

# Seeking Clozure: Robust Hypernym Extraction from BERT with Anchored Prompts

Chunhua Liu Trevor Cohn Lea Frermann

## Motivation

Existing work in zero-shot hypernym prompting shows that pre-trained language models

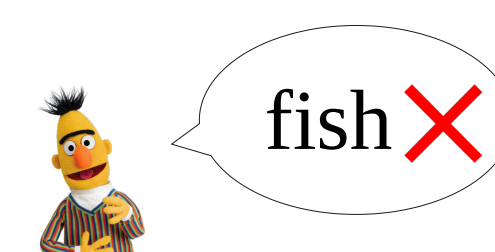
- struggle with **rare or abstract concepts**
- **perform inconsistently** with paraphrased prompts

Can *established patterns* from corpus-mining bring benefits?

## Anchored Prompts

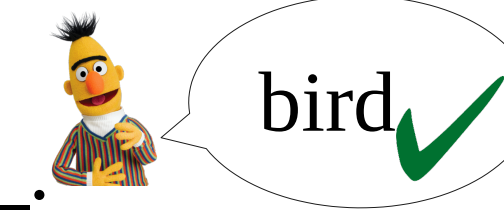
Raw Prompts:

A kea is a \_\_\_\_.



Anchored Prompts:

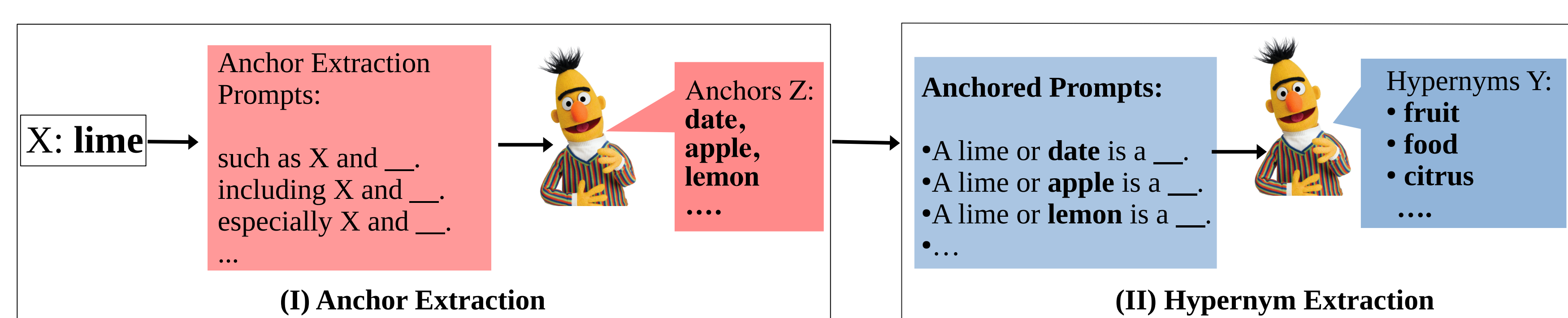
A kea or **parrot** is a \_\_\_\_.



## Two Pattern Sets for Hypernym Extraction

Raw Patterns	+Anchor
A(n) X is a Y.	A(n) X or <b>Z</b> is a Y.
A(n) X is a type of Y.	A(n) X or <b>Z</b> is a type of Y.
A(n) X is a kind of Y.	A(n) X or <b>Z</b> is a kind of Y.

