Sentence Representation Learning:
Theories of Semantic Representation

Ellie Pavlick
Department of Computer Science, Brown University
General Purpose Sentence Representation Learning Team, JSALT
Long Term Goal...
Very Long Term Goal...
If there were machines which bore a resemblance to our bodies and imitated our actions as closely as possible for all practical purposes, we should still have two very certain means of recognizing that they were not real men. The first is that they could never use words, or put together signs, as we do in order to declare our thoughts to others. For we can certainly conceive of a machine so constructed that it utters words, and even utters words that correspond to bodily actions causing a change in its organs...But it is not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as the dullest of men can do... 

as quoted in https://plato.stanford.edu/entries/turing-test/
I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about $10^9$, to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning. … I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.

as quoted in https://plato.stanford.edu/entries/turing-test/
The Media (c. now)

MIT's artificial intelligence passes key Turing test

A computer has passed the Turing test for humanity - should we be worried?

What is the Turing test? And are we all doomed now?

Computer Posing as Teenager Achieves Artificial-Intelligence Milestone

Computer AI passes Turing test in 'world first'
Passing the Turing Test...?
Passing the Turing Test...?

Grover Cleveland

John Tyler, whose wife Letitia had died a year and nine months before, wed Julia Gardiner in New York City on June 25, 1844; bachelor Grover Cleveland (the only president to be married in a White House ceremony) wed Frances Folsom, the daughter of a former law partner, on June 2, 1886; Woodrow Wilson, whose wife Ellen ...

What are some interesting facts about presidents and first ladies...
https://www.whitehousehistory.org/.../what-are-some-interesting-facts-about-presidents-f...
Passing the Turing Test...?
Passing the Turing Test...?

Grover Cleveland

John Tyler, whose wife Letitia had died a year and nine months before, wed Julia Gardiner in New York City on June 25, 1844; bachelor Grover Cleveland (the only president to be married in a White House ceremony) wed Frances Folsom, the daughter of a former law partner, on June 2, 1886; Woodrow Wilson, whose wife Ellen ...

What are some interesting facts about presidents and first ladies ...
https://www.whitehousehistory.org/.../what-are-some-interesting-facts-about-presidents-f...

Message
To: Mom
I'm gay
Cancel
Send

@UnkindledGurg @PooWithEyes chill
im a nice person! i just hate everybody
Passing the Turing Test...?

Syntax and Structure
Passing the Turing Test...?

Common Sense and Social Awareness
Passing the Turing Test...?

Context, Knowledge, and "State"
Representing language is hard.
Crash Course: Non-Computational Linguistics
Crash Course: Non-Computational Linguistics

Pencil-and-Paper NLP
Model Theory
Model Theory

\[ x > 17 \]
Model Theory

\[ x > 17 \]

\[ 'x' \colon= 32 \]
Model Theory

\[ x > 17 \]

\[ \text{}`x` := 14 \]
Model Theory

$x > 17$

'x' := 378
'x' := 18
'x' := 14
'x' := 17
'x' := 32
Model Theory

Language

\[ x > 17 \]

\[ x' := 378 \]

\[ x' := 18 \]

\[ x' := 17 \]

\[ x' := 14 \]

\[ x' := 32 \]
Model Theory

Language

\[ x > 17 \]

The World

\[ 'x' := 378 \]
\[ 'x' := 18 \]
\[ 'x' := 17 \]
\[ 'x' := 14 \]
\[ 'x' := 32 \]
Model Theory

Language

\[ x > 17 \]

The World (TBD)
Model Theory

Language

\[ x > y \]
\[ y > z \]
\[ \frac{}{} \]
\[ x > z \]

The World (TBD)
Model Theory

Language

\[ x > y \]

\[ y > z \]

\[ x > z \]

Variables (to be grounded)

The World (TBD)
Model Theory

Language

\[
\begin{align*}
  x & > y \\
  y & > z \quad \text{(defined)} \\
  \hline \\
  x & > z
\end{align*}
\]

The World (TBD)
Model Theory

Entailment

$x > y$

$z > w$

\[ x > w \]

The World (TBD)
Model Theory

Entailment

\[ x > y \]
\[ z > w \]
\[ \frac{z > w}{x > w} \]

\[ x = 10 \quad y = 5 \quad z = 11 \quad w = 8 \]
Model Theory

Entailment

\[ x > y \]
\[ z > w \]

\[ \frac{x > w}{x > w} \]

\[ x = 10 \quad y = 5 \quad z = 11 \quad w = 8 \]
Model Theory

Entailment

\[ x > y \]
\[ z > w \]

\[ x > w \]

\[ x = 10 \quad y = 5 \quad z = 11 \quad w = 8 \]
Model Theory

Entailment

\[
\begin{align*}
x & > y \\
z & > w \\
\hline
x & > w
\end{align*}
\]

\[
x = 10 \quad y = 5 \quad z = 11 \quad w = 8
\]
Model Theory

Entailment

\[
x > y
\]
\[
z > w
\]
\[
\underline{x > w}
\]

✔

\[
x = 10 \quad y = 5 \quad z = 11 \quad w = 8
\]
Model Theory

Entailment

\[ x > y \]
\[ z > w \]
\[ \underline{x > w} \]

\[ x = 10 \quad y = 5 \quad z = 12 \quad \omega = 11 \]
Model Theory

A premise (p) entails a hypothesis (h) iff, in every possible world in which p is true, h is also true.

\[ \forall \mathcal{I}((\mathcal{I} \models p) \Rightarrow (\mathcal{I} \models h)) \]
Formal Semantics
There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed I consider it possible to comprehend the syntax and semantics of both kinds of languages with a single natural and mathematically precise theory.

(Richard Montague)
Formal Semantics

The basic aim of semantics is to characterize the notion of a true sentence (under a given interpretation) and of entailment.

(Richard Montague)
Formal Semantics

the notion of a true sentence

Broca is a bird
Formal Semantics

the notion of a true sentence

Broca is a bird

Broca
Formal Semantics

the notion of a true sentence

Broca is a bird

Broca
Formal Semantics

the notion of a entailment

No birds are gray
Broca is a bird

Broca is gray
Formal Semantics

the notion of a entailment

All birds are gray
Broca is a bird

Broca is gray
Formal Semantics

the notion of a entailment

All birds are gray
Broca is a bird
Broca is gray

Entities
Formal Semantics

the notion of an entailment

All **birds** are **gray**
Broca is a **bird**

Broca is **gray**
Formal Semantics

the notion of a entailment

**All** birds are gray

Broca is a bird

---

Broca is gray

Higher-Order Relations
Formal Semantics
Formal Semantics

Entities

⟦Broca⟧
Formal Semantics

Entities

⟦Broca⟧
Formal Semantics

Predicates

\[[\text{Broca}]\]

\[[\text{bird}]\]
Formal Semantics

Predicates

[Broca]

[bird]

[cat]
Formal Semantics

Predicates

Broca is a bird

\[ \forall x ([\text{Broca}](x) \implies [\text{bird}](x)) \]
Formal Semantics

Predicates

Broca is a bird

$\forall x (\llbracket Broca \rrbracket(x) \Rightarrow \llbracket bird \rrbracket(x))$

Takes entity as argument.
Returns true if $x$ is “Broca”.
Formal Semantics

Predicates

\[ \forall x (\llbracket \text{Broca} \rrbracket (x) \Rightarrow \llbracket \text{bird} \rrbracket (x)) \]

Takes entity as argument.
Returns true if x is an element of the set referred to by “bird”.

