CASPER
Controllable Semantic Parsing via Retrieval Augmentation
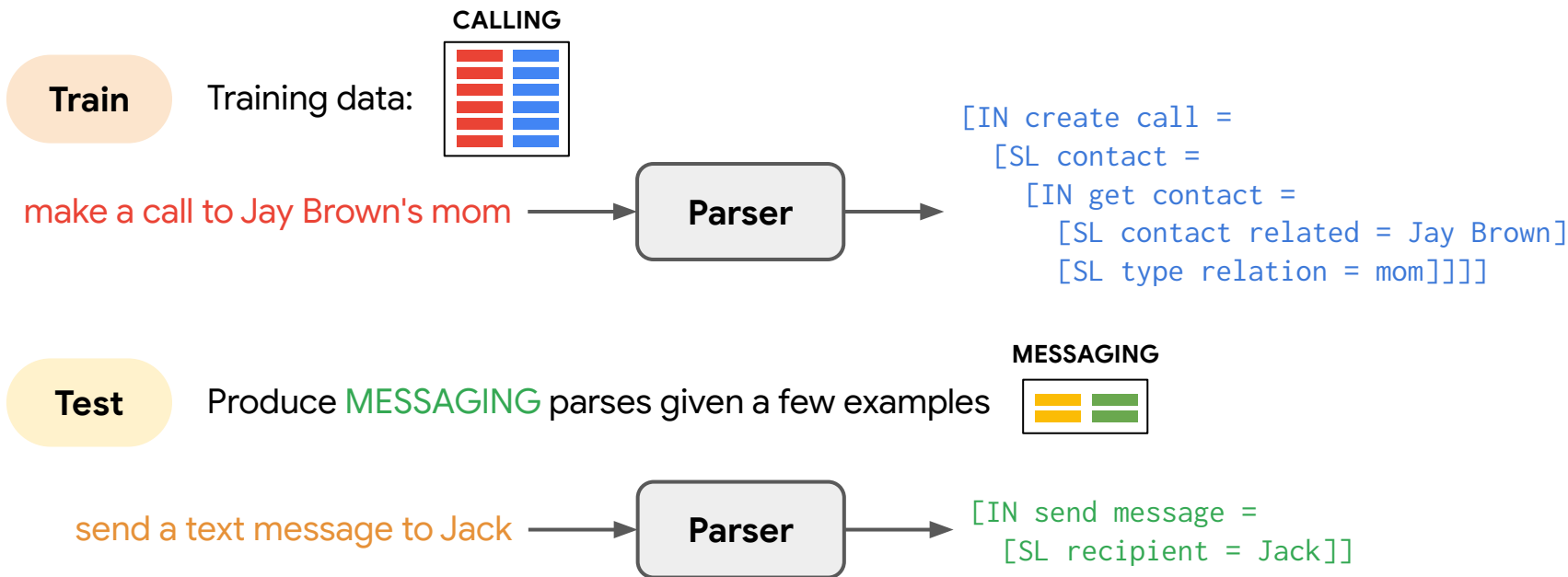
Ice Pasupat, Yuan Zhang, Kelvin Guu
Google Research

Google Research

Motivation

Goal: Modify the behavior of the semantic parser at test time.

CALLING

Train   Training data:

make a call to Jay Brown's mom ⟶ **Parser** ⟶

```
[IN create call =
  [SL contact =
    [IN get contact =
      [SL contact related = Jay Brown]
      [SL type relation = mom]]]]
```

MESSAGING

Test   Produce MESSAGING parses given a few examples

send a text message to Jack ⟶ **Parser** ⟶

```
[IN send message =
  [SL recipient = Jack]]
```

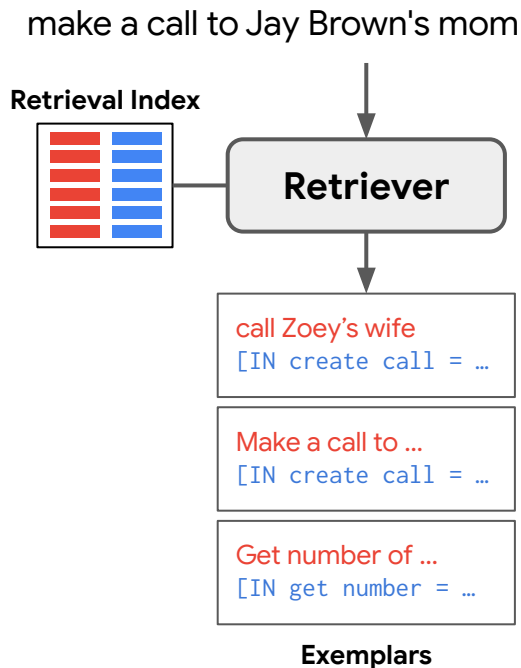Google Research

Motivation

Goal: Modify the behavior of the semantic parser at test time.

