

End-to-End IE without Token-Level Supervision

Rasmus Berg Palm Technical University of Denmark

Dirk Hovy University of Copenhagen

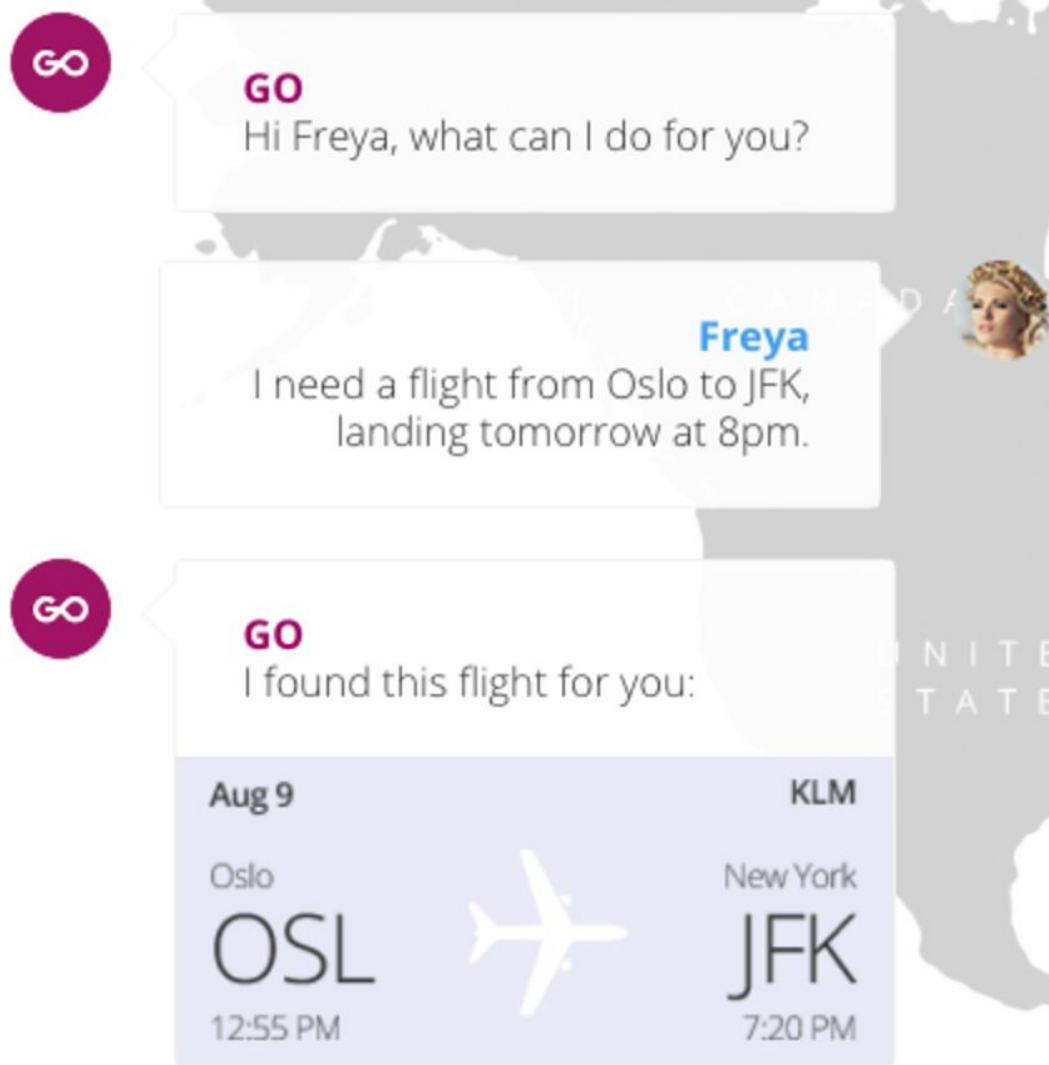
Florian Laws Tradeshift

Ole Winther Technical University of Denmark

Let's create a travel concierge app.

Takes natural language input.

Proposes flights.