Broca is a bird
Formal Semantics

Predicates

[[Broca]]

[[bird]]

[[gray]]
Formal Semantics

\[ [\text{gray}] (\text{bird}) = [\text{gray bird}] \]
Formal Semantics

\[ [\text{gray bird}] (x) = ( [\text{bird}] (x) \land [\text{gray}] (x) ) \]

Takes set as argument. Returns true if \( x \) is an element of the set referred to by “bird”. 
Formal Semantics

$\text{[gray]} (\text{[bird]}) = \text{[gray bird]}$

Broca is a gray bird

entails

Broca is a bird
Formal Semantics

Broca is a gray bird
entails
Broca is a bird
Formal Semantics

$\forall x ([\text{Broca}] (x) \Rightarrow [\text{gray bird}] (x))$

entails

Broca is a bird
Formal Semantics

\( \forall x (\text{⟦Broca⟧}(x) \Rightarrow \text{⟦gray bird⟧}(x)) \)

⇒

Broca is a bird
Formal Semantics

\[ \forall x ([\text{Broca}] (x) \Rightarrow [\text{gray bird}] (x)) \]

\[ \Rightarrow \]

\[ \forall x ([\text{Broca}] (x) \Rightarrow [\text{bird}] (x)) \]
∀x([Broca](x) ⇒ ([bird](x) ∧ [gray](x)))
⇒
∀x([Broca](x) ⇒ [bird](x))
Formal Semantics

∀x([Broca](x) ⇒ [bird](x))

⇒

∀x([Broca](x) ⇒ [bird](x))

✔
Formal Semantics

All gray birds are birds
Broca is a gray bird
Broca is a bird
Formal Semantics

All gray birds are birds
Broca is a gray bird
Broca is a bird

Higher-Order Relations
Formal Semantics

\textbf{All} gray birds are birds

\[
\llbracket \text{all} \rrbracket (x) = \lambda P \lambda Q \forall x (P(x) \Rightarrow Q(x))
\]
Formal Semantics

All gray birds are birds

\[ \text{all}(x) = \lambda P \lambda Q \forall x (P(x) \Rightarrow Q(x)) \]

Takes arbitrary predicates (P and Q) as arguments. Returns true if Q is true whenever P is true.
Formal Semantics

All gray birds are birds

\[ \text{⟦all⟧}(x)(\text{gray\_birds})(\text{birds}) = \forall x(\text{gray\_bid}(x) \Rightarrow \text{bird}(x)) \]

Takes arbitrary predicates (P and Q) as arguments. Returns true if Q is true whenever P is true.
Formal Semantics:
Takeaways

• Language is like logic—we can ground symbols to the world, but we can also reason abstractly using only the ungrounded symbols.

• The world can be represented as sets of entities.

• Word meanings are represented in terms of entities, sets, or functions which operate on entities/sets.

• Words have types (nouns, verbs, adjectives) which determine their representation (i.e. nouns refer to sets, adjectives to functions on sets).
Formal Semantics: Takeaways

• Language is like logic—we can ground symbols to the world, but we can also reason abstractly using only the ungrounded symbols
Formal Semantics: Takeaways

• Language is like logic—we can ground symbols to the world, but we can also reason abstractly using only the ungrounded symbols.

• The world can be represented as sets of entities.
Formal Semantics: Takeaways

• Language is like logic—we can ground symbols to the world, but we can also reason abstractly using only the ungrounded symbols.

• The world can be represented as sets of entities.

• Word meanings are represented in terms of entities, sets, or functions which operate on entities/sets.
Formal Semantics: Takeaways

- Language is like logic—we can **ground** symbols to the world, but we can also **reason abstractly** using only the ungrounded symbols.

- The world can be represented as sets of entities.

- Word meanings are represented in terms of entities, sets, or functions which operate on entities/sets.

- Words have types (nouns, verbs, adjectives) which determine their representation (i.e. nouns refer to sets, adjectives to functions on sets).
Distributional Semantics
Distributional Semantics

The Distributional Hypothesis:

You shall know a word by the company it keeps!
(Firth, 1957)
Distributional Semantics

The Distributional Hypothesis:

The meaning of a word is determined by the contexts where it is used.
BOW Vector Space Models

The domestic cat is a small, typically furry, carnivorous mammal.

Your cat's online owners manual, featuring articles about breed information, cat selection, training, grooming and care for cats and kittens.

Wish you had a secret decoder guide to cat behavior and cat language? Here's a primer to things your cat wishes you understood.


Welcome to the new WebMD Cat Health Center. WebMD veterinary experts provide comprehensive information about cat health care, offer nutrition and feeding …

Yes, they're independent and willful, but felines can be taught certain behaviors—to the benefit of both cat and human.
BOW Vector Space Models

The domestic cat is a small, typically furry, carnivorous mammal.

Your cat's online owners manual, featuring articles about breed information, cat selection, training, grooming and care for cats and kittens.

Wish you had a secret decoder guide to cat behavior and cat language? Here's a primer to things your cat wishes you understood.


Welcome to the new WebMD Cat Health Center. WebMD veterinary experts provide comprehensive information about cat health care, offer nutrition and feeding …

Yes, they're independent and willful, but felines can be taught certain behaviors—to the benefit of both cat and human.
BOW Vector Space Models

The domestic **cat** is a small, typically furry, carnivorous mammal.

Your **cat's** online owners manual, featuring articles about breed information, **cat** selection, training, grooming and care for **cats** and kittens.

Wish you had a secret decoder guide to **cat** behavior and **cat** language? Here's a primer to things your **cat** wishes you understood.

"The **cat** does not offer services," William Burroughs wrote. "The **cat** offers itself." But it does so with unapologetic ambivalence.

Welcome to the new WebMD **Cat** Health Center. WebMD veterinary experts provide comprehensive information about **cat** health care, offer nutrition and feeding …

Yes, they're independent and willful, but felines can be taught certain behaviors—to the benefit of both **cat** and human.
BOW Vector Space Models

The domestic **cat** is a small, typically furry, carnivorous mammal.

Your **cat's** online owners manual, featuring articles about breed information, **cat** selection, training, grooming and care for **cats** and kittens.

Wish you had a secret decoder guide to **cat** behavior and **cat** language? Here's a primer to things your **cat** wishes you understood.

"The **cat** does not offer services," William Burroughs wrote. "The **cat** offers itself." But it does so with unapologetic ambivalence.

Welcome to the new WebMD **Cat** Health Center. WebMD veterinary experts provide comprehensive information about **cat** health care, offer nutrition and feeding …

Yes, they're independent and willful, but felines can be taught certain behaviors—to the benefit of both **cat** and human.