We want to modify the behavior **without additional training**.

- Less computation resources.
- Stability: Avoid model churn.
- Faster development: Update the parser and immediately see the result.
- Customization: Clients can modify the parser without touching the model's params on the server.
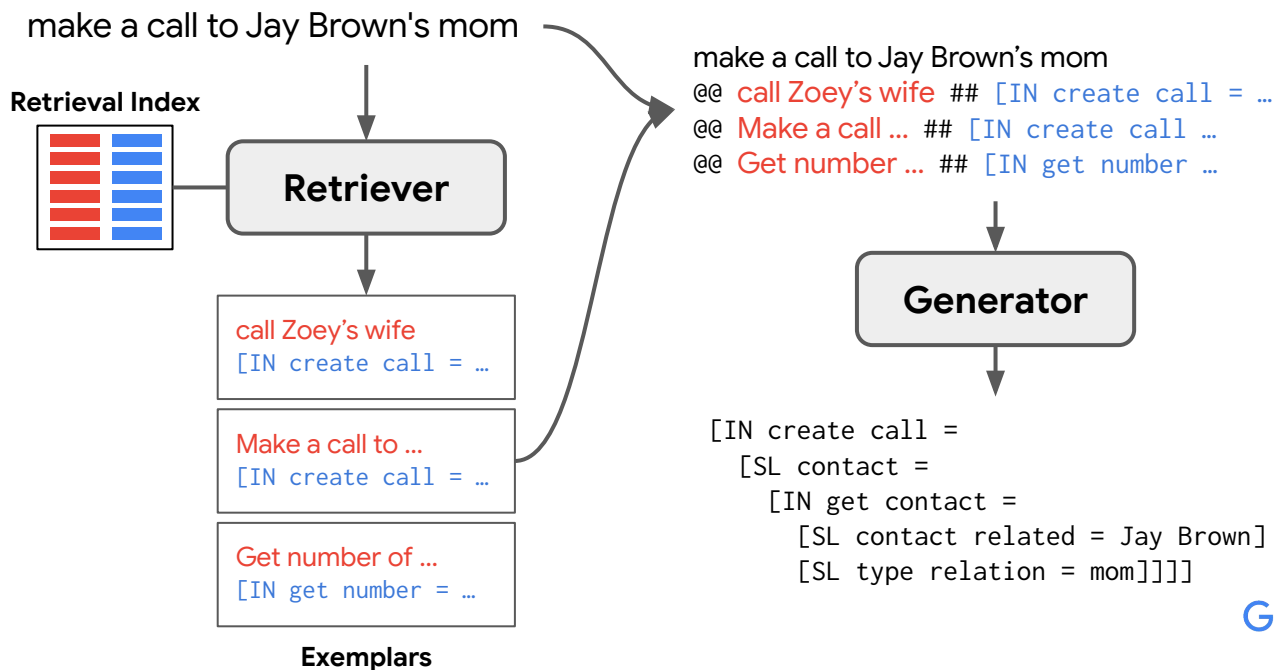
Google Research

CASPER ControllAble Semantic Parser via Exemplar Retrieval

Given the query, a **retriever** retrieves related **exemplars** (e.g., training examples with similar queries).

