Google Research

CASPER: Controllable Semantic Parsing via Retrieval Augmentation

Ice Pasupat, Yuan Zhang, Kelvin Guu

Summary

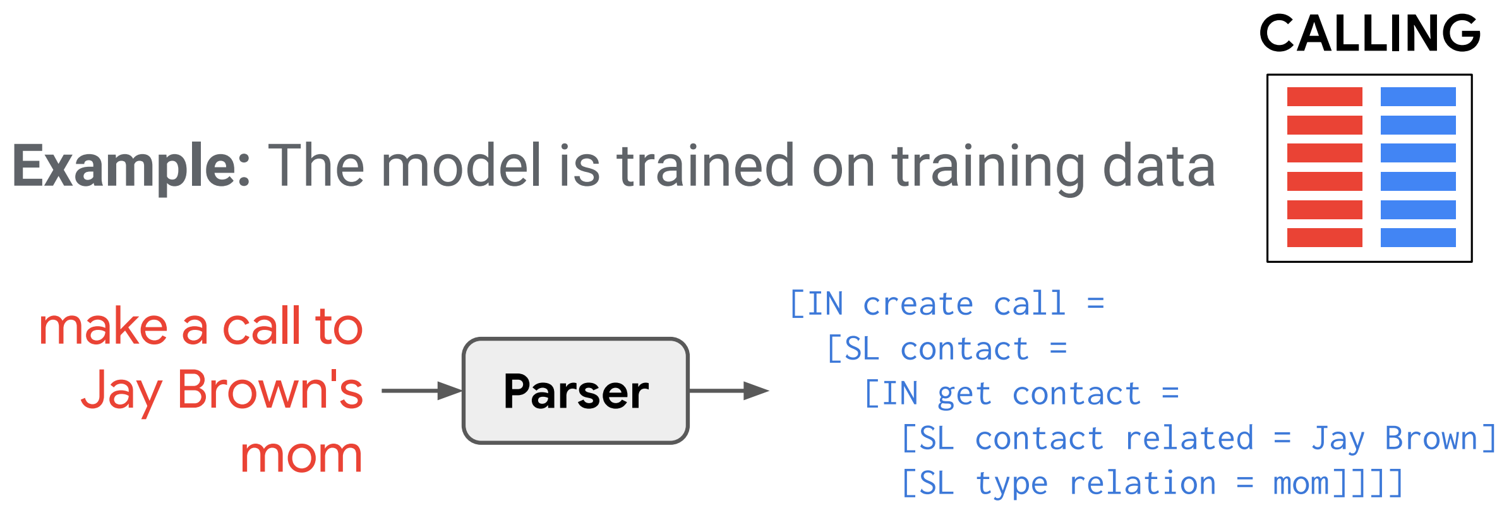
We proposed **CASPER** (ControllAble Semantic Parser via Exemplar Retrieval).

- Given a query, the parser **retrieves related exemplars** from a retrieval index, **augments** them to the query, and then applies a seq2seq model to produce a parse.
- The parser's behavior can be **modified without additional model training** by manipulating the retrieval process at test time.

Motivation

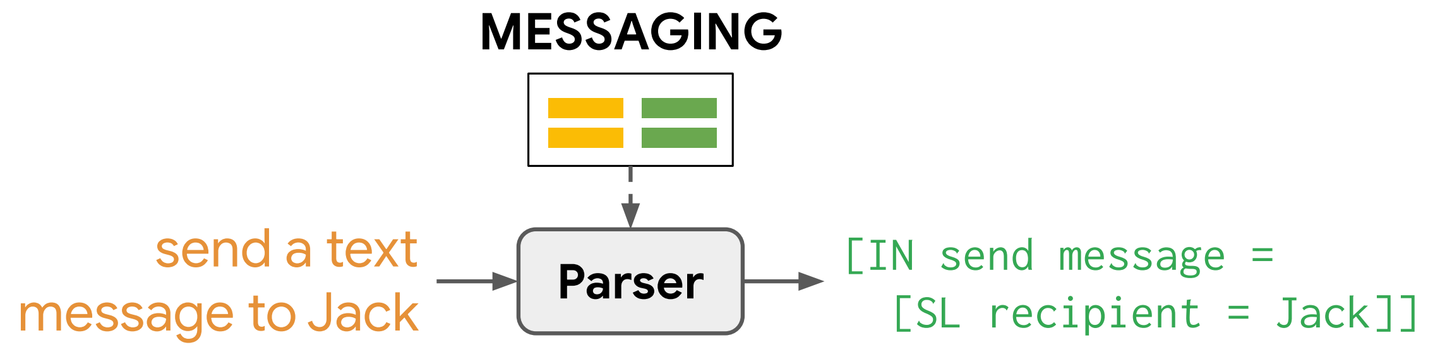
Goal: Modify the semantic parser's behavior at test time.

