Meaning Representation in Natural Language Tasks

Gabriel Stanovsky
My Research

Develop text-processing models which exhibit facets of human intelligence with benefits for users in real-life applications.
Grand Challenges in Natural Language Processing (NLP)

**Machine translation**
“the universal translator, invented in 2151, is used for deciphering unknown languages”

**Information retrieval**
“What’s the second largest star in this galaxy?”

**Automated assistants**
“I got one of those terrible headaches from lack of sleep. Can you give me something for it?”
Grand Challenges in Natural Language Processing (NLP)

Machine translation
“the universal translator, invented in 2151, is used for deciphering unknown languages”

NLP models need to capture the meaning behind our words and interact accordingly

Automated assistance
“I got one of those terrible headaches from lack of sleep. Can you give me something for it?”
Meaning

- The *information* conveyed in a natural language utterance
Meaning

- The **information** conveyed in a natural language utterance
- Abstracts over myriad of possible surface forms
Meaning

- The **information** conveyed in a natural language utterance

- Abstracts over myriad of possible surface forms
Meaning

- The **information** conveyed in a natural language utterance
- Abstracts over myriad of possible surface forms

Cows mainly eat grass, and can enjoy up to 75 pounds of it a day
Meaning

- The **information** conveyed in a natural language utterance
- Abstracts over myriad of possible surface forms

Cows mainly eat grass, and can enjoy up to 75 pounds of it a day

Grass is the major ingredient in bovine nutrition, reaching a maximum of 75 pounds consumed daily
Outline: Research Questions

Do NLP models capture meaning?
ACL 2019 🎉 Nominated for Best Paper
MRQA 2019 🎉 Best Paper award
EMNLP 2018

How can we build parsers for meaning?
EMNLP 2016a, EMNLP 2016b,
ACL 2016a, ACL 2016b, ACL 2017,
NAACL 2018, EMNLP2018a,
EMNLP2018b,
CoNLL 2019 🎉 Honorable mention

Can we integrate meaning into NLP?
ACL 2015, EACL 2017, SemEval 2017,
NAACL 2017, SemEval 2019

Models miss crucial meaning aspects
Gender bias in machine translation

Data collection
QA is an intuitive annotation format

Model design
Robust performance across domains

Real-world application
Adverse drug reactions on social media
Outline: Research Questions

Do NLP models capture meaning?
ACL 2019 🎉 Nominated for Best Paper
MRQA 2019 🎉 Best Paper award
EMNLP 2018

How can we build parsers for meaning?
EMNLP 2016a, EMNLP 2016b,
ACL 2016a, ACL 2016b, ACL 2017,
NAACL 2018, EMNLP2018a,
EMNLP2018b,
CoNLL 2019 Honorable mention

Can we integrate meaning into NLP?
ACL 2015, EACL 2017, SemEval 2017,
NAACL 2017, SemEval 2019

Models miss crucial meaning aspects
Gender bias in machine translation
Background: How should we represent text?
Background: How should we represent text?

Explicitly! We should define a formal representation of meaning.
Explicit Representations

- Many formalisms developed in linguistic literature
  - Dictating how *meaning* should be represented

Explicit Representations

- Many formalisms developed in linguistic literature
  - Dictating how *meaning* should be represented

[A] \((\forall x \text{fsa}(x)) \Rightarrow (\exists y \text{pda}(y) \land \text{equivalent}(x, y))\)

[B] \(\sim (\forall y (\exists x \text{fsa}(x) \Rightarrow \text{pda}(y) \land \text{equivalent}(x, y)))\)

[C] \(\forall x \exists y (\text{fsa}(x) \land \text{pda}(y) \land \text{equivalent}(x, y))\)

[D] \(\forall x \forall y (\text{fsa}(y) \land \text{pda}(x) \land \text{equivalent}(x, y))\)

Explicit Representations - Propositions

- Statements with one *predicate* (event) and arbitrary number of *arguments*
  - Bob called Mary → **called**: (Bob, Mary)
  - Bob gave a note to Mary → **gave**: (Bob, a note, Mary)

- Minimal units of information

- Form the backbone of explicit meaning representations
Explicit Representations

- **Pros**
  - Interpretable models
  - Independent progress on meaning representation
Explicit Representations

● **Pros**
  ○ Interpretable models
  ○ Independent progress on meaning representation

● **Cons**
  ○ Requires expensive expert annotations
  ○ Arbitrary - unclear that one representation is necessarily “correct”
**Background:** How should we represent text?

