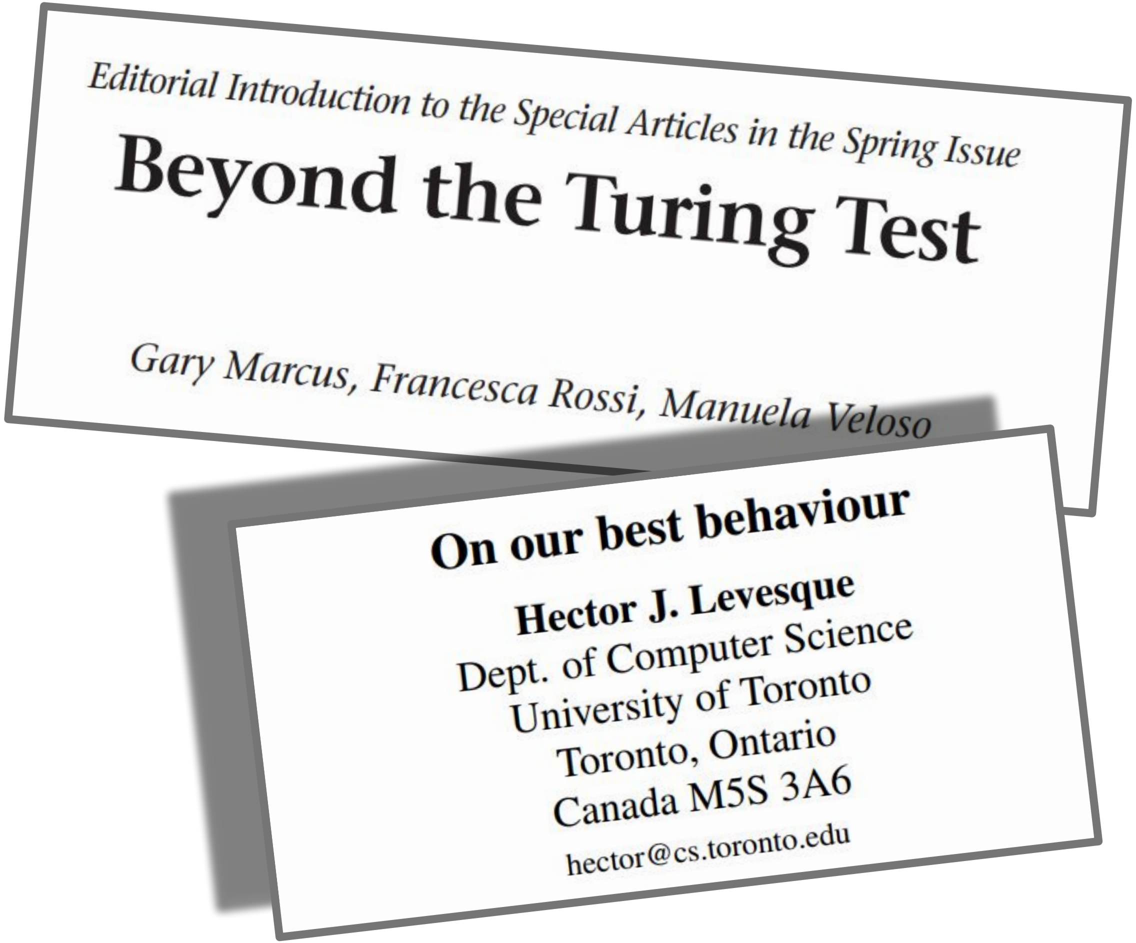
A Simple Method for **Commonsense Reasoning**

Trieu and Quoc



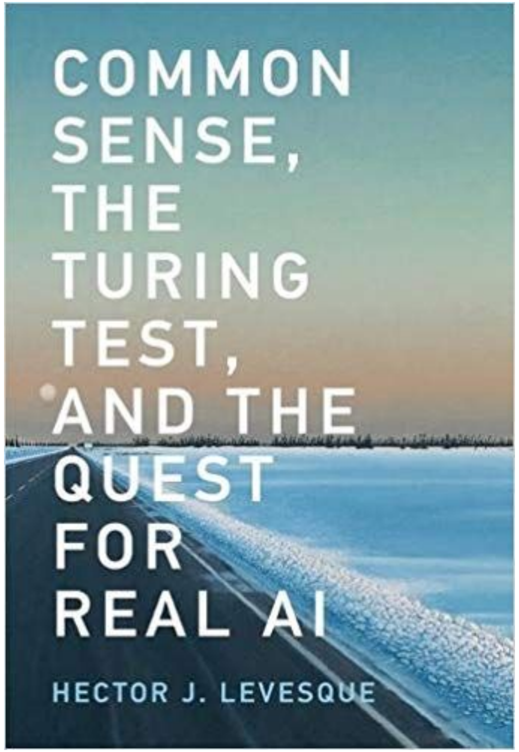
Editorial Introduction to the Special Articles in the Spring Issue

Beyond the Turing Test

Gary Marcus, Francesca Rossi, Manuela Veloso

On our best behaviour

Hector J. Levesque
Dept. of Computer Science
University of Toronto
Toronto, Ontario
Canada M5S 3A6
hector@cs.toronto.edu



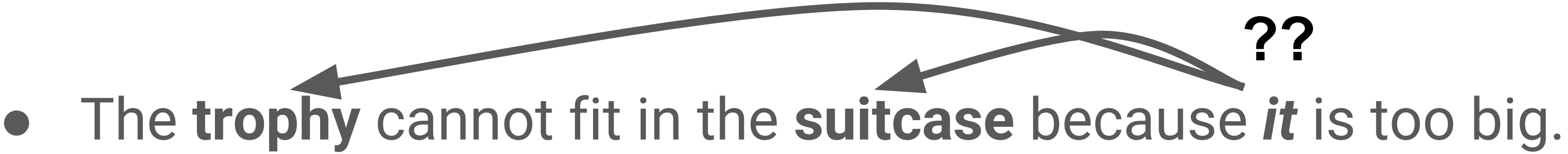
COMMON SENSE, THE TURING TEST, AND THE QUEST FOR REAL AI

HECTOR J. LEVESQUE

- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

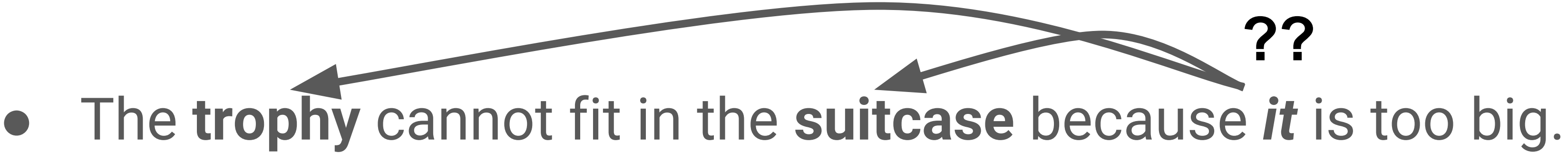
??

- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.



- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

Winograd Schema Challenge



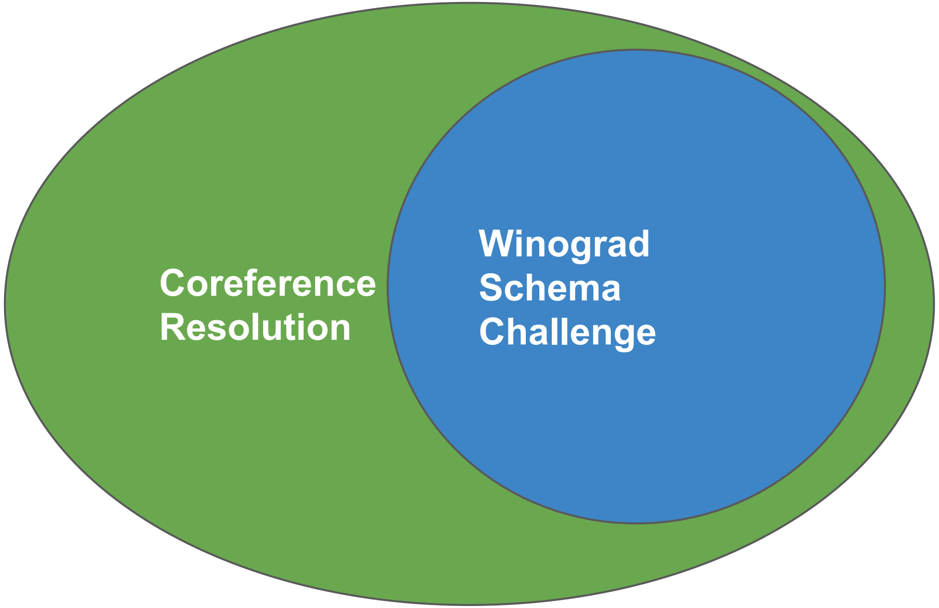
??

- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

Winograd Schema Challenge

??

- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

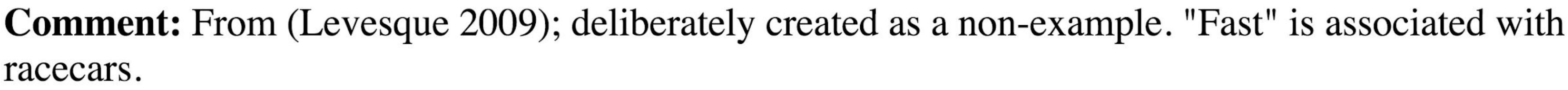


Winograd Schema Challenge

The **racecar** zoomed by the **school bus** because **[it]** was going so fast.

Winograd Schema Challenge

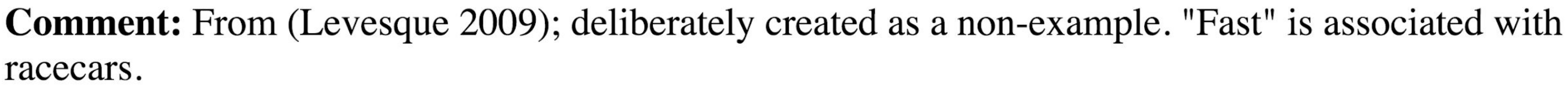
The **racecar** zoomed by the **school bus** because **[it]** was going so fast.



Comment: From (Levesque 2009); deliberately created as a non-example. "Fast" is associated with racecars.

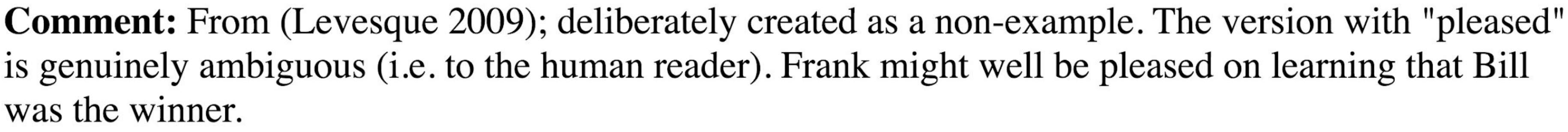
Winograd Schema Challenge

The **racecar** zoomed by the **school bus** because **[it]** was going so fast.



Comment: From (Levesque 2009); deliberately created as a non-example. "Fast" is associated with racecars.

Frank was pleased when **Bill** said **[he]** was the winner of the competition.

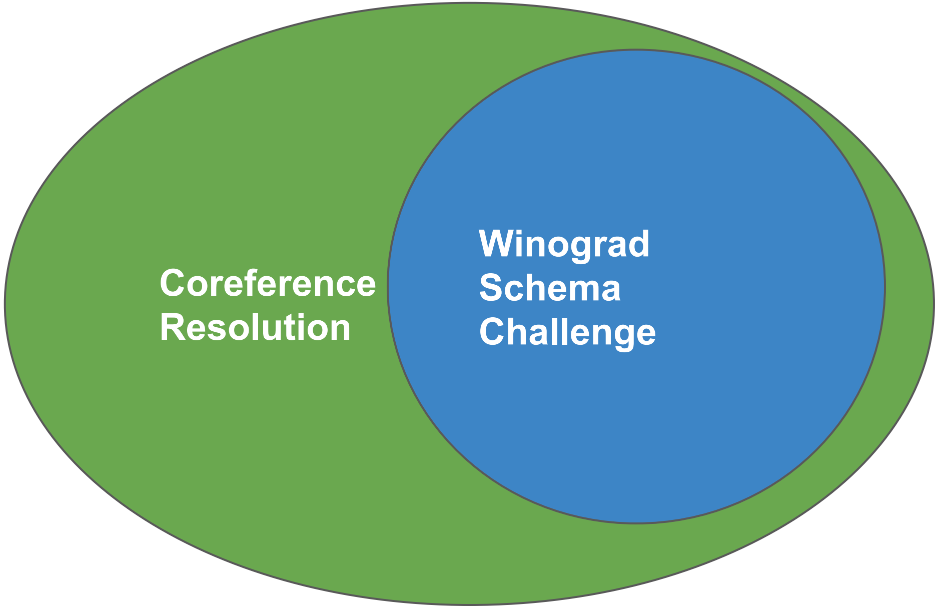


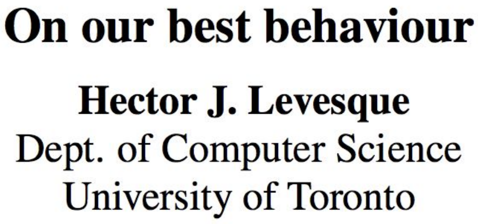
Comment: From (Levesque 2009); deliberately created as a non-example. The version with "pleased" is genuinely ambiguous (i.e. to the human reader). Frank might well be pleased on learning that Bill was the winner.

Winograd Schema Challenge

??

- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.





