Zero and Few-shot Semantic Parsing with Ambiguous Inputs

Elias Stengel-Eskin, Kyle Rawlins, and Benjamin Van Durme

ICLR 2024

THE UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL
UNC NLP
JOHNS HOPKINS UNIVERSITY
CENTER FOR LANGUAGE AND SPEECH PROCESSING

Ambiguous semantic parsing

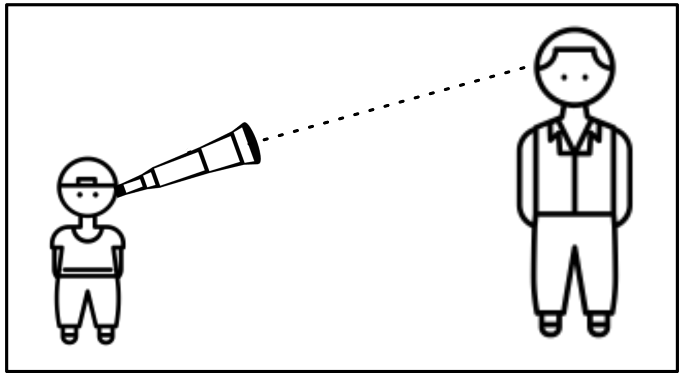
Semantic parsing…

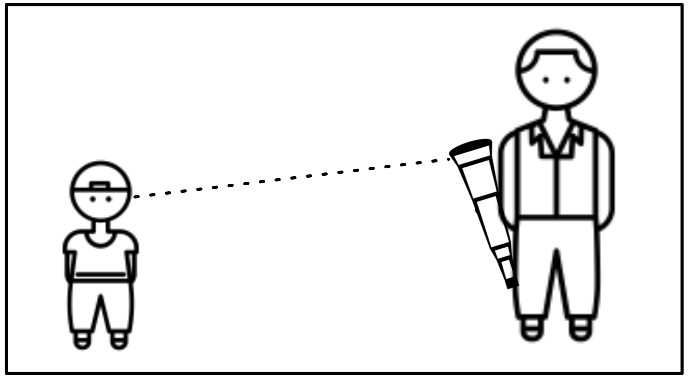
Mapping text to meaning

One input, one meaning

What about ambiguity?

The boy saw the man with the telescope





$\exists x.\exists y.\exists z.\exists a.boy(x) \wedge man(y)$
$\wedge\, telescope(z) \wedge saw(a)\wedge$
$agent(a, x) \wedge patient(a, y)\wedge$
$instrument(a, z)$

$\exists x.\exists y.\exists z.\exists a.\exists e.boy(x)$
$\wedge\, man(y) \wedge telescope(z)$
$\wedge\, saw(a) \wedge agent(a, x)$
$\wedge\, patient(a, y) \wedge have(e)$
$\wedge\, agent(e, x) \wedge patient(e, z)$

Dataset

AmP: Ambiguous Parsing

Templates for 5 ambiguity types

Each sentence has 2 interpretations

Dataset

AmP: Ambiguous Parsing

Templates for 5 ambiguity types

Each sentence has 2 interpretations

Prepositional phrase attachment (pp)

Quantifier scope (scope)

Reversed quantifier scope (revscope)

Conjunctions (conj.)

Bound pronoun coreference (bound)

Dataset

AmP: Ambiguous Parsing

Templates for 5 ambiguity types

Each sentence has 2 interpretations

Prepositional phrase attachment (pp)

Quantifier scope (scope)

Reversed quantifier scope (revscope)

Conjunctions (conj.)

Bound pronoun coreference (bound)

See paper for examples of each!

Settings

Zero-shot: model has no evidence of how to parse ambiguity

Sees “ingredients” but not ambiguity

Few-shot: model sees a few examples of each interpretation

“mixed prompt” setting

Zero-shot results

Does model confidence reflect 2 possible parses?