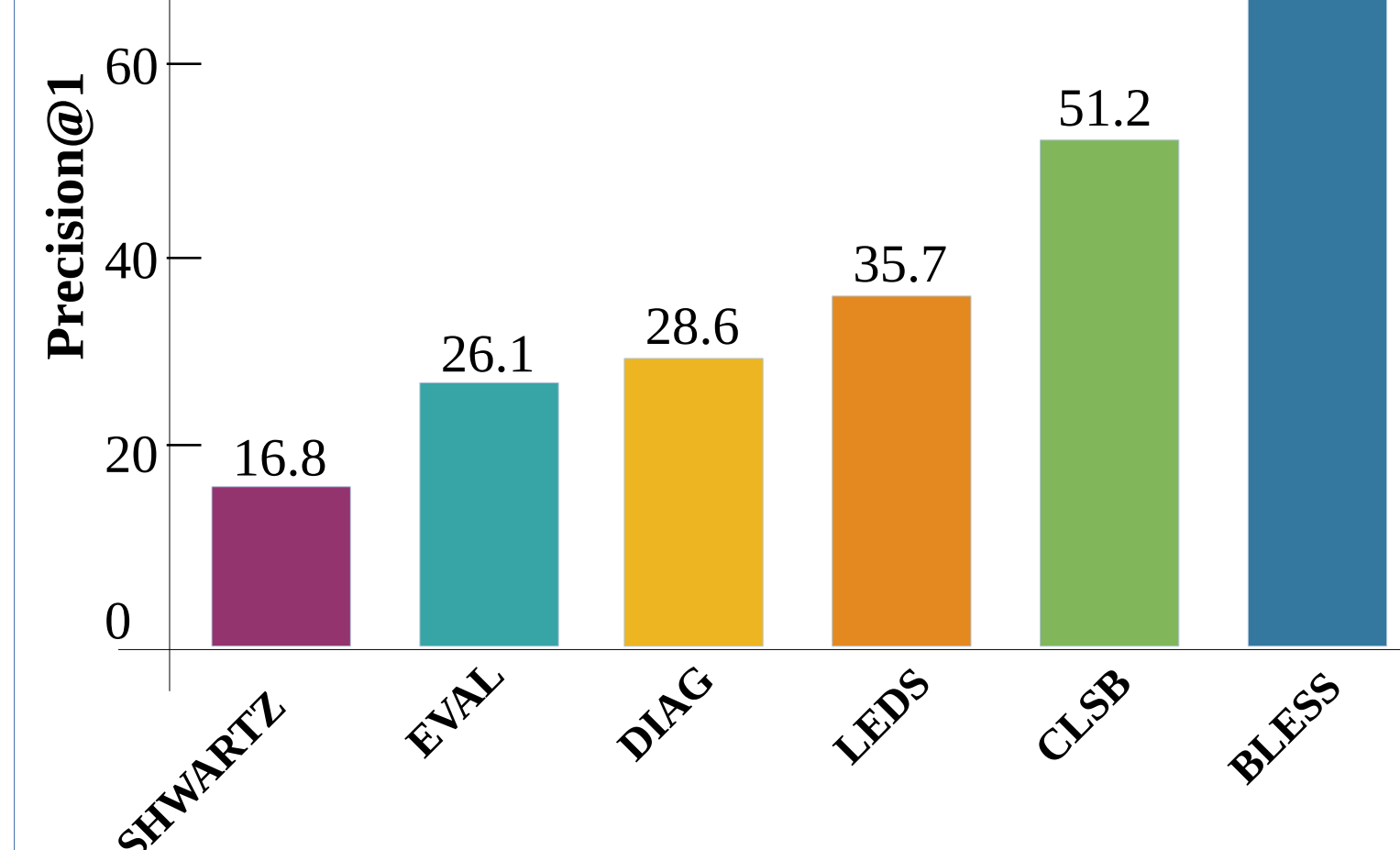
## Framework



## Results I: Anchor Validation

### Quantitative Evaluation with WordNet Siblings

(Schick and Schütz., 2020)



- Lower precision for SHWARTZ, EVAL and DIAG due to concept diversity/low frequency

