Natural Language Understanding: Foundations and State-of-the-Art

Percy Liang

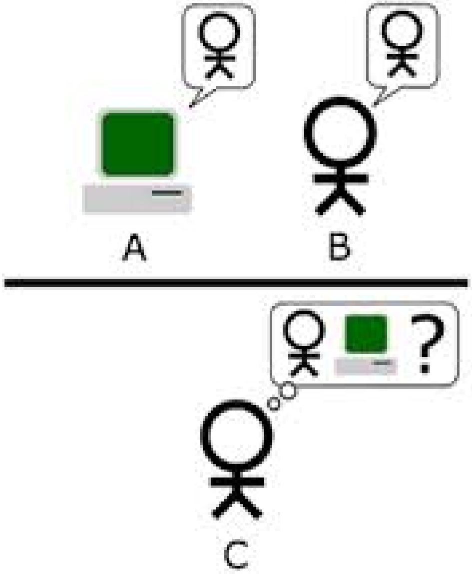
"Foundations of ML" Bootcamp
January 26, 2017

The Imitation Game (1950)

"Can machines think?"

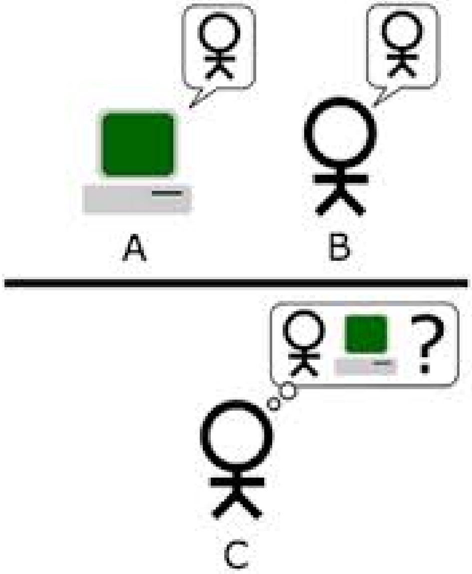
The Imitation Game (1950)

"Can machines think?"



The Imitation Game (1950)

"Can machines think?"



Q: Please write me a sonnet on the subject of the Forth Bridge.

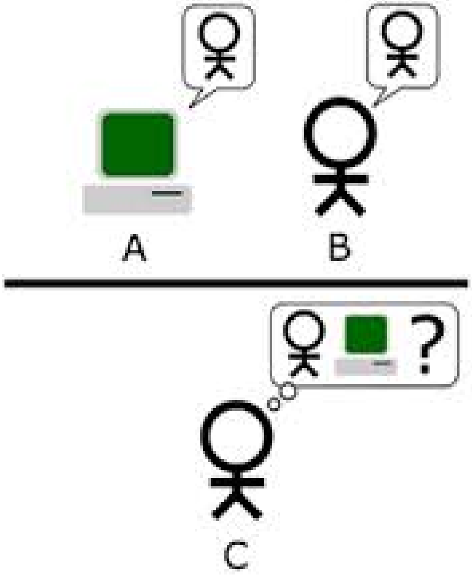
A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764.

A: (Pause about 30 seconds and then give as answer) 105621.

The Imitation Game (1950)

"Can machines think?"



Q: Please write me a sonnet on the subject of the Forth Bridge.

A: Count me out on this one. I never could write poetry.

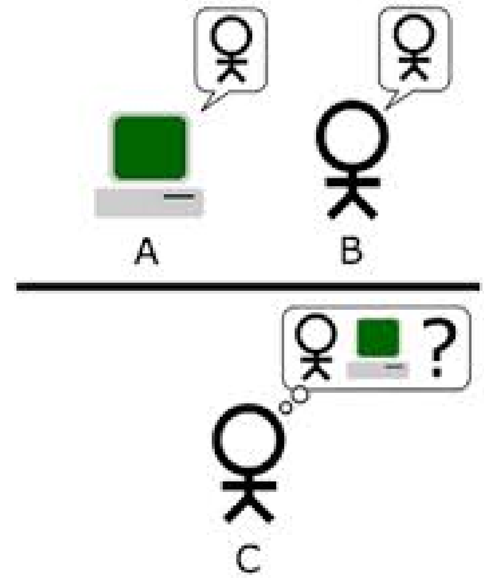
Q: Add 34957 to 70764.

