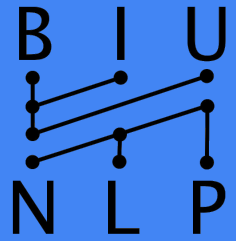
Where's My Head?

Definition, Dataset and Models for Numeric Fused-Heads Identification and Resolution

Yanai Elazar and Yoav Goldberg
NLP Lab, Bar-Ilan University

@TACL, November 7, 2019



This work

I will present:

- A new area of language understanding, that was not well studied
- A new task …
- a new dataset
- and results

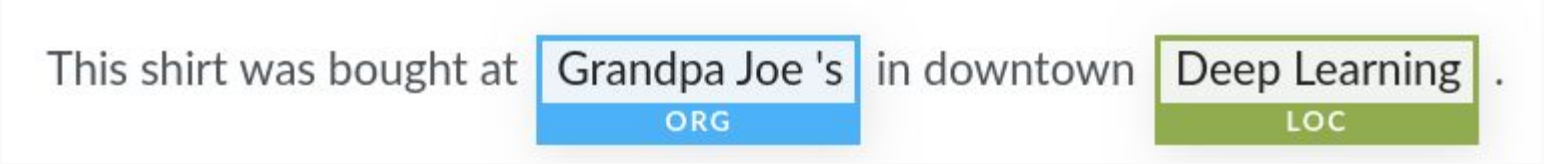
The NLP World Today

We have a bunch of tasks and models

The NLP World Today

We have a bunch of tasks and models

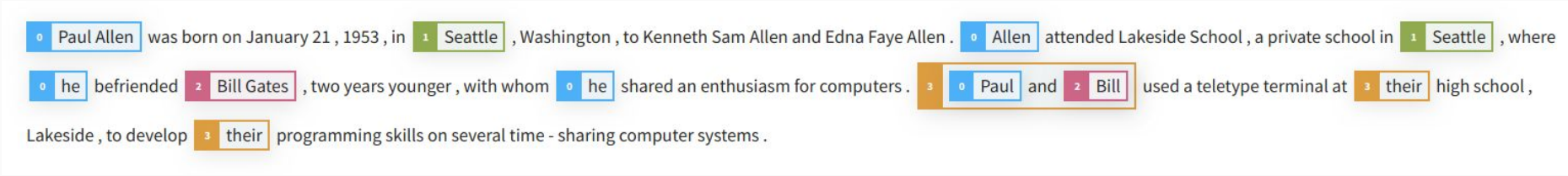
Named Entity Recognition



The NLP World Today

We have a bunch of tasks and models

Coreference Resolution



The NLP World Today

We have a bunch of tasks and models

Textual Entailment

Premise

An interplanetary spacecraft is in orbit around a gas giant's icy moon.

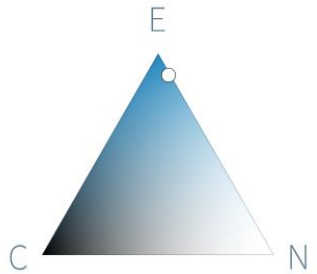
Hypothesis

The spacecraft has the ability to travel between planets.

RUN

Summary

It is **very likely** that the premise **entails** the hypothesis.



Judgment	Probability
Entailment	89.4%
Contradiction	0.8%
Neutral	9.8%

The NLP World Today

And many others…

The NLP World Today

And many others…

And we do well on (some of) them, to some extent.

The NLP World Today

And many others…

And we do well on (some of) them, to some extent.

But they only deal with things which are being said explicitly
And not with what's **not being said**.

The Missing Elements

So what about things which are not being said?

The Missing Elements

So what about things which are not being said?

“Mary is a great programmer but John isn’t”

verb phrase ellipsis (VPE)

The Missing Elements

So what about things which are not being said?

“Mary is a great programmer but John isn’t ___”

verb phrase ellipsis (VPE)

The Missing Elements

So what about things which are not being said?

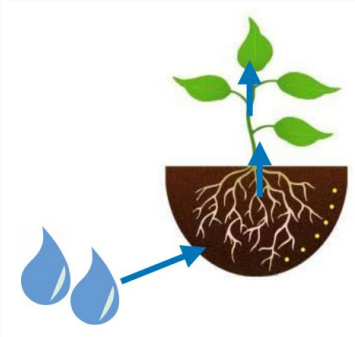
“Water enter the plant through the roots. The water is transported up the stem to the leafs.”

bridging

The Missing Elements

So what about things which are not being said?

“Water enter the plant through the roots. The water is transported up the stem to the leafs.”

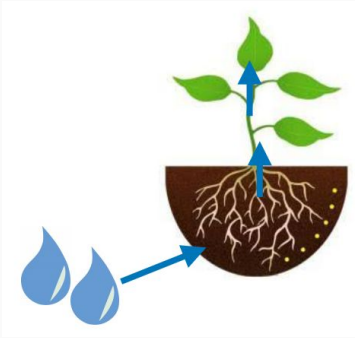


bridging

The Missing Elements

So what about things which are not being said?

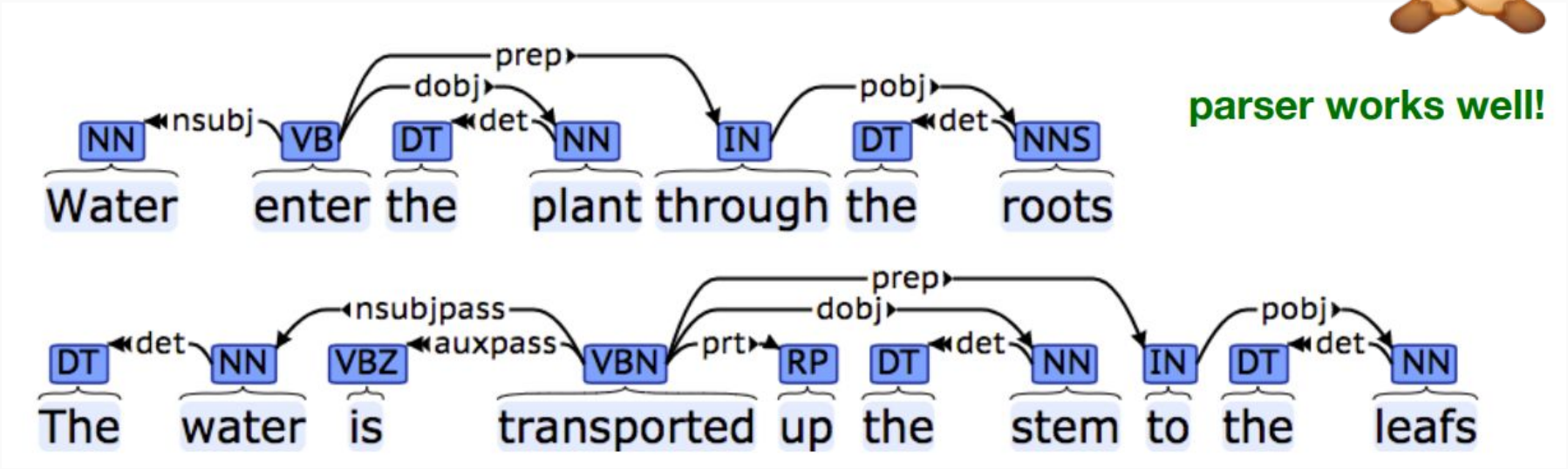
“Water enter the plant through the roots ___. The water is transported up the stem ___ to the leafs ___.”