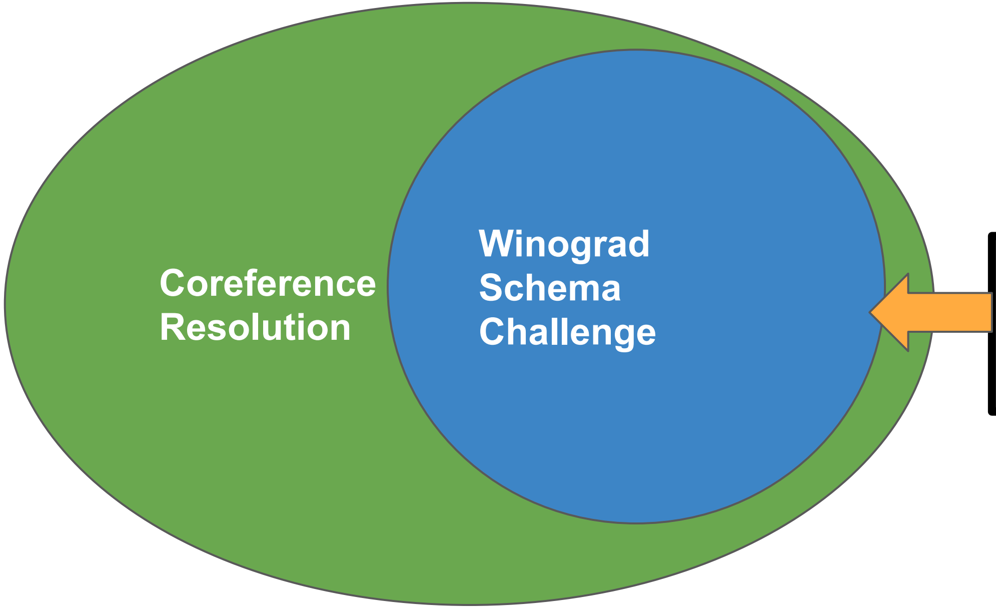
On our best behaviour

Hector J. Levesque
Dept. of Computer Science
University of Toronto

Winograd Schema Challenge

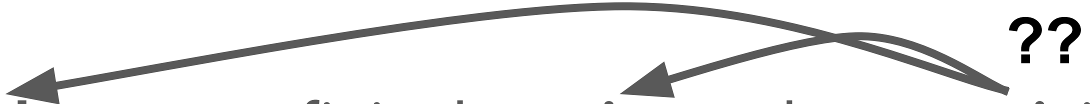
??

- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

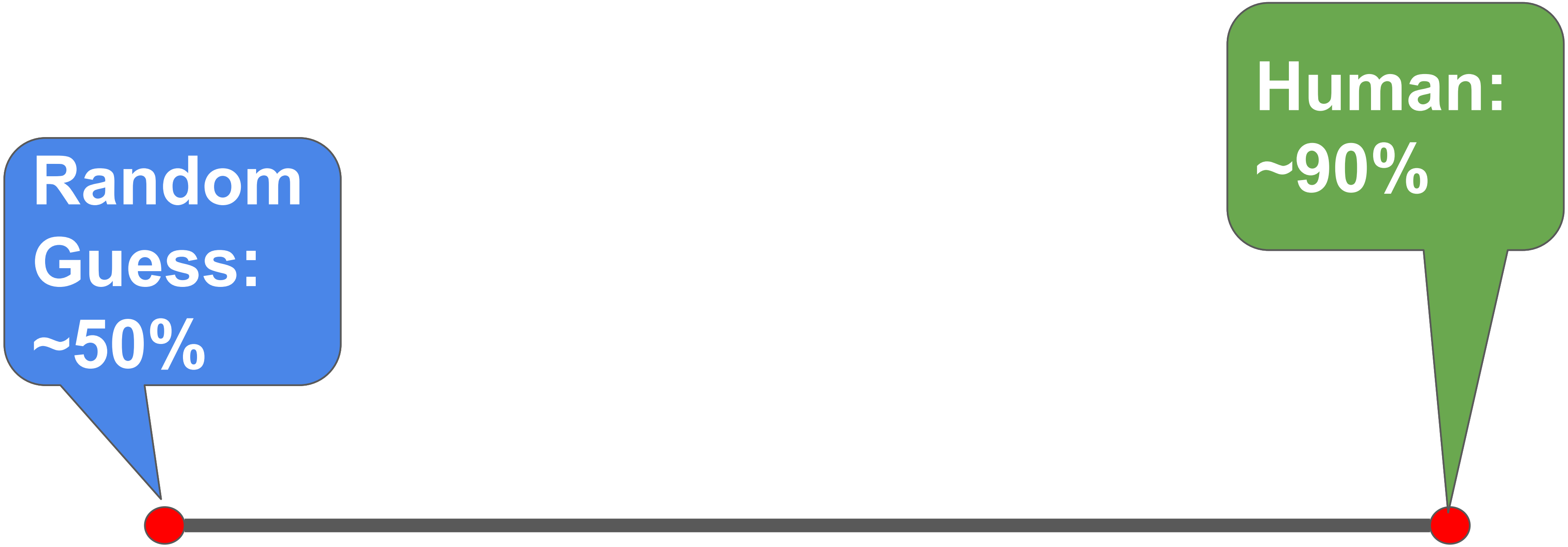


Expensive to construct
commonsensereasoning.org
< 300 samples.

Winograd Schema Challenge



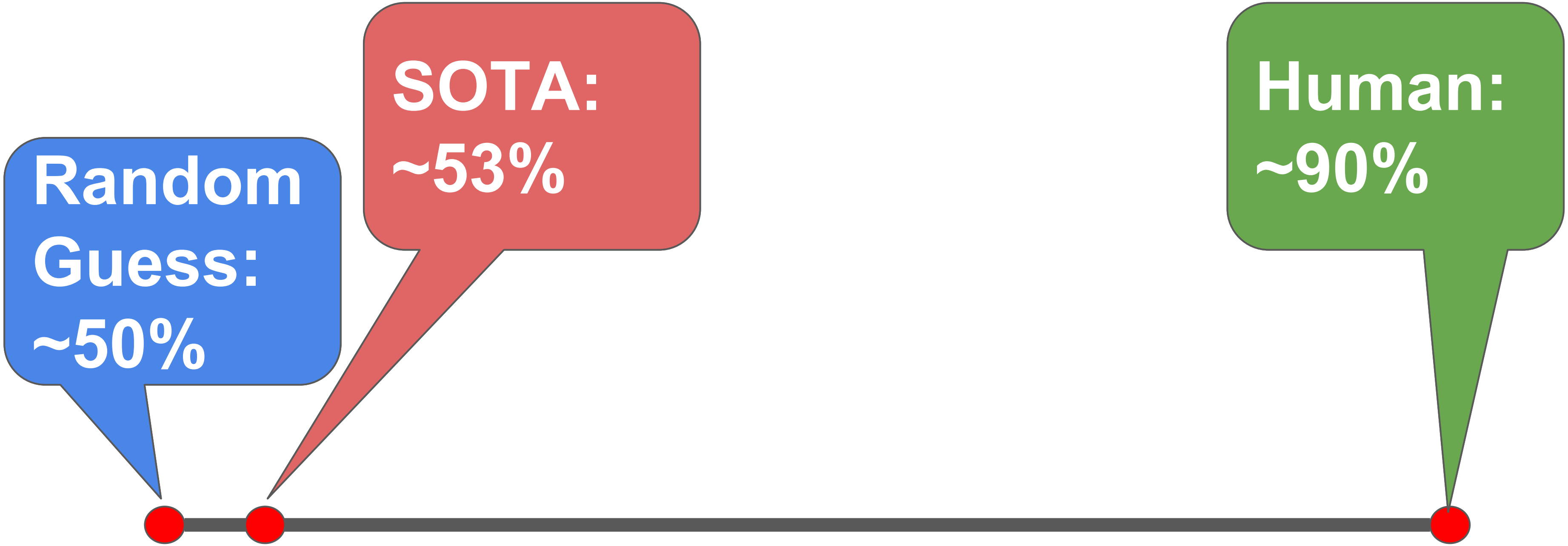
- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.



Winograd Schema Challenge

??

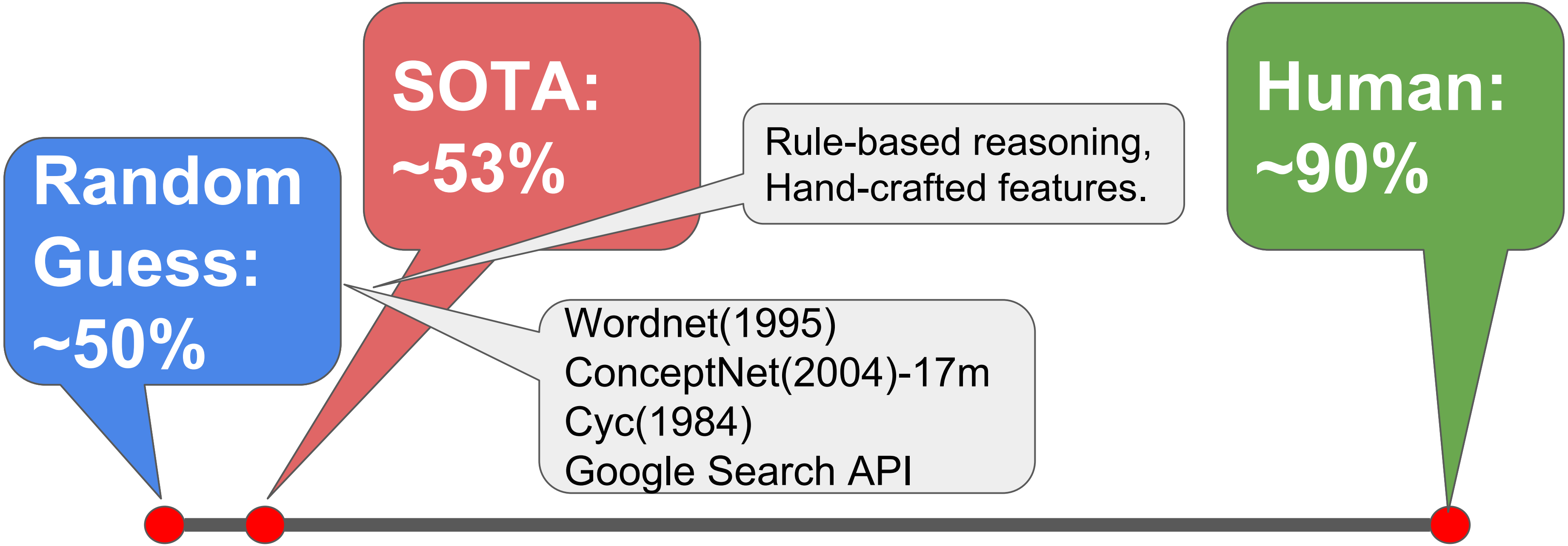
- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.



Winograd Schema Challenge

??

- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.



Random Guess: ~50%

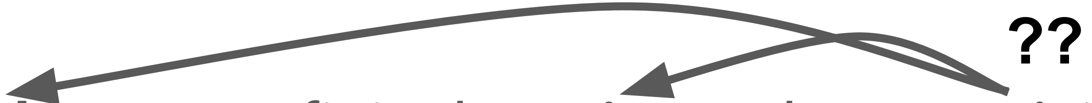
SOTA: ~53%

Rule-based reasoning, Hand-crafted features.

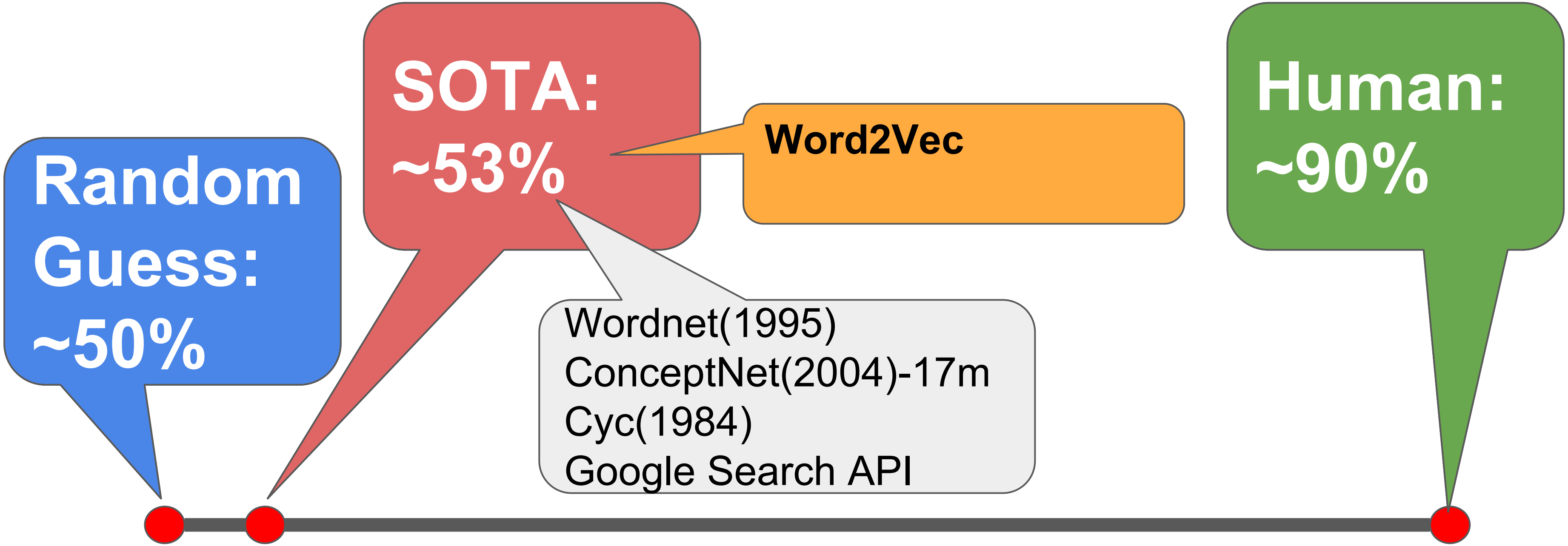
Wordnet(1995)
ConceptNet(2004)-17m
Cyc(1984)
Google Search API

Human: ~90%

Winograd Schema Challenge



- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.



