I will present:

- A new area of language understanding, that was not well studied
- A new task …
- a new dataset
- and results
We have a bunch of tasks and models
We have a bunch of tasks and models

This shirt was bought at **Grandpa Joe's** in downtown **Deep Learning**.
The NLP World Today

We have a bunch of tasks and models

Paul Allen was born on January 21, 1953, in Seattle, Washington, to Kenneth Sam Allen and Edna Faye Allen. Allen attended Lakeside School, a private school in Seattle, where he befriended Bill Gates, two years younger, with whom he shared an enthusiasm for computers. Paul and Bill used a teletype terminal at their high school, Lakeside, to develop their programming skills on several time-sharing computer systems.
We have a bunch of tasks and models

**Premise**

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

**Hypothesis**

The spacecraft has the ability to travel between planets.

**Summary**

It is very likely that the premise entails the hypothesis.

<table>
<thead>
<tr>
<th>Judgment</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entailment</td>
<td>89.4%</td>
</tr>
<tr>
<td>Contradiction</td>
<td>0.8%</td>
</tr>
<tr>
<td>Neutral</td>
<td>9.8%</td>
</tr>
</tbody>
</table>
And many others…
And many others…

And we do well on (some of) them, to some extent.
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?
So what about things which are not being said?

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

verb phrase ellipsis (VPE)
So what about things which are not being said?

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

verb phrase ellipsis (VPE)
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 leaves.”
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.”
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 leaves ___.”
The Missing Elements

Current models can be correct on these cases

parser works well!
Current models can be correct on these cases

But useless for concrete understanding of the situation
The Missing Elements

"I told you to do it for 14 years."

"I told you to do it since you were 14 years old."
Many things in language are not being stated explicitly.

These phenomena stand as obstacles for text understanding.

We call these: “The Missing Elements”
In this work, we focus on a specific type of missing elements: the **Numeric Fused-Head**.
The Numeric Fused-Head
“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
● “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.
Recovering the missing *element* can be useful for many downstream tasks

- Translation
- Information Extraction
- Named Entity Recognition
- etc...
Given text:

- Discover all the fused-heads
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
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
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
Method #1: Rule Based

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

95.5% F1
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

Method #2: Machine Learning

95.5% F1

96.6% F1
Fused Head - Identification Task

Now that we know how to identify,

How many of the Numbers in corpora are NFH?

- 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
● 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”
We can resolve with 4 deterministic rules.

NFH

35% 65%

The hard cases
Fused Head - Resolution Task Candidates

NFH

65%

The hard cases
Fused Head - Resolution Task Candidates

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

NFH

Sampled 10K examples

The hard cases
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**: ...
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)*
- Other *(please specify)*
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
Some hard cases:

- “Mexicans have **fifteen**, Jews have thirteen, rich girls have sweet sixteen…”

- “All her **communications** are to Minnesota **numbers**. There’s not **one** from California.”
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
How to combine the Reference decisions with the implicit ones?

Based on Lee et al. 17
Fused Head - Resolution Model

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

Based on Lee et al. 17
How to combine the *Reference* decisions with the *implicit* ones?
How to combine the **Reference** decisions with the *implicit* ones?

Based on Lee et al. 17
How to combine the *reference* decisions with the *Implicit* ones?
How to combine the *reference* decisions with the *Implicit* ones?
How to combine the *reference* decisions with the *Implicit* ones?

Fused Head - Resolution Model

- Anchor-Candidate representation
- bi-LSTM
- Word & characters embeddings
- Embedding Matrix
How to combine the *Reference* decisions with the *Implicit* ones?

```
token_i reference score

[Diagram showing tokens and scores]

need/one

```

```
token_j reference score

[Diagram showing tokens and scores]

gun/one

```

```
IMPLICIT_k class score

[Diagram showing tokens and scores]

YEAR/one
```
Fused Head - Resolution Results
Fused Head - Resolution Results

![Graph showing accuracy comparison between Most_Common and Linear models]

- **Most_Common** Model: 0.235
- **Linear** Model: 0.28
Fused Head - Resolution Results
Fused Head - Resolution Results
We are on spacy!

```
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_nf is True
assert doc[18].__is_nf is False
assert doc[3].__is_deter_nf is True
assert doc[16].__is_deter_nf is False
assert len(doc.__nf) == 4
```
We are on spacy!

&

We have a demo!
Input Document

He just always comes home by seven.

Possible Head Examples

- Reference
- Year
- Age
- Currency
- People
- Time

Submit

He just always comes home by seven TIME.
Fused Head - A Working Model

Input Document

Got a back-up weapon? Never had the need. Get one. It in your suit unless I tell you to take it out. Get yourself a Glock and lose that nickel - plated sissy pistol.

Possible Head Examples

- REFERENCE
- YEAR
- AGE
- CURRENCY
- PEOPLE
- TIME
- OTHER

Submit

Got a back-up weapon? Never had the need. Get one. It in your suit unless I tell you to take it out. Get yourself a Glock and lose that nickel - plated sissy pistol.
NFH Resolution - Analysis
How well do we separate between the Reference and Implicit classes?

86%
Therefore,

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

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

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

import spacy
from num_fl 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 (times)"
assert doc[10]...is_nfh is True
assert doc[18]...is_nfh is False
assert doc[10]...is_deter_nfh is True
assert doc[10]...is_deter_nfh is False
assert len(doc.nfh) == 4