<table>
<thead>
<tr>
<th></th>
<th>the</th>
<th>domestic</th>
<th>is</th>
<th>a</th>
<th>your</th>
<th>online</th>
<th>owners</th>
<th>breed</th>
<th>information</th>
<th>selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>1000</td>
<td>40</td>
<td>500</td>
<td>700</td>
<td>400</td>
<td>3</td>
<td>80</td>
<td>100</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>dog</td>
<td>1050</td>
<td>50</td>
<td>400</td>
<td>950</td>
<td>500</td>
<td>1</td>
<td>105</td>
<td>160</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>lychee</td>
<td>2000</td>
<td>2</td>
<td>500</td>
<td>1000</td>
<td>25</td>
<td>50</td>
<td>2</td>
<td>3</td>
<td>45</td>
<td>700</td>
</tr>
</tbody>
</table>
BOW Vector Space Models

cat

dog

lychee
Skip-Gram Model (word2vec)
Skip-Gram Model
(word2vec)

<table>
<thead>
<tr>
<th>cat</th>
<th>the</th>
<th>domestic</th>
<th>is</th>
<th>a</th>
<th>your</th>
<th>online</th>
<th>owners</th>
<th>breed</th>
<th>informa-tion</th>
<th>selection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1000</td>
<td>40</td>
<td>500</td>
<td>700</td>
<td>400</td>
<td>3</td>
<td>80</td>
<td>100</td>
<td>15</td>
<td>6</td>
</tr>
</tbody>
</table>
Skip-Gram Model (word2vec)

The skip-gram model uses a neural network to predict the context of a word given its embedding. In this example, the word 'cat' is predicted given the context of 'the', 'domestic', 'is', 'a', 'your', 'online', 'owners', 'breed', 'information', and 'selection'.
Skip-Gram Model (word2vec)

Representing Context

The Distributional Hypothesis:

The meaning of a word is determined by the contexts where it is used.
Representing Context

The domestic cat is a small, typically furry, carnivorous mammal.

Your cat's online owners manual, featuring articles about breed information, cat selection, training, grooming and care for cats and kittens.

Wish you had a secret decoder guide to cat behavior and cat language? Here's a primer to things your cat wishes you understood.


Welcome to the new WebMD Cat Health Center. WebMD veterinary experts provide comprehensive information about cat health care, offer nutrition and feeding ...

Yes, they're independent and willful, but felines can be taught certain behaviors—to the benefit of both cat and human.
Representing Context

Australian scientist discovers star with telescope
Representing Context

Australian scientist discovers star with telescope

Skip-Gram contexts with $n=2$
Representing Context

Australian scientist **discovers** star with telescope

collapsing “prep” links

Australian scientist **discovers** star with telescope
Representing Context

Australian scientist **discovers** star with telescope

- **amod**
- **nsubj**
- **dobj**
- **pobj**

collapsing “prep” links

- **prep**
- **prep_with**

- **scientist/nssubj**
- **star/dobj**
- **telescope/prep_with**

Levy and Goldberg, 2014
## Representing Context

<table>
<thead>
<tr>
<th>Target Word</th>
<th>BOW5</th>
<th>DEPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>batman</td>
<td>nightwing</td>
<td>superman</td>
</tr>
<tr>
<td></td>
<td>aquaman</td>
<td>superboy</td>
</tr>
<tr>
<td></td>
<td>catwoman</td>
<td>supergirl</td>
</tr>
<tr>
<td></td>
<td>superman</td>
<td>catwoman</td>
</tr>
<tr>
<td></td>
<td>manhunter</td>
<td>aquaman</td>
</tr>
<tr>
<td>hogwarts</td>
<td>dumbledore</td>
<td>sunnyvale</td>
</tr>
<tr>
<td></td>
<td>hallows</td>
<td>collinwood</td>
</tr>
<tr>
<td></td>
<td>half-blood</td>
<td>calarts</td>
</tr>
<tr>
<td></td>
<td>malfoy</td>
<td>greendale</td>
</tr>
<tr>
<td></td>
<td>snape</td>
<td>millfield</td>
</tr>
<tr>
<td>turing</td>
<td>nondeterministic</td>
<td>pauling</td>
</tr>
<tr>
<td></td>
<td>non-deterministic</td>
<td>hotelling</td>
</tr>
<tr>
<td></td>
<td>computability</td>
<td>heting</td>
</tr>
<tr>
<td></td>
<td>deterministic</td>
<td>lessing</td>
</tr>
<tr>
<td></td>
<td>finite-state</td>
<td>hamming</td>
</tr>
<tr>
<td>florida</td>
<td>gainesville</td>
<td>texas</td>
</tr>
<tr>
<td></td>
<td>fla</td>
<td>louisiana</td>
</tr>
<tr>
<td></td>
<td>jacksonville</td>
<td>georgia</td>
</tr>
<tr>
<td></td>
<td>tampa</td>
<td>california</td>
</tr>
<tr>
<td></td>
<td>lauderdale</td>
<td>carolina</td>
</tr>
</tbody>
</table>
Representing Context

... 5 farmers were thrown into jail in Ireland ...
... fünf Landwirte festgenommen, weil ...
... oder wurden festgenommen, gefoltert ...
... or have been imprisoned, tortured ...

Bannard and Callison-Burch, 2005
# Representing Context

<table>
<thead>
<tr>
<th>Cosine Similarity</th>
<th>Monolingual (symmetric)</th>
<th>Bilingual</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ shades/the shade</td>
<td>¬ large/small</td>
<td>≡ dad/father</td>
</tr>
<tr>
<td>☐ yard/backyard</td>
<td>≡ few/several</td>
<td>☐ some kid/child</td>
</tr>
<tr>
<td># each other/man</td>
<td>¬ different/same</td>
<td>≡ a lot of/many</td>
</tr>
<tr>
<td>☐ picture/drawing</td>
<td>¬ other/same</td>
<td>≡ female/woman</td>
</tr>
<tr>
<td>~ practice/target</td>
<td>¬ put/take</td>
<td>≡ male/man</td>
</tr>
</tbody>
</table>

Pavlick et al, 2015
Representing Context

$L_{ling}(w_t) = \text{maximize context prediction}$

$L_{vision}(w_t) = \text{maximize similarity}$

$L_{ling}(w_t) + L_{vision}(w_t) = \text{CAT}$

$M^{u \rightarrow v}$
Representing Context

<table>
<thead>
<tr>
<th>Target</th>
<th>Skip-gram</th>
<th>MMSkip-gram-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>donut</td>
<td>fridge, diner, candy</td>
<td>pizza, sushi, sandwich</td>
</tr>
<tr>
<td>owl</td>
<td>pheasant, woodpecker, squirrel</td>
<td>eagle, falcon, hawk</td>
</tr>
<tr>
<td>mural</td>
<td>sculpture, painting, portrait</td>
<td>painting, portrait, sculpture</td>
</tr>
<tr>
<td>tobacco</td>
<td>coffee, cigarette, corn</td>
<td>cigarette, cigar, smoking</td>
</tr>
<tr>
<td>depth</td>
<td>size, bottom, meter</td>
<td>sea, size, underwater</td>
</tr>
<tr>
<td>chaos</td>
<td>anarchy, despair, demon</td>
<td>demon, anarchy, shadow</td>
</tr>
</tbody>
</table>
Distributional Semantics: Takeaways

• The "meaning" of a word is the contexts in which that word can be used.
• We can represent a word as a point in continuous space by using a vector to store all the contexts in which the word has been observed.
• "Embeddings" are just a low-dimensional way of representing the contexts that used to be stored as big sparse vectors.
• We can (and should) be as creative as we want with how we define "contexts".
Distributional Semantics: Takeaways

- The “meaning” of a word is the contexts in which that word can be used.
Distributional Semantics: Takeaways

- The “meaning” of a word is the contexts in which that word can be used.
- We can represent word as a point in continuous space by using a vector to store all the contexts in which the words has been observed.
Distributional Semantics: Takeaways

- The “meaning” of a word is the contexts in which that word can be used.