bridging

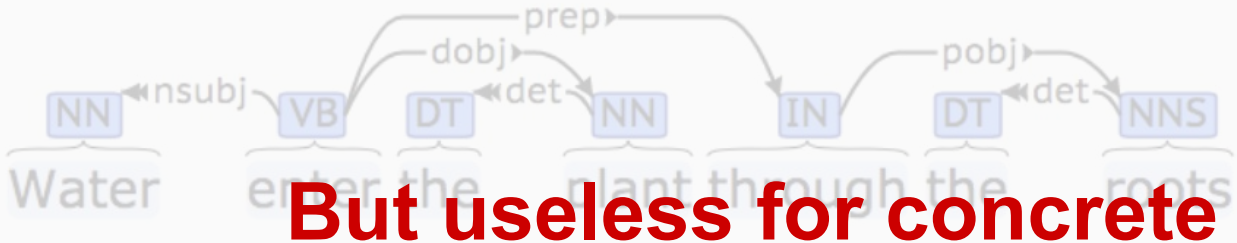
The Missing Elements

Current models can be correct on these cases



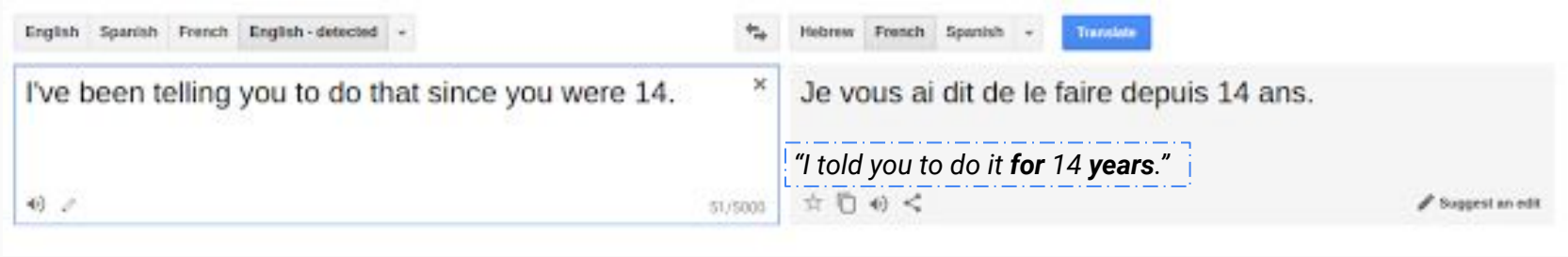
The Missing Elements

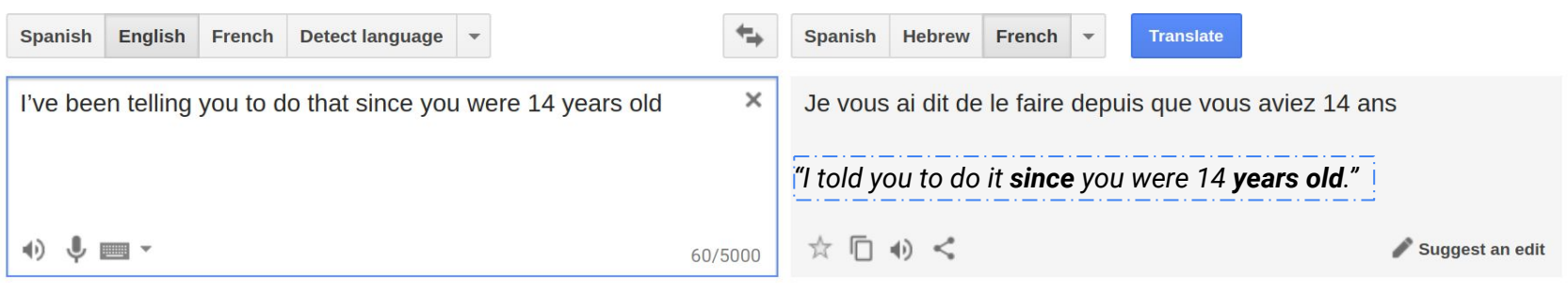
Current models can be correct on these cases



But useless for concrete understanding of the situation

The Missing Elements





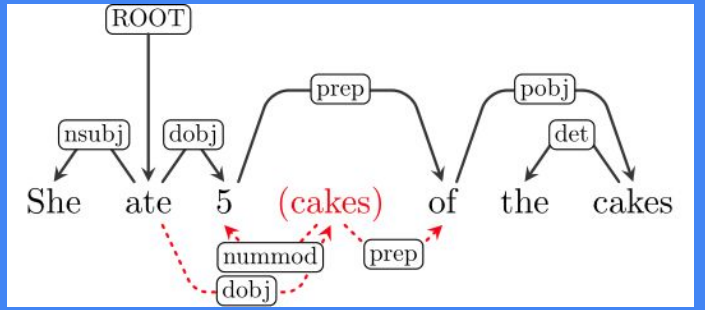
The Missing Elements

- Many things in language are not being stated explicitly
- These phenomena stand as obstacles for text understanding
- We call these: “**The Missing Elements**”

The Missing Elements

In this work, we focus on a specific type of missing elements: the **Numeric Fused-Head**.

The Numeric Fused-Head



Fused Head - Definition

“FHs constructions are noun phrases (NPs) in which the head noun is missing and is said to be “fused” with its dependent modifier.”

--Huddleston et al., 2002

Fused Head - Examples

- *“Only the **rich** __ will benefit.”*
- *“Rolly, didn’t you have **enough** __ to eat?”*
- *Do you want another cruller?*

  *No thanks, I’ve had **two** __ already.*

Fused Head - Examples

- *“Only the **rich** __ will benefit.”*
- *“Rolly, didn’t you have **enough** __ to eat?”*
- *Do you want another cruller?*

  *No thanks, I’ve had **two** __ already.*

In this work, we focus on numbers.

Fused Head - Motivation

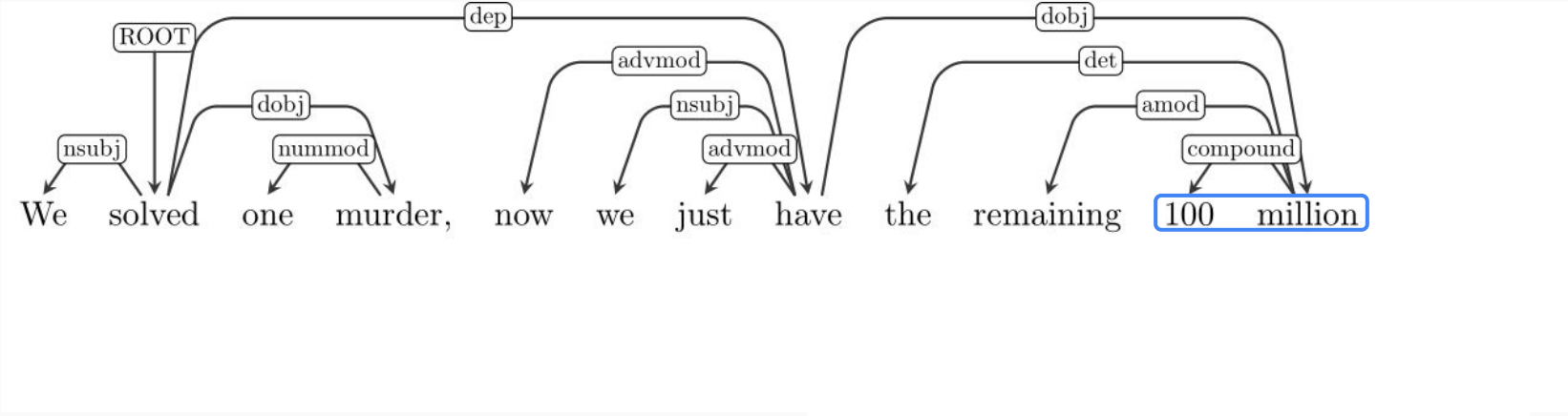
Recovering the missing *element* can be useful for many downstream tasks

- Translation
- Information Extraction
- Named Entity Recognition
- etc...

Fused Head - Task

Given text:

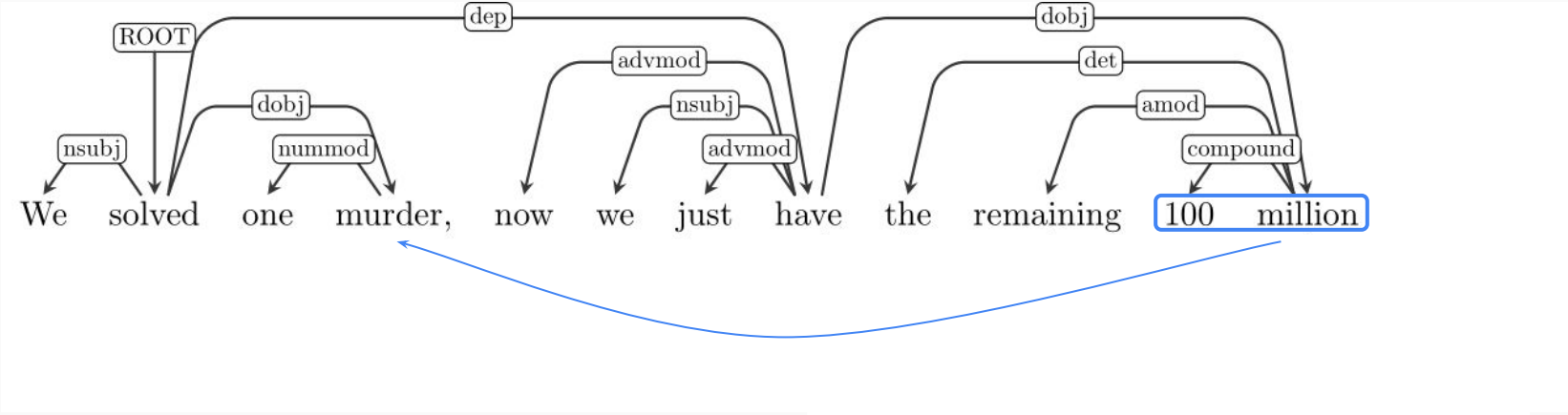
- Discover all the fused-heads



Fused Head - Task

Given text:

- Discover all the fused-heads
- Recover their missing head



Fused Head - Task

Given text:

- Discover all the fused-heads
- Recover their missing head

And by recovering the missing head…

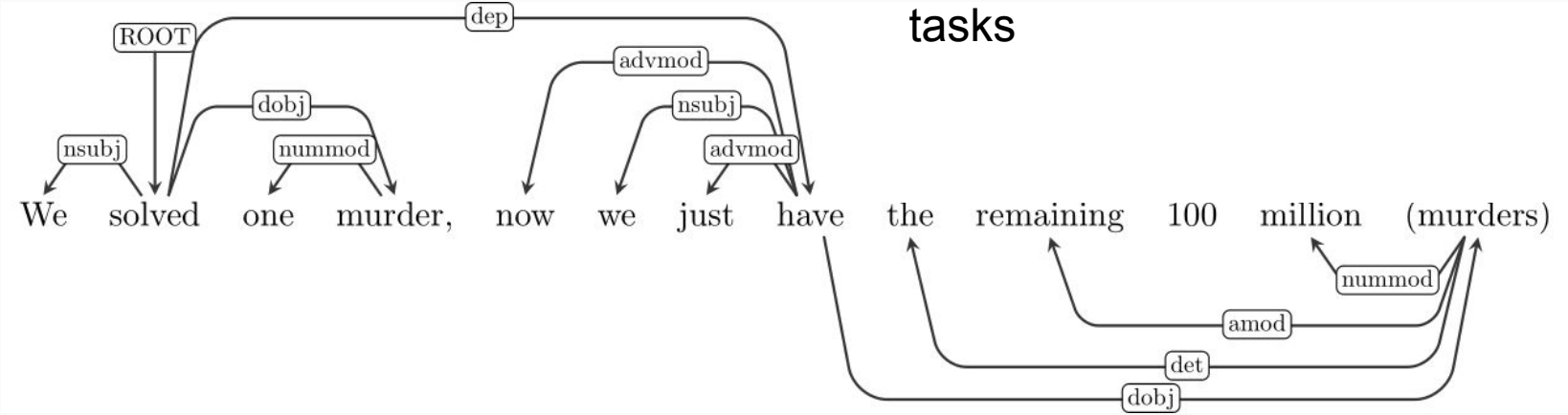
We solved one murder, now we just have the remaining 100 million (murders)

Fused Head - Task

Given text:

- Discover all the fused-heads
- Recover their missing head

And by recovering the missing head…
We can improve downstream tasks



Fused Head - Task

Given text:

- Discover all the fused-heads
- Recover their missing head

We treat the problem as two separate tasks, and tackle each one separately

Fused Head - Task

Given text:

- Discover all the fused-heads (**Identification**)
- Recover their missing head (**Resolution**)

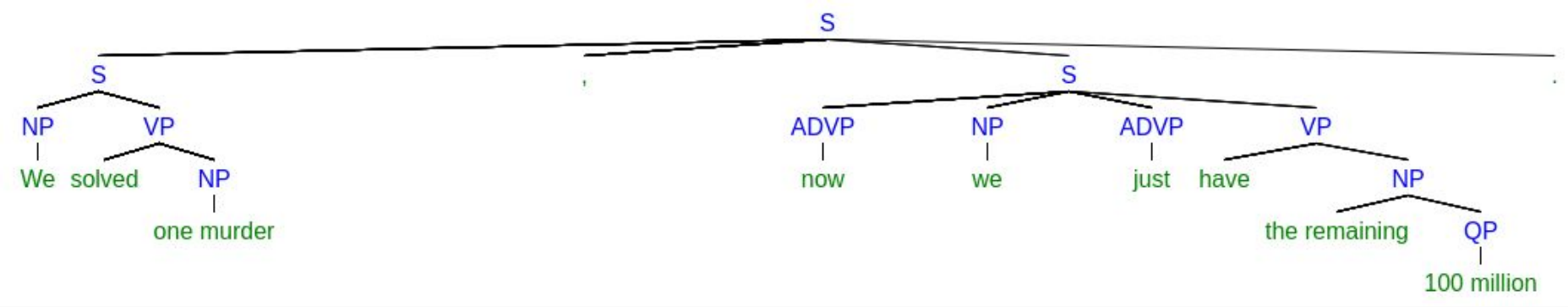
We treat the problem as two separate tasks, and tackle each one separately

NFH Identification

Fused Head - Identification

We make use of the definition:
An NP without a noun

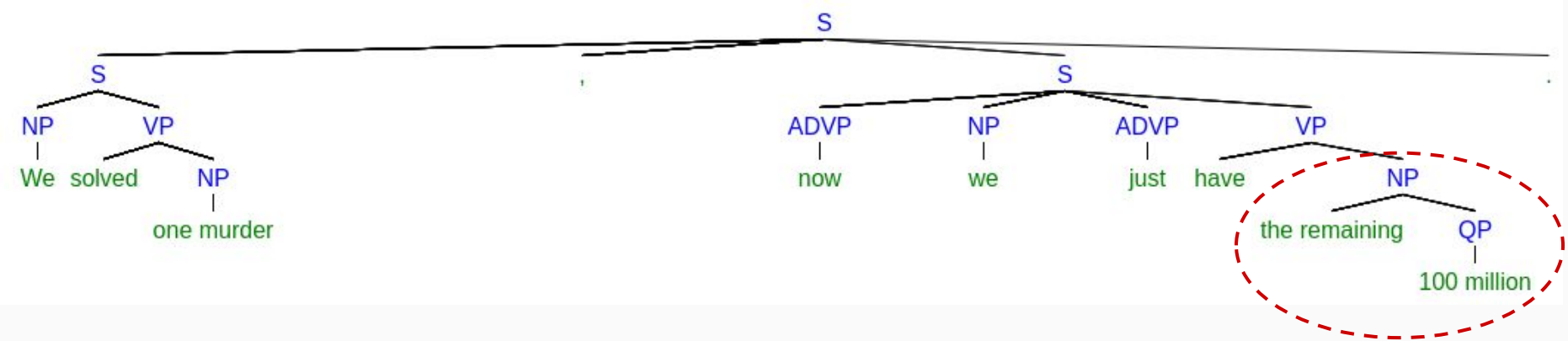
*“We solved one murder, now we just have the remaining **100 million**.”*



Fused Head - Identification

We make use of the definition:
An NP without a noun

*“We solved one murder, now we just have the remaining **100 million**.”*



Fused Head - Identification Task

- Given a sentence, return a list of spans corresponding to NFHs within it
- We created a test-set of 500 annotated numbers
  - For binary Fused-Head or not

Fused Head - Identification Task

Method #1: Rule Based

95.5% F1

- NP with no Nouns
- + Additional 10 text-based patterns. E.g:
  - **Particle** - “Everyone here lives to their **90**'s”
  - …

Fused Head - Identification Task

Method #2: Machine Learning

Check out the paper (or talk to me later)

Fused Head - Identification Task

Method #1: Rule Based

95.5% F1

Fused Head - Identification Task

Method #1: Rule Based

95.5% F1

Method #2: Machine Learning

96.6% F1

Fused Head - Identification Task

Now that we know how to identify,

How many of the Numbers in corpora are NFH?

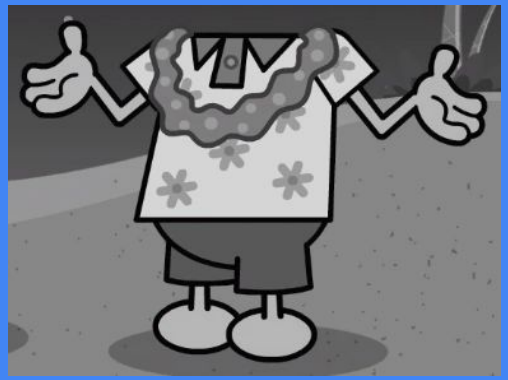
Casual text

Formal text

- 41.2% in IMDB
- 37.5% in Amazon product reviews
- 35.5% in TED talks
- 33.2% in Wikipedia (WikiText2)