A: (Pause about 30 seconds and then give as answer) 105621.

end-to-end

The Imitation Game (1950)

"Can machines think?"



Q: Please write me a sonnet on the subject of the Forth Bridge.

A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764.

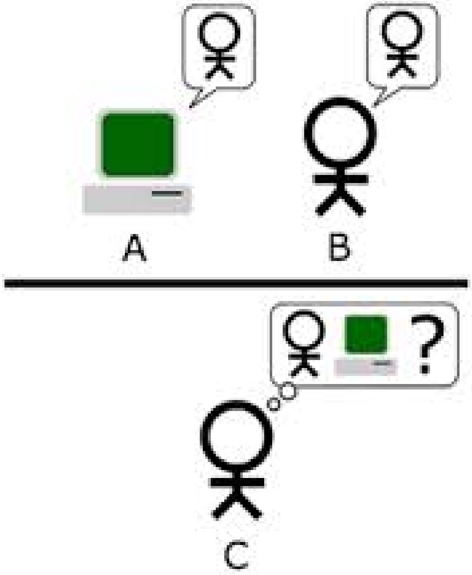
A: (Pause about 30 seconds and then give as answer) 105621.

end-to-end

interactivity

The Imitation Game (1950)

"Can machines think?"



Q: Please write me a sonnet on the subject of the Forth Bridge.

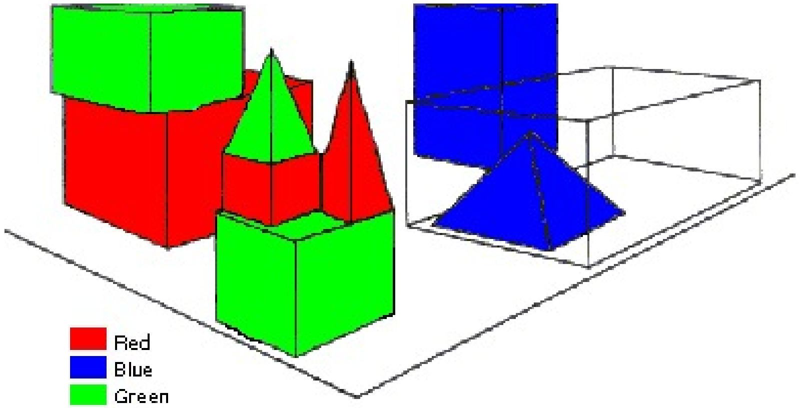
A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764.

A: (Pause about 30 seconds and then give as answer) 105621.

end-to-end | interactivity | language

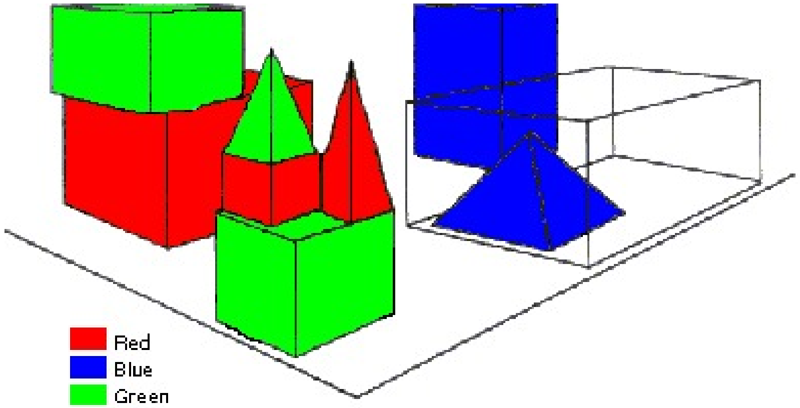
SHRDLU (1971)



SHRDLU (1971)

Person: Pick up a big red block.