- We can represent word as a point in continuous space by using a vector to store all the contexts in which the words has been observed.

- “Embeddings” are just a low-dimensional way of representing the contexts that used to be stored as big sparse vectors.
Distributional Semantics: Takeaways

• The “meaning” of a word is the contexts in which that word can be used.

• We can represent word as a point in continuous space by using a vector to store all the contexts in which the words has been observed.

• “Embeddings” are just a low-dimensional way of representing the contexts that used to be stored as big sparse vectors.

• We can (and should) be as creative as we want with how we define “contexts”.
More General Takeaways

• Formal Semantics
  • Represent language as logic
  • Focus is on representing compositionality: how do word meanings combine?
More General Takeaways

• Formal Semantics
  • Represent language as logic
  • Focus is on representing compositionality: how do word meanings combine?

• Distributional Semantics
  • Represent language as vectors/points in space
  • Focus is on learning meaning from context
More General Takeaways

- Formal Semantics
  - Represent language as logic
  - Focus is on representing compositionality: how do word meanings combine?

- Distributional Semantics
  - Represent language as vectors/points in space
  - Focus is on learning meaning from context
Pause: Questions!
Sentence-Level Semantics
# Sentence-Level Semantics

<table>
<thead>
<tr>
<th>Natural Language Inference</th>
<th>Logical Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Sentence-Level Semantics

<table>
<thead>
<tr>
<th>Natural Language Inference</th>
<th>Logical Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>given a premise $p$ and a hypothesis $h$, predict whether $p$ entails $h$</td>
<td>given a sentence $s$ return an executable representation (e.g. mathematical formula, SQL query…)</td>
</tr>
<tr>
<td>Natural Language Inference</td>
<td>Logical Forms</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>ungrounded—relate text to other text</td>
<td>grounded—relate text to tables in a database, or actions on a robot</td>
</tr>
</tbody>
</table>
Sentence-Level Semantics

Natural Language Inference

ungrounded—relate text to other text

Logical Forms

grounded—relate text to tables in a database, or actions on a robot
Semantic Parsing
What is the largest state? Alaska

<table>
<thead>
<tr>
<th>NAME</th>
<th>Alaska</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbrev.</td>
<td>AK</td>
</tr>
<tr>
<td>SIZE</td>
<td>663,268</td>
</tr>
<tr>
<td>CAPITAL</td>
<td>Juneau</td>
</tr>
</tbody>
</table>
What is the largest state?

<table>
<thead>
<tr>
<th>NAME</th>
<th>Alaska</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbrev.</td>
<td>AK</td>
</tr>
<tr>
<td>SIZE</td>
<td>663,268</td>
</tr>
<tr>
<td>CAPITAL</td>
<td>Juneau</td>
</tr>
</tbody>
</table>
What is the largest state?

- Prolog
- SQL
- FunQL
- First-order logic
- \( \lambda \)-calculus

<table>
<thead>
<tr>
<th>NAME</th>
<th>Alaska</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbrev.</td>
<td>AK</td>
</tr>
<tr>
<td>SIZE</td>
<td>663,268</td>
</tr>
<tr>
<td>CAPITAL</td>
<td>Juneau</td>
</tr>
</tbody>
</table>
What is the largest state?

\[ \text{\lambda-calculus} \]

<table>
<thead>
<tr>
<th>NAME</th>
<th>Alaska</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbrev.</td>
<td>AK</td>
</tr>
<tr>
<td>SIZE</td>
<td>663,268</td>
</tr>
<tr>
<td>CAPITAL</td>
<td>Juneau</td>
</tr>
</tbody>
</table>

Zettlemoyer and Collins (2012)
Supervised Semantic Parsing

What is the largest state?

\[ \text{argmax}(\lambda x. \text{state}(x), \lambda x. \text{size}(x)) \]
Supervised Semantic Parsing

What is the largest state?

\[
\text{argmax} (\lambda x. \text{state}(x), \lambda x. \text{size}(x))
\]
Combinatory Categorial Grammar

\[
\text{state} \ := \ \text{NP} : \text{state}
\]

Zettlemoyer and Collins (2012)
Combinatory Categorial Grammar

\[
\text{state} \ := \ NP : \text{state}
\]

Syntax

Zettlemoyer and Collins (2012)
Combinatory Categorial Grammar

\[
\text{state} \ := \ NP \ : \ state
\]

Zettlemoyer and Collins (2012)
Combinatory Categorial Grammar

\[ \text{borders} := (S\backslash NP) / NP : \lambda x. \lambda y. \text{borders}(y, x) \]

Zettlemoyer and Collins (2012)
Combinatory Categorial Grammar

\[
\text{borders} := (S\backslash NP)/NP : \lambda x.\lambda y.\text{borders}(y,x)
\]

\[
\begin{array}{ccc}
\text{utah} & \text{borders} & \text{idaho} \\
\text{NP} & (S\backslash NP)/NP & \text{NP}
\end{array}
\]

Zettlemoyer and Collins (2012)
Combinatory Categorial Grammar

\[
\text{borders} := (S\backslash\text{NP})/\text{NP} : \lambda x.\lambda y. \text{borders}(y,x)
\]

\[
\text{borders(utah, idaho)}
\]

Zettlemoyer and Collins (2012)
Combinatory Categorial Grammar

<table>
<thead>
<tr>
<th>utah</th>
<th>borders</th>
<th>idaho</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>(S\NP)/NP</td>
<td>NP</td>
</tr>
<tr>
<td>utah</td>
<td>$\lambda x. \lambda y. \text{borders}(y,x)$</td>
<td>idaho</td>
</tr>
</tbody>
</table>
Combinatory Categorial Grammar

\[
\text{utah} \quad \text{borders} \quad \text{idaho} \\
\text{NP} \quad (S \setminus \text{NP}) / \text{NP} \quad \text{NP} \\
\text{utah} \quad \lambda x.\lambda y.\text{borders}(y, x) \quad \text{idaho}
\]
Combinatory Categorial Grammar

\[
\begin{align*}
\text{utah} & \quad \text{borders idaho} \\
\text{NP} & \quad (S \backslash \text{NP}) \\
\text{utah} & \quad \lambda y.\text{borders}(y, \text{idaho})
\end{align*}
\]
Combinatory Categorial Grammar

\[
\begin{align*}
\text{utah} & \quad \text{borders idaho} \\
\text{NP} & \quad (S \backslash \text{NP}) \\
\text{utah} & \quad \lambda y. \text{borders}(y, \text{idaho})
\end{align*}
\]
Combinatory Categorial Grammar

\[ S \]

\[ \text{borders}(\text{utah}, \text{idaho}) \]
Supervised Semantic Parsing

Does Utah border Idaho?

\[ \text{borders(utah, idaho)} \]

<table>
<thead>
<tr>
<th>State Pair</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>(utah, idaho)</td>
<td>TRUE</td>
</tr>
<tr>
<td>(utah, michigan)</td>
<td>FALSE</td>
</tr>
<tr>
<td>(idaho, michigan)</td>
<td>FALSE</td>
</tr>
<tr>
<td>(idaho, montana)</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

Zettlemoyer and Collins (2012)
Supervised Semantic Parsing

Does Utah border Idaho?  

training data

Zettlemoyer and Collins (2012)
Weakly-supervised Semantic Parsing

Does Utah border Idaho? Yes

<table>
<thead>
<tr>
<th>Borders</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(utah, idaho)</td>
<td>TRUE</td>
</tr>
<tr>
<td>(utah, michigan)</td>
<td>FALSE</td>
</tr>
<tr>
<td>(idaho, michigan)</td>
<td>FALSE</td>
</tr>
<tr>
<td>(idaho, montanana)</td>
<td>TRUE</td>
</tr>
</tbody>
</table>
Weakly-supervised Semantic Parsing

Does Utah border Idaho?