*The **NFH** phenomenon is very common*

NFH Resolution



Fused Head - Resolution Task

- Given an NFH, resolve its missing head

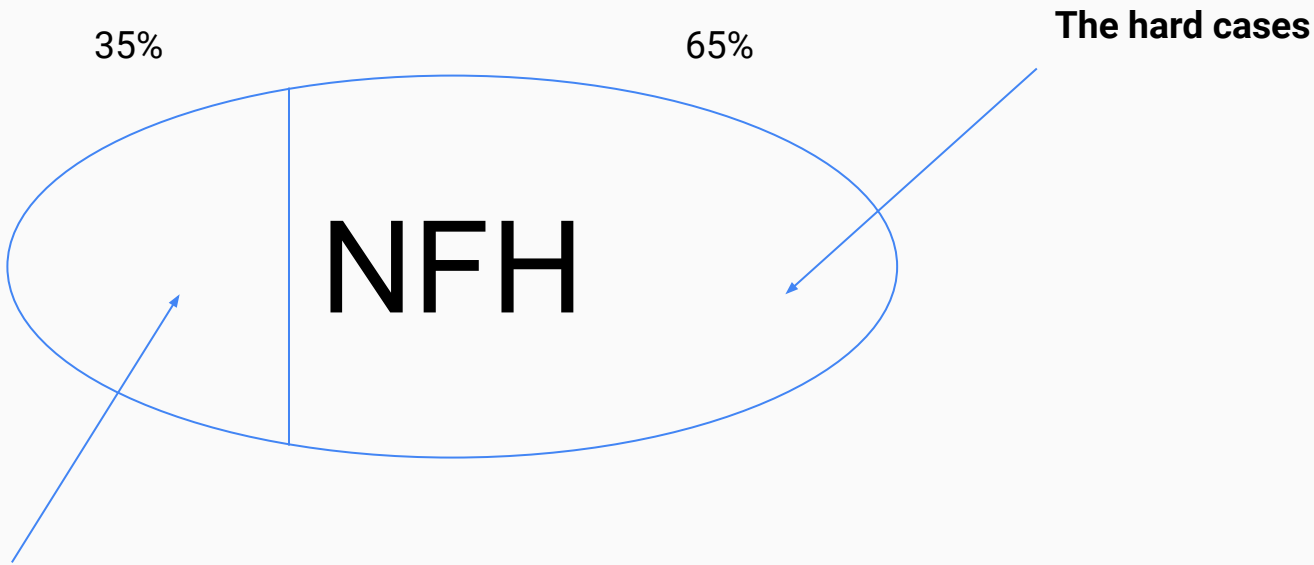
- The head can be within the context (***Reference***)
  - *“Have you ever had an unexpressed thought? I’m having **one** __ now.”*
- or it doesn’t appear in the text at all (***Implicit***)

Fused Head - Resolution Task

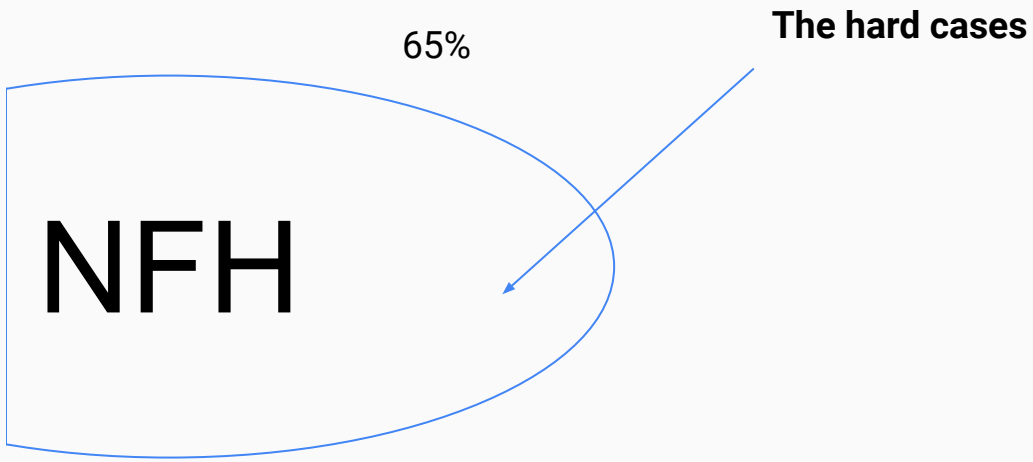
- Given an NFH, resolve its missing head

- The head can be within the context (***Reference***)
  - *“Have you ever had an unexpressed thought? I’m having **one** __ now.”*
- or it doesn’t appear in the text at all (***Implicit***)
  - *“When the clock strikes **one**... the Ghost of Christmas Past”*

Fused Head - Resolution Task Candidates

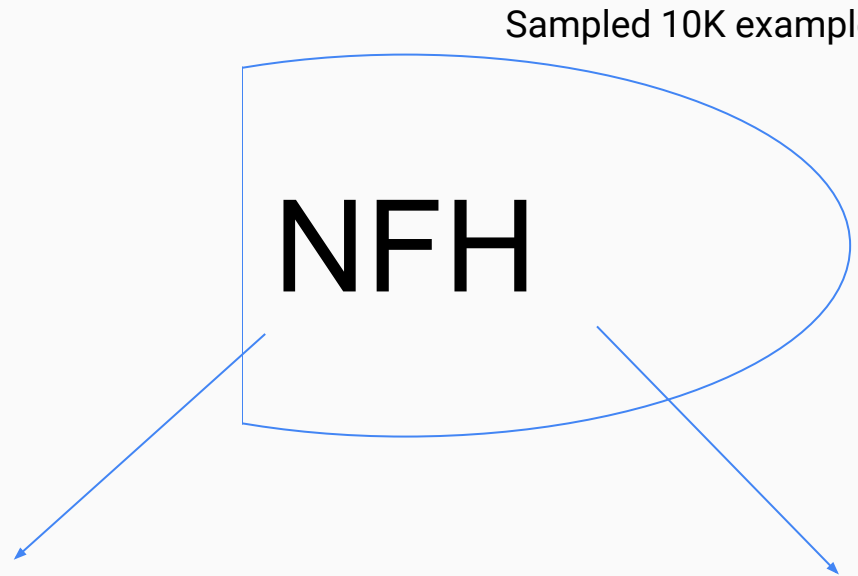


Fused Head - Resolution Task Candidates



Fused Head - Resolution Task Candidates

The hard cases



- In case of ***Reference***, need to choose a span within the context
- In case of ***Implicit***, choose one of most common categories

Fused Head - Resolution Task

The most common ***Implicit*** examples of the dataset

- Year: *“I’d like to wish a happy **1969** to our new President.”*
- Age: *“I probably feel worse than Debi Moore did when she turned **50.**”*
- People: *...*
- Currency: *...*
- Time: *...*
- Other: ...

Fused Head - Resolution Task Annotation

We solved one murder, now we just have the remaining **100 million**

What does the number 100 million refer to?

- ○ Reference **(please mark with your mouse)**
- ○ Year
- ○ Age
- ○ Currency
- ○ Other **(please specify)**
- ○ Person/People
- ○ Time
- ○ Unknown

Fused Head - Resolution Task Annotation

We solved one murder, now we just have the remaining **100 million**

What does the number 100 million refer to?

- (●) Reference **(please mark with your mouse)**
- ( ) Year
- ( ) Age
- ( ) Currency
- ( ) Other **(please specify)**
- ( ) Person/People
- ( ) Time
- ( ) Unknown

Fused Head - Resolution Task Annotation

- 10K examples
- Crowdsourcing
- 3 annotators per task
- 0.73 Fleiss *kappa* (substantial agreement)
- Using majority as correct label

Fused Head - Resolution Task Annotation

Some hard cases:

- *“Mexicans have **fifteen**, Jews have thirteen, rich girls have sweet sixteen…"* *[YEARS-OLD, celebration]?*
- *“All her communications are to Minnesota numbers. There’s not **one** from California.”* *[communications, numbers]?*

Fused Head - Resolution Modeling

Given a numeric anchor and its surrounding context, assign its head, from:

- The surrounding text (***Reference***)
- The most common categories (***Implicit***)

In this work, we use:

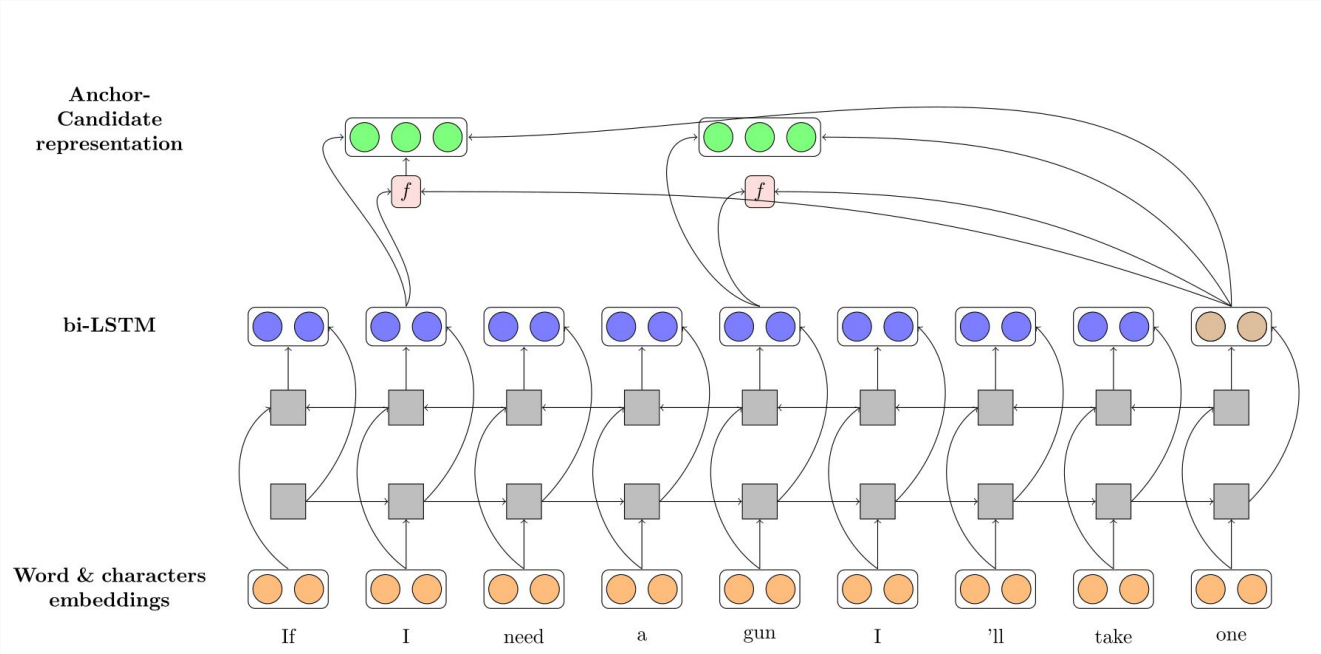
- A single head per NFH
- A single token head