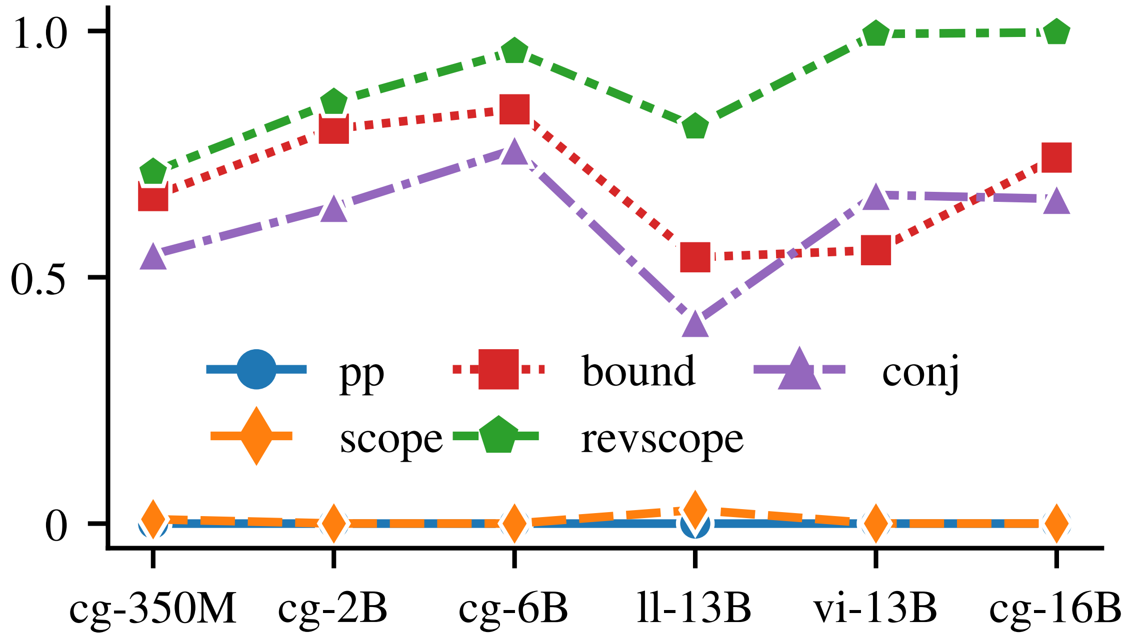
$$\hat{p} = \min_{i=1}^{N} P_\theta(t_i | x; t_{1:i-1})$$

$$P_\theta(LF_0) = \frac{\hat{p}_{LF_0}}{\hat{p}_{LF_0} + \hat{p}_{LF_1}}$$

Sometimes!