Yes

<table>
<thead>
<tr>
<th>(utah, idaho)</th>
<th>TRUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(utah, michigan)</td>
<td>FALSE</td>
</tr>
<tr>
<td>(idaho, michigan)</td>
<td>FALSE</td>
</tr>
<tr>
<td>(idaho, montana)</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

Liang et al. (2012)
Does Utah border Idaho?

borders(utah, idaho)

<table>
<thead>
<tr>
<th>(utah, idaho)</th>
<th>TRUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(utah, michigan)</td>
<td>FALSE</td>
</tr>
<tr>
<td>(idaho, michigan)</td>
<td>FALSE</td>
</tr>
<tr>
<td>(idaho, montana)</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

Liang et al. (2012)
Weakly-supervised Semantic Parsing

Does Utah border Idaho? → Yes

Liang et al. (2012)
Weakly-supervised Semantic Parsing

Does Utah border Idaho? Yes

What is the largest state? Alaska
Utah borders Idaho TRUE
highest point in the US Mt. McKinley
city in California Los Angeles, San Francisco...

training data

Liang et al. (2012)
Dependency-based compositional semantics

Utah borders Idaho

Liang et al. (2012)
Dependency-based compositional semantics

Utah borders Idaho

Liang et al. (2012)
Dependency-based compositional semantics

Utah borders Idaho

Liang et al. (2012)
Dependency-based compositional semantics

Utah borders Idaho

Liang et al. (2012)
Dependency-based compositional semantics

Utah borders Idaho

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(UT,ID)</td>
<td>TRUE</td>
<td></td>
</tr>
<tr>
<td>(UT,MI)</td>
<td>FALSE</td>
<td></td>
</tr>
<tr>
<td>(ID,MI)</td>
<td>FALSE</td>
<td></td>
</tr>
<tr>
<td>(ID,MT)</td>
<td>TRUE</td>
<td></td>
</tr>
</tbody>
</table>
Dependency-based compositional semantics

Utah borders Idaho

| (UT, ID) | TRUE |
| (UT, MI) | FALSE |
| (ID, MI) | FALSE |
| (ID, MT) | TRUE |

Liang et al. (2012)
Dependency-based compositional semantics

Utah borders Idaho

\{ (UT, ID), (ID, MT), (UT, AZ), (UT, WY) \}

Liang et al. (2012)
Dependency-based compositional semantics

Utah borders Idaho

\{(\text{UT}, \text{ID}), (\text{ID}, \text{MT}), (\text{UT}, \text{AZ}), (\text{UT}, \text{WY})\}

<table>
<thead>
<tr>
<th>Relation</th>
<th>Truth Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{UT}, \text{ID})</td>
<td>TRUE</td>
</tr>
<tr>
<td>(\text{UT}, \text{MI})</td>
<td>FALSE</td>
</tr>
<tr>
<td>(\text{ID}, \text{MI})</td>
<td>FALSE</td>
</tr>
<tr>
<td>(\text{ID}, \text{MT})</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

Liang et al. (2012)
Dependency-based compositional semantics

Utah borders Idaho

\{ (UT, ID), (UT, AZ), (UT, WY) \}

Liang et al. (2012)
Dependency-based compositional semantics

Utah borders Idaho

\{(UT, ID), (UT, AZ), (UT, WY)\}

Liang et al. (2012)
Dependency-based compositional semantics

Utah borders Idaho

\{(UT, ID)\}

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(UT, ID)</td>
<td>TRUE</td>
</tr>
<tr>
<td>(UT, MI)</td>
<td>FALSE</td>
</tr>
<tr>
<td>(ID, MI)</td>
<td>FALSE</td>
</tr>
<tr>
<td>(ID, MT)</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

Liang et al. (2012)
Dependency-based compositional semantics

Utah borders Idaho

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(UT,ID)</td>
<td>TRUE</td>
</tr>
<tr>
<td>(UT,MI)</td>
<td>FALSE</td>
</tr>
<tr>
<td>(ID,MI)</td>
<td>FALSE</td>
</tr>
<tr>
<td>(ID,MT)</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

Liang et al. (2012)
Does Utah border Idaho?

Dependency-based compositional semantics

Borders(utah, idaho) → Yes

| (utah, idaho)   | TRUE |
| (utah, michigan) | FALSE |
| (idaho, michigan) | FALSE |
| (idaho, montana)  | TRUE |

Liang et al. (2012)
Does Utah border Idaho?

Liang et al. (2012)
Sentence-Level Semantics

<table>
<thead>
<tr>
<th>Natural Language Inference</th>
<th>Logical Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>ungrounded—relate text to other text</td>
<td>grounded—relate text to tables in a database, or actions on a robot</td>
</tr>
</tbody>
</table>
Compositional Distributional Semantics
Compositional Distributional Semantics

Grefenstette and Sadrzadeh (2011)
Compositional Distributional Semantics

Grefenstette and Sadrzadeh (2011)
Compositional Distributional Semantics

Grefenstette and Sadrzadeh (2011)
Compositional Distributional Semantics

Grefenstette and Sadrzadeh (2011)
Compositional Distributional Semantics

Grefenstette and Sadrzadeh (2011)
Compositional Distributional Semantics

Grefenstette and Sadrzadeh (2011)
Compositional Distributional Semantics

Grefenstette and Sadrzadeh (2011)
Compositional Distributional Semantics

borders ⊗ idaho ⊗ utah    is near ⊗ boise ⊗ ogden

Grefenstette and Sadrzadeh (2011)
## Compositional Distributional Semantics

<table>
<thead>
<tr>
<th>bad luck</th>
<th>electronic communication</th>
<th>historical map</th>
</tr>
</thead>
<tbody>
<tr>
<td>bad</td>
<td>elec. storage</td>
<td>topographical atlas</td>
</tr>
<tr>
<td>bad weekend</td>
<td>elec. transmission purpose</td>
<td>hist. material</td>
</tr>
<tr>
<td>good spirit</td>
<td>nice girl</td>
<td>little war</td>
</tr>
<tr>
<td>important route</td>
<td>great war</td>
<td>major war</td>
</tr>
<tr>
<td>important transport</td>
<td>big girl</td>
<td>small war</td>
</tr>
<tr>
<td>important road</td>
<td>guy</td>
<td></td>
</tr>
<tr>
<td>major road</td>
<td>special collection</td>
<td>young husband</td>
</tr>
<tr>
<td>red cover</td>
<td>general collection archives</td>
<td>small son</td>
</tr>
<tr>
<td>black cover</td>
<td>small collection archives</td>
<td>small daughter</td>
</tr>
<tr>
<td>hardback</td>
<td>red label</td>
<td>mistress</td>
</tr>
</tbody>
</table>

Table 2: Nearest 3 neighbors of specific ANs.