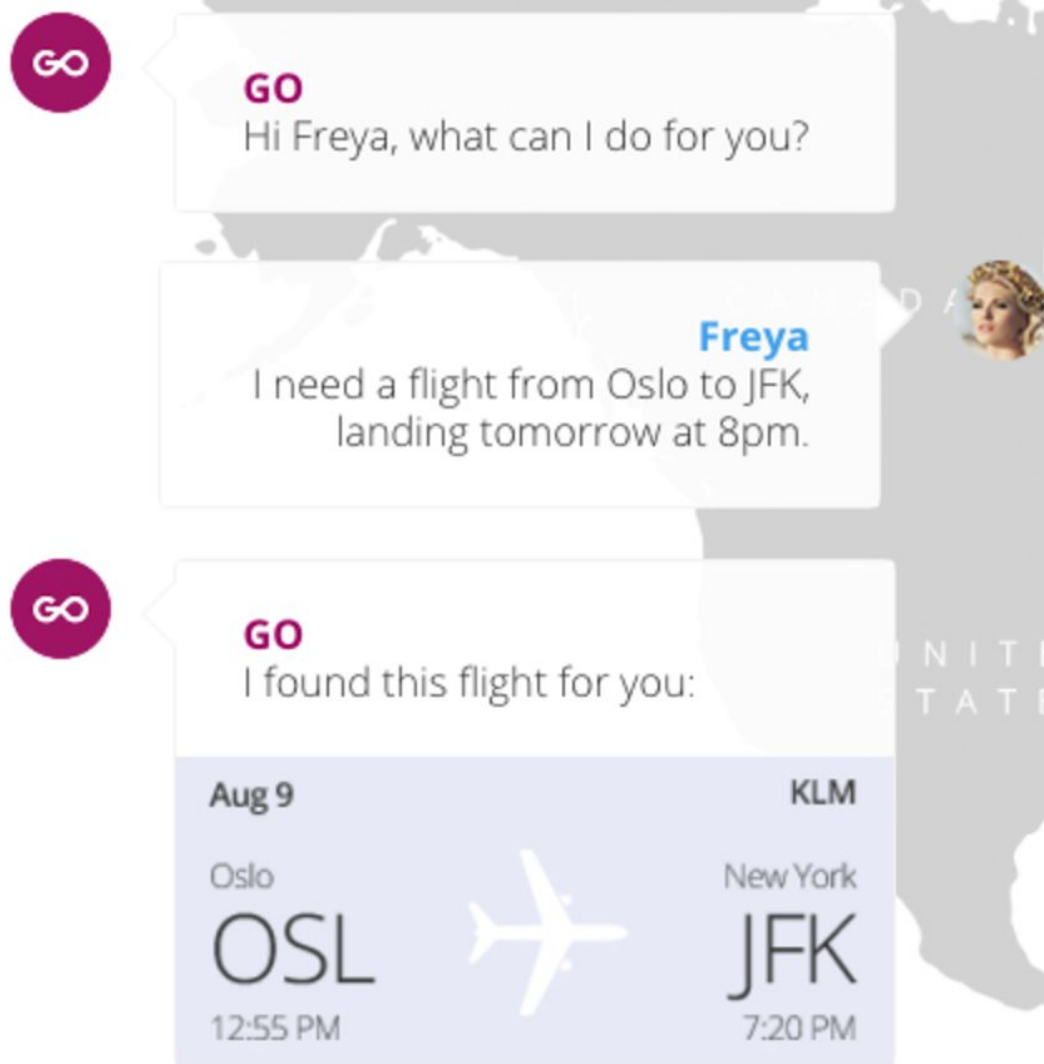


Behind the covers..

We use a flight search engine

The search engine accepts a fixed set of fields, e.g. “from”, “to”, “day”, etc.

We return the top-1 hit



Behind the covers..

We hire human operators to extract the values for the search engine fields.

This labor is tedious for the operators and costly for us

GO
Hi Freya, what can I do for you?

Freya
I need a flight from Oslo to JFK, landing tomorrow at 8pm.

GO
I found this flight for you:

Aug 9	KLM
Oslo	New York
OSL	JFK
12:55 PM	7:20 PM

So let's automate this extraction of information

Token level

Token	Label
cheapest	B-PRICE
airfare	O
from	O
tacoma	B-FROM
to	O
st.	B-TO
louis	I-TO

Token level

Token	Label
cheapest	B-PRICE
airfare	O
from	O
tacoma	B-FROM
to	O
st.	B-TO
louis	I-TO

Token level

Token	Label
cheapest	B-PRICE
airfare	O
from	O
tacoma	B-FROM
to	O
st.	B-TO
louis	I-TO

Field level

Field	Value
FROM	tacoma
TO	st. louis
PRICE	cheapest
DAY	-
MONTH	-
YEAR	-
AIRLINE	-

Token level

Token	Label
cheapest	B-PRICE
airfare	O
from	O
tacoma	B-FROM
to	O
st.	B-TO
louis	I-TO