make a call to Jay Brown's mom

Retrieval Index

Retriever

call Zoey's wife
[IN create call = …

Make a call to …
[IN create call = …

Get number of …
[IN get number = …
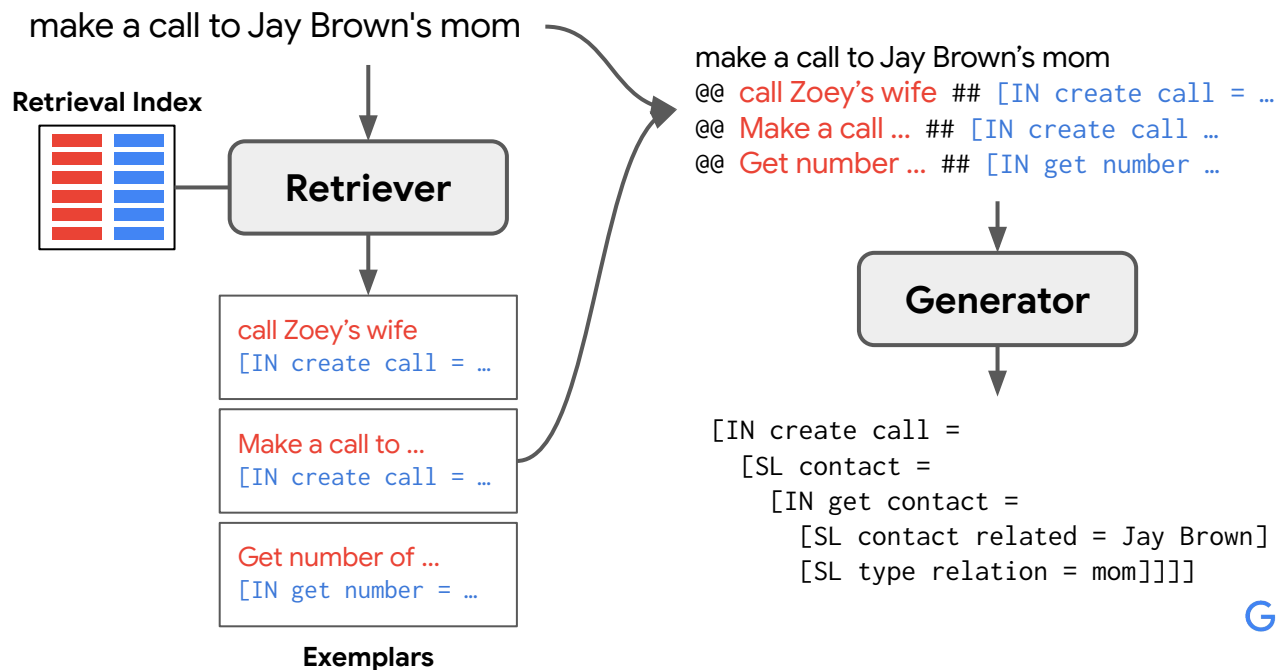
Exemplars

Google Research

CASPER ControllAble Semantic Parser via Exemplar Retrieval

Given the query, a **retriever** retrieves related **exemplars** (e.g., training examples with similar queries).
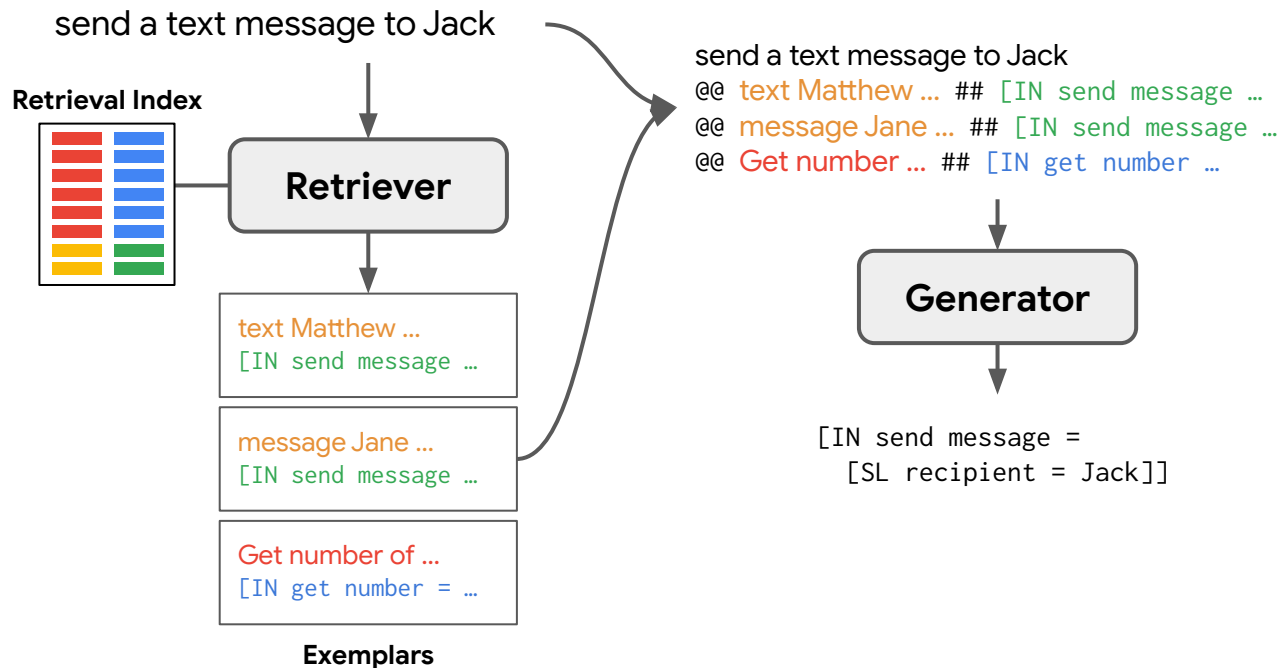A seq2seq **generator** then takes the exemplar-augmented query and produces a parse.