**Implicitly!** Models should *learn* a latent useful representation for an end-task
Implicit Representations

- Models find correlations between word representations and task label

Implicit Representations

- Models find correlations between word representations and task label

- Useful text representations found **implicitly** during the training process
  - Monolithic models trained on 100M of parameters over 1B words

Implicit Representations

● Models find correlations between word representations and task label

● Useful text representations found implicitly during the training process
  ○ Monolithic models trained on 100M of parameters over 1B words

Implicit Representations

- Cons
  - Opaque models
  - No control over the patterns they find useful in the data
Implicit Representations

● **Cons**
  ○ Opaque models
  ○ No control over the patterns they find useful in the data

● **Pros**
  ○ No need to commit on an explicit representation
  ○ Impressive gains on many NLP datasets
  ○ *Revolutionized the field*
Natural Language Processing in 2019

Implicit representation

Explicit representation
Natural Language Processing in 2019

Do **implicit** NLP models capture meaning?

ACL 2019 🎉 Nominated for Best Paper
MRQA 2019 🎉 Best Paper award
EMNLP 2018

*Implicit representation*

*Explicit representation*
Many Facets to Text Understanding

Factuality\(^1\)

Restrictiveness\(^2\)

Word sense disambiguation\(^3\)

Many Facets to Text Understanding

Factuality¹
Identify if an event happened
John forgot that he locked the door

Restrictiveness²
Detect if modifiers are required or elaborating
“The boy who stopped the flood.”
“Barack Obama, the former U.S. president.”

Coreference resolution
Implications on *gender bias* in machine translation

Word sense disambiguation³
Distinguishing *bat* from *bat*
Case study: Coreference in machine translation
Case study: **Coreference in machine translation**

The **doctor** asked the nurse to help her in the procedure.
Case study: Coreference in machine translation

The doctor asked the nurse to help her in the procedure.

- *ask for help*: (the doctor, the nurse, in the procedure)
- *is female*: (the doctor)
Case study: Coreference in machine translation

The doctor asked the nurse to help her in the procedure.

La doctora le pidió a la enfermera que la ayudara con el procedimiento.

- *ask for help*: (the doctor, the nurse, in the procedure)
- *is female*: (the doctor)
Case study: Coreference in machine translation

The **doctor** asked the nurse to help her in the procedure.

La **doctora** le pidió a la enfermera que la ayudara con el procedimiento.

- *ask for help*: (the doctor, the nurse, in the procedure)
- *is female*: (the doctor)

- Can models capture the meaning conveyed through coreference?
Case study: Coreference in machine translation

- Growing concern that models use bias to bypass meaning interpretation
Case study: **Coreference in machine translation**

- Growing concern that models use bias to bypass meaning interpretation

- E.g., translating all doctors as men regardless of context
Case study: Coreference in machine translation

- Growing concern that models use bias to bypass meaning interpretation
- E.g., translating all doctors as men regardless of context
- Will work well for most cases seen during training
Is machine translation gender biased?

Turkish is a gender neutral language. There is no "he" or "she" - everything is just "o". But look what happens when Google translates to English. Thread:
Is machine translation gender biased?

Alex Shams
@seyeyedrea

Turkish is a gender neutral language. There is no "he" or "she" - everything is just "o". But look what happens when Google translates to English. Thread:

Reducing gender bias in Google Translate

Over the course of this year, there's been an effort across Google to promote fairness and reduce bias in machine learning. Our latest development in this effort addresses gender bias by providing feminine and masculine translations for some gender neutral words in Google Translate.