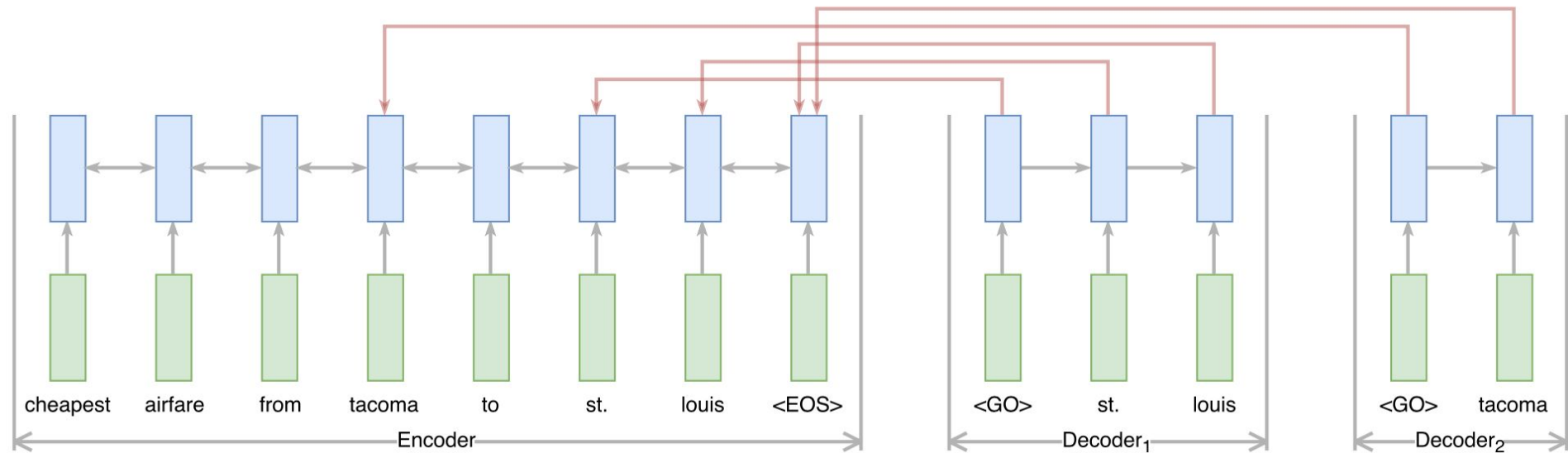
Chunking



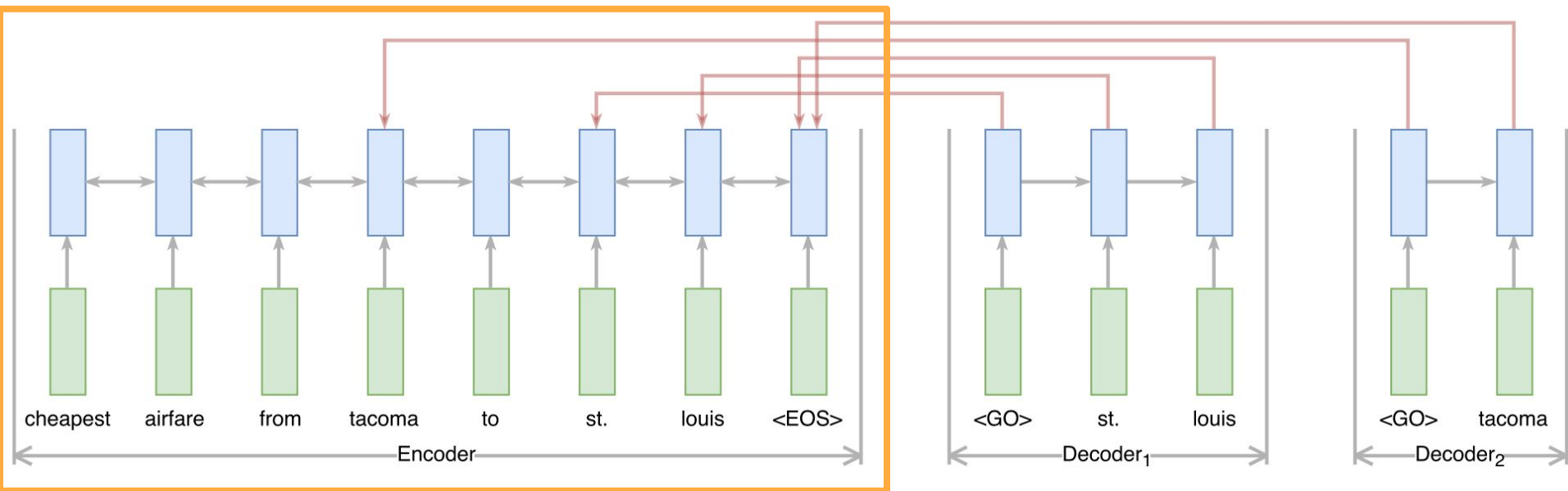
Field level

Field	Value
FROM	tacoma
TO	st. louis
PRICE	cheapest
DAY	-
MONTH	-
YEAR	-
AIRLINE	-