Random Guess: ~50%

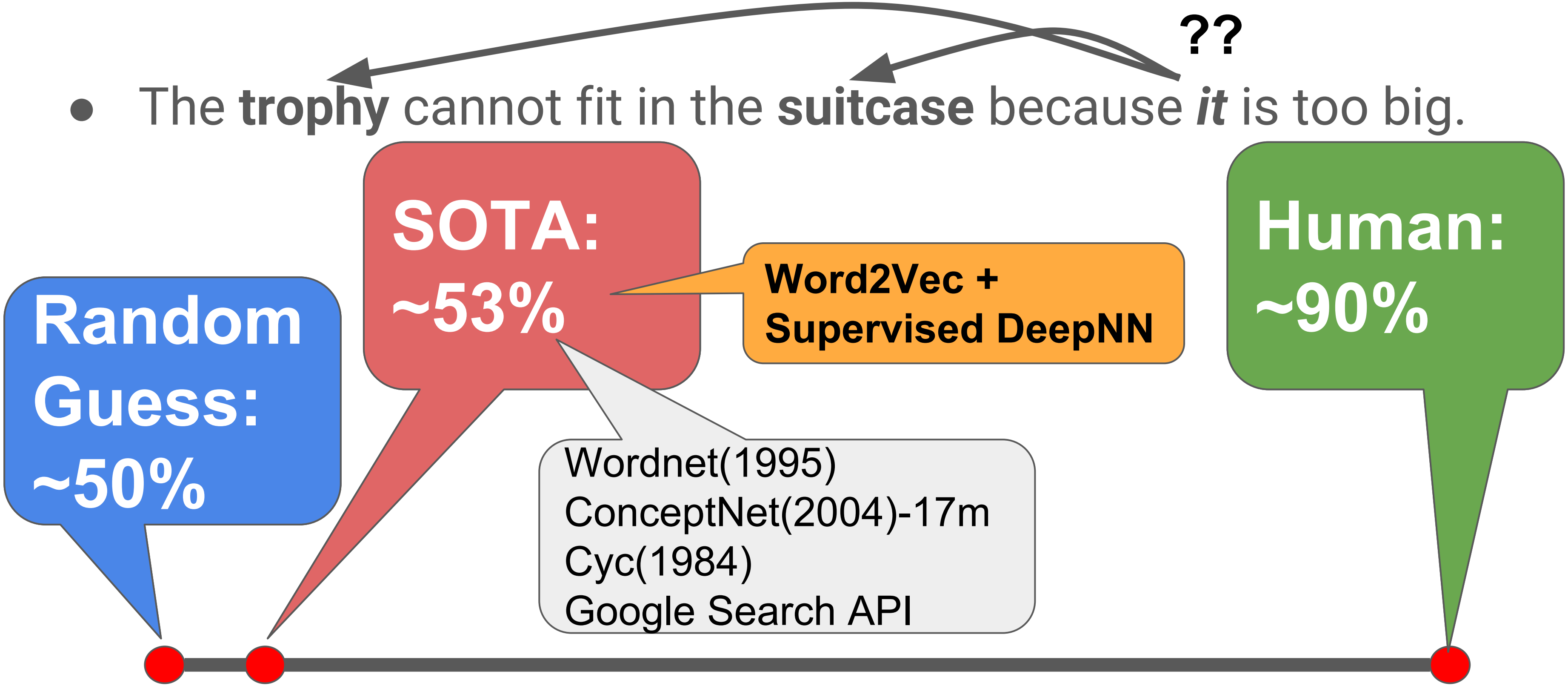
SOTA: ~53%

Word2Vec

Wordnet(1995)
ConceptNet(2004)-17m
Cyc(1984)
Google Search API

Human: ~90%

Winograd Schema Challenge

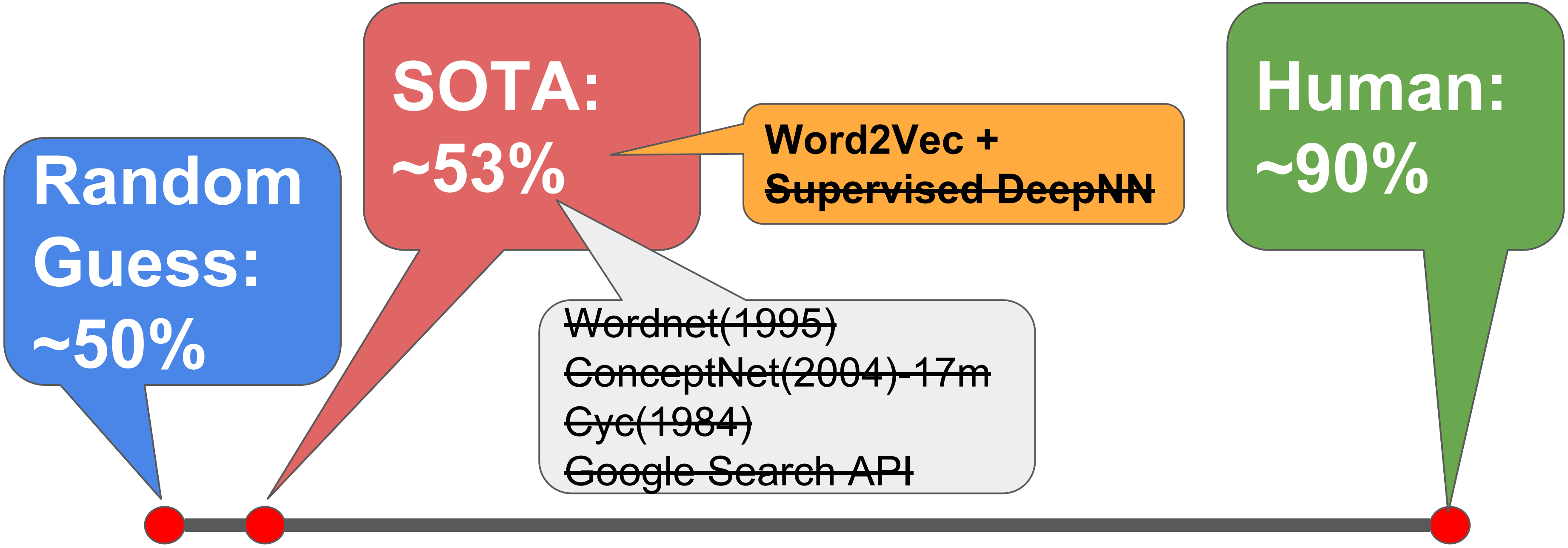


- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

Winograd Schema Challenge

??

- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.



Random Guess: ~50%

SOTA: ~53%

Word2Vec +
~~Supervised DeepNN~~

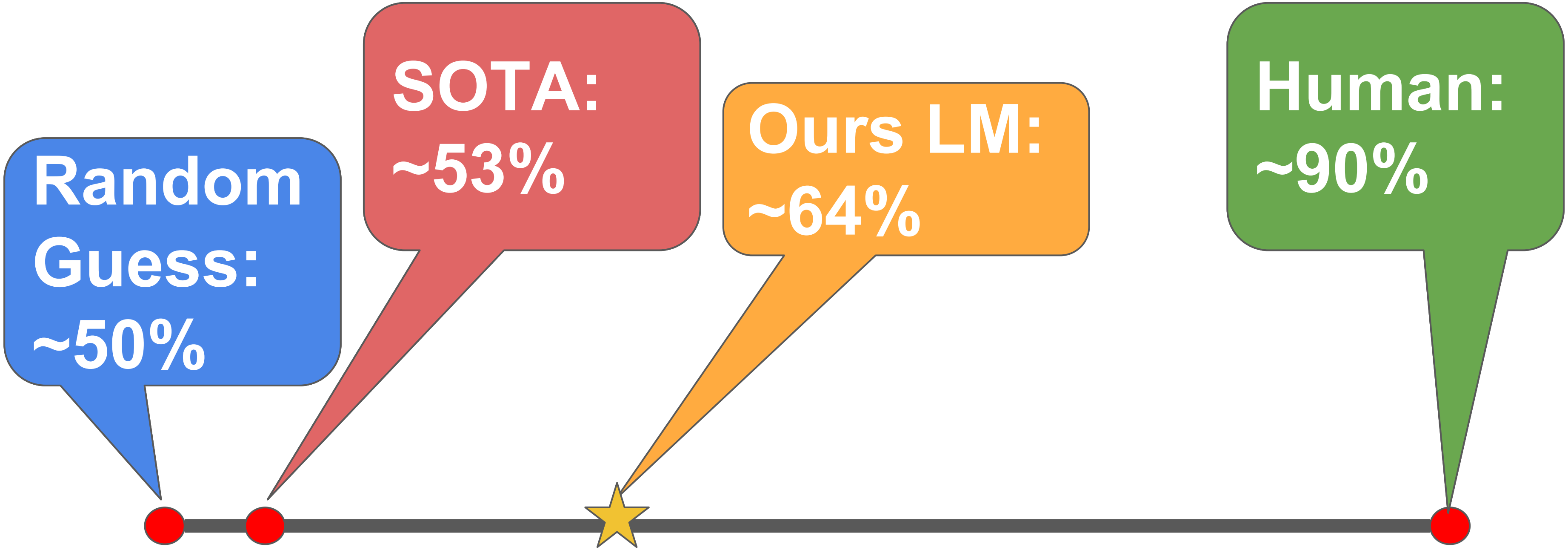
~~Wordnet(1995)~~
~~ConceptNet(2004)-17m~~
~~Cyc(1984)~~
~~Google Search API~~

Human: ~90%

Winograd Schema Challenge

??

- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.



Random Guess: ~50%

SOTA: ~53%

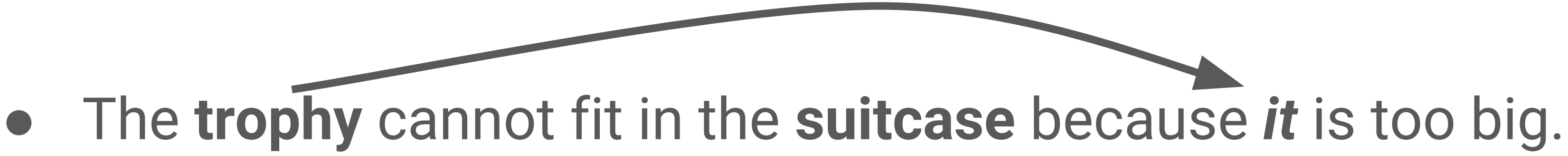
Ours LM: ~64%

Human: ~90%

Method

- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

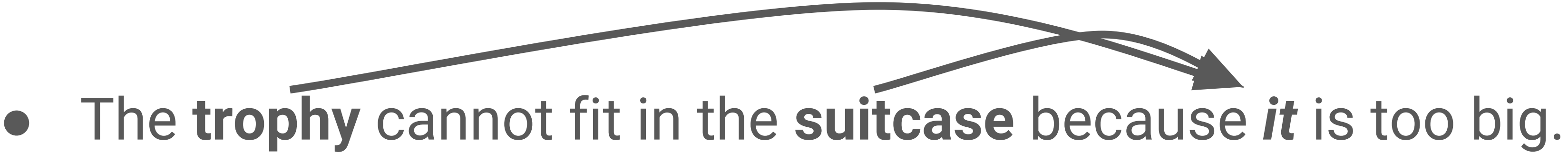
Method



- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

Method

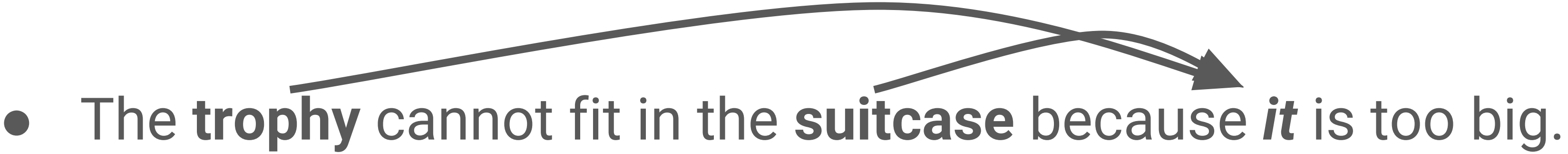


- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method



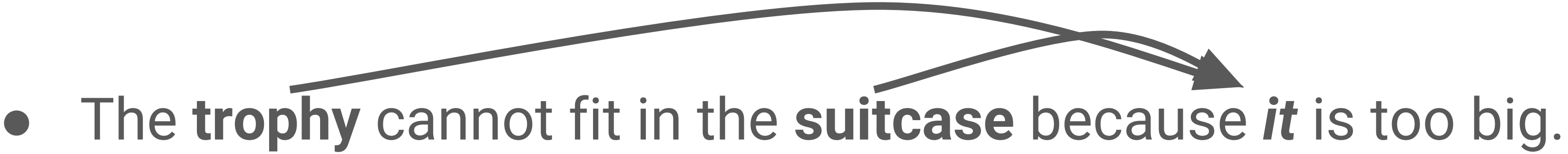
- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

More **probable**?

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

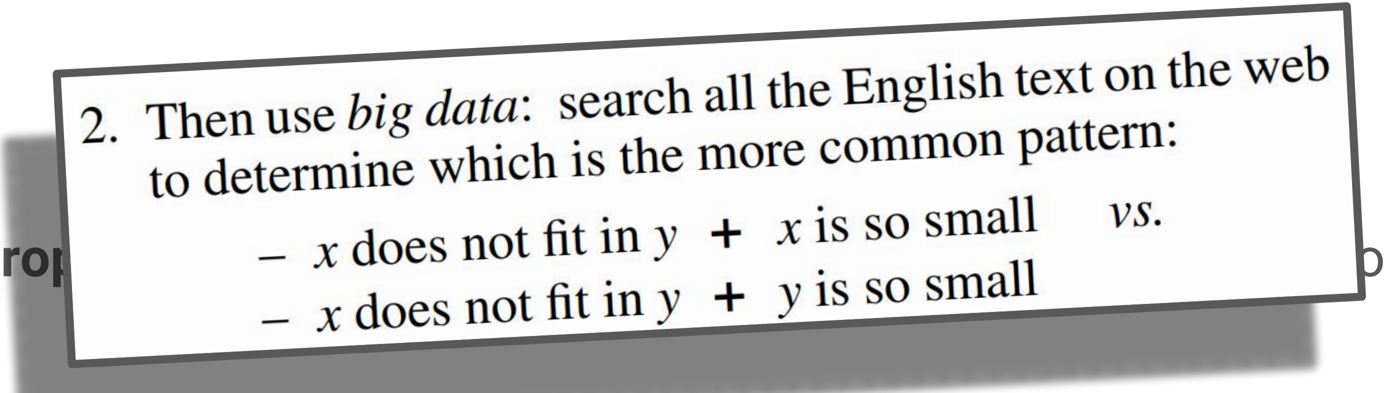
The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method



- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

The **trop** o big .