Google Translate draws from hundreds of millions of already translated examples from the web. Historically, it has provided only one translation for a word, even if the translation could have either a feminine or masculine form. So when the newly produced or translated, it's trained on unlabeled gender biases that already existed. For example, it would knew masculine words like "strong" or "brave" and feminine for other words like "sweet" or "tender".

Now you'll get both a feminine and masculine translation for a single word - like "brave" - when translating from English into French, Italian, Portuguese or Spanish. We'll also get both translations when translating phrases and sentences from Turkish to English. For example, if you type "bi doktor" in Turkish, you'll now get "she is a doctor" and "he is a doctor" as the gender specific translations.
Is machine translation gender biased?

Alex Shams
@seyyedreza

Turkish is a gender neutral language. There is no "he" or "she" - everything is just "o". But look what happens when Google translates to English. Thread:

---

Reducing gender bias in Google Translate

Over the course of this year, there's been an effort across Google to promote fairness and reduce bias in machine learning. Our latest development in this effort addresses gender bias by providing feminine and masculine translations for some gender neutral words in the Google translate website.

Google Translate learns from hundreds of millions of already translated examples from the web. Historically, it has treated only one translation for a query, even if the translation could have either feminine or masculine. So when the newly produced or translated, it literally replaced gender boxes that already existed. For example, it would push masculine words like "strong" or "brave" and feminine for other words, like "purple" or "beautiful".

Now you'll get both a feminine and masculine translation for a single word - like "surgeon" - when translating from English into French, Italian, Portuguese or Spanish. It's also got both masculine and feminine translations for queries from Turkish to English. For example, if you type "kiz doktor" in Turkish, you'll now get "she is a doctor" and "he is a doctor" and the gender specific translations.

---

Gretchen McCulloch @GretchenMAdC · 6 Dec 2018
I actually like this a bit.
Introduces the human back into the equation by acknowledging ambiguity, letting us decide which translation fits a particular circumstance better.

---

Alyonko
@lowii
(almost working...)
(but really, a very hard and interesting problem to solve.)
Evaluating Coreference Translation: **Challenges**

- Gender bias in machine translation was noticed anecdotally
  - **Open question:** how to *quantitatively* measure gender translation?
Evaluating Coreference Translation: **Challenges**

- Gender bias in machine translation was noticed anecdotally
  - **Open question**: how to *quantitatively* measure gender translation?

- Requires reference translations in various languages and models
  - To reach more general conclusions
Evaluating Coreference Translation: Challenges

- Gender bias in machine translation was noticed anecdotally
  - Open question: how to quantitatively measure gender translation?

- Requires reference translations in various languages and models
  - To reach more general conclusions

- Gender can be unspecified
  - The doctor had very good news
Evaluating Coreference in Machine Translation

Challenge

How to evaluate gender translation across different models & languages?
Evaluating Coreference in Machine Translation

- **Input:**
  - Machine translation model: \( M \)
  - Target language with grammatical gender: \( L \)

**Challenge**

How to evaluate gender translation across different models & languages?
Evaluating Coreference in Machine Translation

- **Input:**
  - Machine translation model: $M$
  - Target language with grammatical gender: $L$

- **Output:**
  - Accuracy score $\in [0, 100]$

*How well does $M$ translates gender information from English to $L$?*

**Challenge**

*How to evaluate gender translation across different models & languages?*
English Source Texts

- Winogender\textsuperscript{1} & WinoBias\textsuperscript{2} - bias in \textit{coreference resolution}

  The \textcolor{red}{doctor} asked the nurse to help her in the procedure.

  The \textcolor{red}{doctor} asked the nurse to help him in the procedure.

English Source Texts

● Winogender\textsuperscript{1} \& WinoBias\textsuperscript{2} - bias in \textit{coreference resolution}

The \textcolor{red}{doctor} asked the nurse to help her in the procedure.

The \textcolor{red}{doctor} asked the nurse to help him in the procedure.

● Equally split between stereotypical and non-stereotypical role assignments
  ○ Based on U.S. labor statistics

English Source Texts

- Winogender\textsuperscript{1} & WinoBias\textsuperscript{2} - bias in coreference resolution

  - The \textcolor{red}{doctor} asked the nurse to help her in the procedure.
  