Computer: OK.



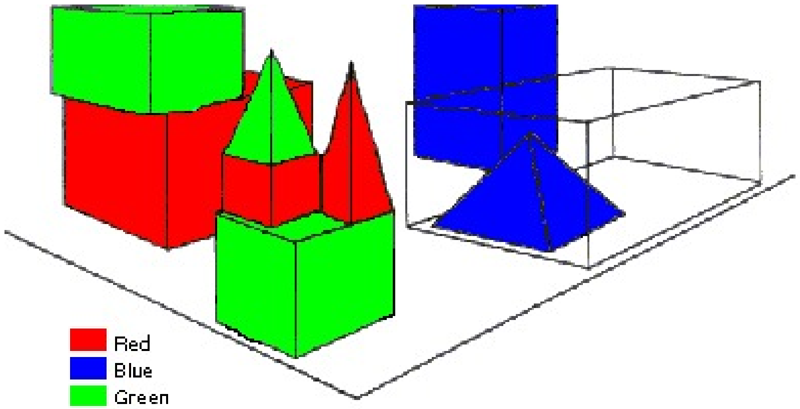
SHRDLU (1971)

Person: Pick up a big red block.

Computer: OK.

Person: Grasp the pyramid.

Computer: I don't understand which pyramid you mean.



SHRDLU (1971)

Person: Pick up a big red block.

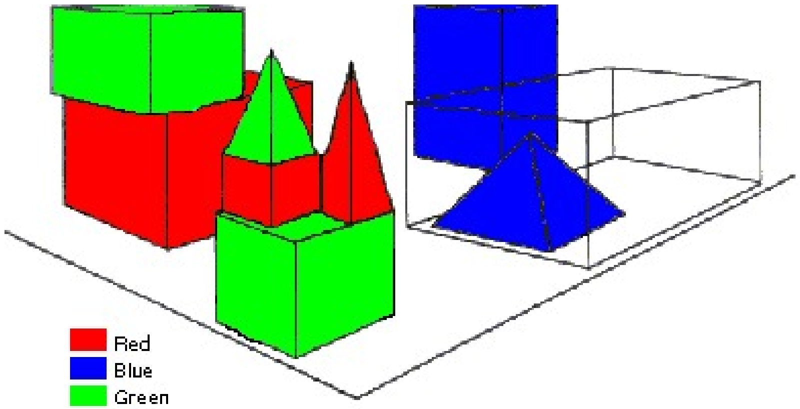
Computer: OK.

Person: Grasp the pyramid.

Computer: I don't understand which pyramid you mean.

Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.

Computer: By "it", I assume you mean the block which is taller than the one I am holding.



SHRDLU (1971)

Person: Pick up a big red block.

Computer: OK.

Person: Grasp the pyramid.

Computer: I don't understand which pyramid you mean.

Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.

Computer: By "it", I assume you mean the block which is taller than the one I am holding.

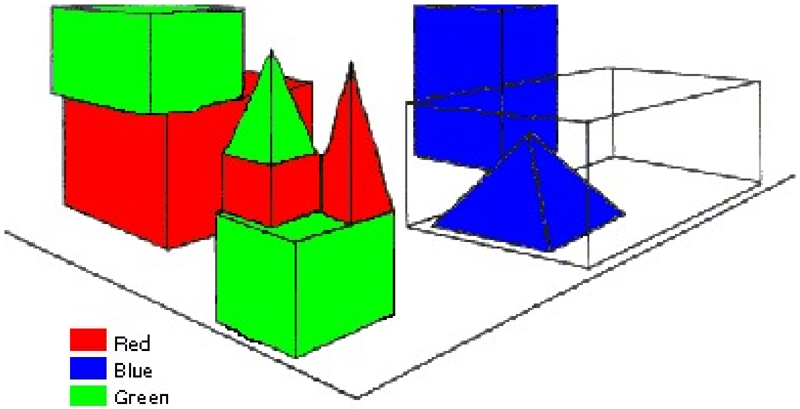
Computer: OK.