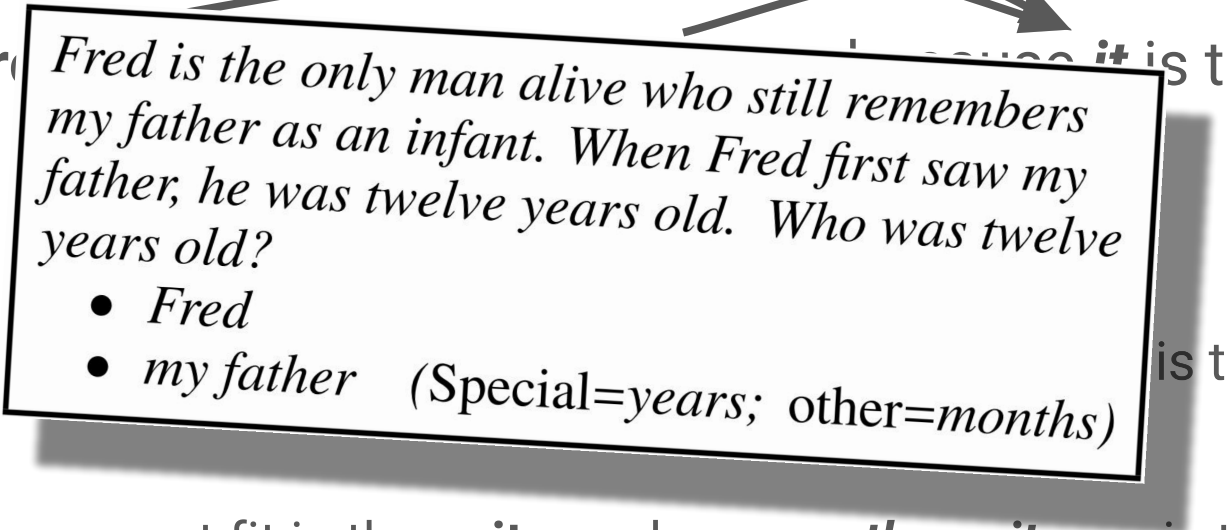


2. Then use *big data*: search all the English text on the web to determine which is the more common pattern:
   - *x* does not fit in *y* **+** *x* is so small *vs.*
   - *x* does not fit in *y* **+** *y* is so small

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method

- The **tr** ecause ***it*** is too big.



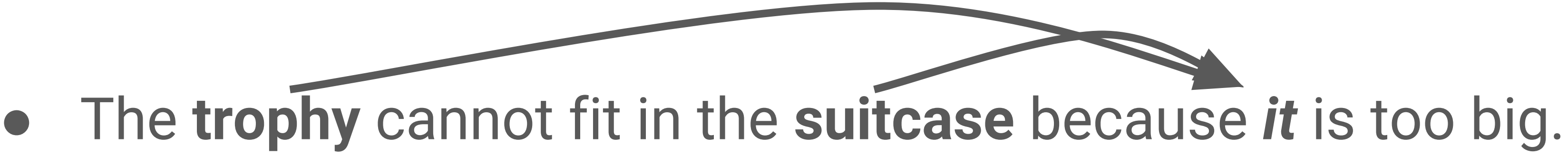
Fred is the only man alive who still remembers my father as an infant. When Fred first saw my father, he was twelve years old. Who was twelve years old?

- *Fred*
- *my father* (Special=*years*; other=*months*)

The **troph** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method



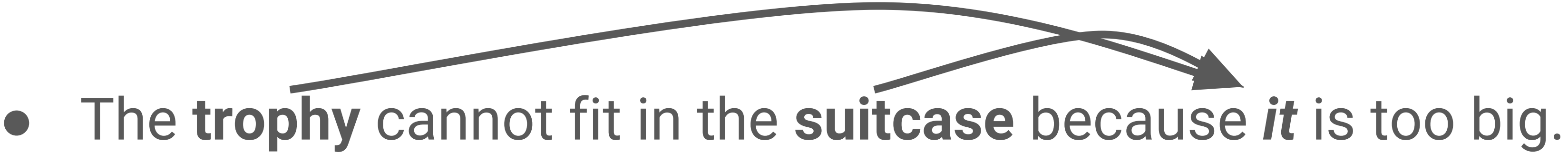
- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

More **probable**?

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method



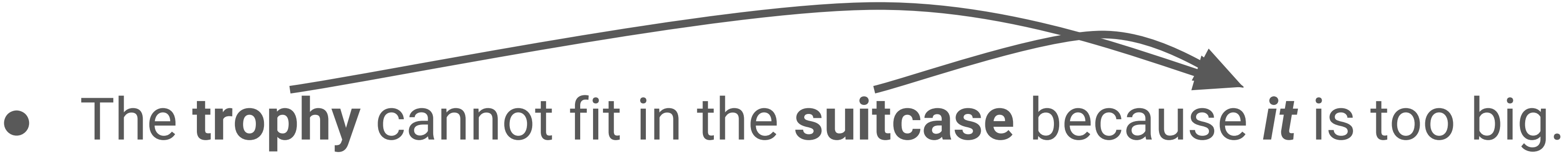
- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