### Qualitative Inspection

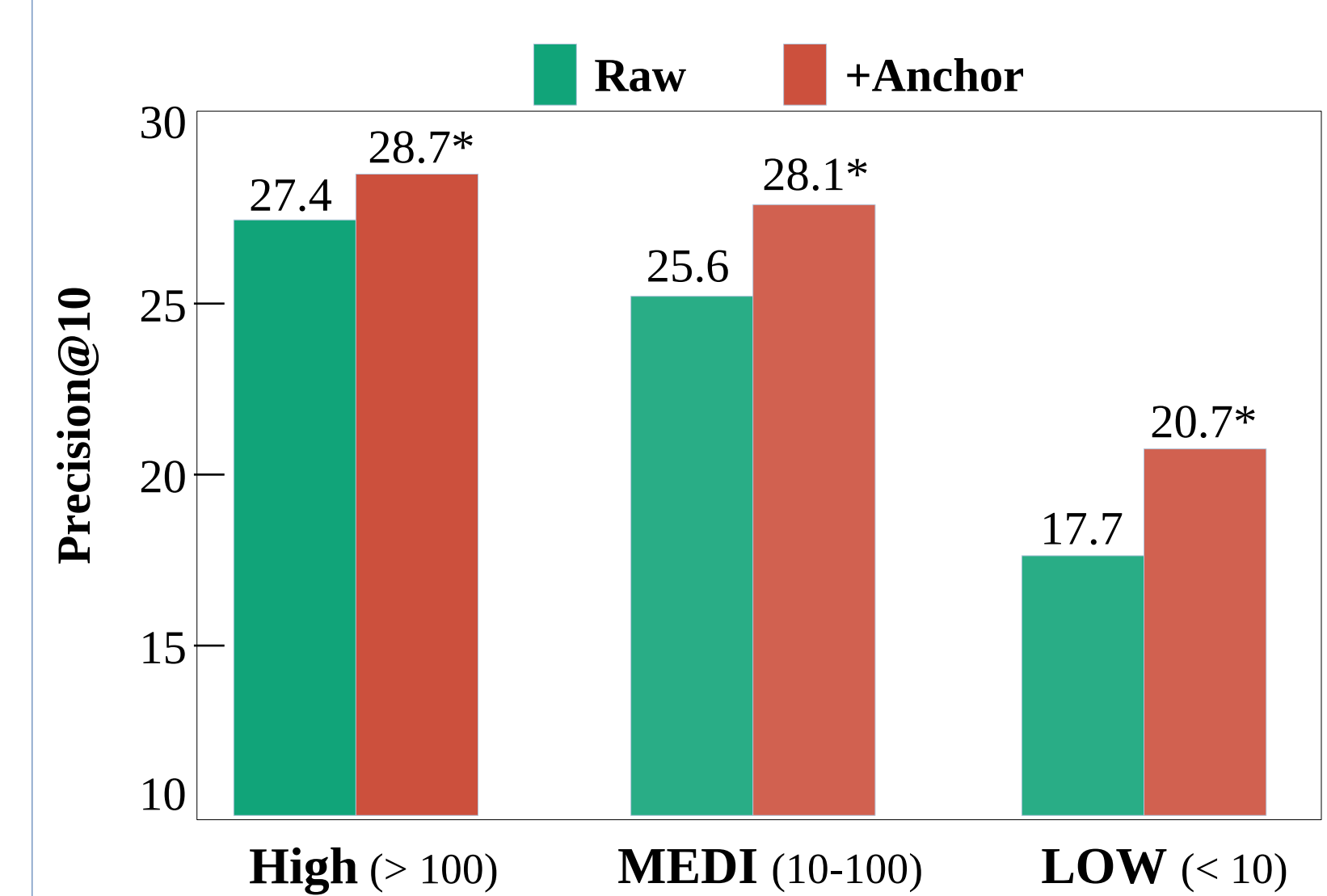
Z in WN Siblings	X	Top 5 predicted anchors (Z)
Yes	car	truck, motorcycle, boat, yes, bike
	apple	grape, pear, nuts, vegetable, date
	train	bus, plane, car, tram, truck
NO	motel	hotel, yes, sure, restaurant, actually
	daisy	rose, yes, lavender, rush, fern
	murre	dog, bird, fox, crow, rabbit

- Retrieved anchors are often reasonable, even not in WordNet siblings.