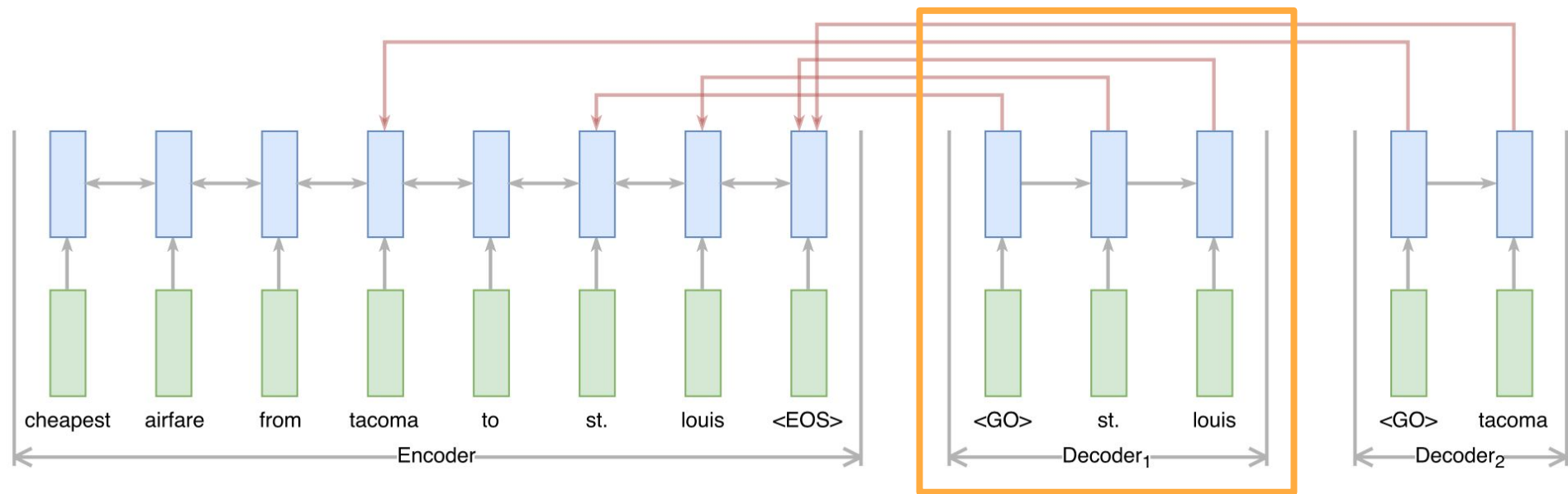
Our model



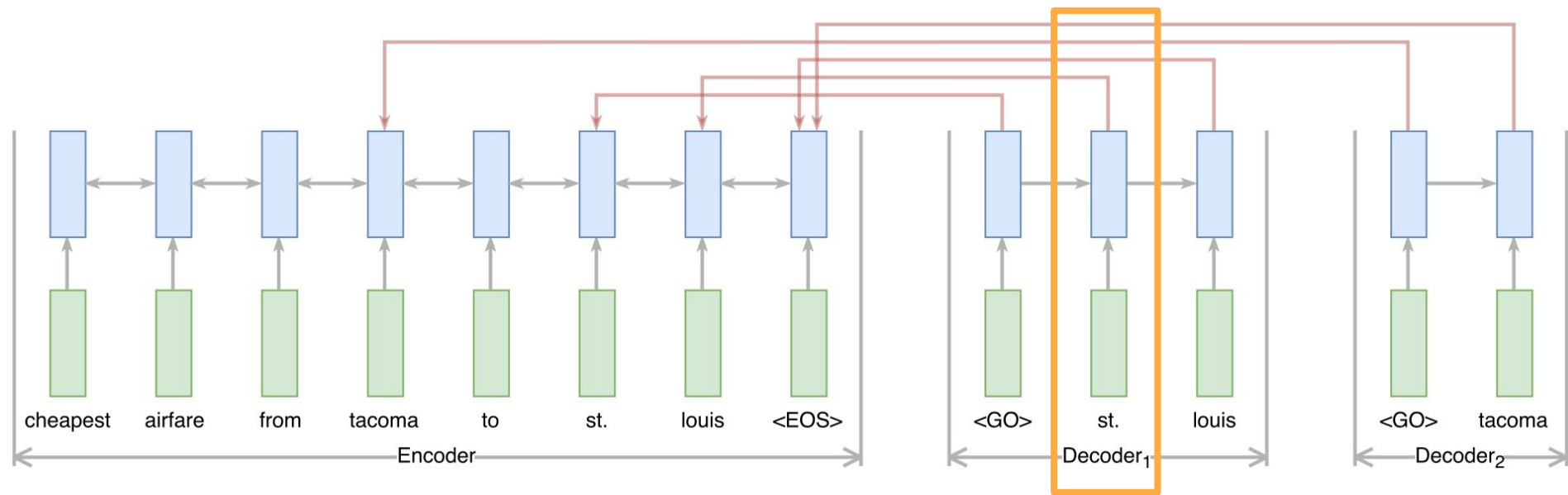
Our model



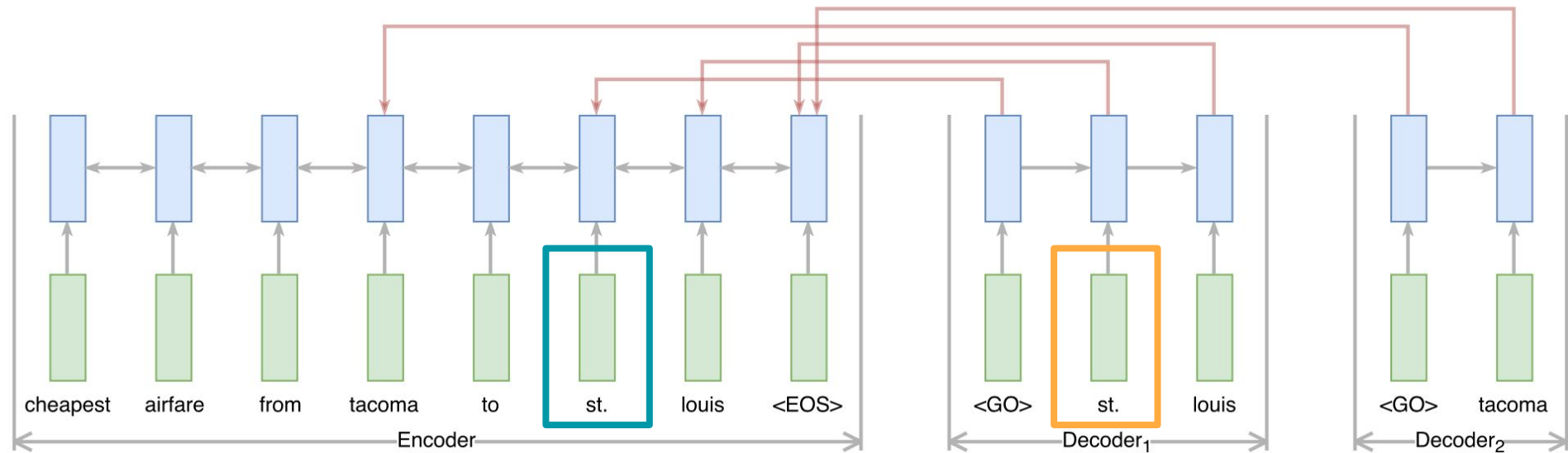
Our model