$$P(substitution|Human\ knowledge)$$

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method



- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

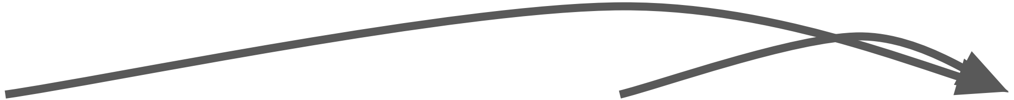
Language Model

$$P(substitution|Human\ knowledge)$$

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method



- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

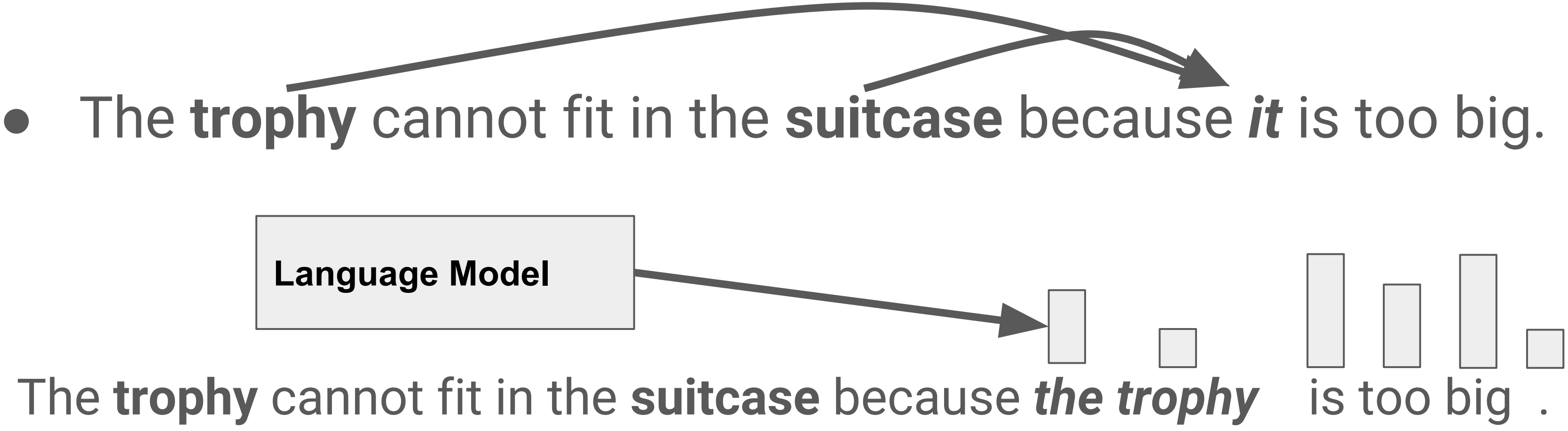
Language Model

$$P(substitution| \qquad \hat{\theta} \qquad)$$

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method



- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

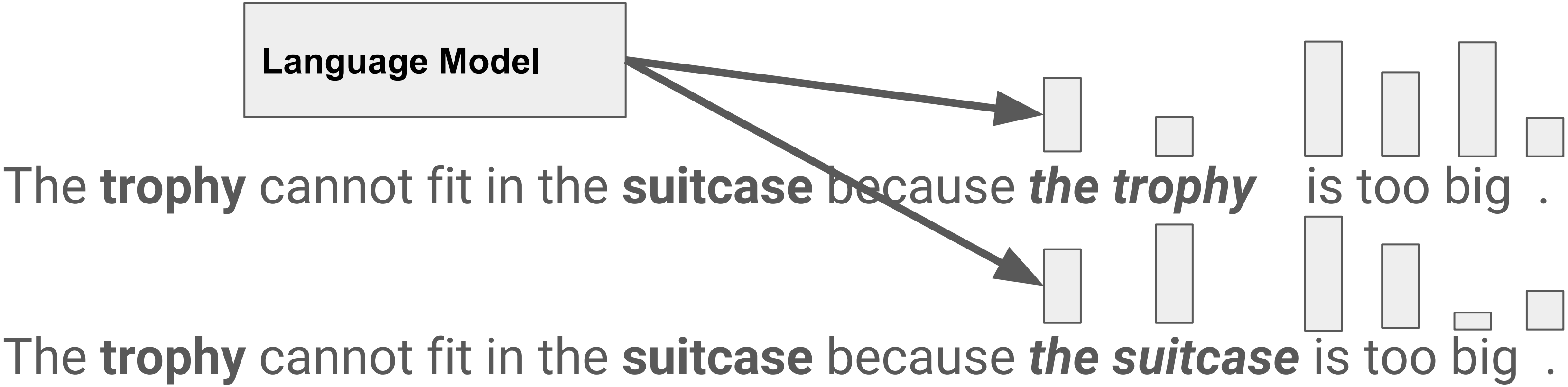
Language Model

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method

- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

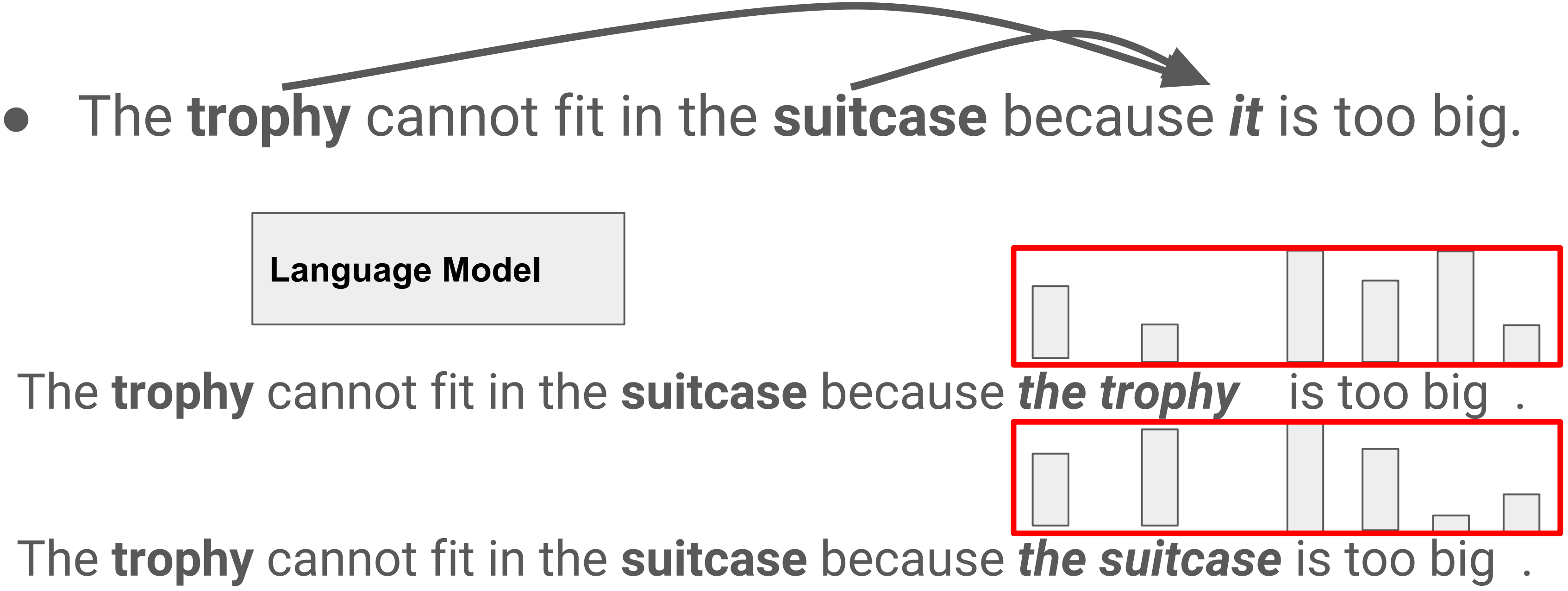


Language Model

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method



- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

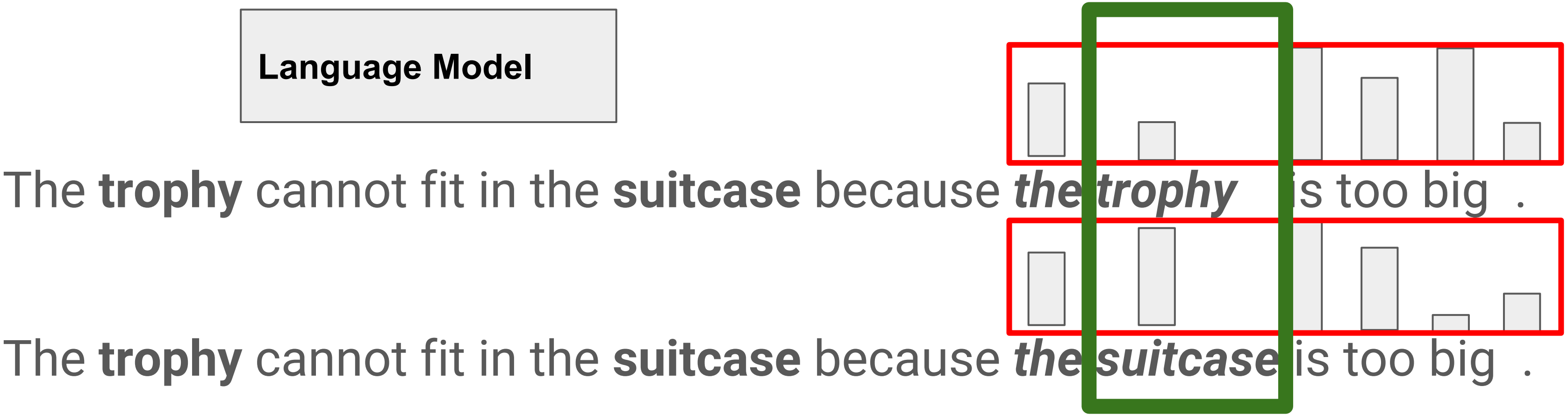
Language Model

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method

- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

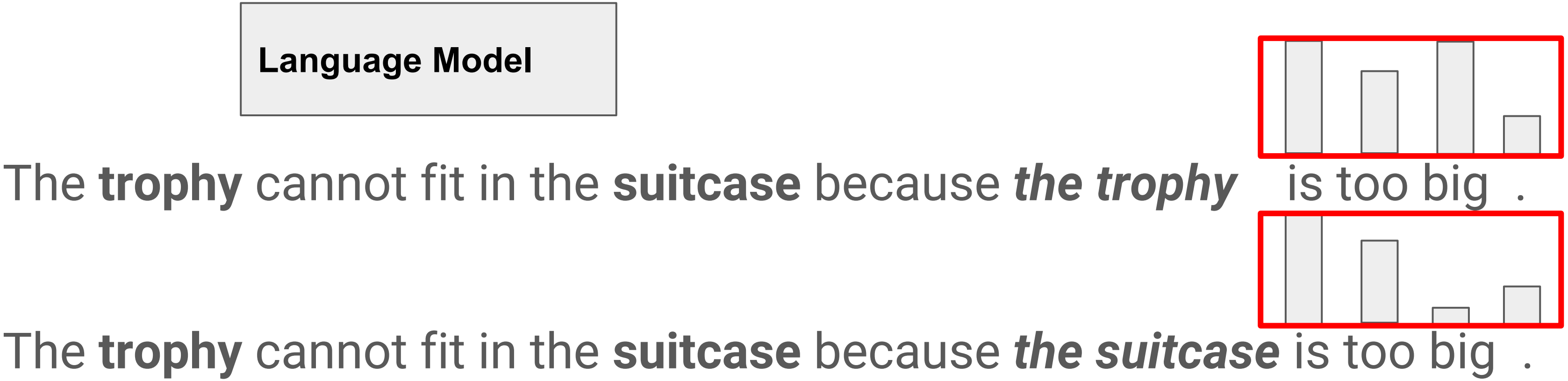


The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method

- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

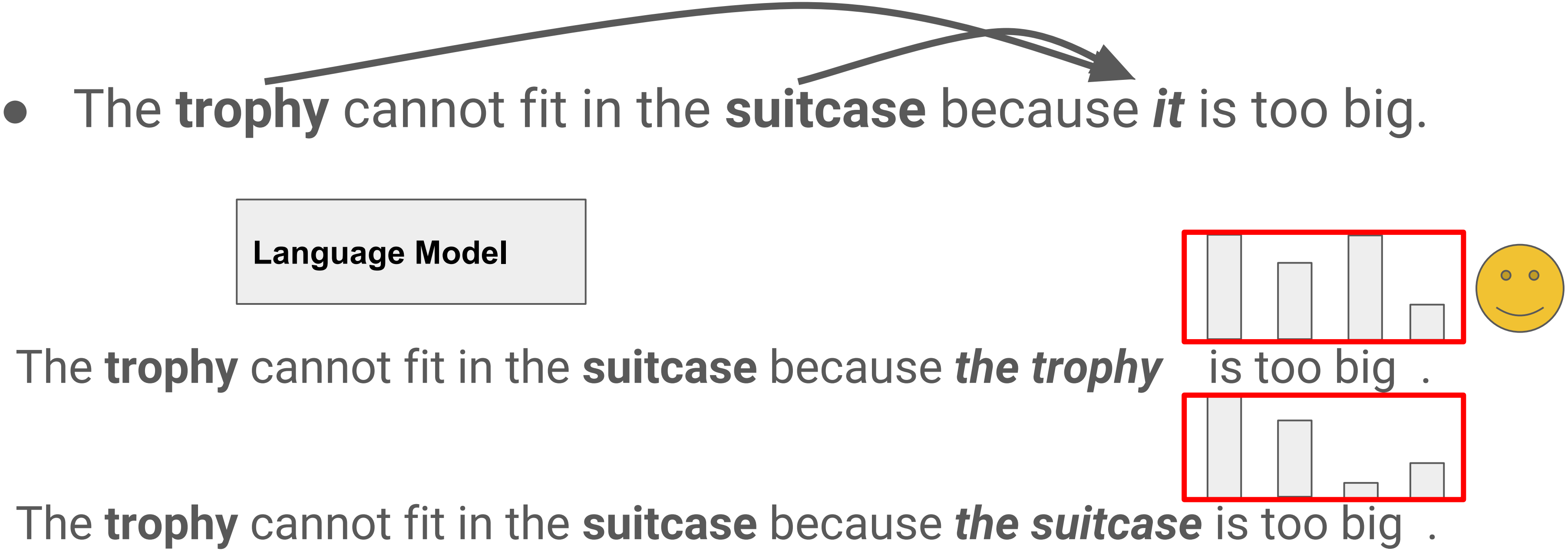


Language Model

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method



- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

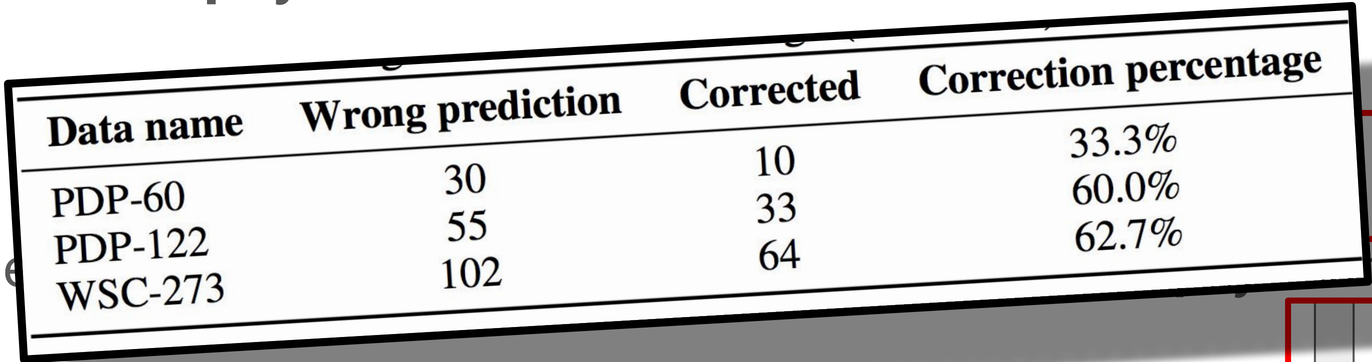
Language Model

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

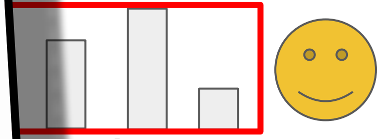
The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method

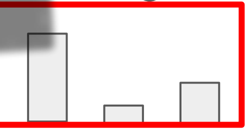
- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.



Data name	Wrong prediction	Corrected	Correction percentage
PDP-60	30	10	33.3%
PDP-122	55	33	60.0%
WSC-273	102	64	62.7%

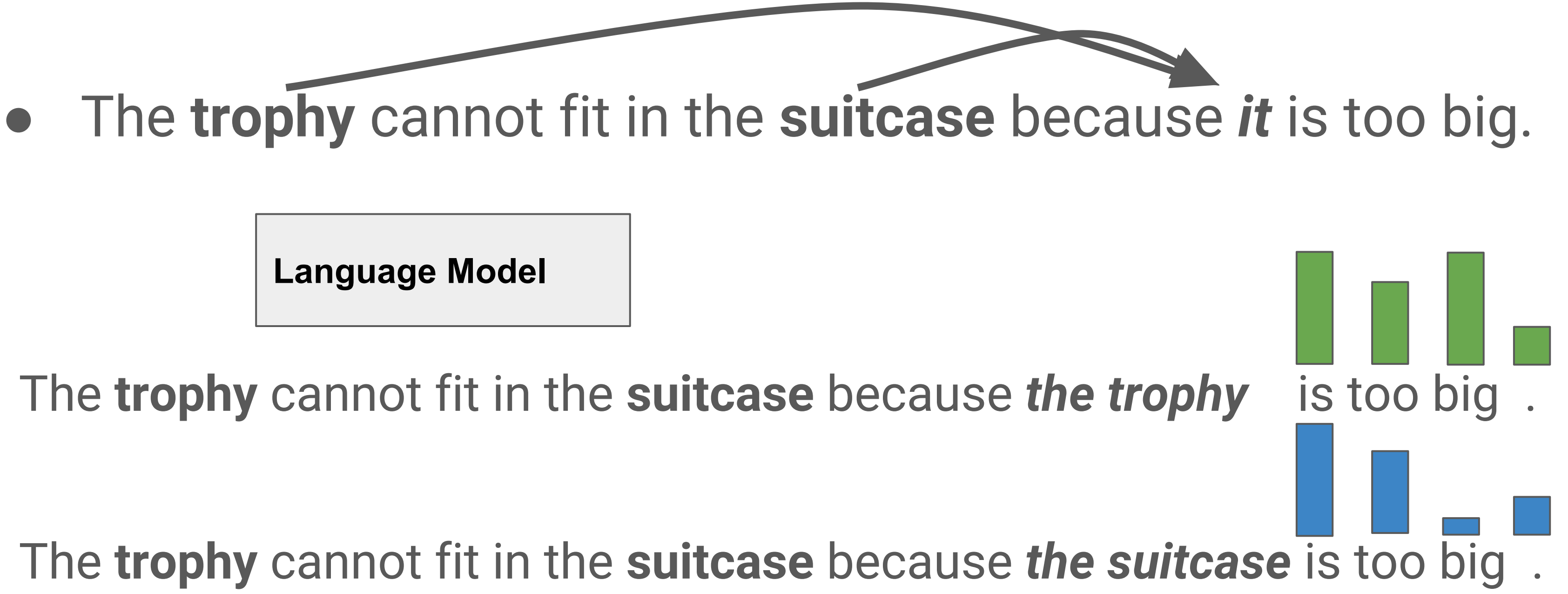


The too big .



The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method



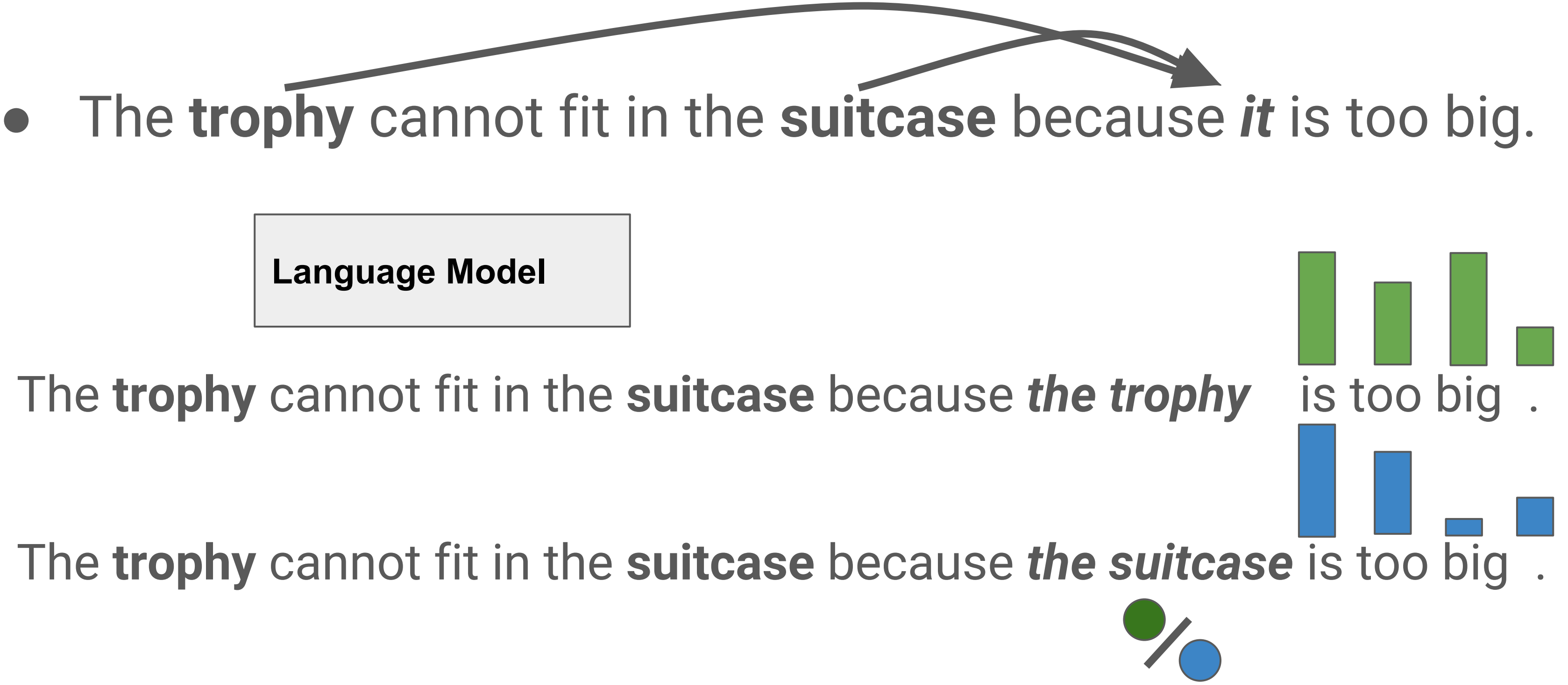
- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