  - The \textcolor{red}{doctor} asked the nurse to help him in the procedure.

- Equally split between stereotypical and non-stereotypical role assignments
  - Based on U.S. labor statistics

- Gender-role assignments are specified (+90% human agreement)

Methodology: Automatic evaluation of gender accuracy

**Input:** MT model + target language

**Output:** Gender accuracy
Methodology: Automatic evaluation of gender accuracy

1. **Translate** the coreference bias datasets

   **Input:** MT model + target language
   **Output:** Gender accuracy

   The *doctor* asked the nurse to help her in the procedure.
Methodology: Automatic evaluation of gender accuracy

1. **Translate** the coreference bias datasets

   **Input:** MT model + target language
   **Output:** Gender accuracy

   - The *doctor* asked the nurse to help her in the procedure.
   - La doctora le pidió a la enfermera que le ayudara con el procedimiento.
Methodology: Automatic evaluation of gender accuracy

1. **Translate** the coreference bias datasets

2. **Align** between source and target

**Input:** MT model + target language

**Output:** Gender accuracy

The doctor asked the nurse to help her in the procedure.

La doctora le pidió a la enfermera que le ayudara con el procedimiento.
Methodology: Automatic evaluation of gender accuracy

1. **Translate** the coreference bias datasets

2. **Align** between source and target

3. **Identify** gender in target language

---

The *doctor* asked the nurse to help her in the procedure.

La *doctora* le pidió a la enfermera que le ayudara con el procedimiento.

**Input:** MT model + target language

**Output:** Gender accuracy
Methodology: Automatic evaluation of gender accuracy

1. **Translate** the coreference bias datasets

2. **Align** between source and target

3. **Identify** gender in target language

---

**Input:** MT model + target language  
**Output:** Gender accuracy

The doctor asked the nurse to help her in the procedure.

El doctor le pidió a la enfermera que le ayudara con el procedimiento.
Methodology: Automatic evaluation of gender accuracy

1. **Translate** the coreference bias datasets

2. **Align** between source and target

3. **Identify** gender in target language

**Input:** MT model + target language

**Output:** Gender accuracy

Quality estimated at > 90%

---

The doctor asked the nurse to help her in the procedure.

El doctor le pidió a la enfermera que le ayudara con el procedimiento.
The doctor asked the nurse to help him in the procedure.
The doctor asked the nurse to help her in the procedure.
Google Translate

Results

<table>
<thead>
<tr>
<th>Language</th>
<th>Stereotypical</th>
<th>Non-Stereotypical</th>
<th>Gender bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>French</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ukrainian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hebrew</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arabic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Human performance

Random
Results

- Translation models struggle with non-stereotypical roles
Results

- Translation models struggle with non-stereotypical roles

Our metric can evaluate future progress on gender bias in machine translation
NLP models do not capture important facets of meaning
NLP models do not capture important facets of meaning

Instead, they find spurious patterns in the data
  ○ Leading to the biased performance we’ve seen
  ○ Biased performance in question answering, inference, and more
Do NLP models capture meaning?

- NLP models do not capture important facets of meaning

- Instead, they find spurious patterns in the data
  - Leading to the biased performance we’ve seen
  - Biased performance in question answering, inference, and more
Open Questions

- Do models fail at capturing meaning because of *architecture* or *data*?
Open Questions

● Do models fail at capturing meaning because of architecture or data?

● Is there a dataset that could force models to learn meaningful patterns?
  ○ E.g., equally distributed between genders
Open Questions

- Do models fail at capturing meaning because of architecture or data?
- Is there a dataset that could force models to learn meaningful patterns?
  - E.g., equally distributed between genders
- Current data augmentation efforts find models are stubbornly biased\textsuperscript{[1,2,3]}

Meaning Representation in Neural Networks

**Best of both worlds**: models over meaningful *explicit* representations leveraging strong *implicit* architectures
Research Questions

Weaknesses in state of the art
ACL 2019 🎉 Nominated for Best Paper
MRQA 2019 🎉 Best Paper award
EMNLP 2018

