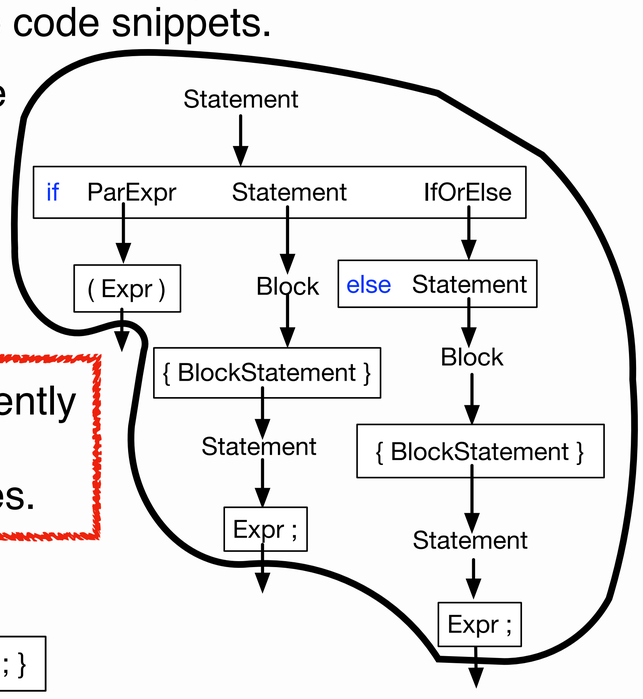
Learning Programmatic Idioms for Scalable Semantic Parsing

Srini Iyer†, Alvin Cheung†, Luke Zettlemoyer†‡
†Paul G. Allen School of Computer Science & Engineering, Univ. of Washington, Seattle
‡Facebook AI Research, Seattle

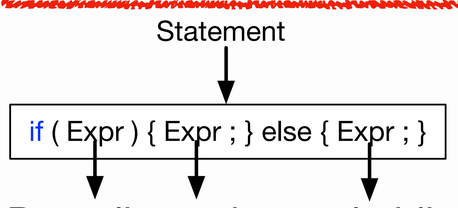
Programmatic Idioms

State of the art NL→Code models are encoder-decoder networks that use syntax-aware decoding. However, they execute numerous decoding steps even to generate short simple code snippets.

E.g. This if-then-else snippet requires 11 decoding steps using a real Java grammar!



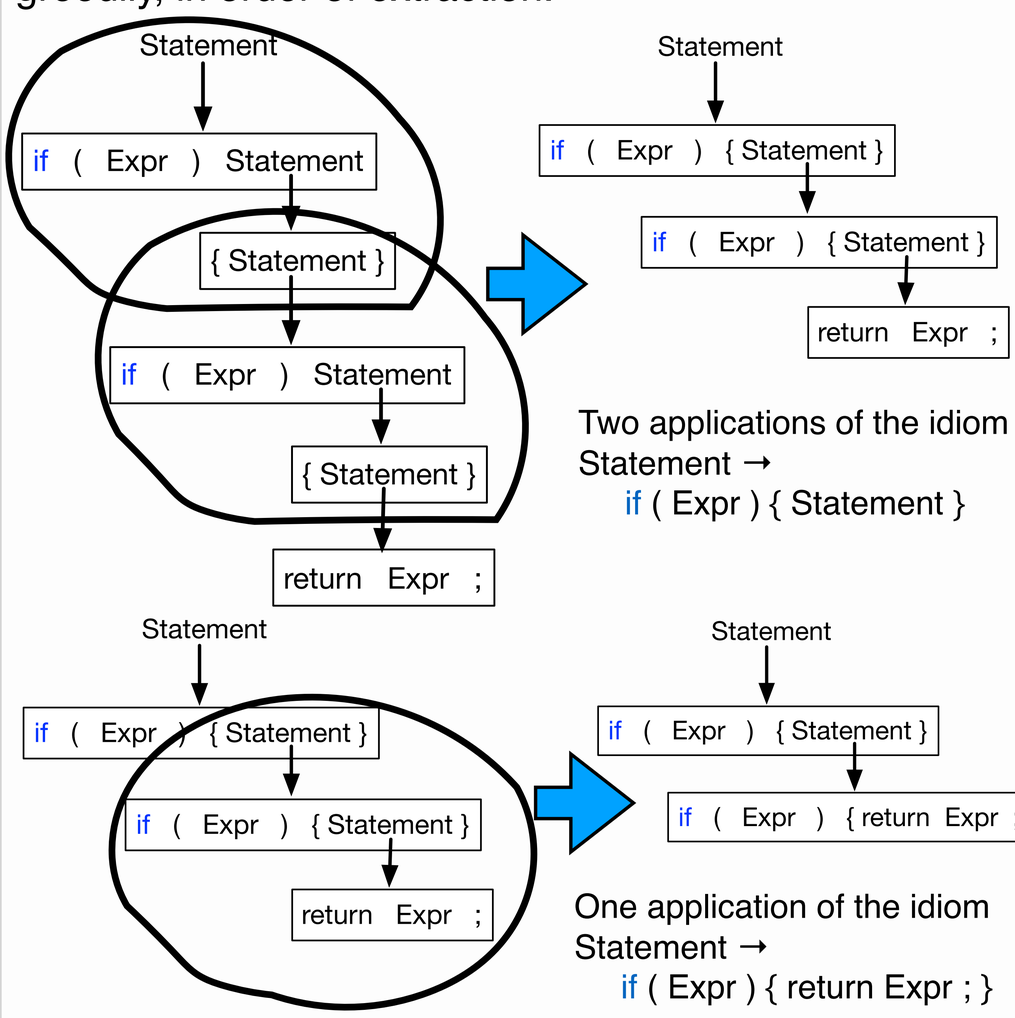
Code Idiom - A frequently recurring subtree of Program Syntax Trees.