## Results II: Hypernym Extraction

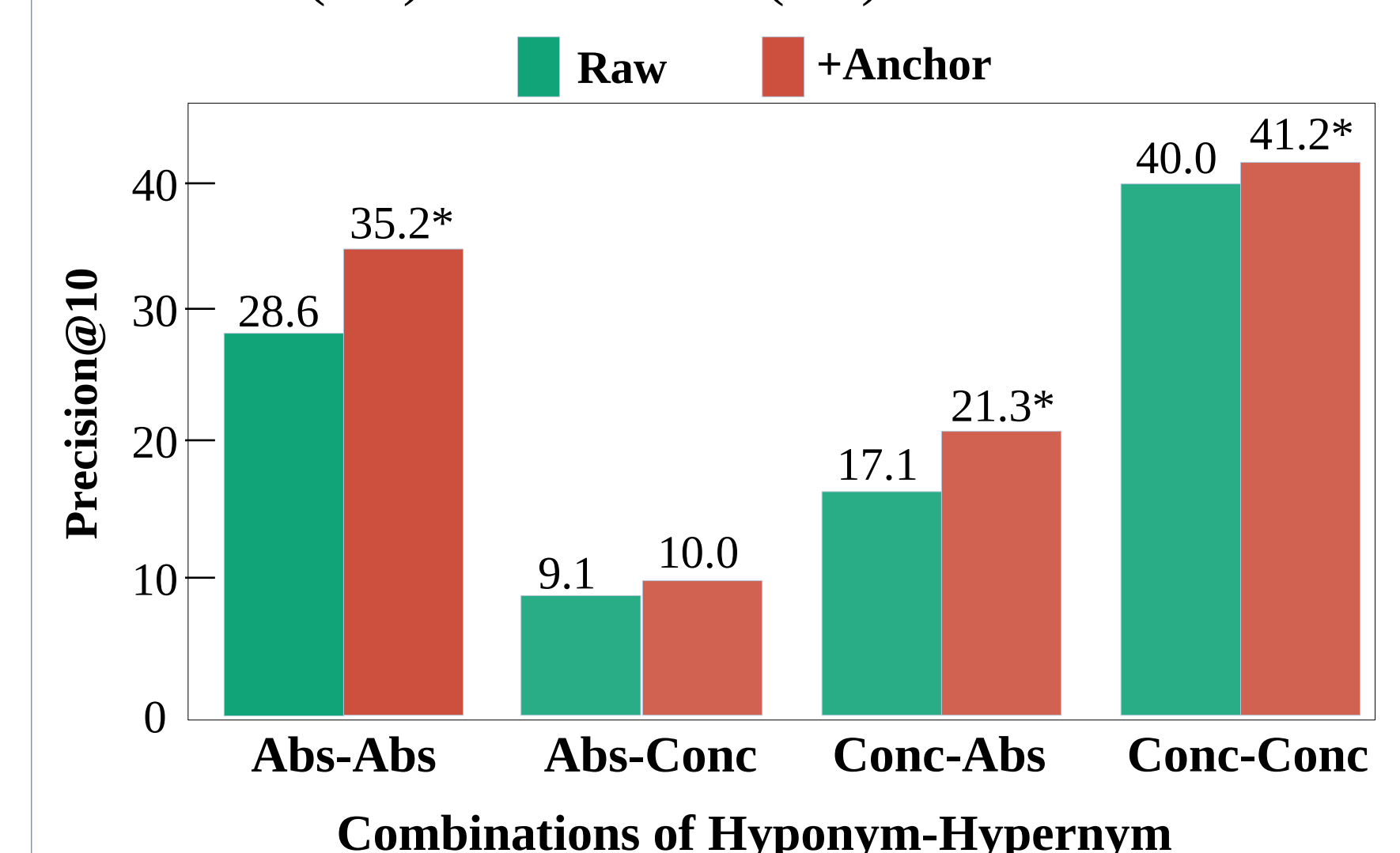
### Anchors improve performance for rare hyponyms

Word frequency from WorldLex (Gimenes and New, 2015)