How can we build parsers for meaning?
EMNLP 2016a, EMNLP 2016b,
ACL 2016a, ACL 2016b, ACL 2017,
NAACL 2018, EMNLP2018a,
EMNLP2018b, CoNLL 2019 🎉 Honorable mention

Can we integrate meaning into NLP?
ACL 2015, EACL 2017, SemEval 2017,
NAACL 2017, SemEval 2019

Data collection
QA is an intuitive annotation format

Model design
Robust performance across domains

Real-world application
Adverse drug reactions on social media
Open Information Extraction (Open IE)

- Extracts **stand-alone propositions** from text
  - *Barack Obama, a former U.S president, was born in Hawaii*
  - (Barack Obama, **was born in**, Hawaii)
  - (a former U.S president, **was born in**, Hawaii)
  - (Barack Obama, **is**, a former U.S. president)

Banko et al, 2007
Open Information Extraction (Open IE)

- Extracts **stand-alone propositions** from text
  - *Barack Obama, a former U.S president, was born in Hawaii*
    - Barack Obama, **was born in**, Hawaii
    - (a former U.S president, **was born in**, Hawaii)
    - (Barack Obama, **is**, a former U.S. president)
  
  - *Obama and Bush were born in America*
    - Obama, **born in**, America
    - (Bush, **born in**, America)

Banko et al, 2007
Mr. Pratt, head of marketing, thinks that lower wine prices have come about because producers don’t like it when hit wines dramatically increase in price.
Mr. Pratt, head of marketing, thinks that lower wine prices have come about because producers don’t like it when hit wines dramatically increase in price.

1. Mr Pratt is the head of marketing
Mr. Pratt, head of marketing, thinks that lower wine prices have come about because producers don’t like it when hit wines dramatically increase in price.

1. Mr Pratt is the head of marketing
2. lower wine prices have come about
Mr. Pratt, head of marketing, thinks that lower wine prices have come about because producers don’t like it when hit wines dramatically increase in price.

1. Mr Pratt is the head of marketing
2. lower wine prices have come about
3. hit wines dramatically increase in price
Mr. Pratt, head of marketing, thinks that lower wine prices have come about because producers don’t like it when hit wines dramatically increase in price.

1. Mr Pratt is the head of marketing
2. lower wine prices have come about
3. hit wines dramatically increase in price
4. producers don’t like (3)
Mr. Pratt, head of marketing, thinks that lower wine prices have come about because producers don’t like it when hit wines dramatically increase in price.

1. Mr Pratt is the head of marketing
2. lower wine prices have come about
3. hit wines dramatically increase in price
4. producers don’t like (3)
5. (2) happens because of (4)
Mr. Pratt, head of marketing, thinks that lower wine prices have come about because producers don’t like it when hit wines dramatically increase in price.
Parsers for Meaning Representation

- **Goal** - Build Open Information Extraction parsers from raw text

- **Challenges**
  - **Obtaining data for the task**
    - Expensive and non-trivial manual annotation
  - **Designing a parser**
    - Which works well for real-world texts
Parsers for Meaning Representation

● **Goal** - Build Open Information Extraction parsers from raw text

● **Challenges**
  ○ **Obtaining data for the task**
    Expensive and non-trivial manual annotation

  ○ **Designing a parser**
    Which works well for real-world texts
Data Collection: Challenges

- Direct annotation requires linguistic expertise
  - Formal definitions for predicates and arguments
Data Collection: Challenges

- Direct annotation requires linguistic expertise
  - Formal definitions for predicates and arguments

- Existing datasets annotated only hundreds of sentences
  - Conflicting guidelines between different works
  - Do not support training
QA is an intuitive interface for data collection

- QA pairs can be **deterministically converted to Open IE propositions**

Where was **Obama** born? **Hawaii**

(Obama, was born in, Hawaii)

Converted based on question template

EMNLP 2016
QA is an intuitive interface for data collection

- QA pairs can be **deterministically converted to Open IE propositions**

  Where was **Obama** born? **Hawaii**
  Who was born in **Hawaii**? **Obama**

  
  \[(\text{Obama}, \text{was born in}, \text{Hawaii})\]
“Mr. Pratt, head of marketing, thinks that lower wine prices have come about because producers don’t like it when hit wines dramatically increase in price.”