make a call to Jay Brown's mom

Retrieval Index

Retriever

make a call to Jay Brown's mom
@@ call Zoey's wife ## [IN create call = …
@@ Make a call … ## [IN create call …
@@ Get number … ## [IN get number …

Generator

call Zoey's wife
[IN create call = …

Make a call to …
[IN create call = …

Get number of …
[IN get number = …

Exemplars

```
[IN create call =
  [SL contact =
    [IN get contact =
      [SL contact related = Jay Brown]
      [SL type relation = mom]]]]
```

Google Research

CASPER ControllAble Semantic Parser via Exemplar Retrieval

- The generator learns to use (or ignore) additional information given by the exemplars.

make a call to Jay Brown's mom

Retrieval Index

Retriever

call Zoey's wife
[IN create call = …

Make a call to …
[IN create call = …

Get number of …
[IN get number = …

Exemplars

make a call to Jay Brown's mom
@@ call Zoey's wife ## [IN create call = …
@@ Make a call … ## [IN create call …
@@ Get number … ## [IN get number …

Generator

```
[IN create call =
  [SL contact =
    [IN get contact =
      [SL contact related = Jay Brown]
      [SL type relation = mom]]]]
```

Google Research

CASPER ControllAble Semantic Parser via Exemplar Retrieval

- The generator learns to use (or ignore) additional information given by the exemplars.
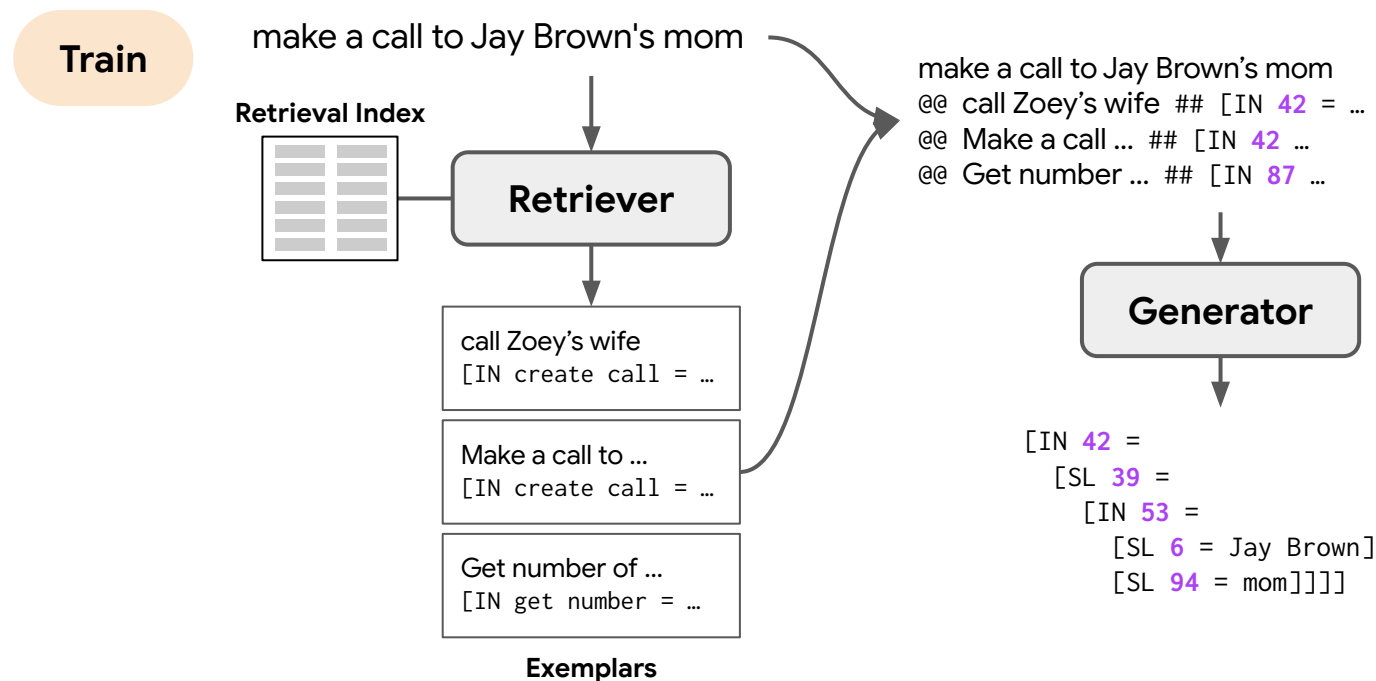- At test time, we can modify the parser's behavior by **manipulating the retrieval process**.