coreference

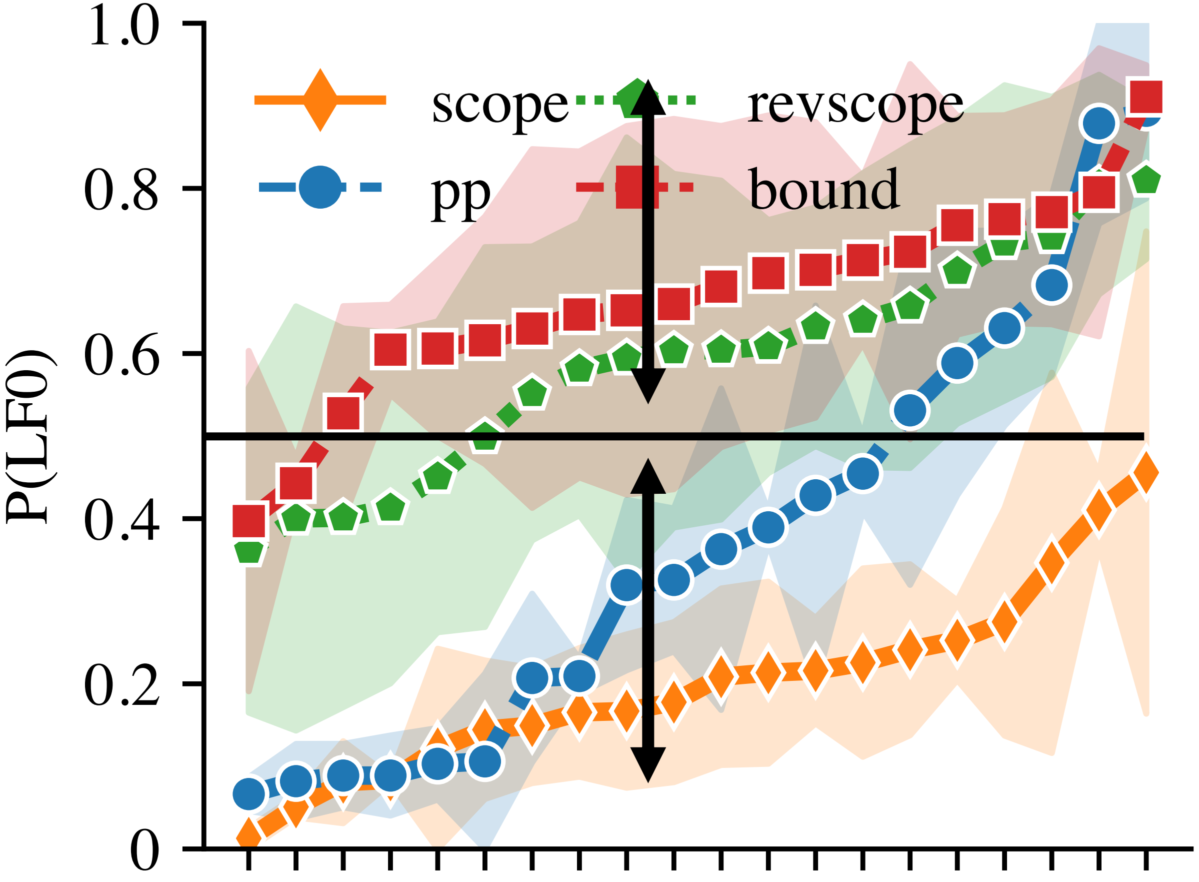
conjunctions



Human results

General preferences

Align with past literature



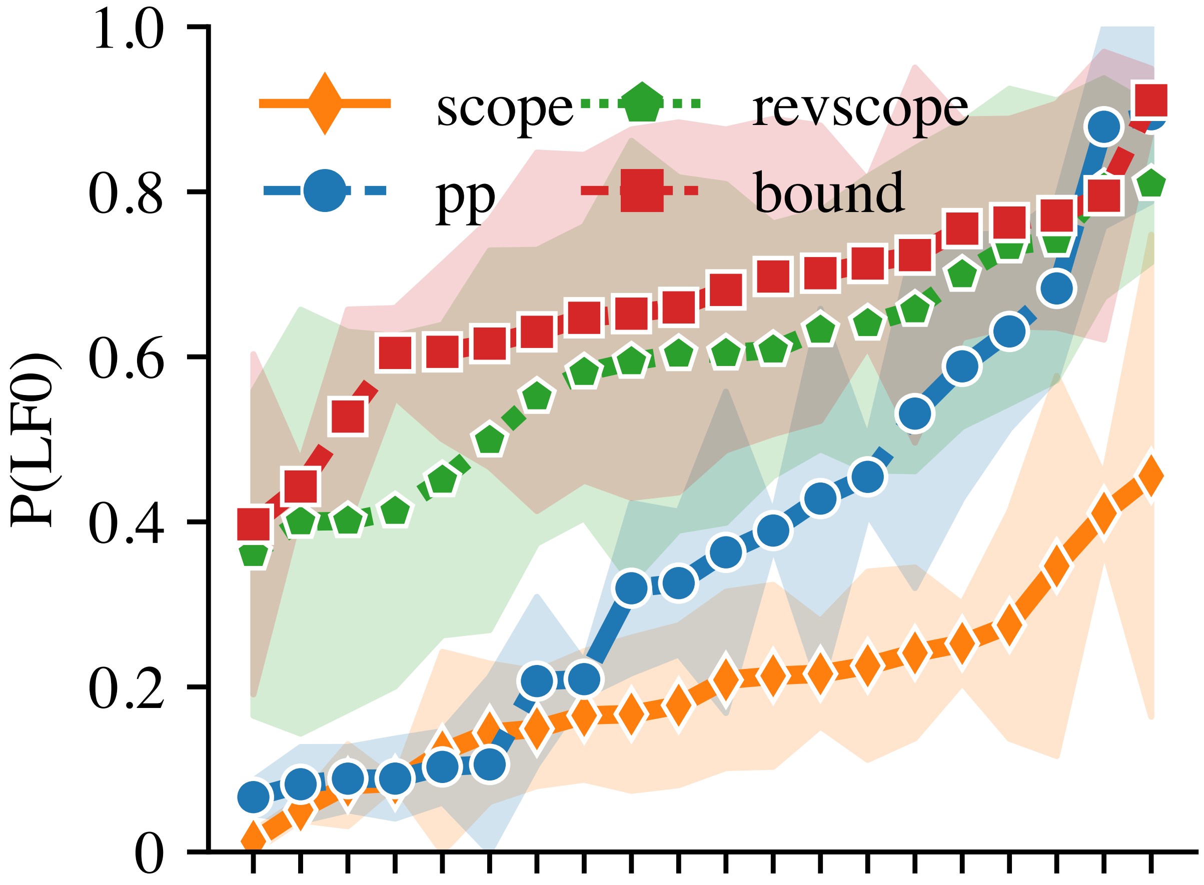
Human results

General preferences

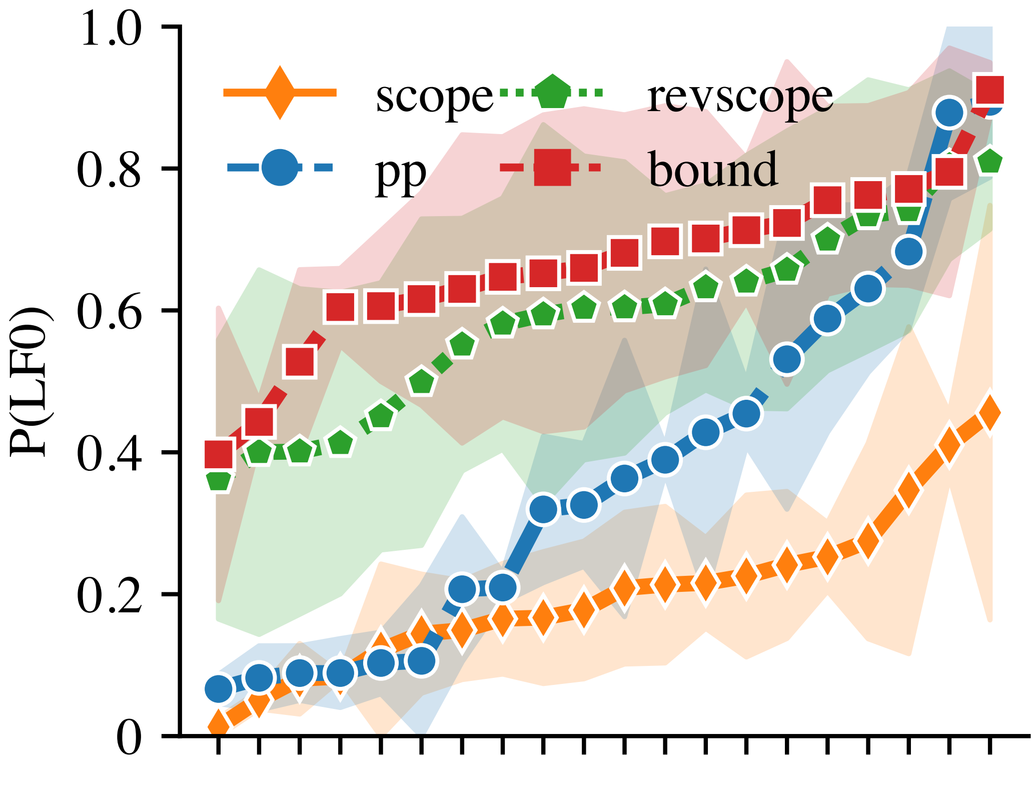
Align with past literature

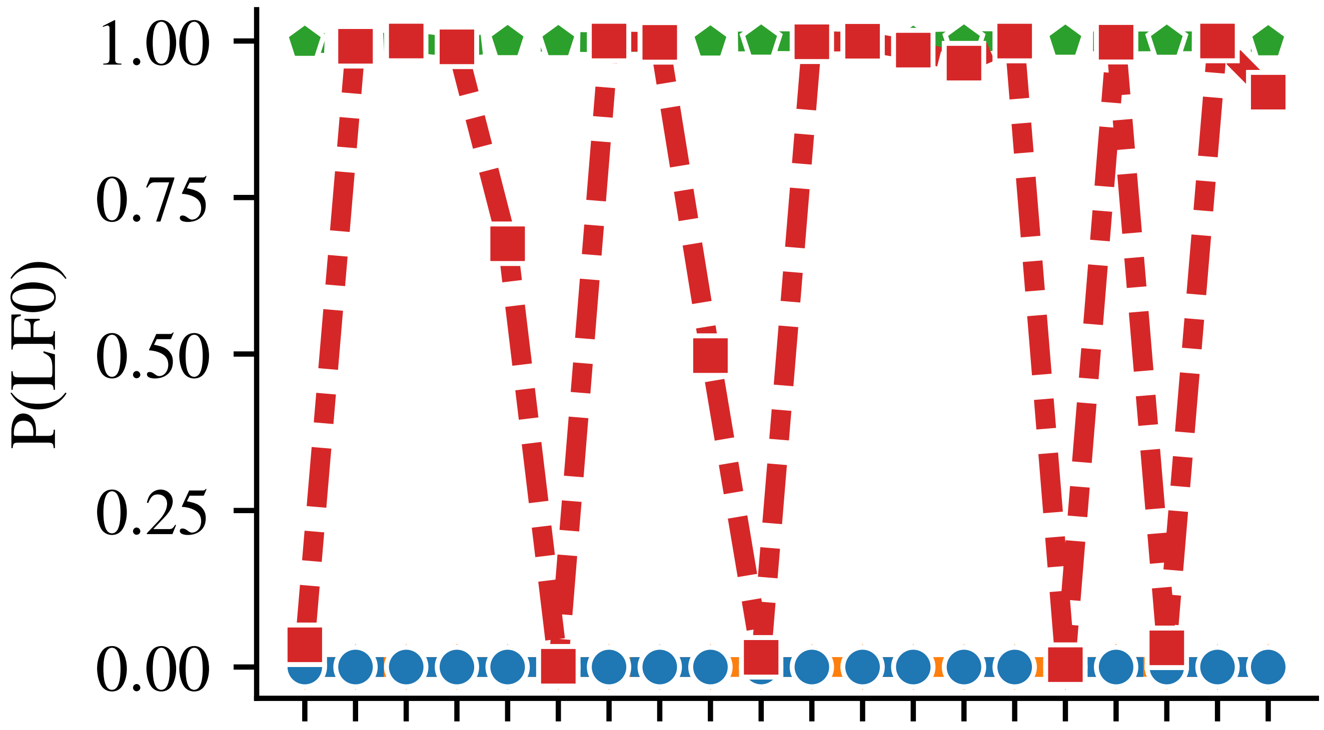
Lexical effects

Gradient of ambiguity



Model comparison





Few-shot parsing

```
Let's translate what a human user says into what a computer
might say.

Human: the boy observed Adele with the spyglass
Computer: exists x . exists y . exists a . exists e . boy
(x) AND spyglass(y) AND observed(a) AND agent(a, x) AND
patient(a, Adele) AND have(e) AND agent(e, Adele) AND
patient(e, y)
```

LF0

```
Human: Sherlock spotted Galileo with the binoculars
Computer: exists x . exists a . exists e . binoculars(x) AND
spotted(a) AND agent(a, Sherlock) AND patient(a, Galileo) AND
have(e) AND agent(e, Galileo) AND patient(e, x)
```