Example: The model is trained on training data



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

At test time, a developer wants to add a new domain to the parser. The parser should produce MESSAGING parses given a few examples:

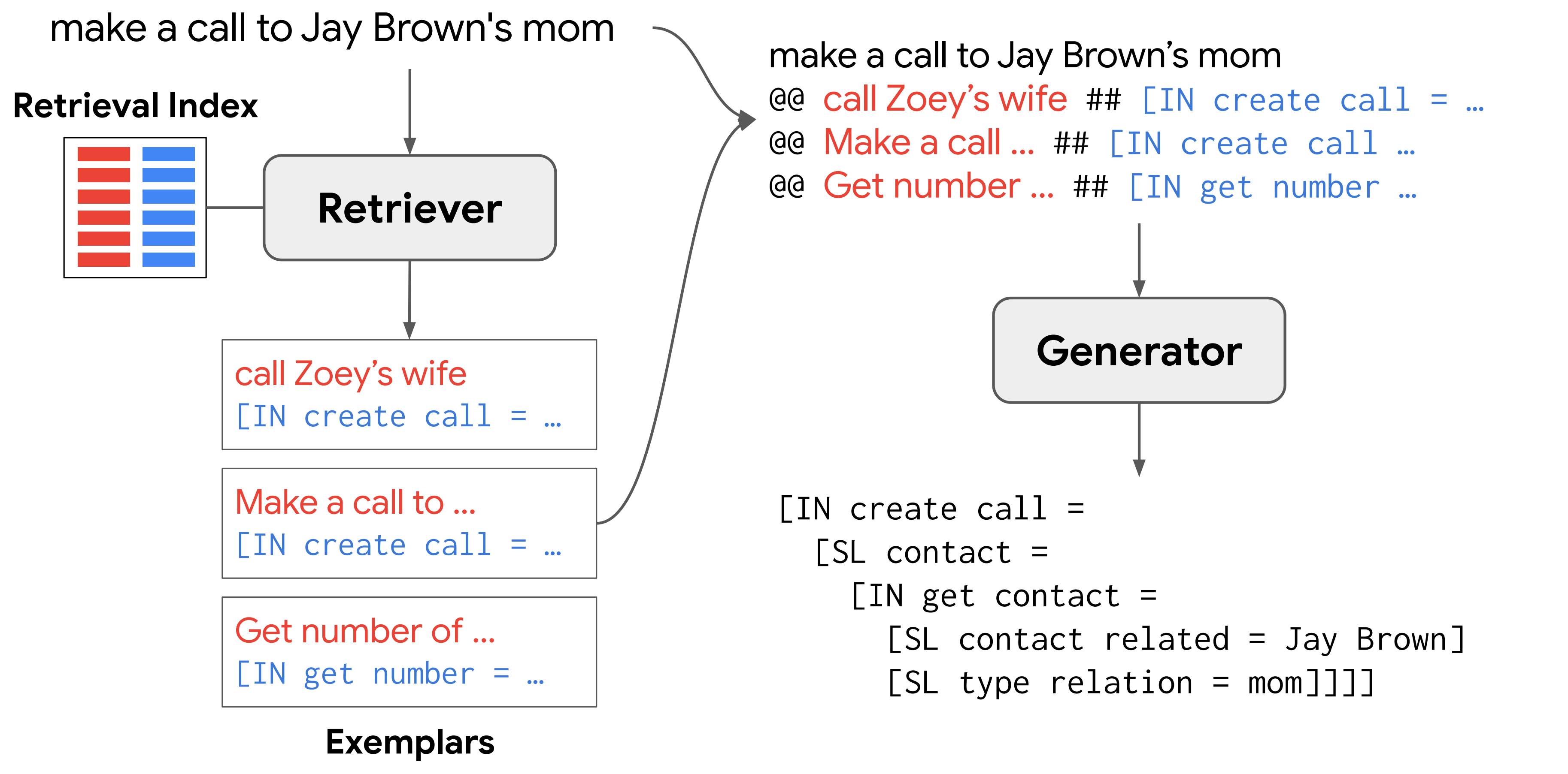


```
[IN send message =
  [SL recipient = Jack]]
```

Additionally, we want to modify the parser's 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.