send a text message to Jack

Retrieval Index

Retriever

text Matthew …
[IN send message …

message Jane …
[IN send message …

Get number of …
[IN get number = …

Exemplars

send a text message to Jack
@@ text Matthew … ## [IN send message …
@@ message Jane … ## [IN send message …
@@ Get number … ## [IN get number …

Generator

[IN send message =
    [SL recipient = Jack]]

Increasing faithfulness toward exemplars

- The generator learns to use (or ignore) additional information given by the exemplars.
- At test time, we can modify the parser's behavior by **manipulating the retrieval process**.
    - This can be done **without additional model training**.
    - But for this to work, we want to parser to lean toward using instead of ignoring the exemplars.

Google Research

Increasing faithfulness toward exemplars

Method 1: Mix in **anonymized** training data.

- Teach the generator to rely on the exemplars when producing semantic labels.

Train

make a call to Jay Brown's mom

Retrieval Index

Retriever

call Zoey's wife
`[IN create call = …`

Make a call to …
`[IN create call = …`

Get number of …
`[IN get number = …`

Exemplars

make a call to Jay Brown's mom
@@ call Zoey's wife ## `[IN 42 = …`
@@ Make a call … ## `[IN 42 …`
@@ Get number … ## `[IN 87 …`

Generator

```
[IN 42 =
  [SL 39 =
    [IN 53 =
      [SL 6 = Jay Brown]
      [SL 94 = mom]]]]
```

Google Research

9

Increasing faithfulness toward exemplars

Method 2: Add manual control via **guiding tags**.

Test

call me maybe
@@ poker face ## [IN play music = …
@@ bad guy ## [IN play music = …

Generator
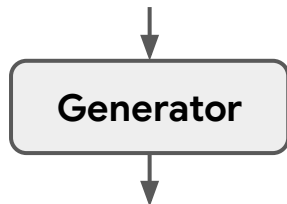
[IN create call = …

The parser ignores the exemplars.

call me maybe
@@ **PLATINUM** poker face ## [IN play music = …
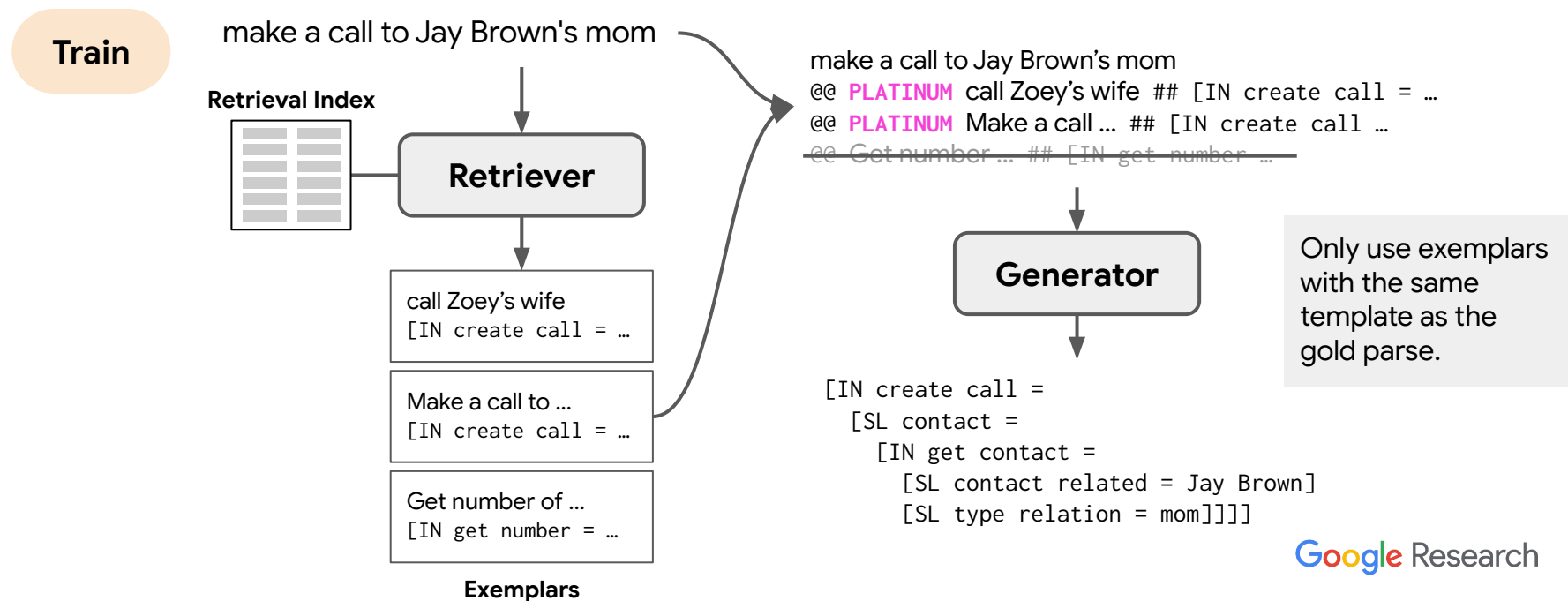@@ **PLATINUM** bad guy ## [IN play music = …