NFH Resolution Model

Fused Head - Resolution Model

Based on Lee et al. 17

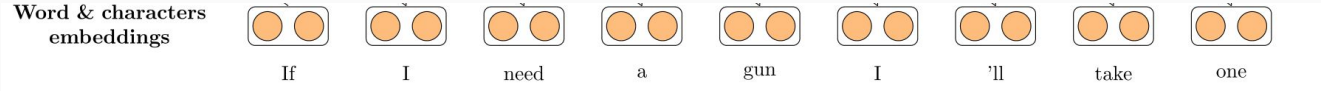
How to combine the ***Reference*** decisions with the *implicit* ones?



Fused Head - Resolution Model

Based on Lee et al. 17

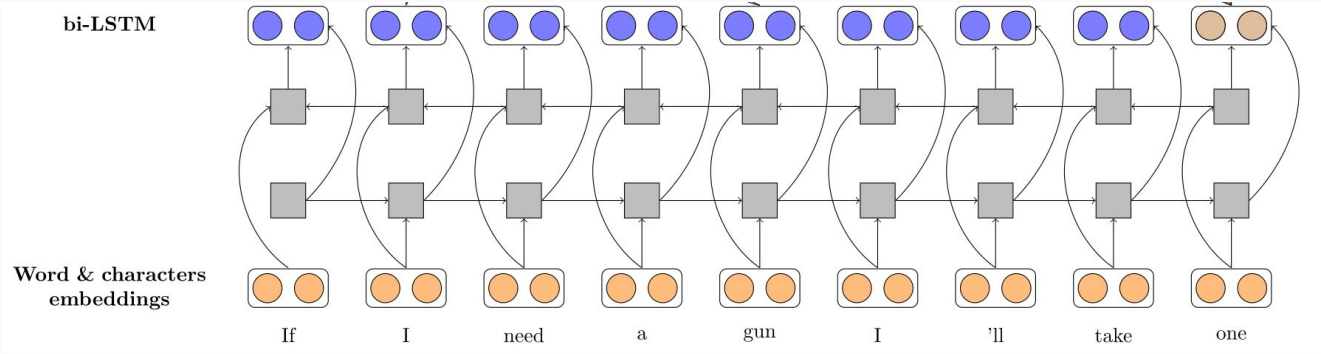
How to combine the ***Reference*** decisions with the *implicit* ones?



Fused Head - Resolution Model

Based on Lee et al. 17

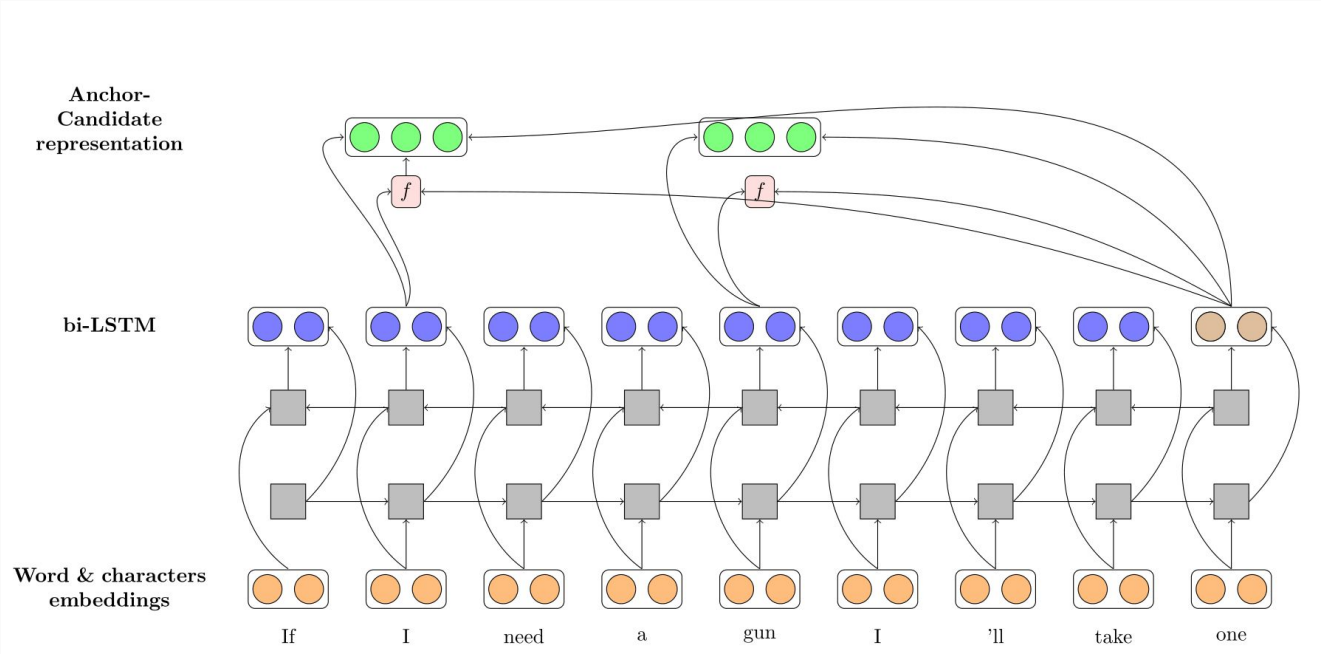
How to combine the ***Reference*** decisions with the *implicit* ones?



Fused Head - Resolution Model

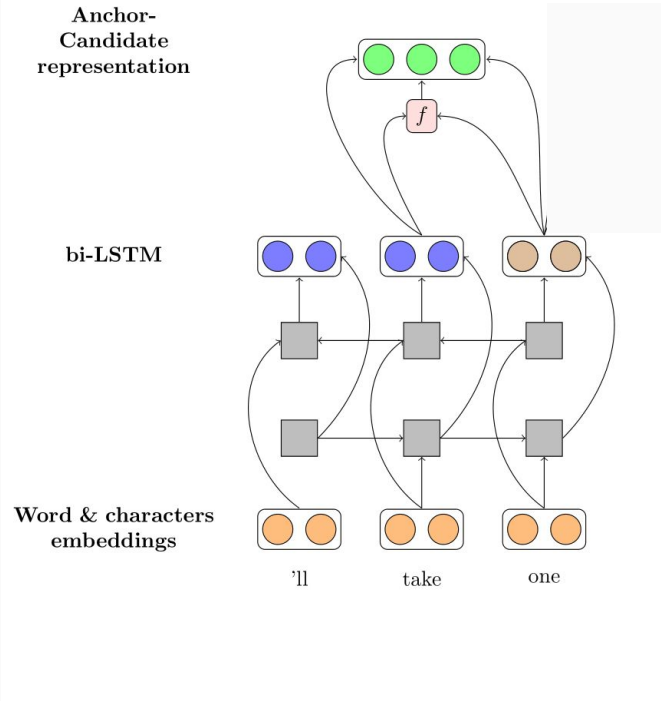
Based on Lee et al. 17

How to combine the ***Reference*** decisions with the *implicit* ones?