Language Model

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method



- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

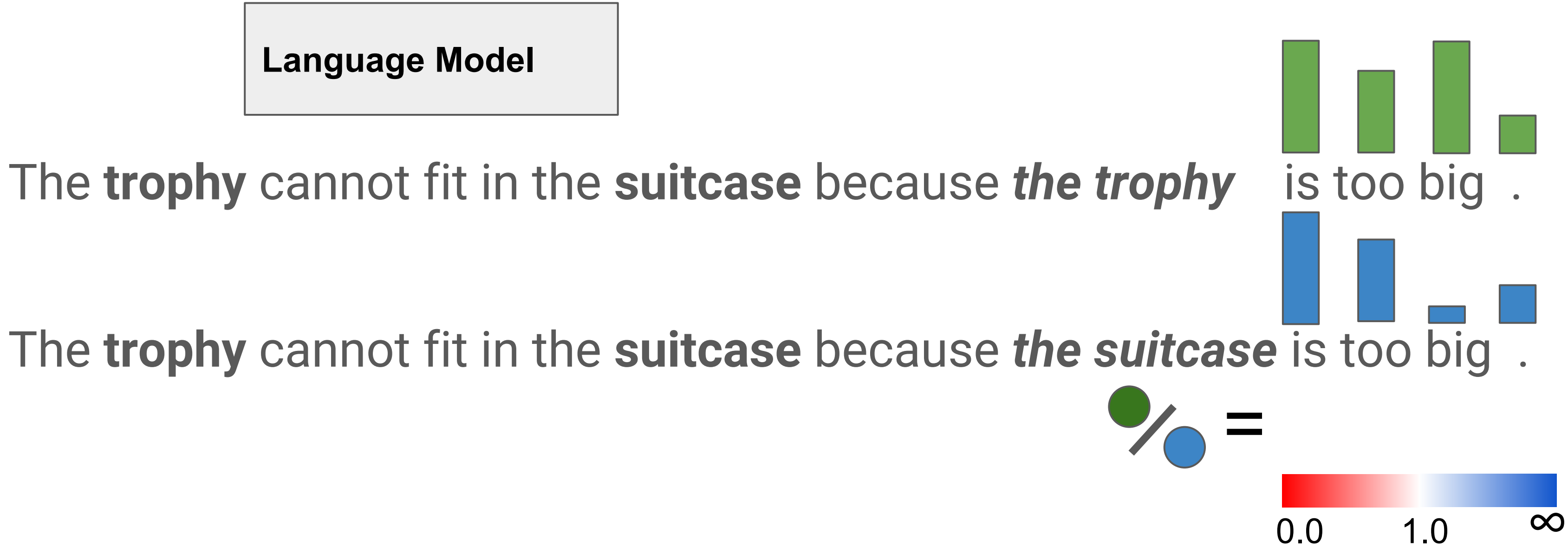
Language Model

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method

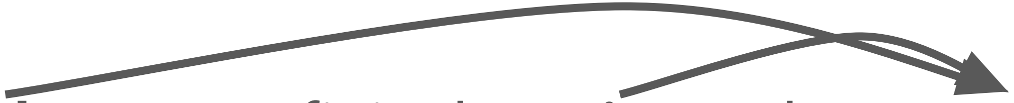
- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.



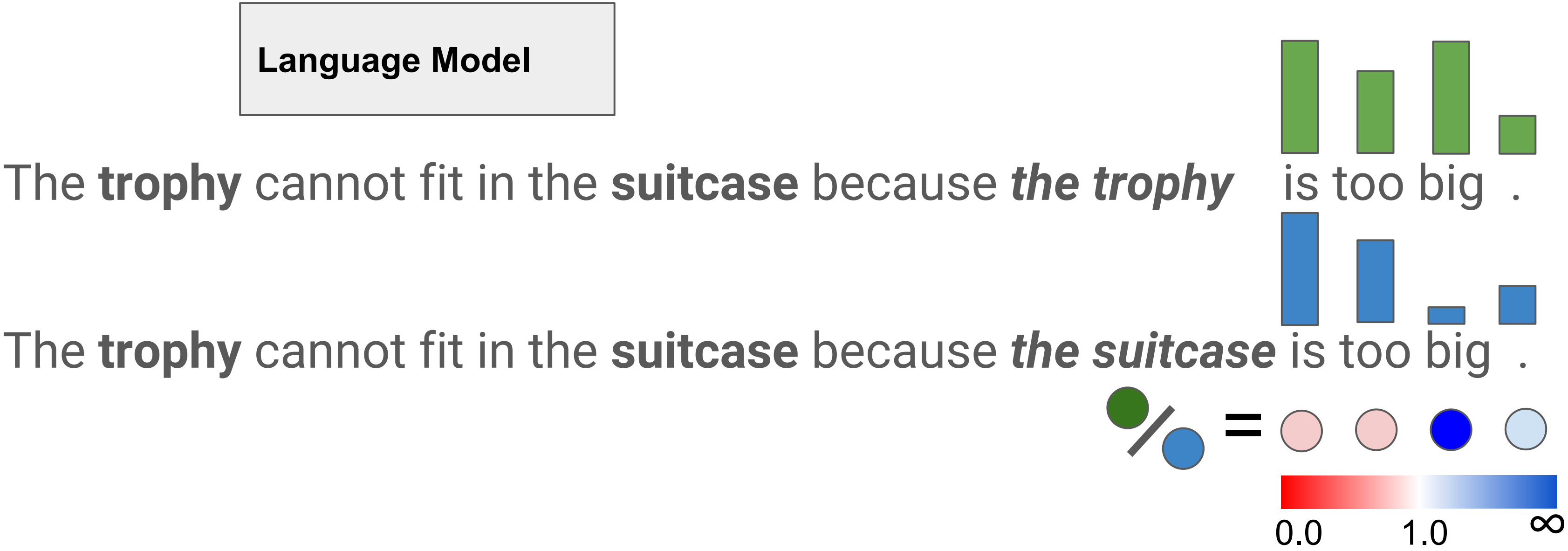
The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method



- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

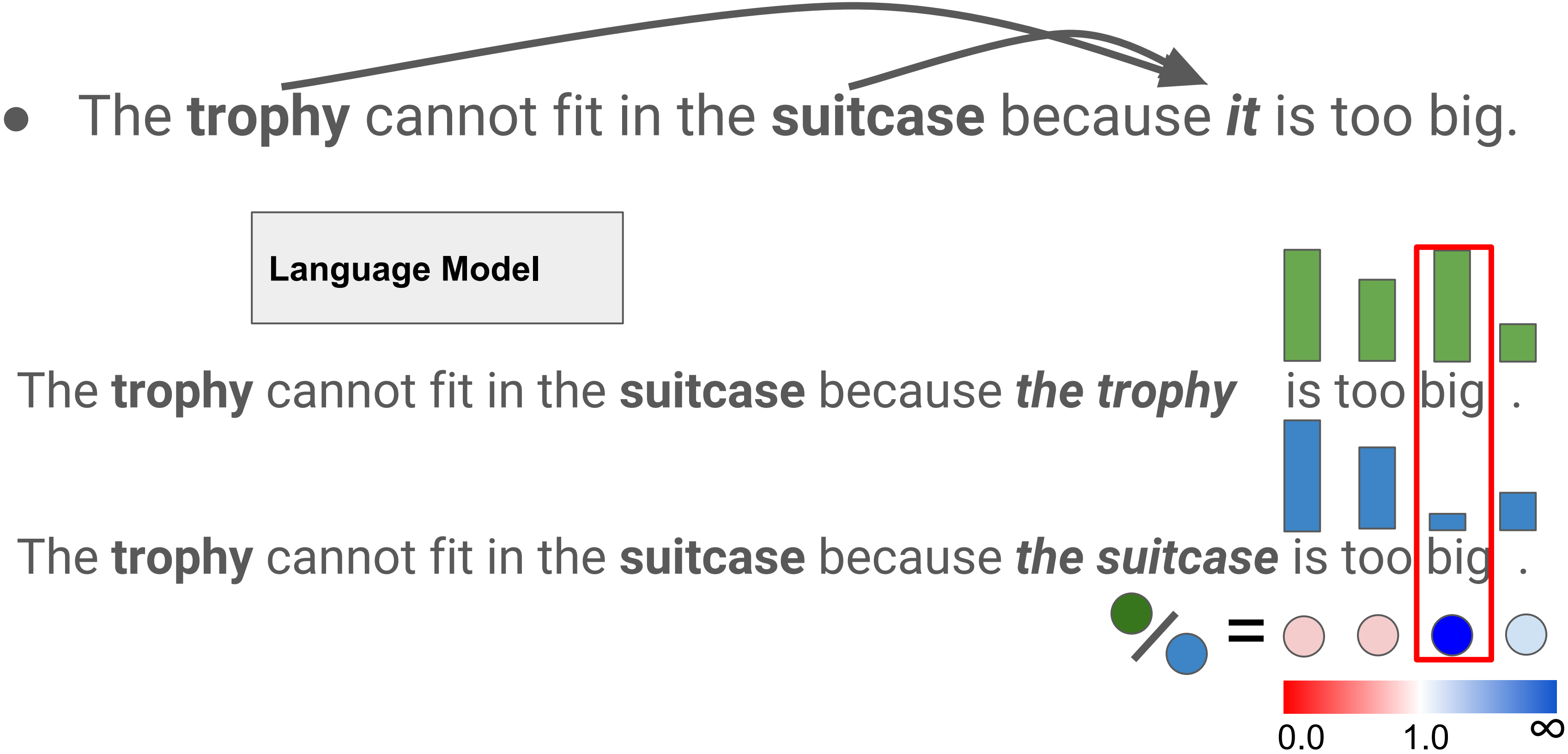


Language Model

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method



- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

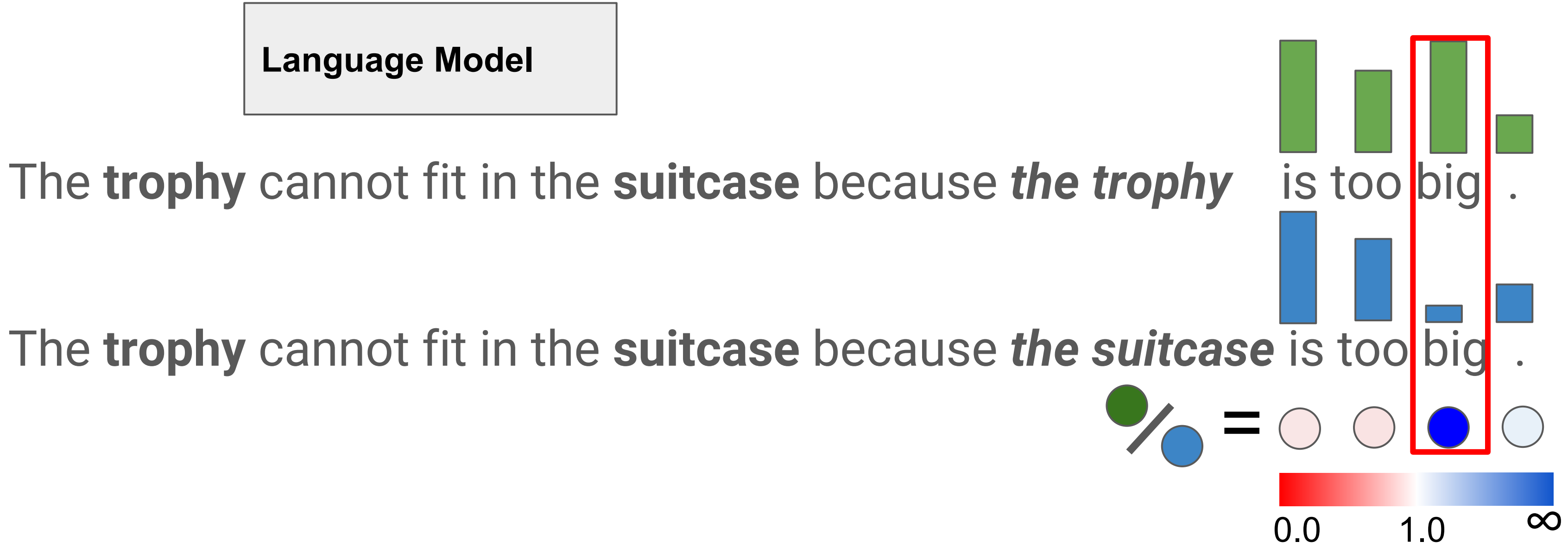
Language Model

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method

- The **trophy** cannot fit in the **suitcase** because ***it*** is too small.
- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.



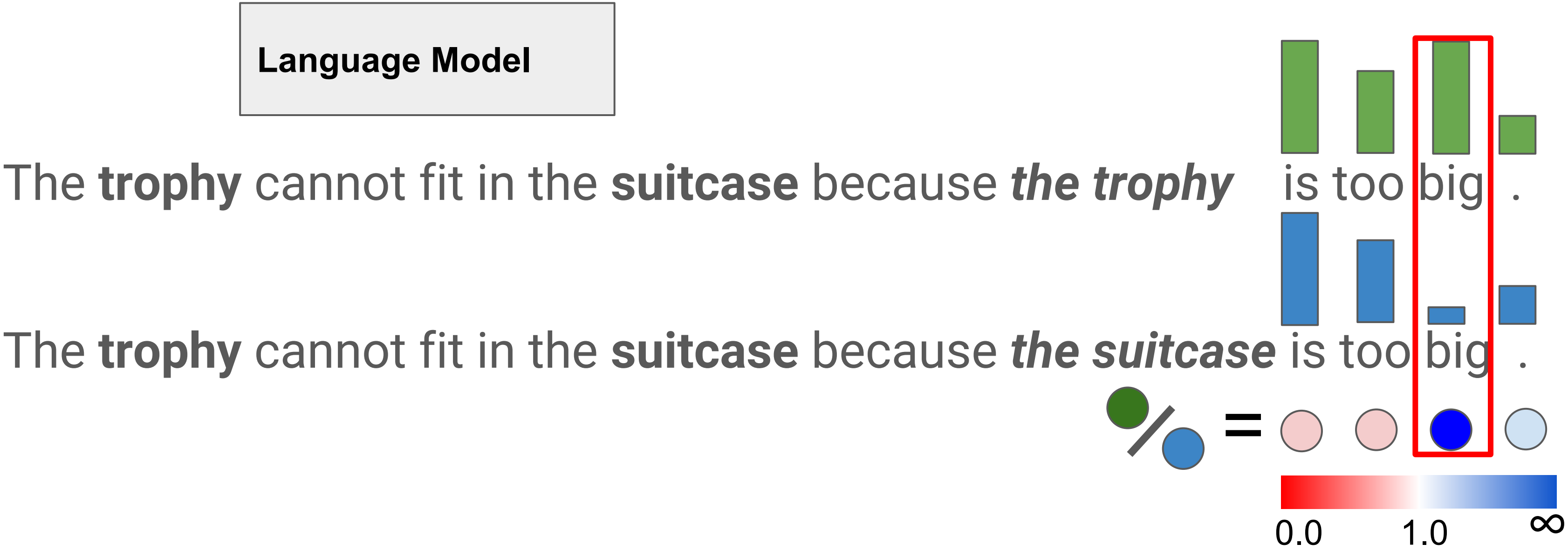
Language Model

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method

- The **trophy** cannot fit in the **suitcase** because ***it*** is too small.
- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.