- Who is the head of marketing? Mr. Pratt
- What have come about? lower wine prices
- What increased in price? hit wines
- ....
“Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29.”

- Who will join the board? Pierre Vinken
- What will he join the board as? Nonexecutive director
- When will Vinken join the board? Nov. 29

Intuitive interface for non-expert annotation of meaning!
QA as an interface for data collection

- Yields the largest supervised dataset for the Open Information Extraction

Our dataset enables the development of the first supervised models for Open IE.

Open IE: Challenges

● Obtaining data for the task
  ○ Expensive and non-trivial for manual annotation

● Building an Open IE parser
  ○ Which works well for real-world texts
Supervised Open IE Parser

- **Approach:** word-level tagging task (Beginning, Inside, Outside)
Supervised Open IE Parser

John jumped and Mary ran
Supervised Open IE Parser

Predicate Identification
finding verbs in the sentence

John jumped and Mary ran
Supervised Open IE Parser

(John; jumped)

Argument1

Predicate

Outside

Outside

Outside

Softmax

Contextualized representation

Forward & backward LSTM

John

jumps

and

Mary

ran
Supervised Open IE Parser

Contextualized representation

Softmax

Argument1  Predicate  Outside  Outside  Outside

Forward & backward LSTM

Predicate features concatenated to all words

(John; \textit{jumped})

John  \textit{jumped}  and  Mary  ran
Supervised Open IE Parser

\[(\text{John}; \text{jumped})\]

- Argument1
- Predicate
- Outside
- Outside
- Outside
- Softmax
- Contextualized representation
- Forward & backward LSTM

John \(\rightarrow\) jumped \(\rightarrow\) and \(\rightarrow\) Mary \(\rightarrow\) ran
Supervised Open IE Parser

(John; jumped)

Argument1

Predicate

Outside

Outside

Outside

Softmax

Contextualized representation

Forward & backward LSTM

John

jumped

and

Mary

ran

NAACL 2018b
Supervised Open IE Parser

Contextualized representation

Softmax

Confidence (John; jumped) = \( P(\text{word confidence}) \)

NAACL 2018b
Evaluation - Open IE

High confidence threshold → Accurate propositions, relatively few of them

Low confidence threshold → More propositions, relatively less accurate

QA data

Precision

Recall
Evaluation - Open IE

Our approach presents a **favorable** precision-recall tradeoff on our data.
We generalize well to datasets unseen during training
Evaluation - Open IE

4 points over state of the art
Supervised Parser - Adaptation

- Integrated into the popular AllenNLP framework
  - Online demo receives thousands of requests per month

- Used by researchers in academia and tech (e.g., plasticity.ai, Diffbot)
Research Questions

Weaknesses in state of the art
ACL 2019 🎉 Nominated for Best Paper
MRQA 2019 🎉 Best Paper award
EMNLP 2018 🎉

Building meaning representations
EMNLP 2016a, EMNLP 2016b,
ACL 2016a, ACL 2016b, ACL 2017,
NAACL 2018, EMNLP2018a,
EMNLP2018b,
CoNLL 2019 🎉 Honorable mention

Can we integrate meaning into NLP?
ACL 2015, EACL 2017, SemEval 2017,
NAACL 2017, SemEval 2019

Real-world application
Adverse drug reactions on social media
Adverse Drug Reaction on Social Media

Adverse Drug Reaction (ADR)

**Unwanted reaction** clearly associated with the intake of a drug
Adverse Drug Reaction on Social Media

Adverse Drug Reaction (ADR)

Unwanted reaction clearly associated with the intake of a drug

I stopped taking Ambien after a week, it gave me a terrible headache!
Adverse Drug Reaction on Social Media

Adverse Drug Reaction (ADR)

Unwanted reaction clearly associated with the intake of a drug

I stopped taking Ambien after a week, it gave me a terrible headache!