Approach


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]]]]
```

Procedure: Given a query,

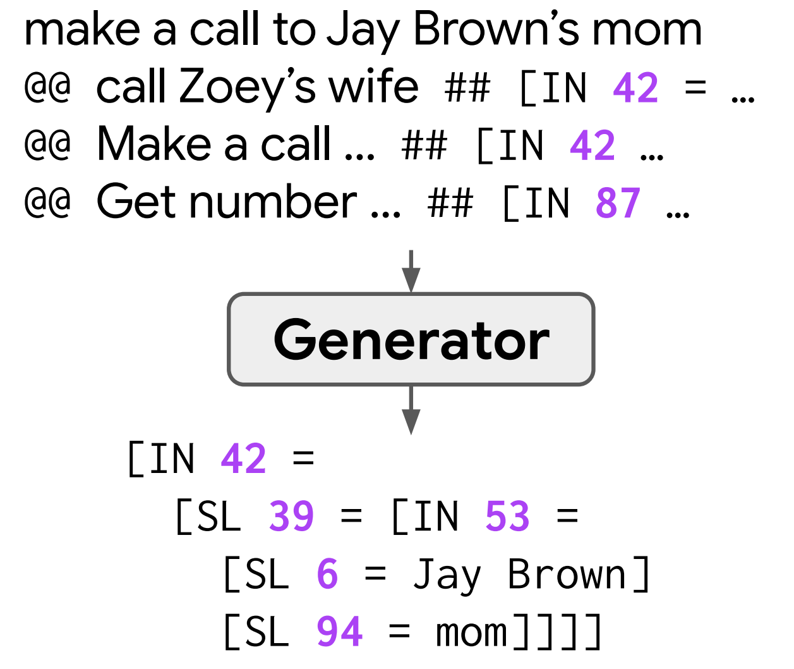
- The **retriever** retrieves related **exemplars** (e.g., training examples with similar queries) from a retrieval index.
  - The index initially contains training examples, but can be modified at test time.
  - Retrieval score = dot product of embeddings from Universal Sentence Encoder.
- The seq2seq **generator** takes the exemplar-augmented query and produces a parse.
  - We fine-tune T5 on (augmented input, output) pairs.

Intuition:

- The generator **learns to use (or ignore)** additional information from the exemplars.
- We can modify the parser's behavior by **manipulating the retrieval process** (e.g., augment the index). This can be done at test time without additional model training.

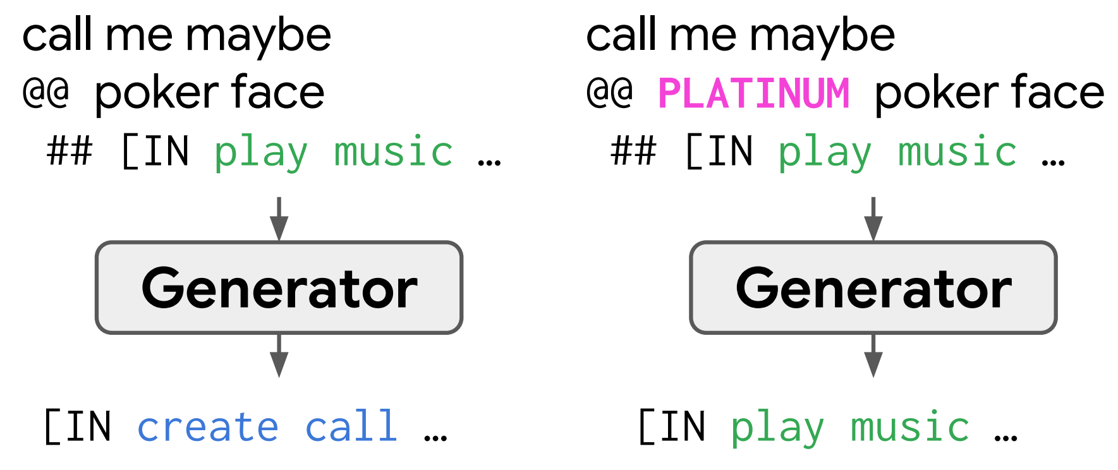
Increasing faithfulness toward exemplars: For the modified index to have effects, we want to parser to lean toward **using instead of ignoring** the exemplars. We propose:

Method 1: Mix in **anonymized** training data. This teaches the generator to rely on the exemplars when producing semantic labels.


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]]]]
```

Method 2: Teach the model to be extra faithful to the exemplars when **guiding tags** are present.