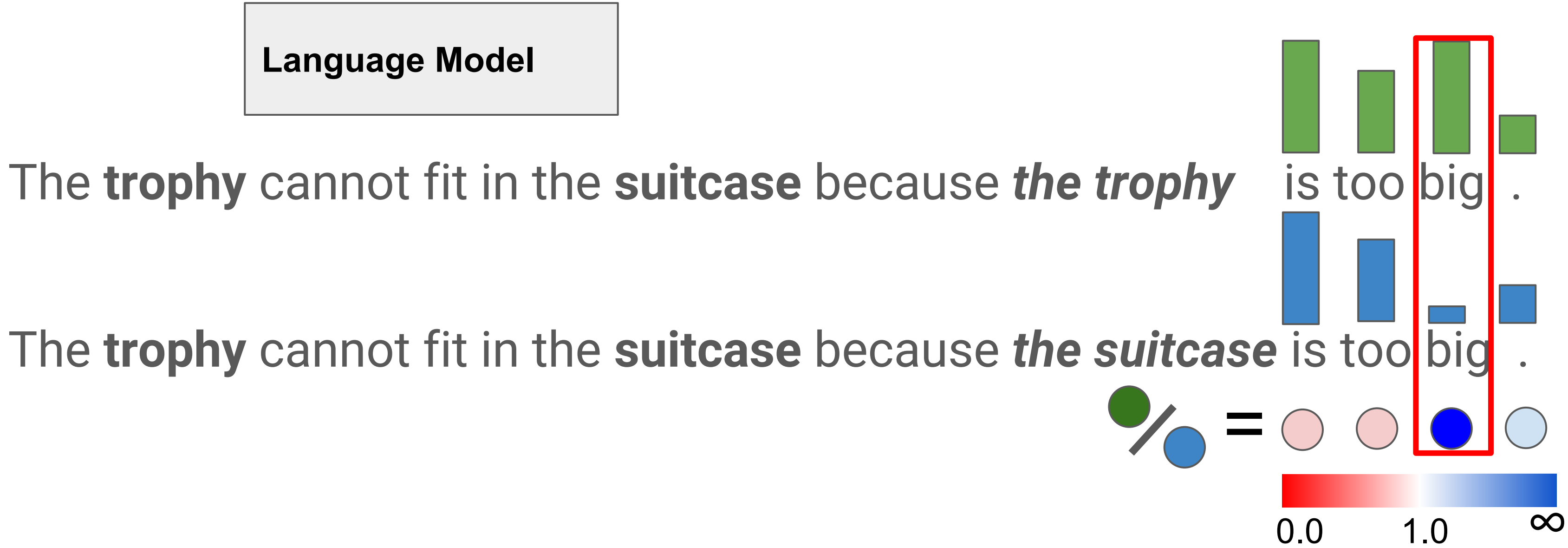
Language Model

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Method

- The **trophy** cannot fit in the **suitcase** because ***it*** is too small.
- The **trophy** cannot fit in the **suitcase** because ***it*** is too big.

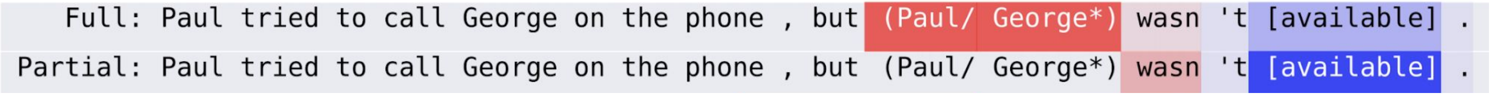


Language Model

The **trophy** cannot fit in the **suitcase** because ***the trophy*** is too big .

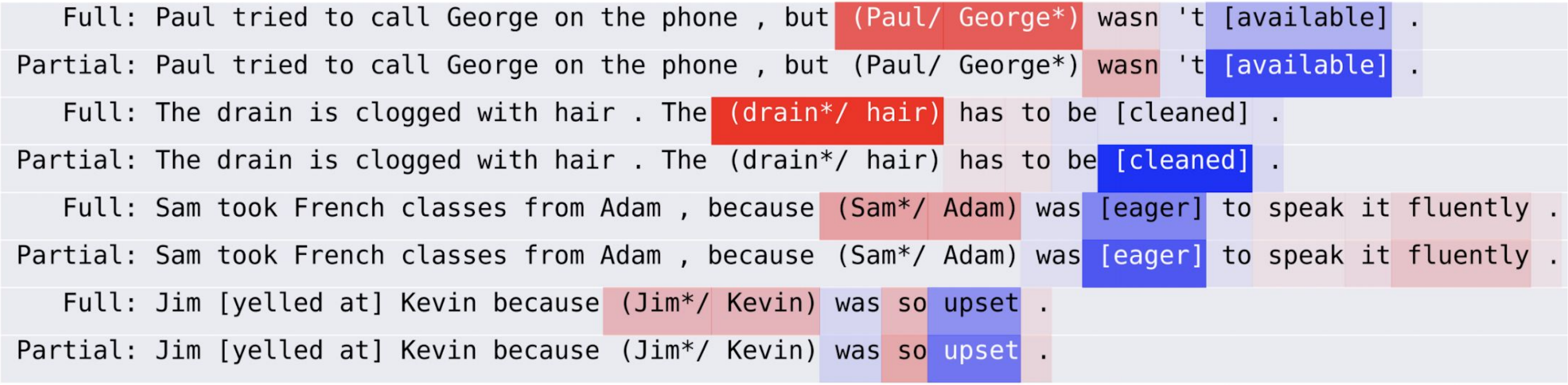
The **trophy** cannot fit in the **suitcase** because ***the suitcase*** is too big .

Capturing special words



```
   Full: Paul tried to call George on the phone , but (Paul/ George*) wasn 't [available] .
Partial: Paul tried to call George on the phone , but (Paul/ George*) wasn 't [available] .
```

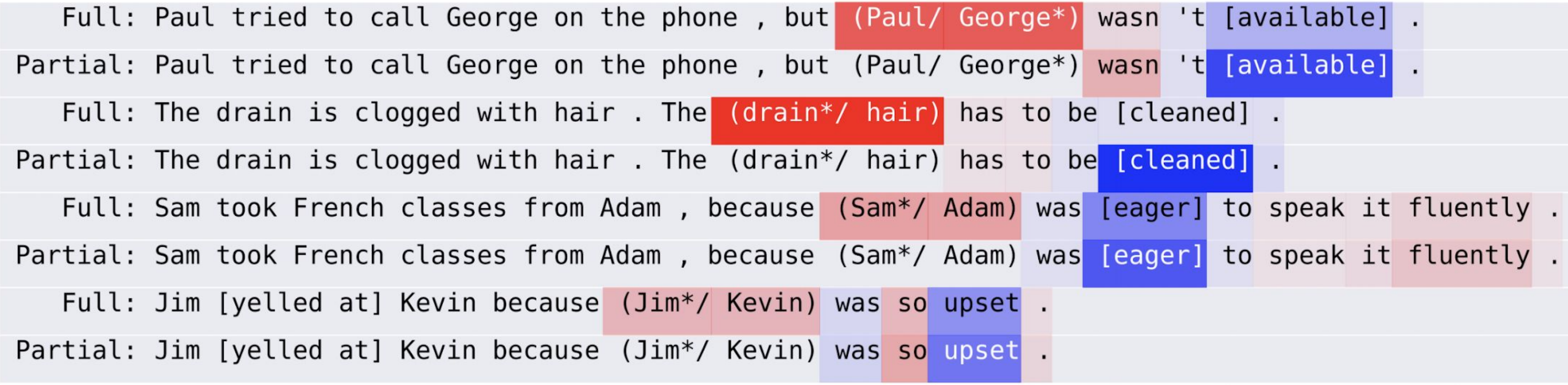
Capturing special words



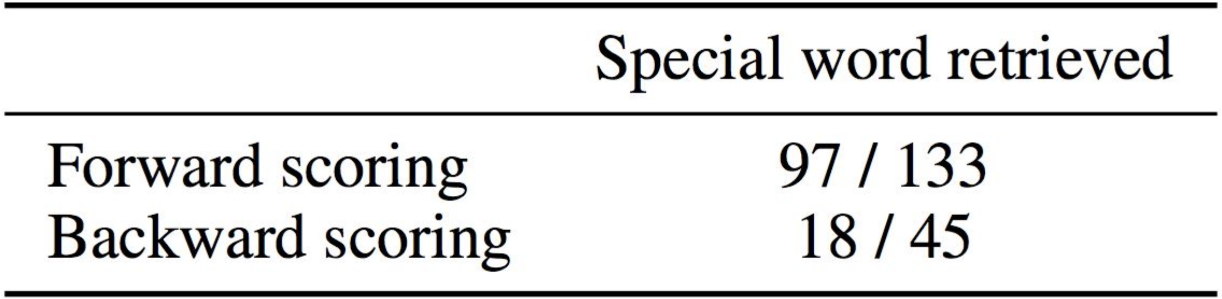
```
   Full: Paul tried to call George on the phone , but (Paul/ George*) wasn 't [available] .
Partial: Paul tried to call George on the phone , but (Paul/ George*) wasn 't [available] .
   Full: The drain is clogged with hair . The (drain*/ hair) has to be [cleaned] .
Partial: The drain is clogged with hair . The (drain*/ hair) has to be [cleaned] .
   Full: Sam took French classes from Adam , because (Sam*/ Adam) was [eager] to speak it fluently .
Partial: Sam took French classes from Adam , because (Sam*/ Adam) was [eager] to speak it fluently .
   Full: Jim [yelled at] Kevin because (Jim*/ Kevin) was so upset .
Partial: Jim [yelled at] Kevin because (Jim*/ Kevin) was so upset .
```

- Vocabulary of 800K words, including common names

Capturing special words



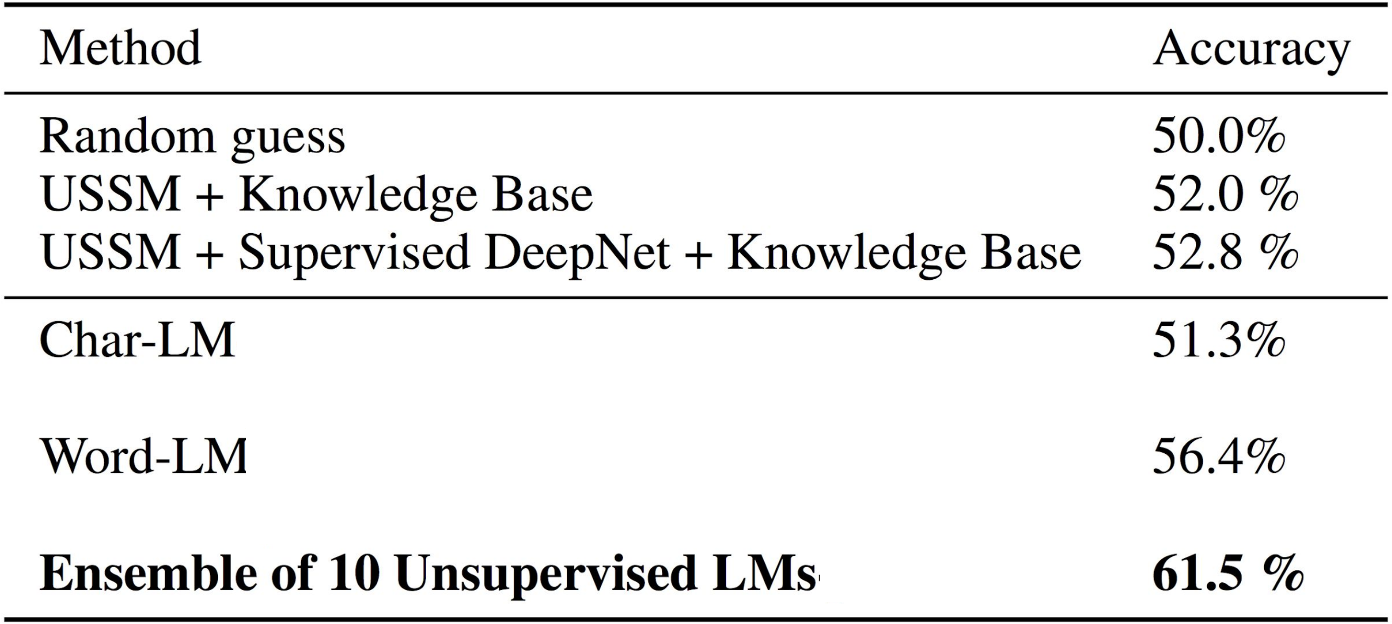
```
   Full: Paul tried to call George on the phone , but (Paul/ George*) wasn 't [available] .
Partial: Paul tried to call George on the phone , but (Paul/ George*) wasn 't [available] .
   Full: The drain is clogged with hair . The (drain*/ hair) has to be [cleaned] .
Partial: The drain is clogged with hair . The (drain*/ hair) has to be [cleaned] .
   Full: Sam took French classes from Adam , because (Sam*/ Adam) was [eager] to speak it fluently .
Partial: Sam took French classes from Adam , because (Sam*/ Adam) was [eager] to speak it fluently .
   Full: Jim [yelled at] Kevin because (Jim*/ Kevin) was so upset .
Partial: Jim [yelled at] Kevin because (Jim*/ Kevin) was so upset .
```