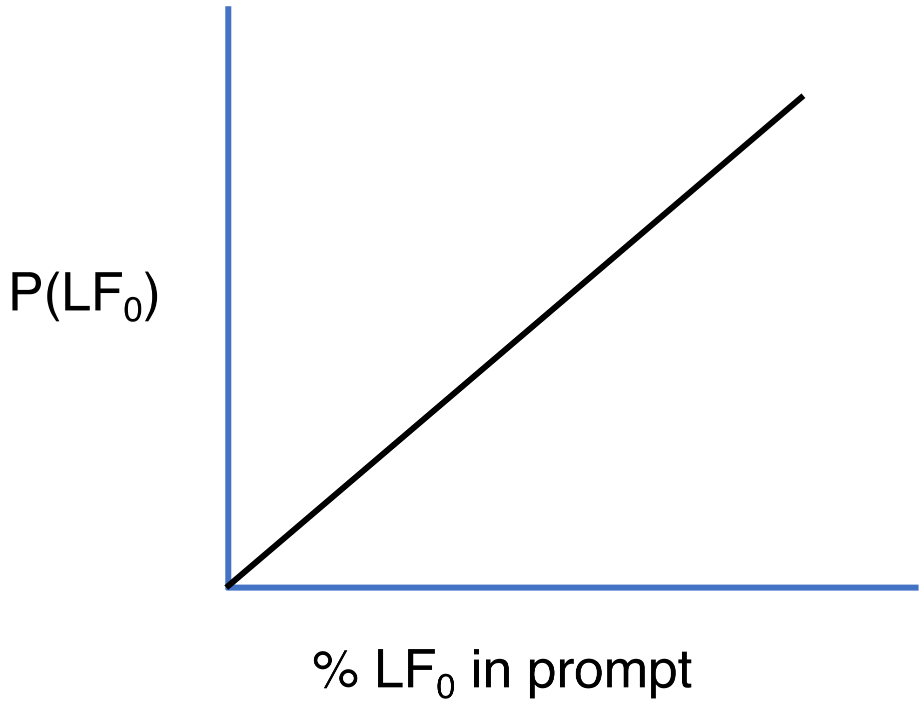
LF0

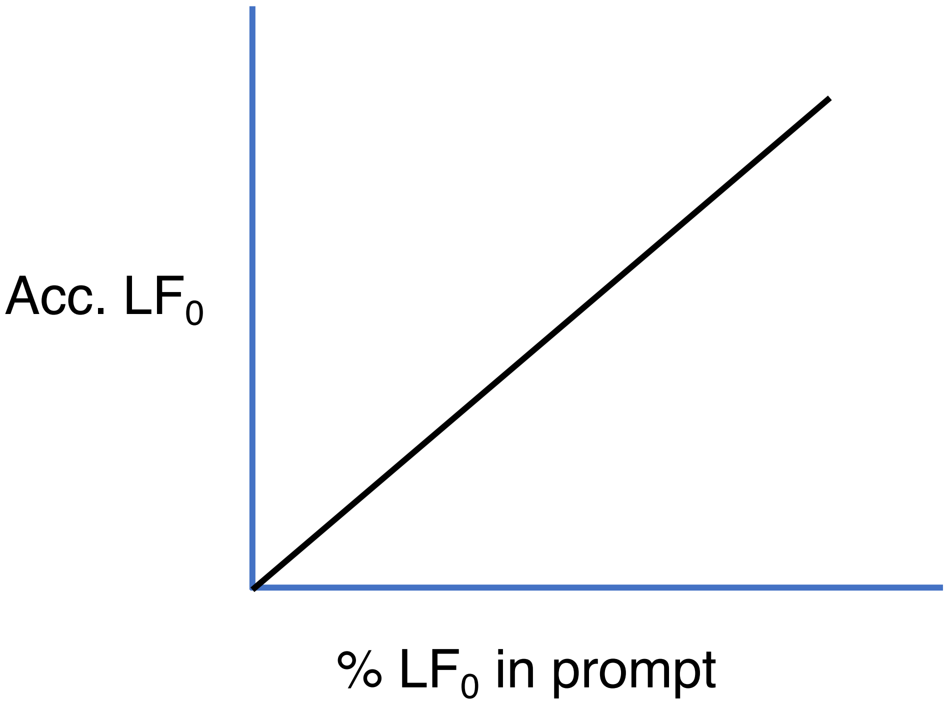
```
Human: the boy spied Mary with the camera
Computer: exists x . exists y . exists a . boy(x) AND
camera(y) AND spied(a) AND agent(a, x) AND patient(a, Mary)
AND instrument(a, y)
```