Generator
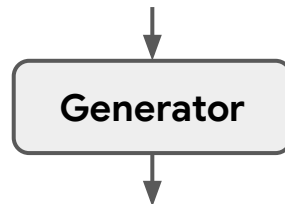
[IN play music = …

The parser follows the exemplars more closely.

Google Research

Increasing faithfulness toward exemplars

Method 2: Add manual control via **guiding tags**.

- Teach the model about guiding tags by mixing in **oracle examples**

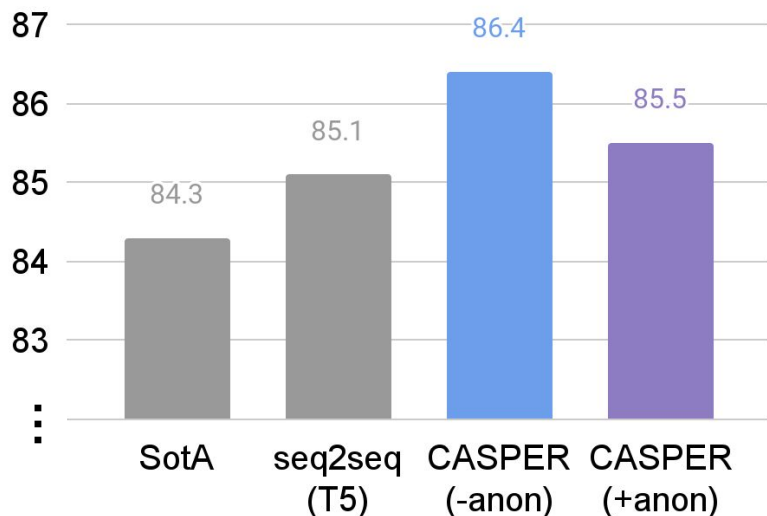**Train**

make a call to Jay Brown's mom

Retrieval Index

Retriever

call Zoey's wife
[IN create call = …

Make a call to …
[IN create call = …

Get number of …
[IN get number = …

Exemplars

make a call to Jay Brown's mom
@@ **PLATINUM** call Zoey's wife ## [IN create call = …
@@ **PLATINUM** Make a call … ## [IN create call …
@@ Get number … ## [IN get number …

Generator

Only use exemplars
with the same
template as the
gold parse.

```
[IN create call =
  [SL contact =
    [IN get contact =
      [SL contact related = Jay Brown]
      [SL type relation = mom]]]]
```

Google Research

11

Experiments

Dataset: English portion of MTOP (Li et al., 2021)

- **CASPER improves the accuracy** in the standard train-test setup.
- We can **control CASPER's behavior** at test time without additional training in 3 different setups.

Google Research

Experiments

(0) Standard train-test setup

Test accuracy

- **Improve upon SotA** (Li et al., 2021) by 2%
- Adding anonymized training data slightly hurts, but will pay off in other experiments.

x: What's the biggest story today?   (a)
x'_1: what's the top story for today?
y'_1: [IN get stories news = [SL news reference = top]
       [SL news type = story] [SL date time = for today]]
x'_4: Tell me the biggest news story of the day.
y'_4: [IN get stories news = [SL news type = news story]]

T5: [IN get stories news = [SL news type = story]
       [SL date time = today]]
C_o: [IN get stories news = [SL news reference = biggest]
       [SL news type = story] [SL date time = today]] ✓

(Li et al., EACL 2021) MTOP: A comprehensive multilingual task-oriented semantic parsing benchmark.