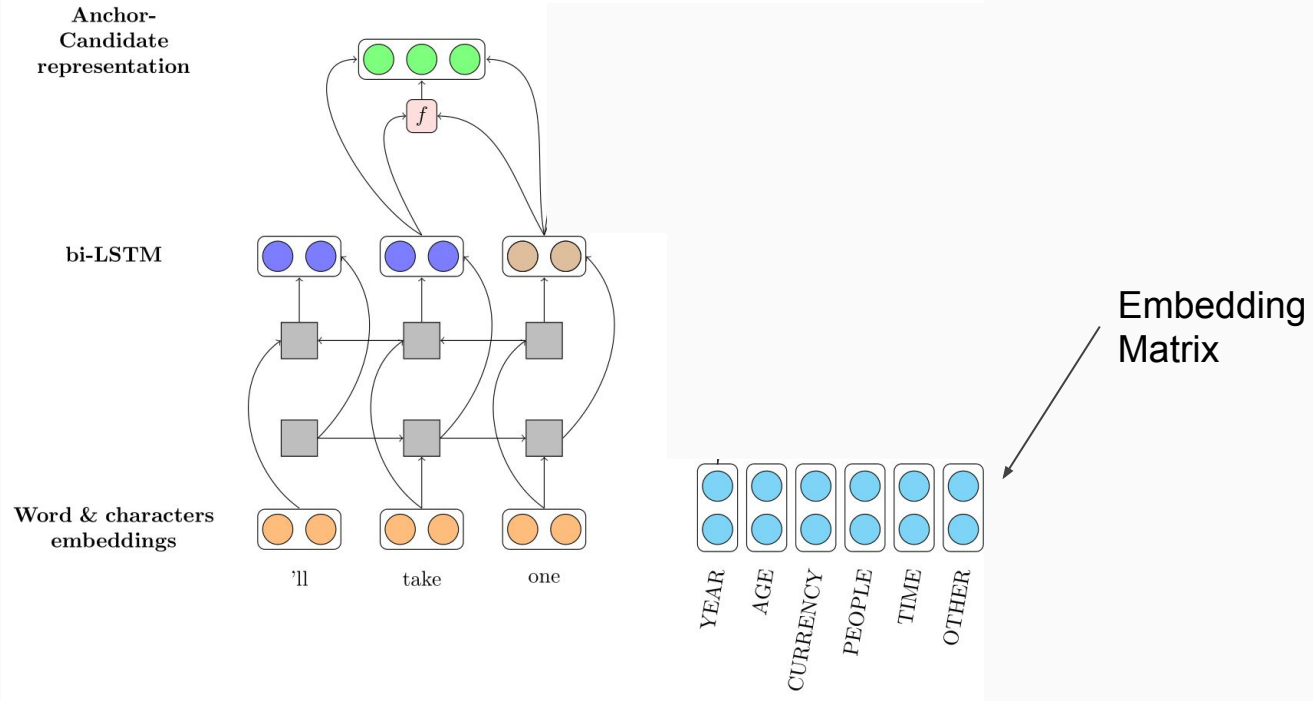
Fused Head - Resolution Model

How to combine the *reference* decisions with the ***Implicit*** ones?



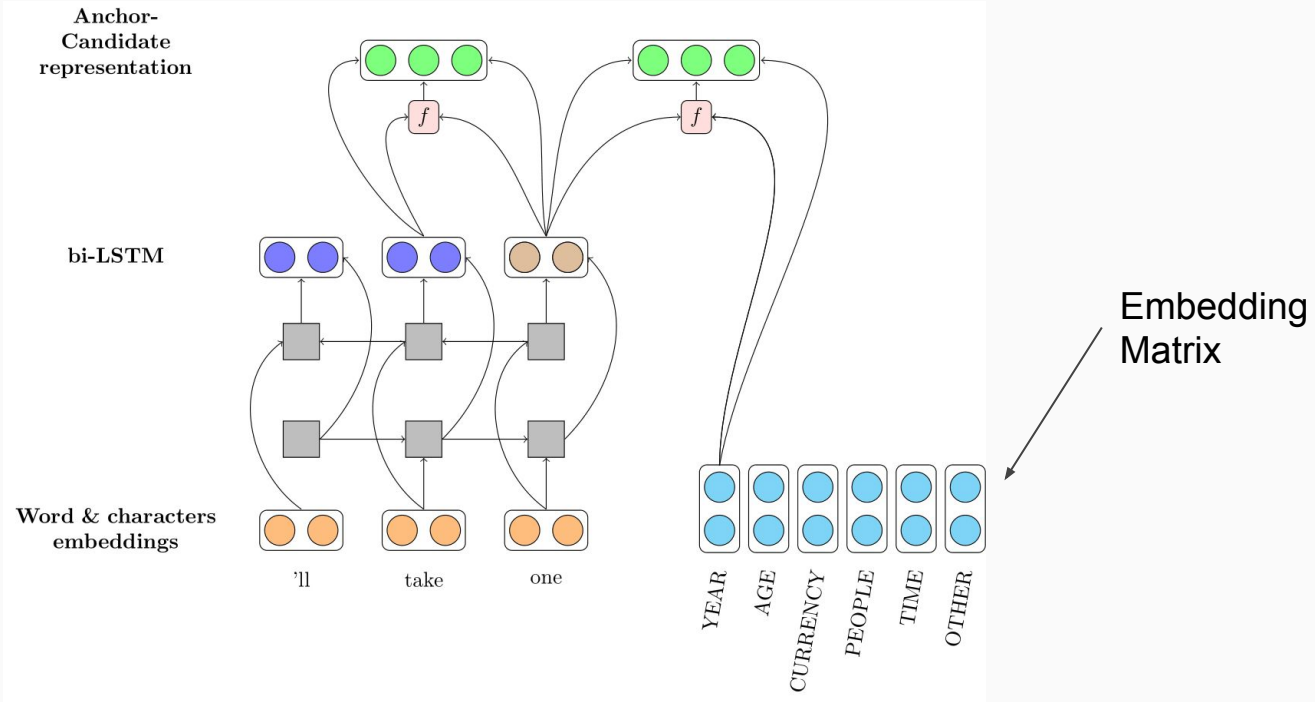
Fused Head - Resolution Model

How to combine the *reference* decisions with the ***Implicit*** ones?



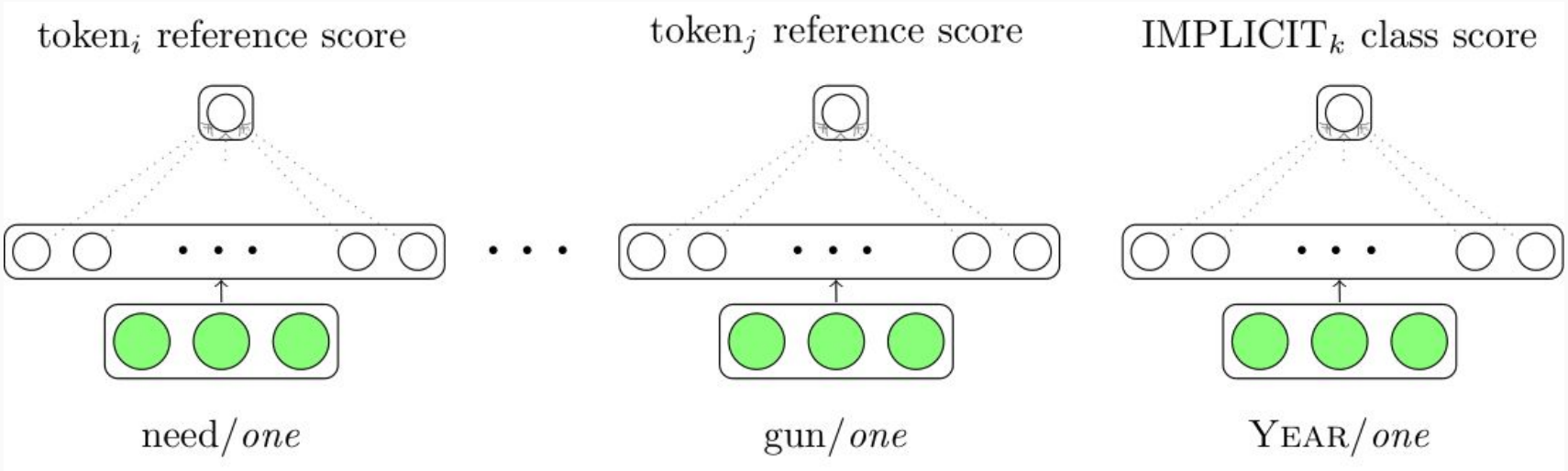
Fused Head - Resolution Model

How to combine the *reference* decisions with the ***Implicit*** ones?