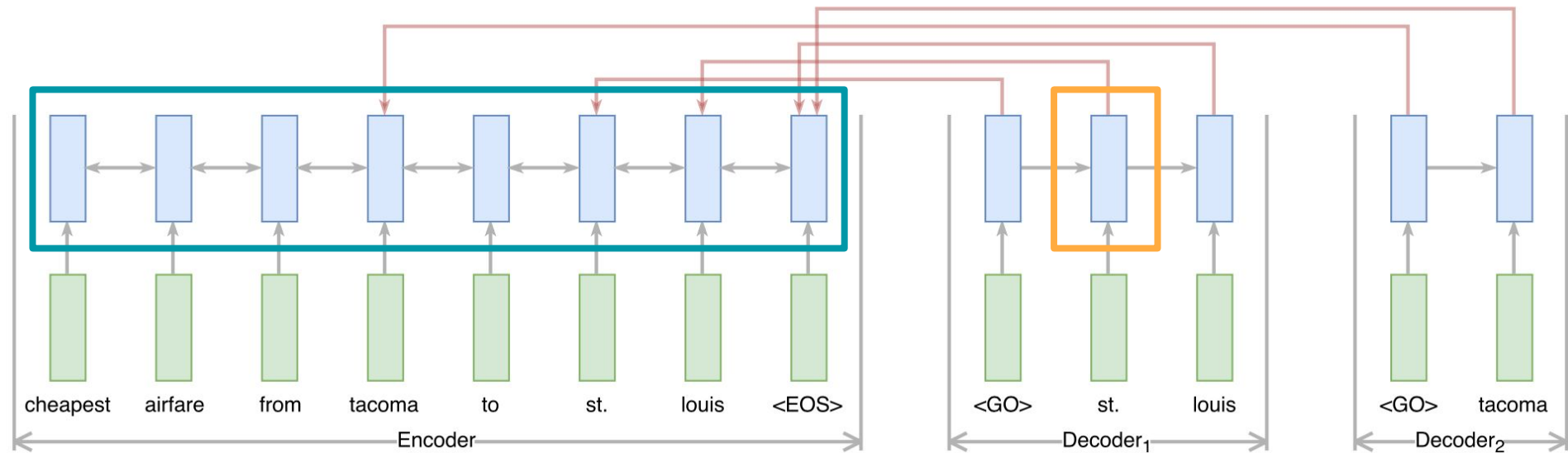
Our model



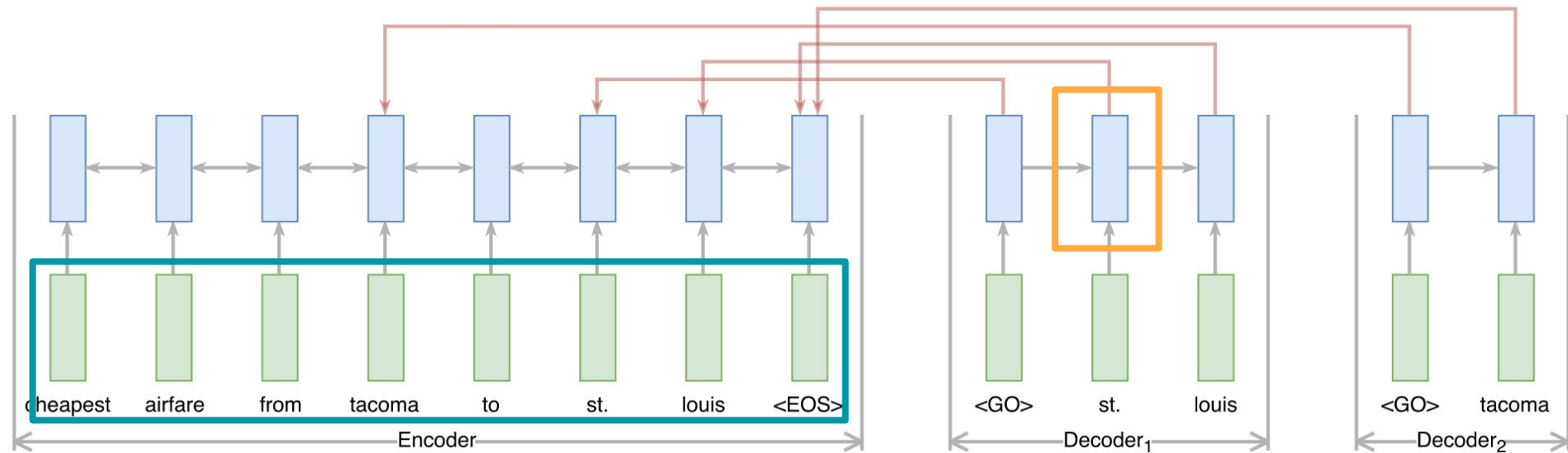
Our model



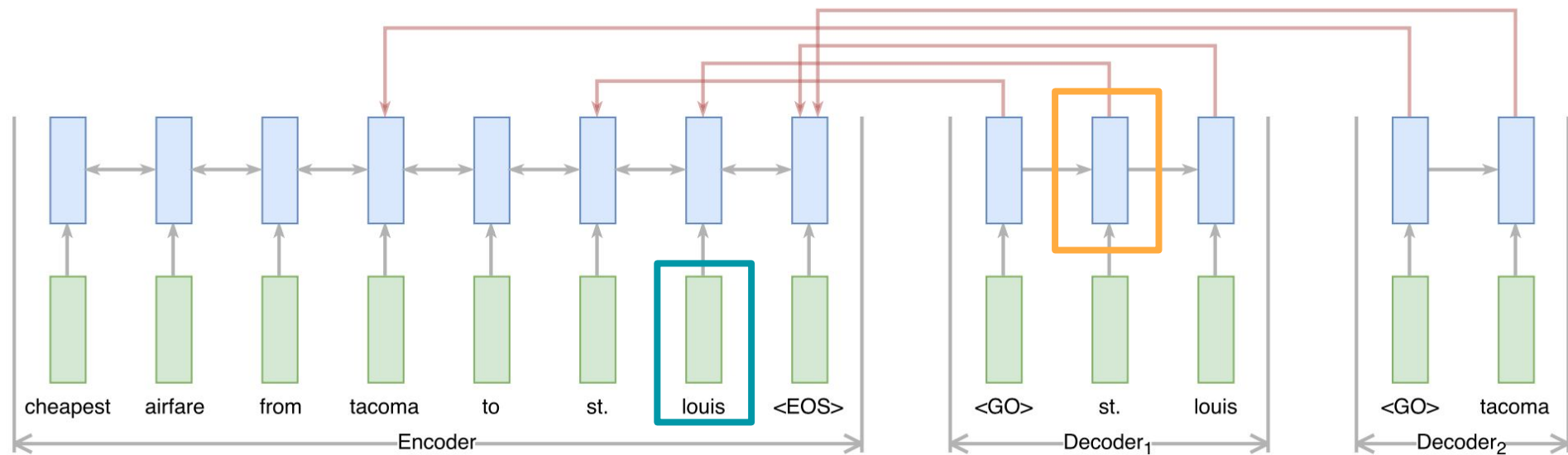
Our model



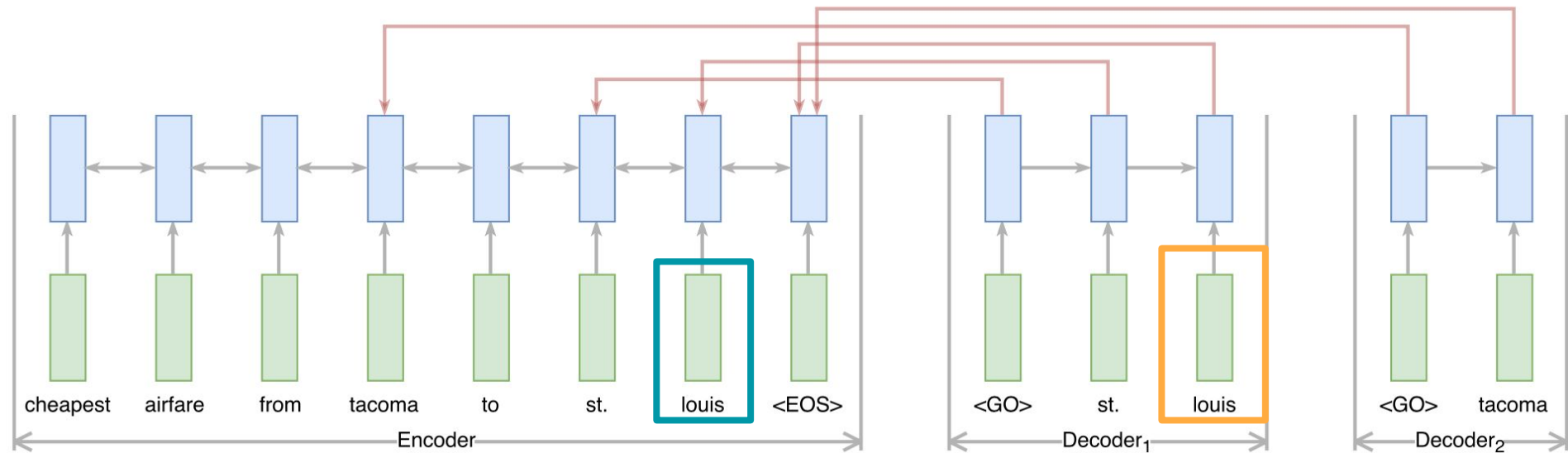