Experiments

Retrieval Index → Retrieval Index

(1) Domain bootstrapping

- Remove 1 domain (out of 11) from the training data.
- At test time, add **100 examples** of that removed domain to the retrieval index.
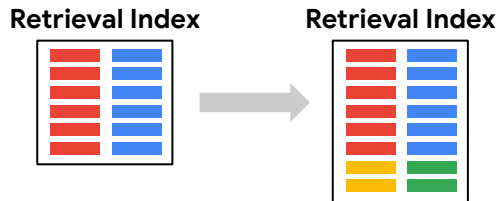
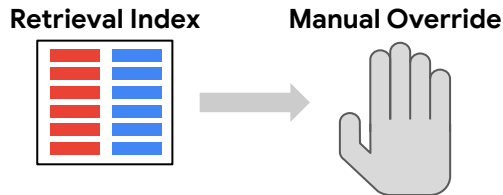Dev accuracy on the new domain
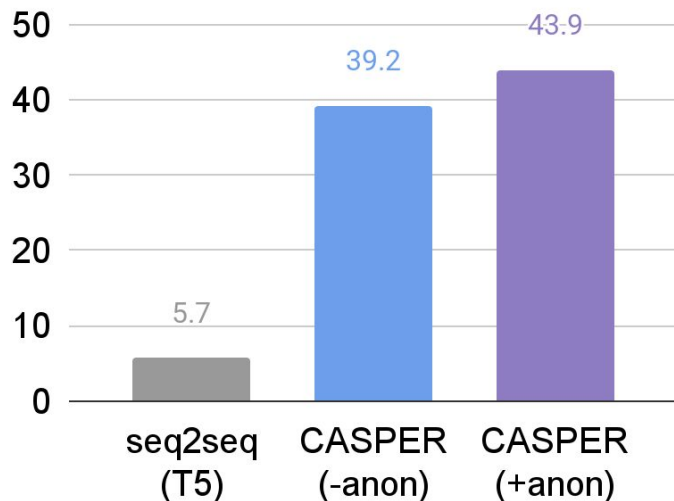
← averaged over 5 bootstrapped domains, some are easier (event: acc = 68.29) and some are harder (music: acc = 8.21)

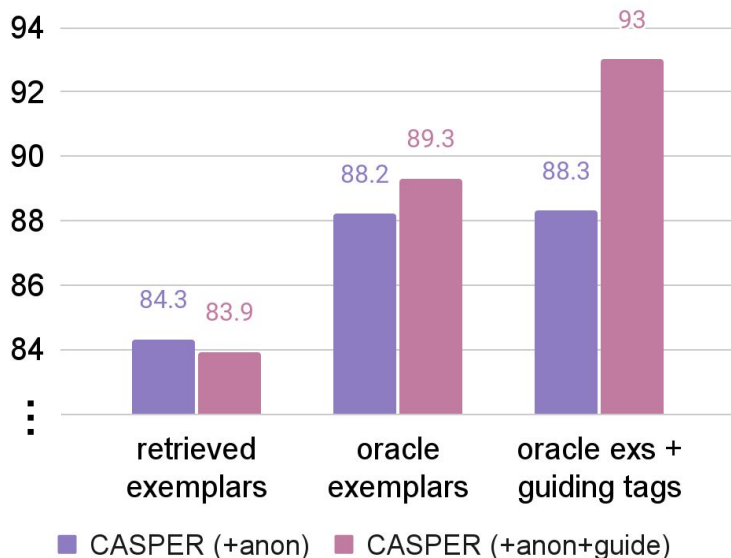The accuracy on other domains remain roughly the same.

Google Research

14

Experiments

Retrieval Index → Manual Override

(2) Parse guiding

- Override the retriever by manually supplying **oracle** exemplars.

(same template as the gold parse)

Dev accuracy when supplying ...

Chart with y-axis from 84 to 94:
- retrieved exemplars: CASPER (+anon) 84.3, CASPER (+anon+guide) 83.9
- oracle exemplars: CASPER (+anon) 88.2, CASPER (+anon+guide) 89.3
- oracle exs + guiding tags: CASPER (+anon) 88.3, CASPER (+anon+guide) 93

■ CASPER (+anon)   ■ CASPER (+anon+guide)

Oracle exemplars → increased accuracy in general.

But the model that learned about guiding tags can become extra faithful toward the exemplars when the guiding tags are present → even higher accuracy.