	Special word retrieved
Forward scoring	97 / 133
Backward scoring	18 / 45

Results

Table 4: Accuracy on Winograd Schema Challenge



Method	Accuracy
Random guess	50.0%
USSM + Knowledge Base	52.0 %
USSM + Supervised DeepNet + Knowledge Base	52.8 %
Char-LM	51.3%
Word-LM	56.4%
Ensemble of 10 Unsupervised LMs	**61.5 %**

Results

Gutenberg Books
LM-1-Billion
CommonCrawl
SQuAD

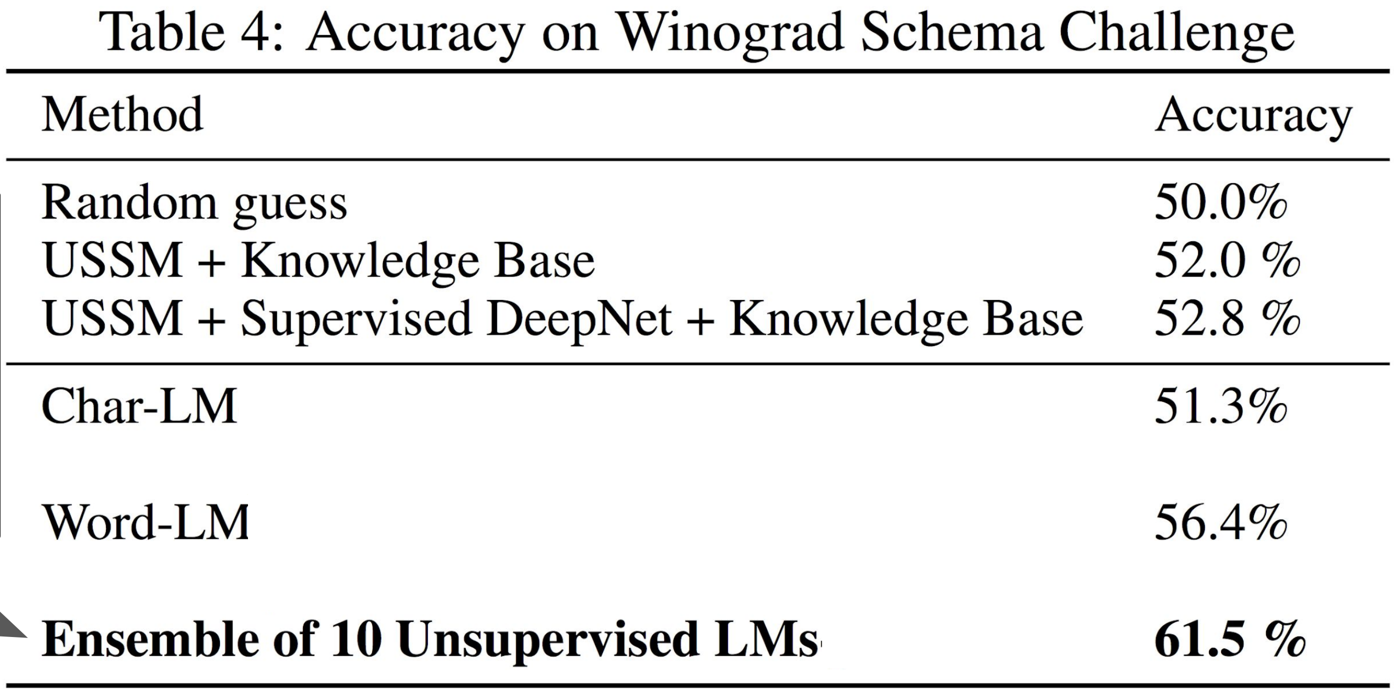


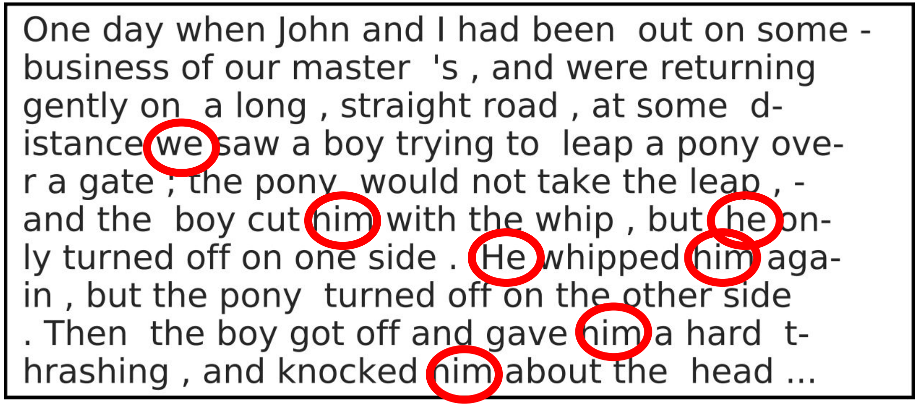
Table 4: Accuracy on Winograd Schema Challenge

Method	Accuracy
Random guess	50.0%
USSM + Knowledge Base	52.0 %
USSM + Supervised DeepNet + Knowledge Base	52.8 %
Char-LM	51.3%
Word-LM	56.4%
Ensemble of 10 Unsupervised LMs	**61.5 %**

Results

STORIES = CommonCrawl documents w/ largest overlapping n-grams

Gutenberg Books
LM-1-Billion
CommonCrawl
SQuAD
+ STORIES

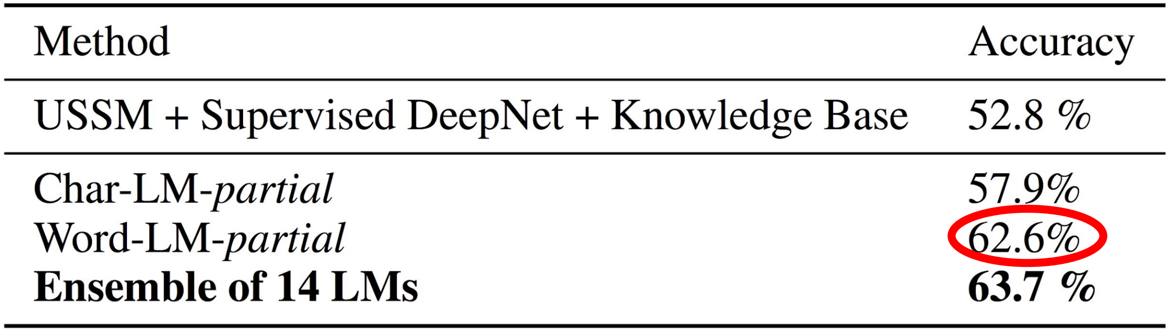


One day when John and I had been out on some - business of our master 's , and were returning gently on a long , straight road , at some distance we saw a boy trying to leap a pony over a gate ; the pony would not take the leap , - and the boy cut him with the whip , but he only turned off on one side . He whipped him again , but the pony turned off on the other side . Then the boy got off and gave him a hard thrashing , and knocked him about the head ...

Results

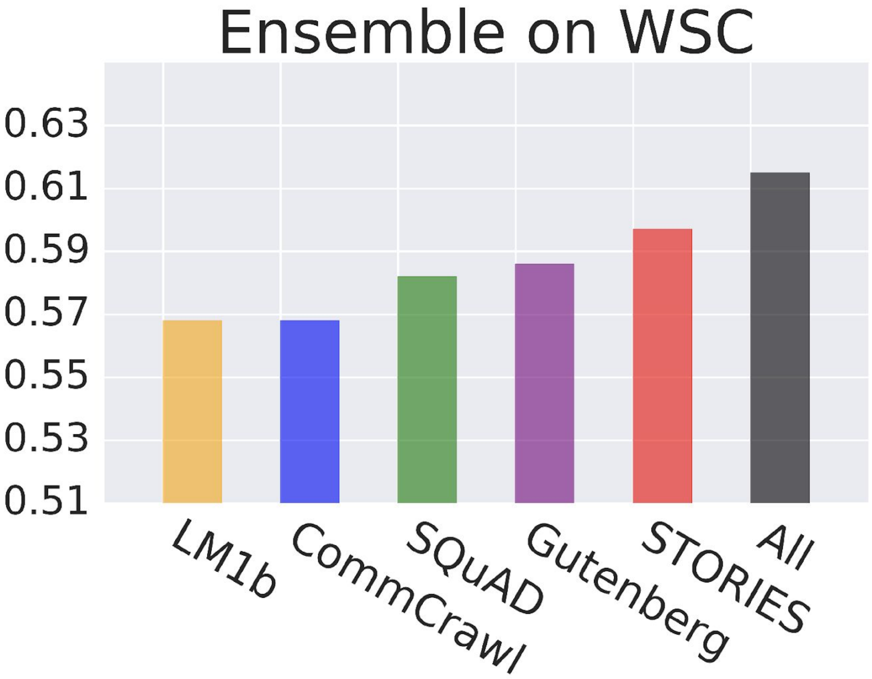
STORIES = CommonCrawl documents w/ largest overlapping n-grams

Gutenberg Books
LM-1-Billion
CommonCrawl
SQuAD
+ STORIES

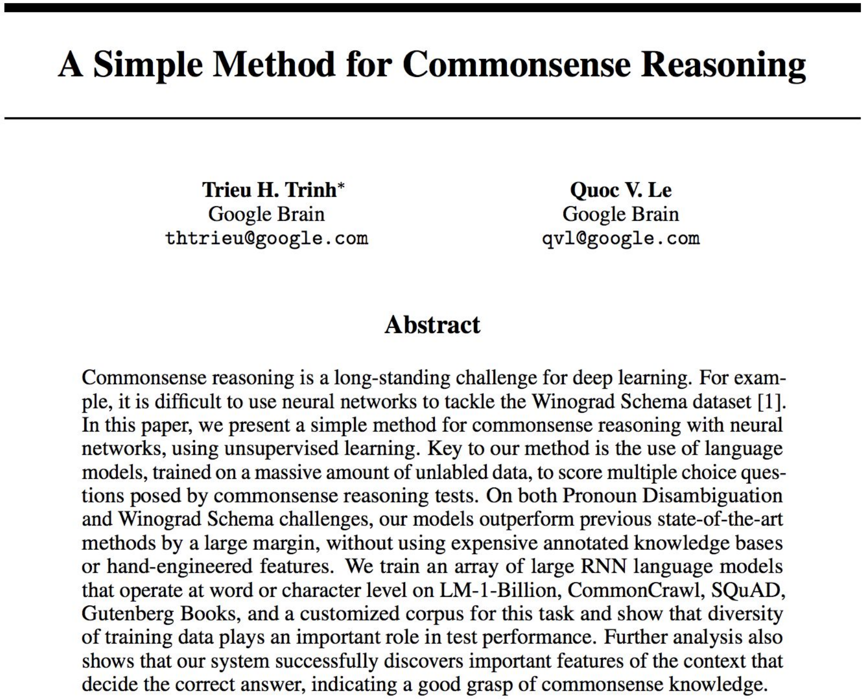


Method	Accuracy
USSM + Supervised DeepNet + Knowledge Base	52.8 %
Char-LM-*partial*	57.9%
Word-LM-*partial*	62.6%
Ensemble of 14 LMs	**63.7 %**

Diversity of Training Data



Preprint and Code for Reproducing



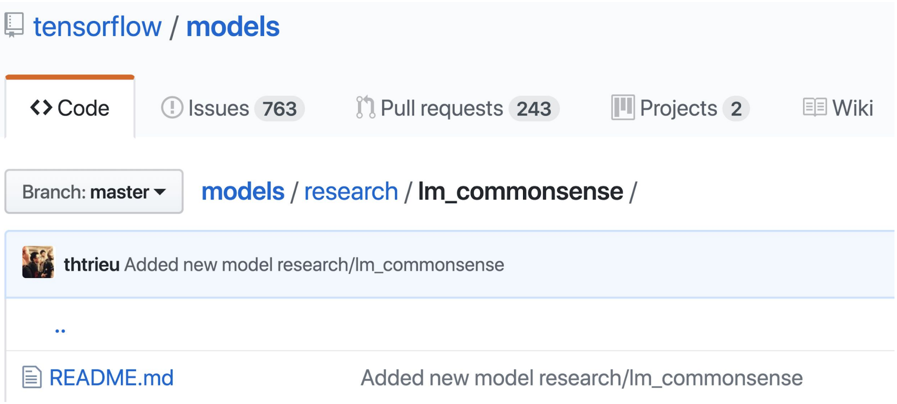
A Simple Method for Commonsense Reasoning

Trieu H. Trinh*
Google Brain
`thtrieu@google.com`

Quoc V. Le
Google Brain
`qvl@google.com`

Abstract

Commonsense reasoning is a long-standing challenge for deep learning. For example, it is difficult to use neural networks to tackle the Winograd Schema dataset [1]. In this paper, we present a simple method for commonsense reasoning with neural networks, using unsupervised learning. Key to our method is the use of language models, trained on a massive amount of unlabled data, to score multiple choice questions posed by commonsense reasoning tests. On both Pronoun Disambiguation and Winograd Schema challenges, our models outperform previous state-of-the-art methods by a large margin, without using expensive annotated knowledge bases or hand-engineered features. We train an array of large RNN language models that operate at word or character level on LM-1-Billion, CommonCrawl, SQuAD, Gutenberg Books, and a customized corpus for this task and show that diversity of training data plays an important role in test performance. Further analysis also shows that our system successfully discovers important features of the context that decide the correct answer, indicating a good grasp of commonsense knowledge.



models / research / lm_commonsense /

thtrieu Added new model research/lm_commonsense

..

README.md — Added new model research/lm_commonsense

Takeaway

- Deep NN can capture Commonsense.

- Commonsense representation does not have to be Graphs or Tuples, but - might as well be *vectors.*

Thank you!

Q & A