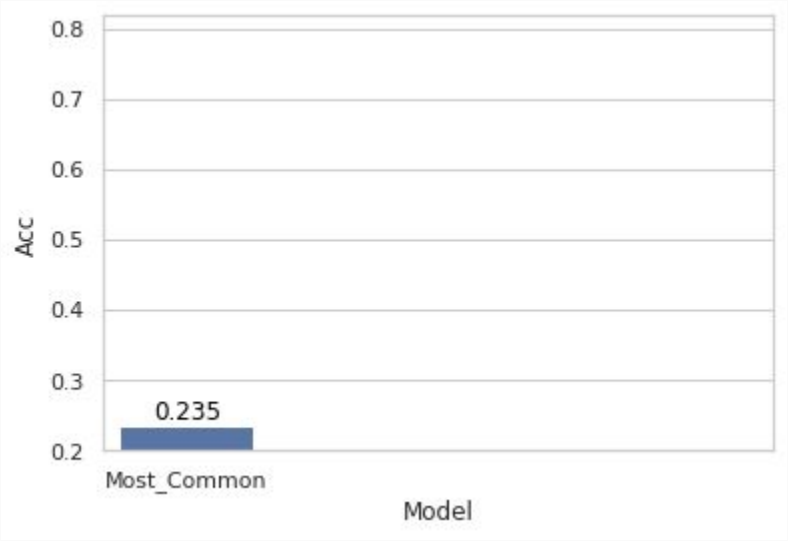


Fused Head - Resolution Model

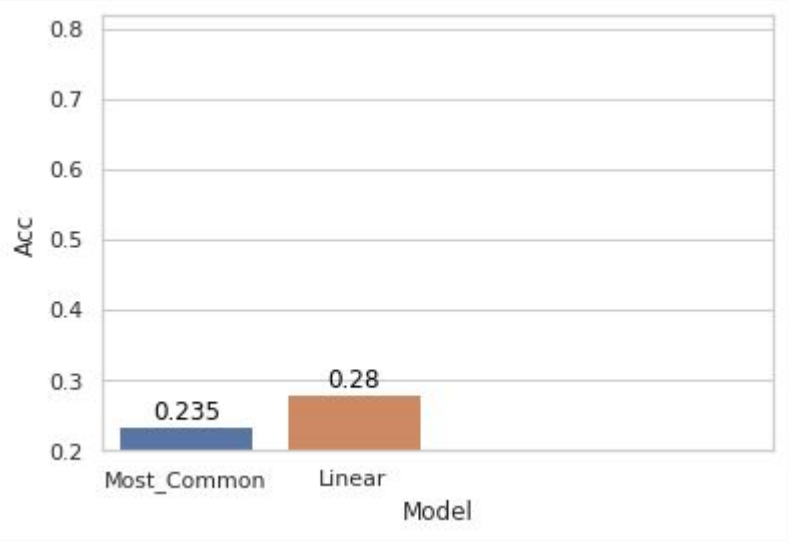
How to combine the ***Reference*** decisions with the ***Implicit*** ones?



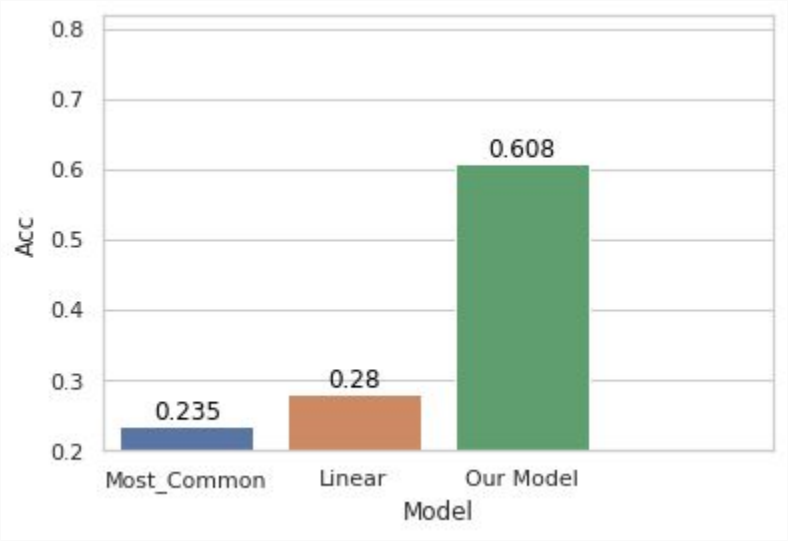
Fused Head - Resolution Results



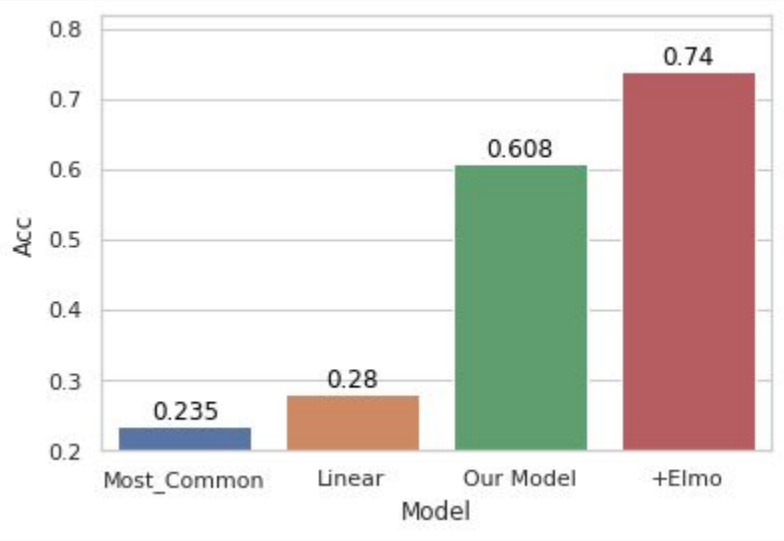
Fused Head - Resolution Results



Fused Head - Resolution Results

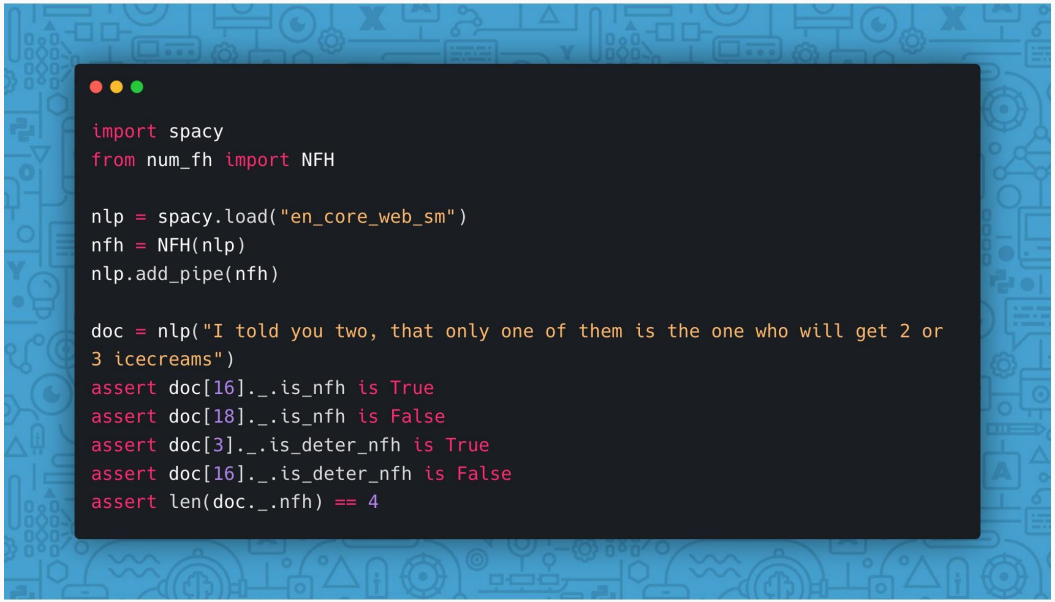


Fused Head - Resolution Results



Fused Head - Resolution Results

We are on spacy!



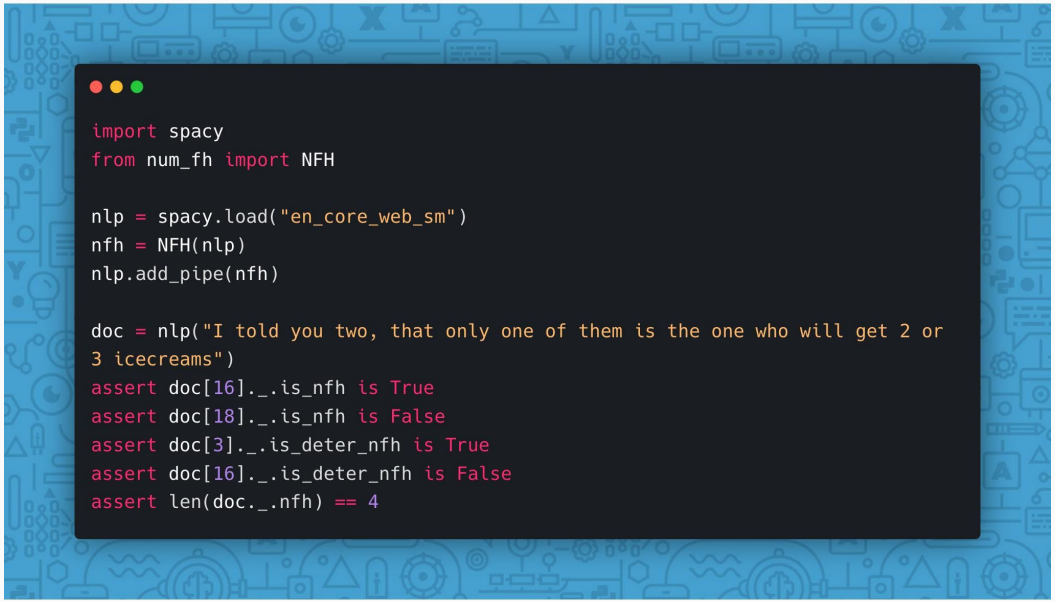
```python
import spacy
from num_fh import NFH

nlp = spacy.load("en_core_web_sm")
nfh = NFH(nlp)
nlp.add_pipe(nfh)

doc = nlp("I told you two, that only one of them is the one who will get 2 or
3 icecreams")
assert doc[16]._.is_nfh is True
assert doc[18]._.is_nfh is False
assert doc[3]._.is_deter_nfh is True
assert doc[16]._.is_deter_nfh is False
assert len(doc._.nfh) == 4
```

Fused Head - Resolution Results

We are on spacy!
&
We have a demo!