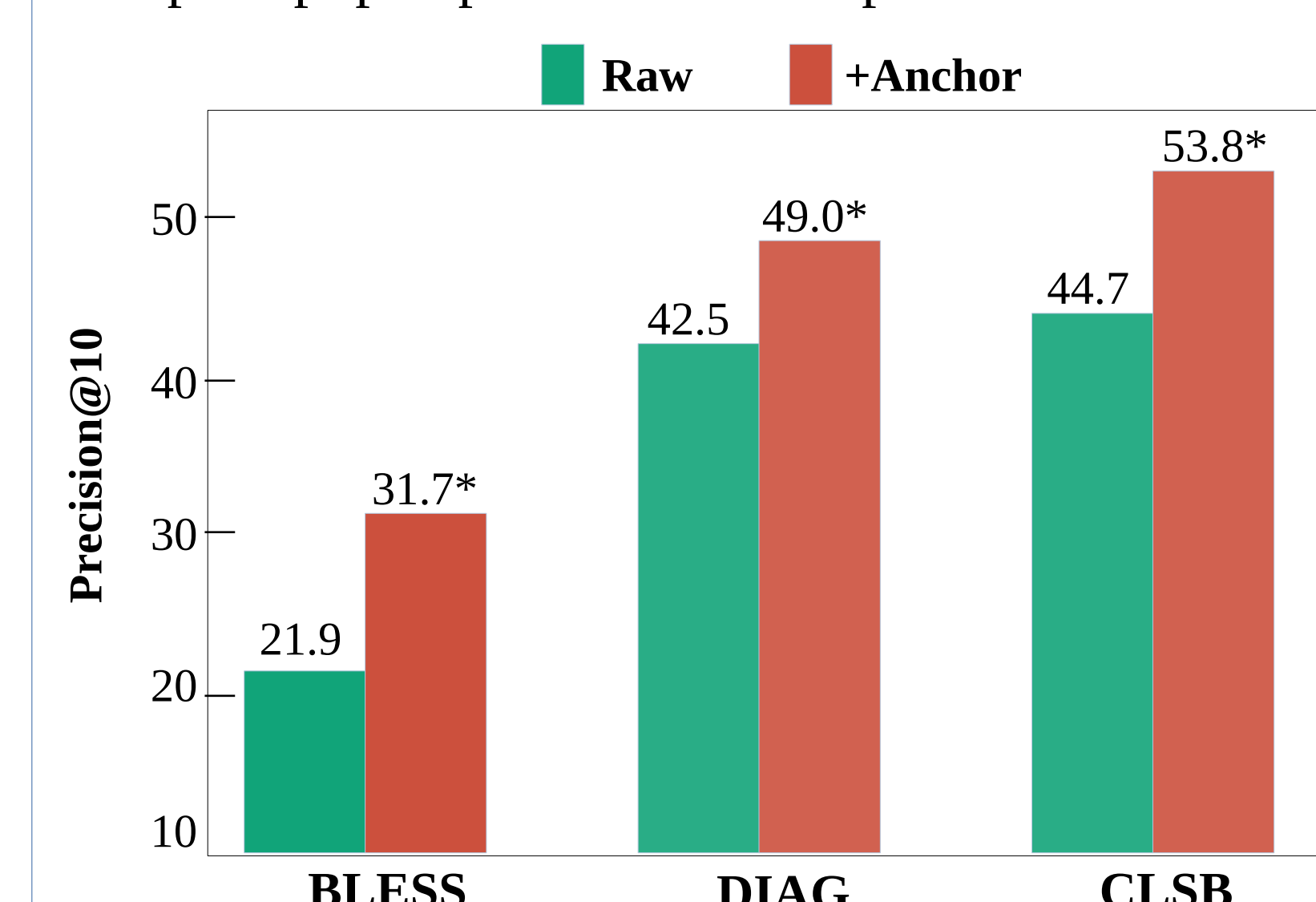
### Anchors largely boost performance for abstract hypernyms

Brysbaert's Abstractness Score (2014): Scale of 1-5 Abstract (< 3) and Concrete (≥ 3)



### Anchors raise hypernyms consistency of paraphrased prompts

Consistent: if a hypernym was correctly predicted by all prompt paraphrases within a pattern set



\* indicates significant difference (p < 0.05) of RAW vs. Anchored prompts

## Hypernym shift with anchors

Examples of rare and abstract hyponyms

### Hypernym Predictions (Y)

X	Raw Prompts	Anchored Prompts	Anchors
dray	boat machine took	<b>vehicle</b> cart tool	wagon tractor cart horse yes
happiness	joy life pleasure	joy feeling <b>emotion</b>	love joy good personal maybe



## Error Analysis

### Why do non-anchored patterns produce inconsistent results?

- Too generic predictions (e.g., things, items)
- Hypernym confusion (e.g., A dog is a cat.)

### When does anchoring hurt?

- Sense ambiguity:
  - e.g., fan is a person or an appliance
  - BERT anchors might skew to one sense
- Noisy anchors: irrelevant words (e.g., actually, yes) and relevant non-siblings (e.g., wood and lake)

## Contribution

- Thorough evaluation over six hypernym datasets
- Anchored prompts improve performance on rare/abstract concepts
- Anchored prompts improve answer consistency

