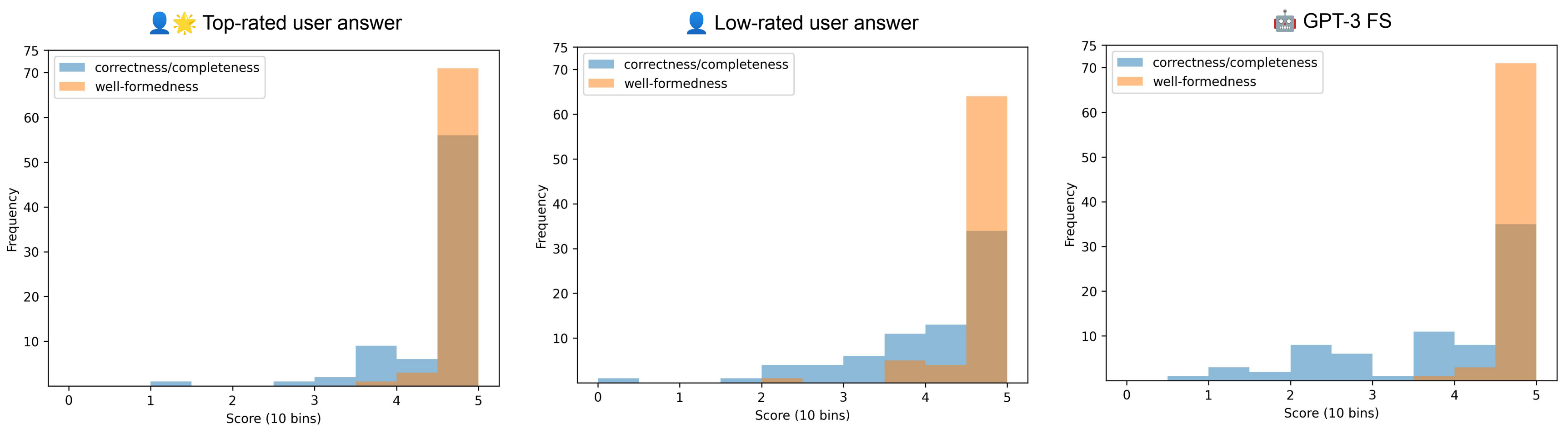
Automatic evaluation scores



Human evaluation scores (out of 5)

🤯 Different impression from auto metrics!

GPT3-FS output is **well-formed** and **often reasonably correct**, but on average *not as good as user-authored answers*.



Histograms of the ratings from our human evaluation of answers to 75 questions: the human-authored answer preferred by site users, the human-authored answer least preferred by site users, and our best model (GPT-3 FS). Each value is an average of two annotators' ratings.

Explain the verb tense in "All that glitters is not gold"
I am confused by the maxim **All that glitters is not gold**. May I know why the word is used there? In my known according the English grammar **glitters** is plural, So after plural we add verb+s. As in above example use auxiliary verb so why not **are** – why **is**?

User Chinmay235 on ell.stackexchange

GPT3-FS

The phrase *all that glitters is not gold* is an idiom, which means that things that appear to be valuable may not actually be so. It is an old proverb, and the verb tense used in the phrase is an example of the so-called "historic present". The historic present is a literary device used to give a sense of immediacy to a narrative. It is used to describe events that happened in the past, but are being described as if they are happening in the present. In this case, the phrase is used to emphasize the timelessness of the proverb. So, while the verb *glitters* is indeed plural, the verb *is* is used in the historic present to give the phrase a sense of immediacy and timelessness.

Avg. human eval scores: Well-formedness: 5/5; Correctness/completeness: 2/5

Outlook

- ❖ Good quality answers from GPT-3 FS: many would likely have received high user ratings if submitted as answers on the site.
- ❖ But the model is not to the point where we would want to trust its answers without human review.
- ❖ Possible future direction: Answer confidence estimation—so that users can be shown only the best model-generated answers—is an important direction for using learner QA models in practice.