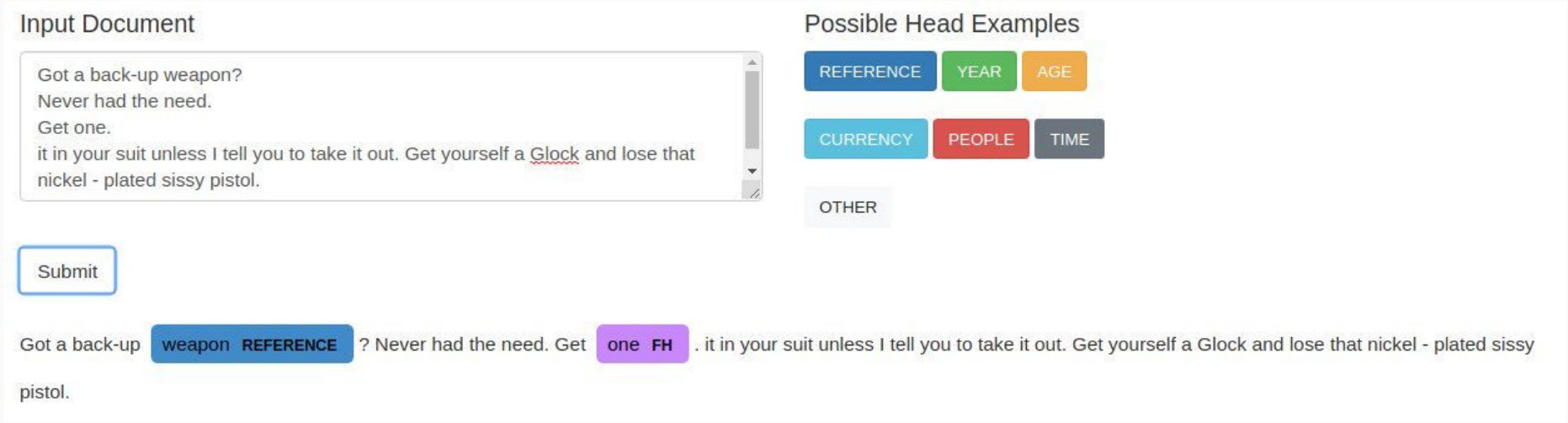
```python
import spacy
from num_fh import NFH

nlp = spacy.load("en_core_web_sm")
nfh = NFH(nlp)
nlp.add_pipe(nfh)

doc = nlp("I told you two, that only one of them is the one who will get 2 or
3 icecreams")
assert doc[16]._.is_nfh is True
assert doc[18]._.is_nfh is False
assert doc[3]._.is_deter_nfh is True
assert doc[16]._.is_deter_nfh is False
assert len(doc._.nfh) == 4
```

Fused Head - A Working Model

Fused Head - A Working Model



NFH Resolution - Analysis

Fused Head - Analysis

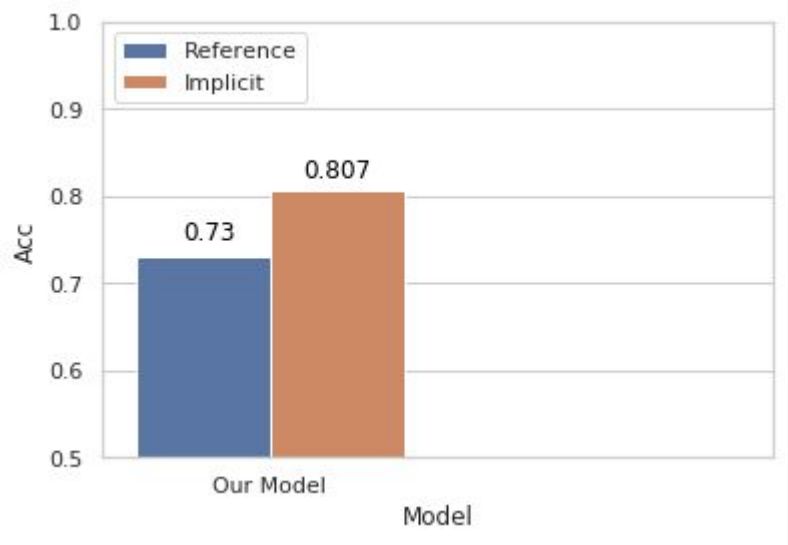
How well do we separate between the ***Reference*** and ***Implicit*** classes?

86%

Fused Head - Analysis

Therefore,

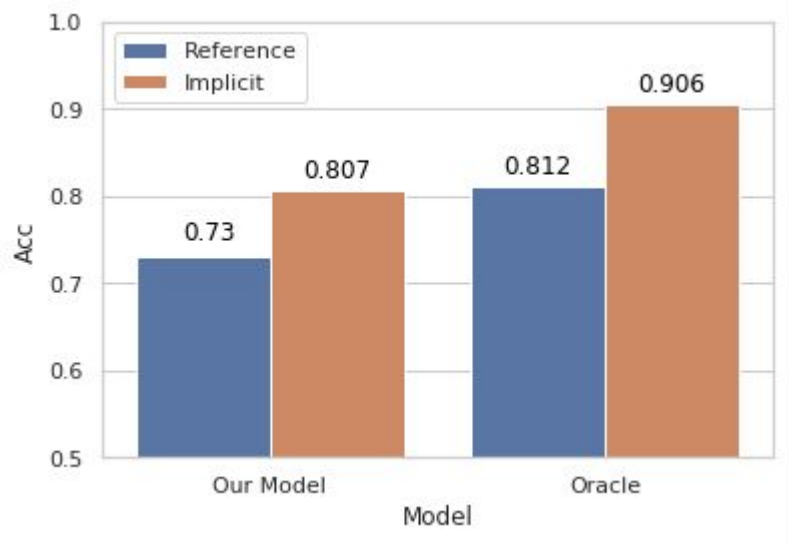
If we had gold separation between ***Reference*** and ***Implicit*** how well can we predict?



Fused Head - Analysis

Therefore,

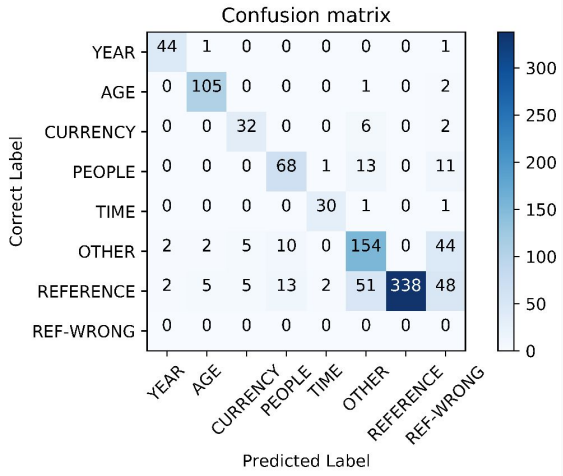
If we had gold separation between ***Reference*** and ***Implicit*** how well can we predict?



Fused Head - Analysis

Therefore,

- There appears to be a gap both in the binary separation of the classes
- as well in the correct label identification

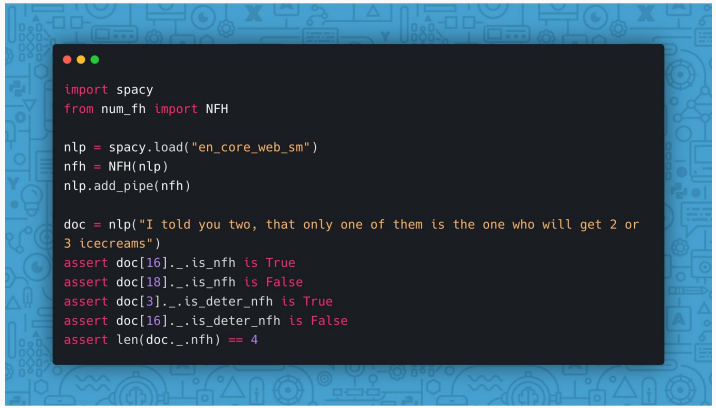


Summary

- Missing Elements are missing in today's NLP!
- They are prevalent and important
- We provide the first "large" scale dataset for handling one aspect of them: the Numeric Fused-Heads
- We provide a simple method for locating them
- and a baseline for resolving them

Thanks!

We're on spaCy Universe!



```python
import spacy
from num_fh import NFH

nlp = spacy.load("en_core_web_sm")
nfh = NFH(nlp)
nlp.add_pipe(nfh)

doc = nlp("I told you two, that only one of them is the one who will get 2 or
3 icecreams")
assert doc[16]._.is_nfh is True
assert doc[18]._.is_nfh is False
assert doc[3]._.is_deter_nfh is True
assert doc[16]._.is_deter_nfh is False
assert len(doc._.nfh) == 4
```

Demo