LF1

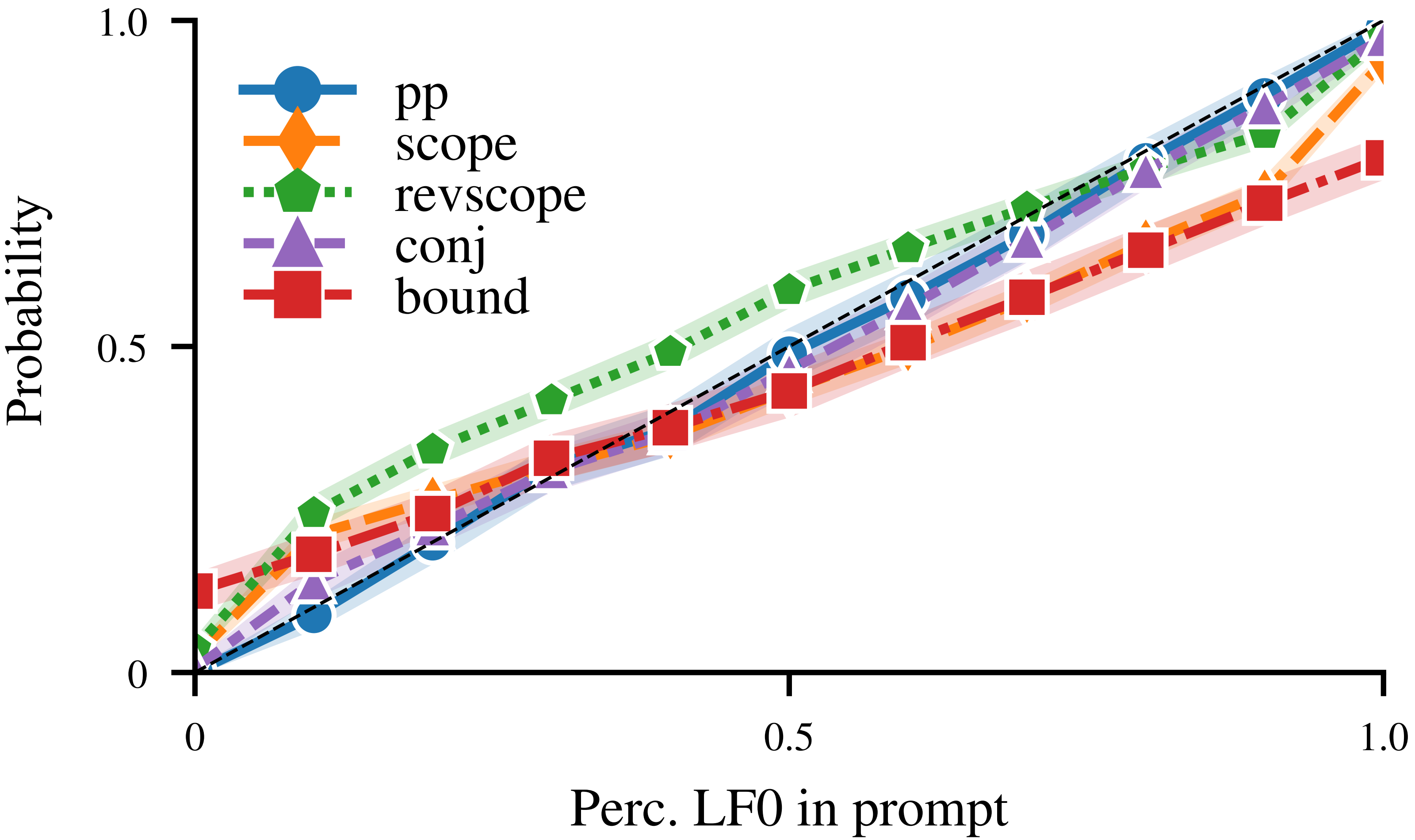
```
Human: Bill saw the girl with the binoculars
Computer:
```