Our model



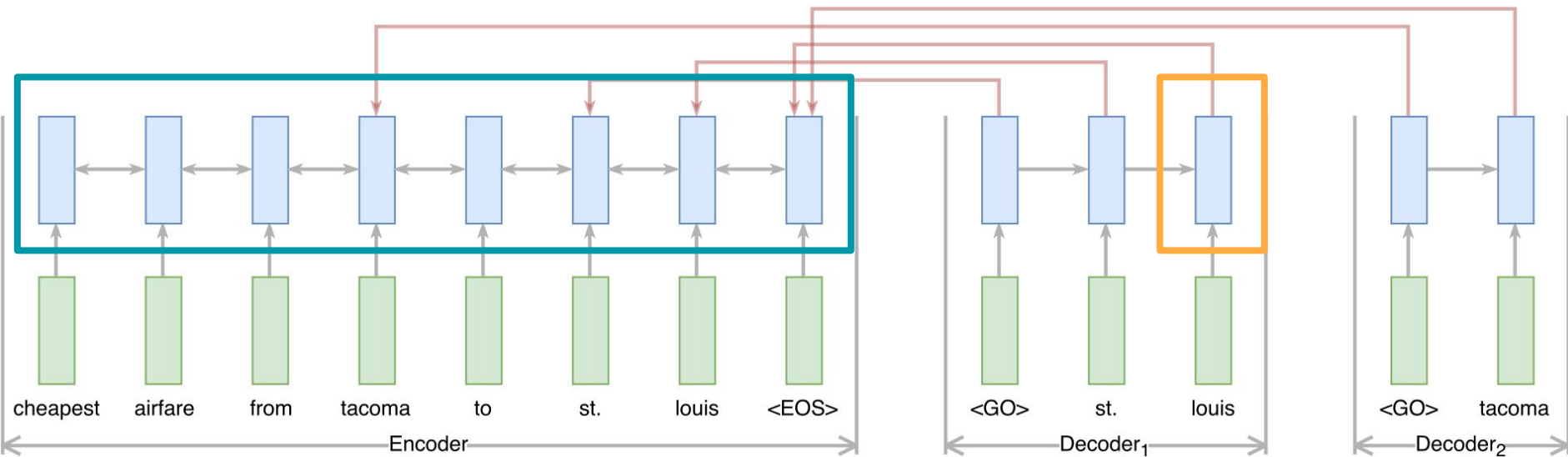
Our model



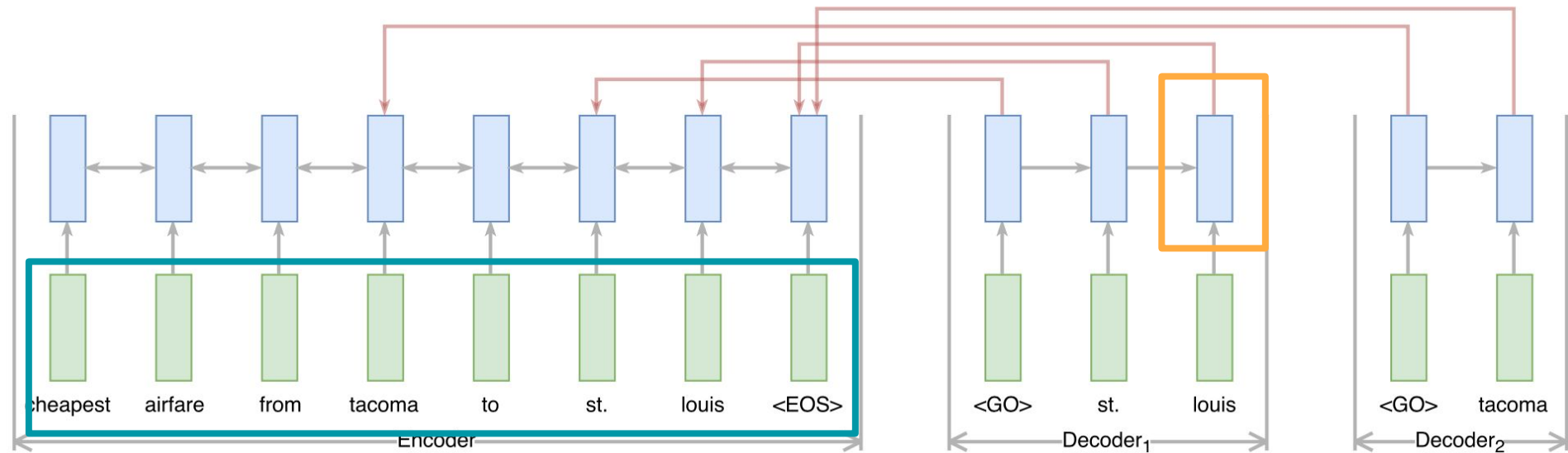
Our model



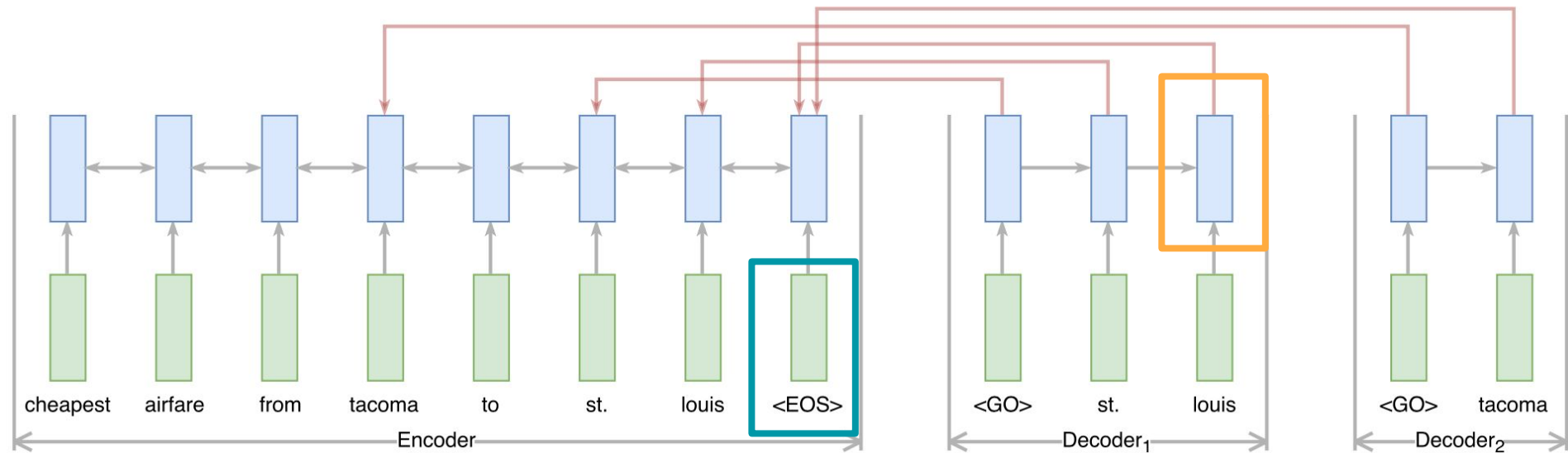