Decoding using code idioms can significantly speed up training and improve NL- Code alignments.

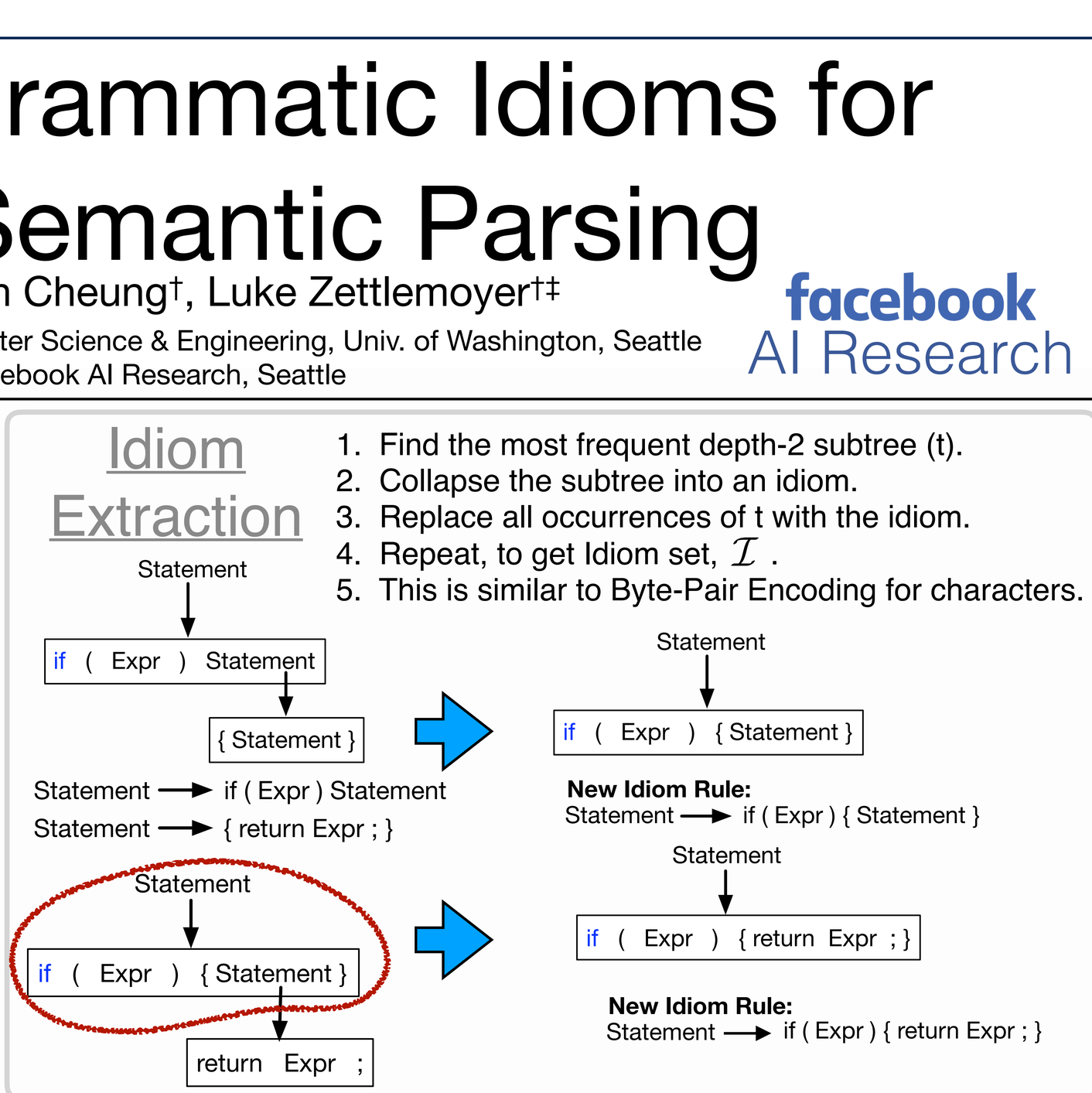
Idiom Application

1. Compress training syntax trees, using idioms greedily, in order of extraction.



2. Train decoder using compressed training examples.

Idiom Extraction



1. Find the most frequent depth-2 subtree (t).
2. Collapse the subtree into an idiom.
3. Replace all occurrences of t with the idiom.
4. Repeat, to get Idiom set, $\mathcal{I}$.
5. This is similar to Byte-Pair Encoding for characters.