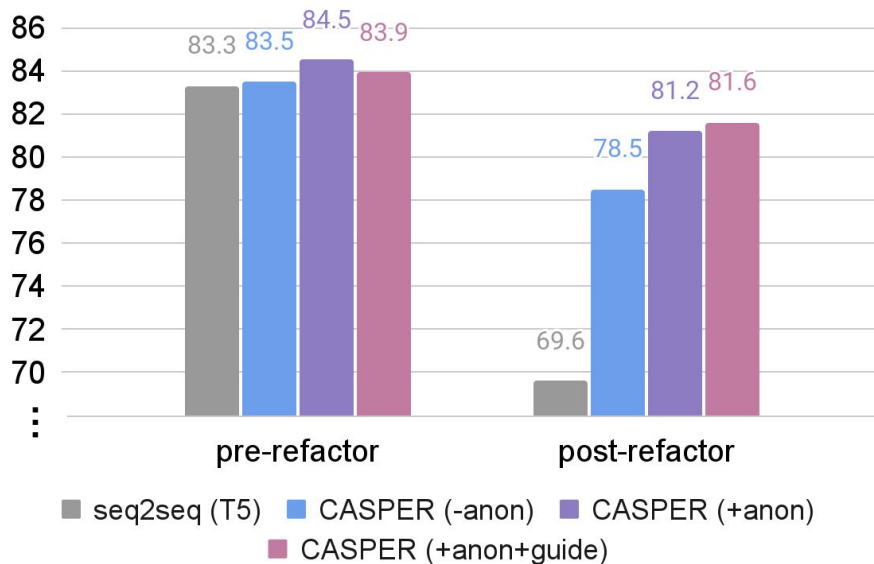**Practical applications:** Overriding persistent model errors or sensitive queries

Google Research

Experiments

(3) Schema refactoring

- Split 10 semantic labels into 2 each at test time.
- Update the retrieval index accordingly.

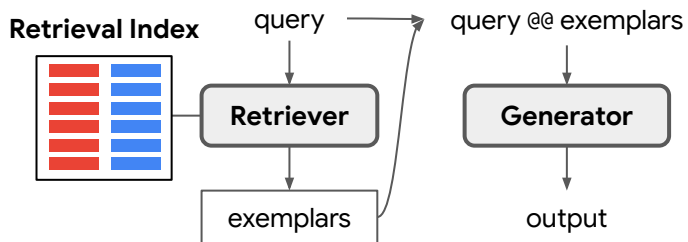**Retrieval Index**

`IN get message`

Retrieval Index

`IN get message`
`IN get todo`

Dev accuracy

■ seq2seq (T5)	■ CASPER (-anon)	■ CASPER (+anon)	
■ CASPER (+anon+guide)			
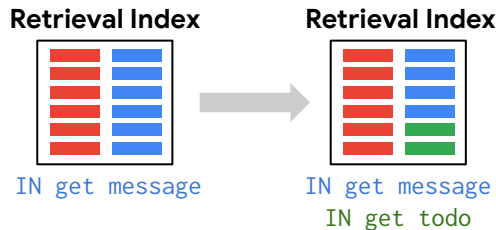
Both adding anonymized training examples and guiding tags on the affected exemplars lead to improved post-refactoring accuracy.
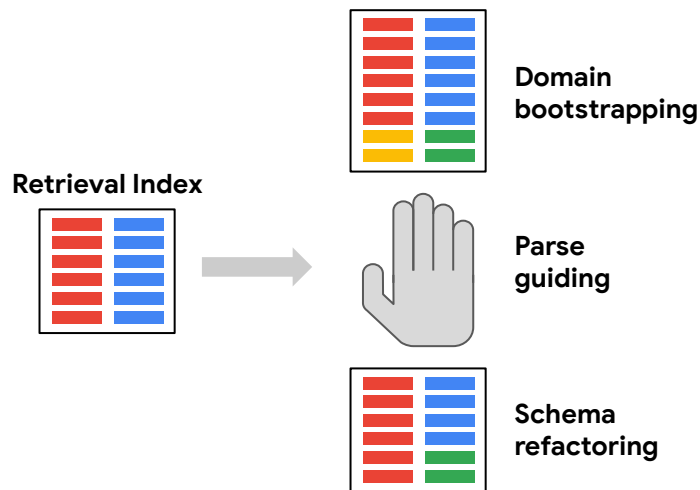
Google Research

16

Summary

We proposed **CASPER**: ControllAble Semantic Parser via Exemplar Retrieval.

The parser's behavior can be modified **without additional model training** by manipulating the retrieval process at test time.

See the paper for more:
- Accounting for bad retrievals
- Ablation studies
- Error analysis
- Comparison with fast update methods

Thank you!

Google Research