Grefenstette and Sadrzadeh (2011)
Compositional Distributional Semantics

<table>
<thead>
<tr>
<th>bad</th>
<th>electronic</th>
<th>historical</th>
</tr>
</thead>
<tbody>
<tr>
<td>luck</td>
<td>communication</td>
<td>map</td>
</tr>
<tr>
<td>bad</td>
<td>elec. storage</td>
<td>topographical</td>
</tr>
<tr>
<td>bad weekend</td>
<td>elec. transmission</td>
<td>atlas</td>
</tr>
<tr>
<td>good spirit</td>
<td>purpose</td>
<td>hist. material</td>
</tr>
<tr>
<td>important route</td>
<td>nice girl</td>
<td>little war</td>
</tr>
<tr>
<td>important transport</td>
<td>good girl</td>
<td>great war</td>
</tr>
<tr>
<td>important road</td>
<td>big girl</td>
<td>major war</td>
</tr>
<tr>
<td>major road</td>
<td>guy</td>
<td>small war</td>
</tr>
<tr>
<td>red cover</td>
<td>special collection</td>
<td>young husband</td>
</tr>
<tr>
<td>black cover</td>
<td>general collection</td>
<td>small son</td>
</tr>
<tr>
<td>hardback</td>
<td>small collection</td>
<td>small daughter</td>
</tr>
<tr>
<td>red label</td>
<td>archives</td>
<td>mistress</td>
</tr>
</tbody>
</table>

Table 2: Nearest 3 neighbors of specific ANs.

Grefenstette and Sadrzadeh (2011)
Compositional Distributional Semantics

borders $\odot$ idaho $\otimes$ utah

borders $\odot$ idaho $\otimes$ texas

Grefenstette and Sadrzadeh (2011)
Compositional Distributional Semantics

borders $\ominus$ idaho $\otimes$ utah

doesn’t border $\ominus$ idaho $\otimes$ utah

Grefenstette and Sadrzadeh (2011)
Compositional Distributional Semantics

Socher et al (2013)
Compositional Distributional Semantics

\[ p_2 = g(a, p_1) \]

\[ p_1 = g(b, c) \]

Socher et al (2013)
Compositional Distributional Semantics

Socher et al (2013)

Basic RNN

Socher et al (2013)
Compositional Distributional Semantics

$\begin{align*}
p_2 &= g(a, p_1) \\
p_1 &= g(b, c)
\end{align*}$

Matrix-Vector RNN

Socher et al (2013)
Compositional Distributional Semantics

Recursive Neural Tensor Network (2013)
Compositional Distributional Semantics

<table>
<thead>
<tr>
<th>Model</th>
<th>Negated Positive</th>
<th>Negated Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>biNB</td>
<td>19.0</td>
<td>27.3</td>
</tr>
<tr>
<td>RNN</td>
<td>33.3</td>
<td>45.5</td>
</tr>
<tr>
<td>MV-RNN</td>
<td>52.4</td>
<td>54.6</td>
</tr>
<tr>
<td>RNTN</td>
<td>71.4</td>
<td>81.8</td>
</tr>
</tbody>
</table>

Socher et al (2013)
### Compositional Distributional Semantics

It is one of the most/least compelling variations of these theme

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy Negated Positive</th>
<th>Accuracy Negated Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>biNB</td>
<td>19.0</td>
<td>27.3</td>
</tr>
<tr>
<td>RNN</td>
<td>33.3</td>
<td>45.5</td>
</tr>
<tr>
<td>MV-RNN</td>
<td>52.4</td>
<td>54.6</td>
</tr>
<tr>
<td>RNTN</td>
<td>71.4</td>
<td>81.8</td>
</tr>
</tbody>
</table>

Socher et al (2013)
The movie was [not] terrible.

---

### Compositional Distributional Semantics

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy Negated Positive</th>
<th>Accuracy Negated Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>biNB</td>
<td>19.0</td>
<td>27.3</td>
</tr>
<tr>
<td>RNN</td>
<td>33.3</td>
<td>45.5</td>
</tr>
<tr>
<td>MV-RNN</td>
<td>52.4</td>
<td>54.6</td>
</tr>
<tr>
<td>RNTN</td>
<td>71.4</td>
<td>81.8</td>
</tr>
</tbody>
</table>

---

Socher et al (2013)
Natural Logic
Every person danced.

Every young woman danced.
Natural Logic

Every person danced.
\( \forall x ( \text{person}(x) \rightarrow \text{danced}(x)) \)

\( \forall x ( ( \text{woman}(x) \land \text{young}(x)) \rightarrow \text{danced}(x)) \)

Every young woman danced.
Every person danced.
\[ \forall x (\text{person}(x) \rightarrow \text{danced}(x)) \]

Every woman danced.
\[ \forall x (\text{woman}(x) \rightarrow \text{person}(x)) \]

Every person followed the rule.
\[ \forall x (\forall P ((P(x) \land \text{young}(x)) \rightarrow P(x))) \]

Every young woman danced.
\[ \forall x ((\text{woman}(x) \land \text{young}(x)) \rightarrow \text{danced}(x)) \]

Every young woman danced.
Natural Logic

Every person danced.

\[ \forall x ( \text{person}(x) \rightarrow \text{danced}(x)) \]

\[ \forall x (\text{woman}(x) \rightarrow \text{person}(x)) \]

\[ \forall x (\text{danced}(x) \rightarrow \text{moved}(x)) \]

\[ \forall x (\forall P ((P(x) \land \text{young}(x)) \rightarrow P(x))) \]

\[ \forall x ((\text{woman}(x) \land \text{young}(x)) \rightarrow \text{moved}(x)) \]

Every young woman danced.
Every person danced.

\[ \forall x (\text{person}(x) \implies \text{danced}(x)) \]

\[ \forall x (\text{woman}(x) \implies \text{person}(x)) \]

\[ \forall x (\text{danced}(x) \implies \text{moved}(x)) \]

\[ \forall x (\forall P ((P(x) \land \text{young}(x)) \implies P(x))) \]

\[ \forall x ((\text{woman}(x) \land \text{young}(x)) \implies \text{moved}(x)) \]

Every young woman danced.

✔
Natural Logic

Every person danced.
\[ \forall x \, ( \text{person}(x) \rightarrow \text{danced}(x)) \]

\[ \forall x \, ( \text{woman}(x) \rightarrow \text{person}(x)) \]

\[ \forall x \, (\forall P \,( (P(x) \land \text{young}(x)) \rightarrow P(x))) \]

\[ \forall x \, ((\text{woman}(x) \land \text{young}(x)) \rightarrow \text{danced}(x)) \]

Every young woman danced.
Every person danced.
\[ \forall x (\text{person}(x) \rightarrow \text{danced}(x)) \]

\[ \forall x (\text{woman}(x) \rightarrow \text{person}(x)) \]

\[ \forall x (\forall P ( (P(x) \land \text{young}(x)) \rightarrow P(x))) \]

\[ \forall x ( (\text{woman}(x) \land \text{young}(x)) \rightarrow \text{danced}(x)) \]

Every young woman danced.
Natural Logic

Every person danced.

Every young woman danced.
Natural Logic

Every person danced.