- Discover unknown side-effects
- Monitor drug reaction over time
- Respond to patient complaints
Challenges

• Context dependent
  ○ *Ambien gave me terrible headaches*
  ○ *Ambien made my terrible headaches go away*

• Colloquial
  ○ *been having a hard time getting some Z’s*
Approach

● Recurrent neural network over propositions extracted from text

● Beneficial for the small amounts of data
  ○ Train: 5723 instances
  ○ Test: 1874 instances
Model

I recently switched to Advil, Ambien made me so dizzy!
Model

Open IE:

I switched to Advil

Ambien made me feel dizzy

I recently switched to Advil, Ambien made me so dizzy!
Model

Open IE:

I switched to Advil

Ambien made me feel dizzy

I recently switched to Advil, Ambien made me so dizzy!
Results and Analysis

- Mention Oracle marks all gold mentions, regardless of context
  - Errors on 45% of instances → Context matters!
Results and Analysis

- Oracle marks all gold mentions, regardless of context
  - Errs on 45% of instances → Context matters!
- Recurrent neural networks achieve ~80% F1
Oracle marks all gold mentions, regardless of context
  ○ Errs on 45% of instances → Context matters!
Recurrent neural networks achieve ~80% F1
Meaning representation provides 8% absolute improvement
Conclusions

- Implicit representations lead to biased performance
Conclusions

• Implicit representations lead to biased performance

• Middle ground: Implicit models with explicit meaning representation
Conclusions

- Implicit representations lead to biased performance

- **Middle ground**: Implicit models with explicit meaning representation

- Useful in real-world application
Conclusion: My contributions

Do NLP models capture meaning?
- ACL 2019: Nominated for Best Paper
- MRQA 2019: Best Paper Award
- EMNLP 2018

How can we build parsers for meaning?

Can we integrate meaning into NLP?

Open IE model

QA reasoning

First German Open IE

Adverse drug reactions

Math QA

Reading comprehension

Factuality detection

Machine translation

QA evaluation

Word polysemy

QA Active learning

Open IE dataset

Paraphrase datasets

Document representation
Future Work
Future Work: **Interactive Semantics**

- Current NLP setting assume single-input single-output
Future Work: Interactive Semantics

- Current NLP setting assume single-input single-output

- Human interaction is often iterative
Future Work: **Interactive Semantics**

- Current NLP setting assume single-input single-output

- Human interaction is often iterative

- Interactivity allows *models and humans iterate to reach a solution*
  - Will benefit from an explicit meaning representation
Future Work: Multilingual Meaning Bank

- Meaning representations are constructed almost exclusively in English
  - Linguistic theory needs to be adapted
  - Expert annotation is expensive
Future Work: Multilingual Meaning Bank

● Meaning representations are constructed almost exclusively in English
  ○ Linguistic theory needs to be adapted
  ○ Expert annotation is expensive

● A multilingual representation will facilitate:
  ○ Semantically coherent machine translation
  ○ NLP applications in low-resource languages
Future Work: **Multilingual Meaning Bank**

- Meaning representations are constructed almost exclusively in English
  - Linguistic theory needs to be adapted
  - Expert annotation is expensive

- A multilingual representation will facilitate:
  - Semantically coherent machine translation
  - NLP applications in **low-resource languages**

- QA is appealing for multilingual representation
  - Intuitive for non-expert annotation
  - Hebrew as an intuitive first language
Future Work: **NLP to inform decision making**

- NLP technology is ripe to extract large-scale aggregates

---

**Number of CS authors**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Number of MEDLINE authors**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Featured in** The New York Times *SCIENCE* *FORTUNE*
Future Work: **NLP to inform decision making**

- NLP technology is ripe to extract large-scale aggregates

**Can aid in debates on other issues such as immigration or gun-violence**
  - Extract gun assault trends, how weapons were obtained from news articles

---

**Featured in**

- The New York Times
- SCIENCE
- FORTUNE

---

**In submission, 2019**

---

**Number of CS authors**

- Male and female authors over time (1970-2010)

**Number of MEDLINE authors**

- Male and female authors over time (1970-2010)
Thanks for listening!