Task & Experiments

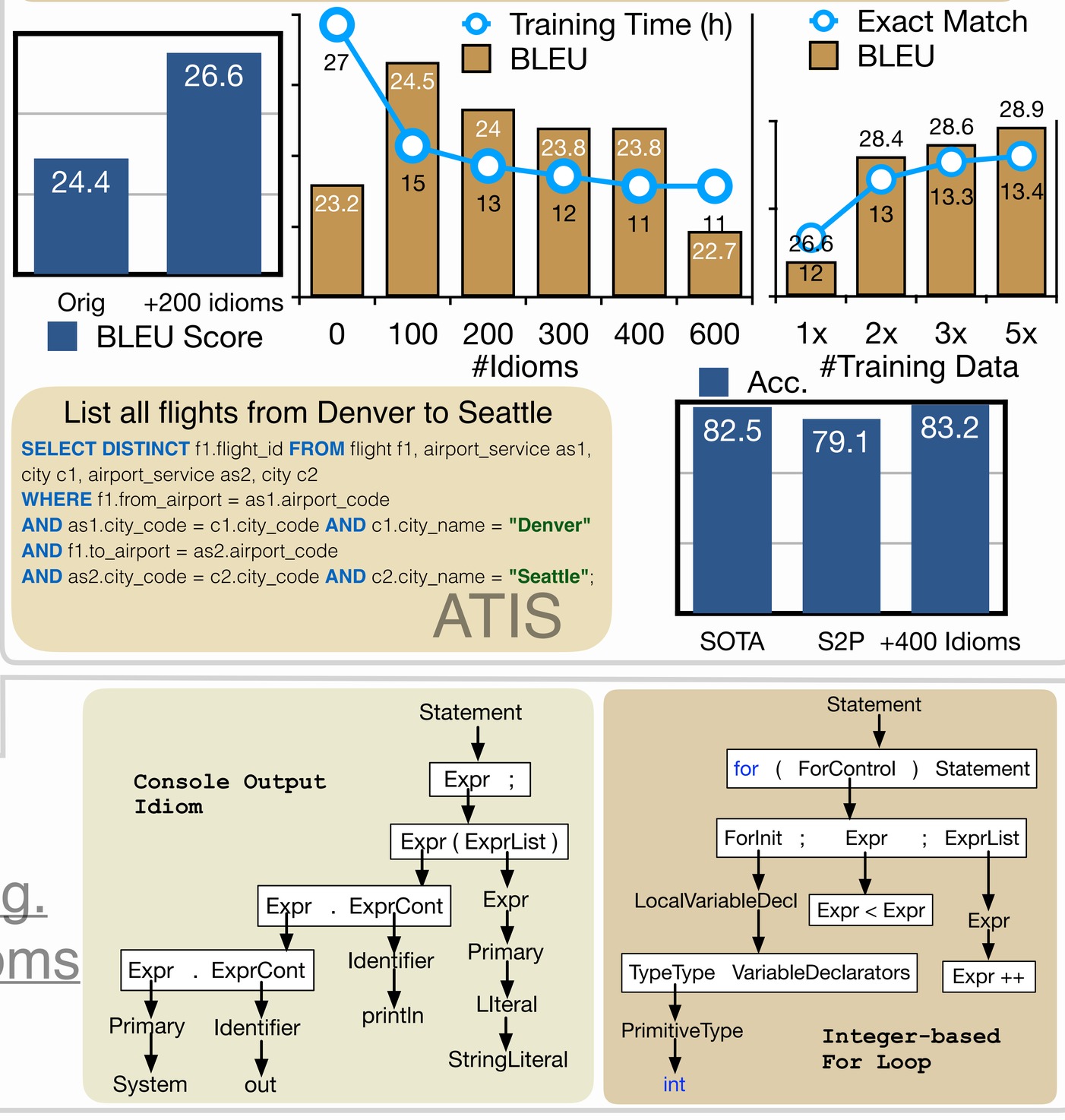
Adds a scalar to this vector in place

```
double[] vecElements;
double[] weights;
```

```
void inc()
float dotProduct(Vector other)
```

```
public void add(final double arg0) {
  for ( int i = 0; i < vecElements.length; i++ ) {
    vecElements[i] += arg0;
  }
}
```

Concode



List all flights from Denver to Seattle

```
SELECT DISTINCT f1.flight_id FROM flight f1, airport_service as1,
city c1, airport_service as2, city c2
WHERE f1.from_airport = as1.airport_code
AND as1.city_code = c1.city_code AND c1.city_name = "Denver"
AND f1.to_airport = as2.airport_code
AND as2.city_code = c2.city_code AND c2.city_name = "Seattle";
```

ATIS

E.g. Idioms