Every young woman danced.

semantic inclusion
Natural Logic

Every person danced.

Every young woman danced.

downward monotone
Natural Logic

Every person danced.

Every young woman danced.

downward monotone
Natural Logic

Every person danced.

Every young woman danced.

downward monotone
The NatLog System

Every person danced.

Every young woman moved.
The NatLog System

Alignment

Every person danced.

Every young woman moved.

McCartney (2009)
The NatLog System

Alignment

Every person danced.

SUB(person,woman) Every woman danced.

INS(young) Every young woman danced.

SUB(danced,moved) Every young woman moved.
The NatLog System

Entailment Classification

Every person danced.

SUB(person,woman) Every woman danced.

INS(young) Every young woman danced.

SUB(danced,moved) Every young woman moved.
The NatLog System

<table>
<thead>
<tr>
<th>equivalence</th>
<th>couch = sofa</th>
</tr>
</thead>
<tbody>
<tr>
<td>forward entailment</td>
<td>woman ⊂ person</td>
</tr>
<tr>
<td>reverse entailment</td>
<td>move ⊃ dance</td>
</tr>
<tr>
<td>negation</td>
<td>able ^ unable</td>
</tr>
</tbody>
</table>

| alternation | cat \ dog |
| independence | happy \ tall |
The NatLog System

Entailment Classification

Every person danced.

SUB(person,woman) Every woman danced.

INS(young) Every young woman danced.

SUB(danced,moved) Every young woman moved.
The NatLog System

Entailment Classification

Every person danced.

Every woman danced.

Every young woman danced.

Every young woman moved.

McCartney (2009)
The NatLog System

Entailment Classification

Every person danced.

\( \text{SUB}(\text{person, woman}) \quad \text{Every woman danced.} \)

\( \text{INS}(\text{young}) \quad \text{Every young woman danced.} \)

\( \text{SUB}(\text{danced, moved}) \quad \text{Every young woman moved.} \)

\( \square \quad \text{reverse entailment} \)

McCartney (2009)
The NatLog System

Entailment Classification

Every person danced.

\[ \text{SUB}(\text{person}, \text{woman}) \quad \text{Every woman danced.} \]

\[ \text{INS}(\text{young}) \quad \text{Every young woman danced.} \]

\[ \text{SUB}(\text{danced, moved}) \quad \text{Every young woman moved.} \]

McCartney (2009)
The NatLog System

Entailment Classification

Every person danced.

SUB(person,woman) Every woman danced. ☐

INS(young) Every young woman danced. ☐

SUB(danced,moved) Every young woman moved. ☐
The NatLog System

Projectivity Marking

Every person danced.
The NatLog System

Projectivity Marking

Every person danced.

upward
monotone
The NatLog System

Projectivity Marking

Every person moved.

danced \sqsubseteq moved

every person danced \sqsubseteq every person moved

McCartney (2009)
The NatLog System

Projectivity Marking

Every *person* danced.

downward
monotone

McCartney (2009)
The NatLog System

Projectivity Marking

Every woman danced.

downward monotone

-person ⊑ young woman
every person ⊑ every young woman

McCartney (2009)
The NatLog System

Projectivity Marking

Every person danced.

\[
\text{SUB}(\text{person}, \text{woman}) \quad \text{Every woman danced.}
\]

\[
\text{INS}(\text{young}) \quad \text{Every young woman danced.}
\]

\[
\text{SUB}(\text{danced}, \text{moved}) \quad \text{Every young woman moved.}
\]
The NatLog System

Projectivity Marking

Every person danced.

SUB(person,woman)  Every woman danced.

INS(young)  Every young woman danced.

SUB(danced,moved)  Every young woman moved.
The NatLog System

Projectivity Marking

Every person danced.

SUB(person,woman) Every woman danced.

INS(young) Every young woman danced.

SUB(danced,moved) Every young woman moved.
The NatLog System

Joining Entailment Relations

Every person danced.

$$\text{SUB(}\text{person,woman)}$$ Every woman danced.

$$\text{INS(}\text{young)}$$ Every young woman danced.

$$\text{SUB(}\text{danced,moved)}$$ Every young woman moved.
The NatLog System

Joining Entailment Relations

Every person danced.

Every woman danced.

Every young woman danced.

Every young woman moved.

McCartney (2009)
The NatLog System

Joining Entailment Relations

Every person danced.

SUB(person,woman) Every woman danced.

INS(young) Every young woman danced.

SUB(danced,moved) Every young woman moved.
The NatLog System

Joining Entailment Relations

No person danced.

SUB(person,woman) No woman danced.

INS(young) No young woman danced.

SUB(danced,moved) No young woman moved.

McCartney (2009)
Sentence-Level Semantics

Natural Language Inference

- ungrounded—relate text to other text

Logical Forms

- grounded—relate text to tables in a database, or actions on a robot
Denotational semantics

Idaho \approx Utah
Denotational semantics

Idaho ≠ Utah
Denotational semantics

\[ \text{Idaho} \neq \text{Utah} \]

John lives in Idaho

John lives in Utah
Denotation Graph

Before applying these rules, a number of special cases and exceptions are checked. For example, in the case of the pattern, a separate lexicon is used.

The set of images that depict the objects they describe, denoted as \( K \), which is the set of images with concrete sets of images. The interpretation of an image (Figure 1), we propose to instantiate sentences that can be used to describe descriptive sentences, i.e., non-negative, episodic (Montague, 1974; Dowty et al., 1981; Barwise and Weir, 2003; Szpektor and Dagan, 2008; Clarke, 2010). To compute the similarity of two strings, the definition of vector-based distributional similarities, similar meaning (Harris, 1954). This has led to the creation of reduction functions (MacCartney and Manning, 2008)'s NatLog lexical sub-constituents (NPs, VPs, or simple Ss), and lexical sub-expressions: the removal of optional material (e.g., PPs drop), the removal of modality (e.g., NP articles), and the removal of focus (e.g., VP modifiers).

The denotation graph, taken from Young et al. (2014), is assumed to be the set of all syntactic and lexical operations like derivation and reductions: if \( a \) expresses \( b \), then \( b \) is a truthful description of \( a \). Semantically, entailment is represented by the subsumption relation \( \subseteq \) and we say that \( a \) entails \( b \) if \( \forall s: L(a, s) \Rightarrow L(b, s) \). These operations are akin to the syntactic and lexical operations extracted from counts of words, either via the cosine play on the beach

(a) Two men work in a butcher shop; one cuts the meat from a butchered cow, while the other hoses the floor.
(b) A green-shirted man with a butcher's apron uses a knife to carve out the hanging carcass of a cow.
(c) A man with a yellow tie looks concerned.
(d) A graying man in a suit is perplexed at a business meeting.
(e) Two men work in a butcher shop; one cuts the meat from a butchered cow, while another man hoses away the blood.

Figure 1: Two images from our data set and their five captions.
Gray haired man in black suit and yellow tie working in a financial environment.

A graying man in a suit is perplexed at a business meeting.

A businessman in a yellow tie gives a frustrated look.

A man in a yellow tie is rubbing the back of his neck.

A man with a yellow tie looks concerned.

Young et al. (2014)
A businessman in a yellow tie gives a frustrated look.