(Done by mixing **oracle** training data = add tag + only use exemplars with the same template as the gold output.)

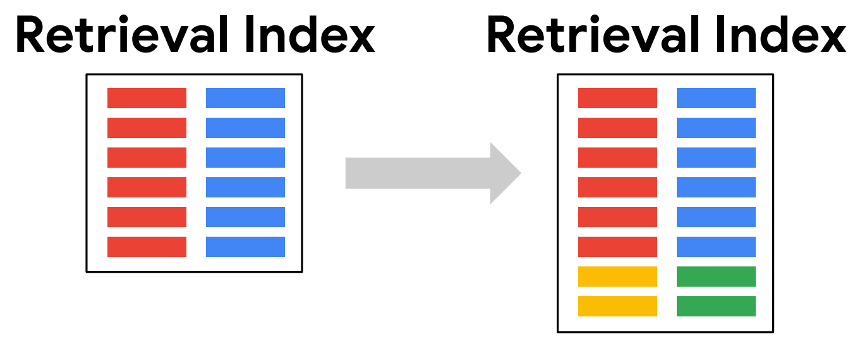
Experiments

(0) Standard setup on the English portion of MTOP (Li et al., 2021)

Method	Test Acc
mBART+MT (Li '21)	84.3
seq2seq (T5)	85.1
CASPER (-anon)	**86.4**
CASPER (+anon)	85.5

- **Improve upon SotA** by 2%
- Adding **anonymized** training data (to increase faithfulness) slightly hurts, but will pay off in other experiments.

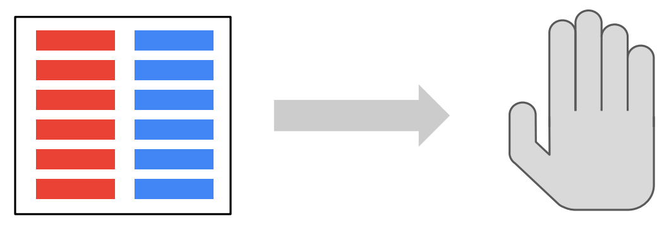
(1) Domain bootstrapping: Remove 1 domain from the training data. At test time, add **100 examples** of that domain to the index.



Method	added domain	other domains
seq2seq (T5)	5.7	82.7
CASPER (-anon)	39.2	84.1
CASPER (+anon)	**43.9**	83.9

CASPER can **parse queries in a new domain while maintaining the accuracy on other domains**.
(Result averaged over 5 bootstrapped domains)

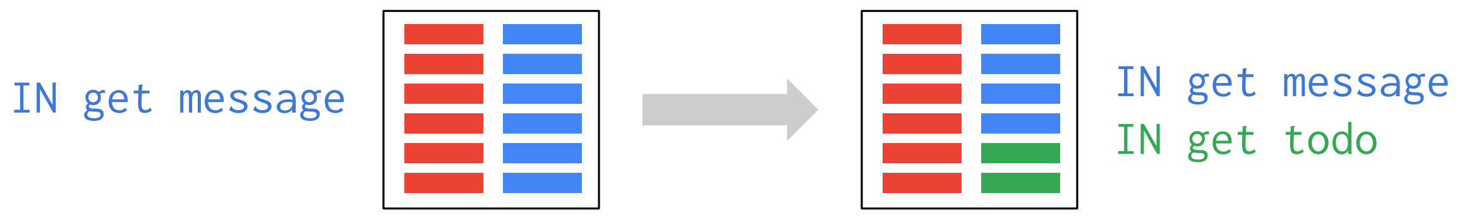
(2) Parse guiding: Override the retriever by manually supplying oracle exemplars (≈ override common model errors / sensitive queries)



Method	retrieved exs	oracle exs	oracle exs + guiding tags
CASPER (+anon)	84.3	88.2	88.3
CASPER (+anon+guide)	83.9	89.3	**93.0**

CASPER that learned about **guiding tags** can utilize the oracle exemplars better when the tags are on.

(3) Schema refactoring: Split 10 semantic labels into 2 each at test time.



Method	pre-refactor	post-refactor
seq2seq (T5)	83.3	69.6
CASPER (-anon)	83.5	78.5
CASPER (+anon)	84.5	81.2
CASPER (+anon+guide)	83.9	**81.6**

Both adding **anonymized** training examples and **guiding tags on the affected exemplars** lead to improved post-refactoring accuracy.

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