Our model



Our model



Our model



See the paper for details

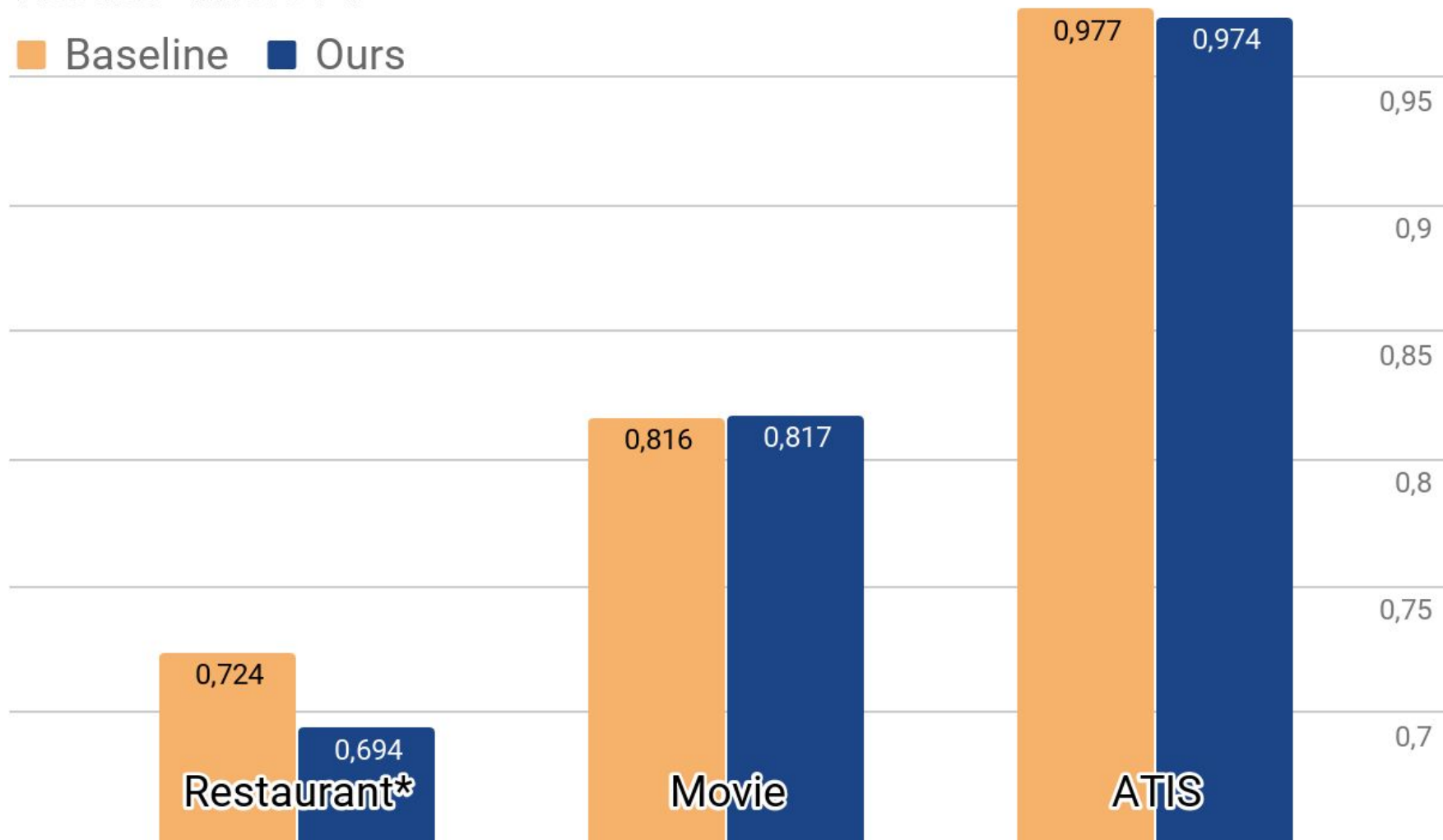
arxiv.org/abs/1707.04913

Code is available at

github.com/rasmusbergpalm/e2e-ie-release

Results - Micro F1

■ Baseline ■ Ours



There's one major limitation

Normalization

'17 Jan 2012' → '2012-01-17'

Thank you!

Questions?