Young et al. (2014)
Denotation Graph

A businessman in a yellow tie

Young et al. (2014)
Denotation Graph

A man in a yellow tie
A man in a tie
Denotation Graph

A man

Young et al. (2014)
Denotation Graph

A man

A businessman in a yellow tie

Young et al. (2014)
## Denotation Graph

<table>
<thead>
<tr>
<th>Distributional Similarity</th>
<th>Denotational Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>play game</td>
<td>tackle person</td>
</tr>
<tr>
<td>play rugby</td>
<td>hold football</td>
</tr>
<tr>
<td>play soccer</td>
<td>run down field</td>
</tr>
<tr>
<td>play on field</td>
<td>wear white jersey</td>
</tr>
<tr>
<td>play ball</td>
<td>avoid</td>
</tr>
</tbody>
</table>
Denotation Graph

yellow tie

red tie

yellow dress

???
Pause: Questions!
So...uh...does this really matter?
Deep Learning is Taking Over NLP!

Titles of ACL Papers, 2017
SOTA on all the benchmark tasks
But what, exactly, are our systems learning?
But what, exactly, are our systems learning?

The attack killed at least 12 civilians.

The deadly attack killed at least 12 civilians.

Pavlick and Callison-Burch (2016)
But what, exactly, are our systems learning?

The past record was held by John Elway…

*Jeff Dean had jersey number 37…*

John Elway

Jia and Liang (2017)
But what, exactly, are our systems learning?

The man is holding a *saxophone*.

⇒

The man is holding an *electric guitar*.

Glockner et al (ACL 2018)
But what, exactly, are our systems learning?

The woman is more cheerful than the man.

↓

The woman is less cheerful than the man.

But what, exactly, are our systems learning?

The woman is **less** cheerful than the man.


---

Probing Tasks Galore!

- White et al (IJCNLP 2017)
- Mahler et al (2017)
- Ettinger et al. (EMNLP 2017)
- Adi et al. (ICLR 2017)
- Poliak et al. (*SEM 2018)
- Conneau et al. (ACL 2018)
- Zhu et al. (ACL 2018)
| Task                          | Metric           | N-gram | MLP | NB | RNN | Graph | Neural | Phrase | Neural | Graph | Neural | SNLI | LexEnt | SNLI | LexEnt | SNLI | LexEnt | SNLI | LexEnt | SNLI | LexEnt |
|-------------------------------|------------------|--------|-----|----|-----|-------|--------|--------|--------|-------|--------|------|-------|------|-------|------|-------|------|-------|------|-------|------|-------|
| Language Modeling             | Perplexity       | 312    | 252 |    |     |       |        |        |        |       |        |      |       |      |       |      |       |      |       |      |       |
| Sentiment Analysis            | Accuracy         | 41.9   | 45.7|    |     |       |        |        |        |       |        |      |       |      |       |      |       |      |       |      |       |
| Dependency Parsing            | Unlabelled Attachment | 90   | 90.7|    |     |       |        |        |        |       |        |      |       |      |       |      |       |      |       |      |       |
| Machine Translation           | BLEU (Ar-En)     | 52.8   | 49.5|    |     |       |        |        |        |       |        |      |       |      |       |      |       |      |       |      |       |
| Modifier Composition          | Accuracy         | 85     | 86  |    |     |       |        |        |        |       |        |      |       |      |       |      |       |      |       |      |       |
| Random Noise                  | Accuracy         | 81.1   | 49.8|    |     |       |        |        |        |       |        |      |       |      |       |      |       |      |       |      |       |
| Lexical Entailments           | Accuracy         | 85     | 52  |    |     |       |        |        |        |       |        |      |       |      |       |      |       |      |       |      |       |
| Negations                     | Accuracy         | 50     | 50  |    |     |       |        |        |        |       |        |      |       |      |       |      |       |      |       |      |       |
Deep Representations = very good at tasks…
Deep Representations = very good at tasks...

...but very bad at language
What do we want our systems to learn?
This workshop deals with the evaluation of general-purpose vector representations for linguistic units (morphemes, words, phrases, sentences, etc). What distinguishes these representations (or embeddings) is that they are not trained with a specific application in mind, but rather to capture broadly useful features of the represented units. Another way to view their usage is through the lens of transfer learning: The embeddings are trained with one objective, but applied on others.

Evaluating general-purpose representation learning systems is fundamentally difficult. They can be trained on a variety of objectives, making simple intrinsic evaluations useless as a means of comparing methods. They are also meant to be applied to a variety of downstream tasks, which will place different demands on them...

RepEval 2017
(Bowman, Goldberg, Hill, Lazaridou, Levy, Reichart, and Søgaard)
This workshop deals with the evaluation of general-purpose vector representations for linguistic units (morphemes, words, phrases, sentences, etc). What distinguishes these representations (or embeddings) is that they are not trained with a specific application in mind, but rather to capture broadly useful features of the represented units. Another way to view their usage is through the lens of transfer learning: The embeddings are trained with one objective, but applied on others.

Evaluating general-purpose representation learning systems is fundamentally difficult. They can be trained on a variety of objectives, making simple intrinsic evaluations useless as a means of comparing methods. They are also meant to be applied to a variety of downstream tasks...

RepEval 2017
(Bowman, Goldberg, Hill, Lazaridou, Levy, Reichart, and Søgaard)
This workshop deals with the evaluation of general-purpose vector representations for linguistic units (morphemes, words, phrases, sentences, etc). What distinguishes these representations (or embeddings) is that they are not trained with a specific application in mind, but rather to capture broadly useful features of the represented units. Another way to view their usage is through the lens of transfer learning: The embeddings are trained with one objective, but applied on others.

Evaluating general-purpose representation learning systems is fundamentally difficult. They can be trained on a variety of objectives, making simple intrinsic evaluations useless as a means of comparing methods. They are also meant to be applied to a variety of downstream tasks, which will place different demands on them…

RepEval 2017
(Bowman, Goldberg, Hill, Lazaridou, Levy, Reichart, and Søgaard)
This workshop deals with the evaluation of general-purpose vector representations for linguistic units (morphemes, words, phrases, sentences, etc). What distinguishes these representations (or embeddings) is that they are not trained with a specific application in mind, but rather to capture broadly useful features of the represented units. Another way to view their usage is through the lens of transfer learning: The embeddings are trained with one objective, but applied on others.

Evaluating general-purpose representation learning systems is fundamentally difficult. They can be trained on a variety of objectives, making simple intrinsic evaluations useless as a means of comparing methods. They are also meant to be applied to a variety of downstream tasks, which will place different demands on them…

RepEval 2017
(Bowman, Goldberg, Hill, Lazaridou, Levy, Reichart, and Søgaard)
“There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed I consider it possible to comprehend the syntax and semantics of both kinds of languages with a single natural and mathematically precise theory.”

—Richard Montague
“There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed I consider it possible to comprehend the syntax and semantics of both kinds of languages with a single natural and mathematically precise theory.”

—Richard Montague
“There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed I consider it possible to comprehend the syntax and semantics of both kinds of languages with a single natural and mathematically precise theory.”

—Richard Montague
Language

\downarrow

Math
Language

\[ \forall x \forall y (P(f(x)) \rightarrow \neg(Q(f(y), x))) \]
Language

\[ \lambda x. f(y, g(x)) \land h(y) \]
\[ o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \]
Language

010110110101001010101
010100011010111011001
010101101010010101100
000110000001011111010
Questions!