Few-shot results

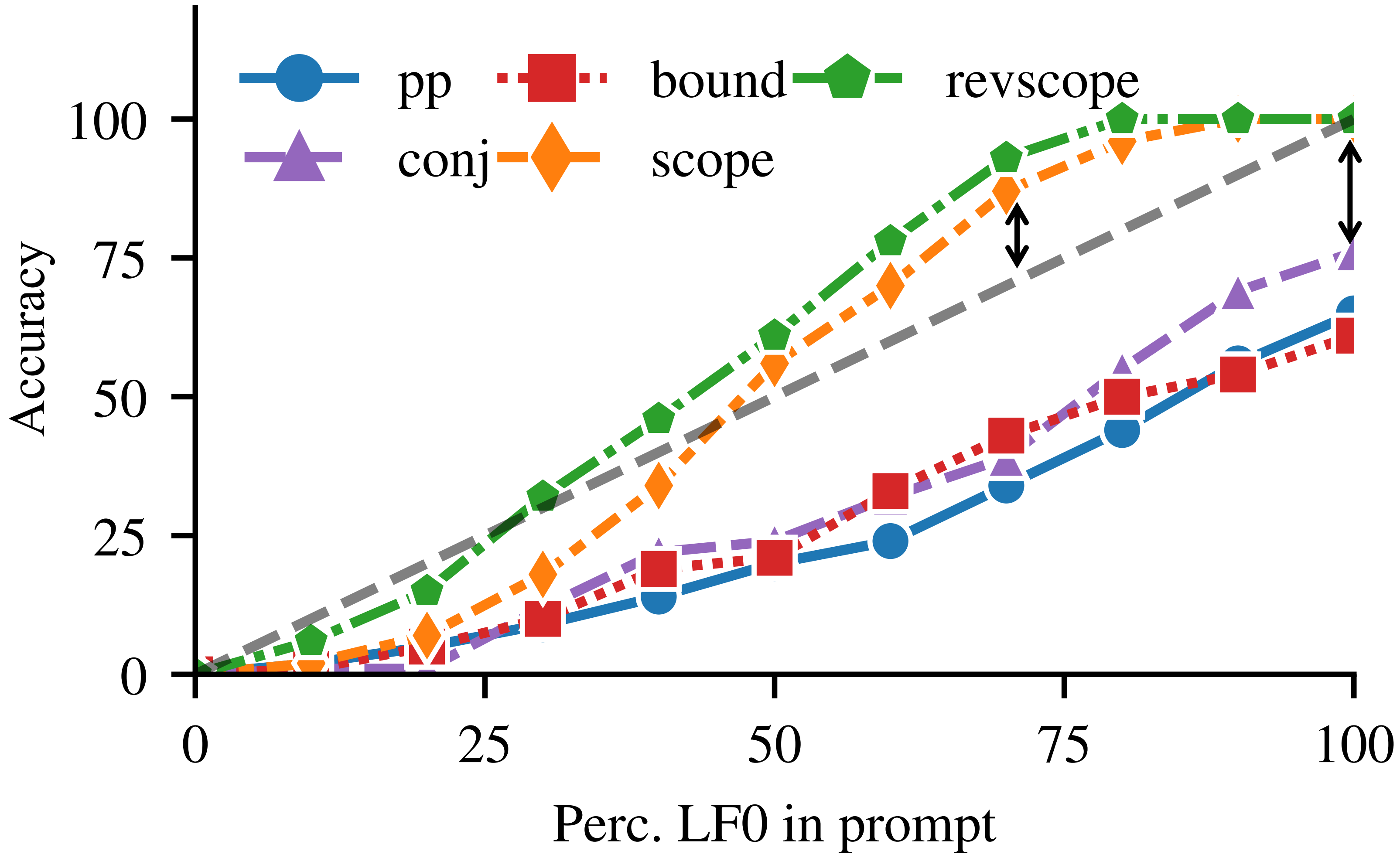




Few-shot results



Few-shot results



Conclusions

Models do not capture zero-shot ambiguity

Not ideal: over-commit to one interpretation

Not sensitive to lexical changes

Conclusions

Models do not capture zero-shot ambiguity

Not ideal: over-commit to one interpretation

Not sensitive to lexical changes

Few-shot results are more promising

If given evidence of ambiguity, models capture it well

Almost perfectly calibrated

Conclusions

Models do not capture zero-shot ambiguity

Not ideal: over-commit to one interpretation

Not sensitive to lexical changes

Few-shot results are more promising

If given evidence of ambiguity, models capture it well

Almost perfectly calibrated

Many more results/analysis in the paper