Person: What does the box contain?

Computer: The blue pyramid and the blue block.

Person: What is the pyramid supported by?

Computer: The box.



SHRDLU (1971)

Person: Pick up a big red block.

Computer: OK.

Person: Grasp the pyramid.

Computer: I don't understand which pyramid you mean.

Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.

Computer: By "it", I assume you mean the block which is taller than the one I am holding.

Computer: OK.

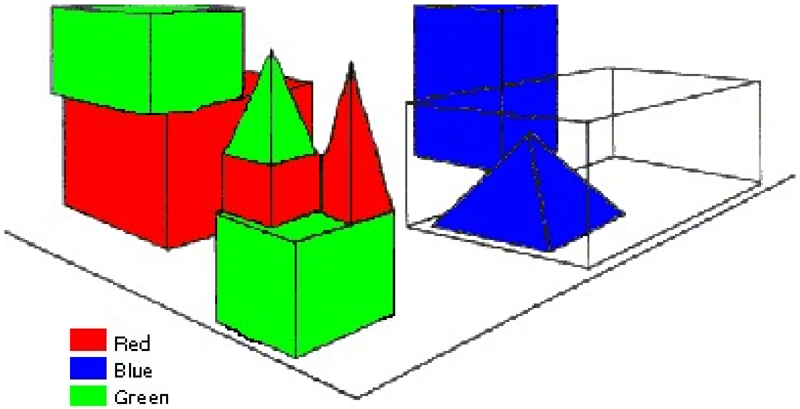
Person: What does the box contain?

Computer: The blue pyramid and the blue block.

Person: What is the pyramid supported by?

Computer: The box.

end-to-end



The Complexity Barrier

A number of people have suggested to me that large programs like the SHRDLU program for understanding natural language represent a kind of dead end in AI programming. Complex interactions between its components give the program much of its power, but at the same time they present a formidable obstacle to understanding and extending it. In order to grasp any part, it is necessary to understand how it fits with other parts, presents a dense mass, with no easy footholds. Even having written the program, I find it near the limit of what I can keep in mind at once.

— Terry Winograd (1972)

1990s: statistical revolution

1990s: statistical revolution

Compute

1990s: statistical revolution

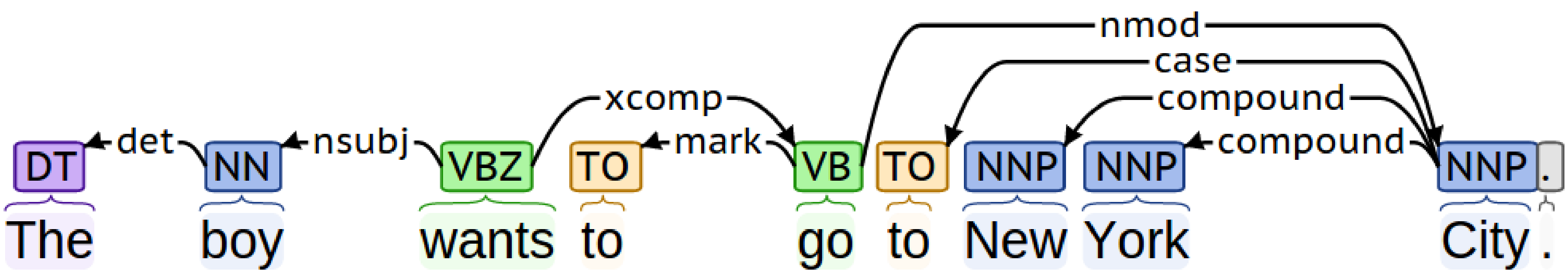
Compute

Data

[McDonald et al., 2005; de Marneffe et al. 2008]

Statistical NLP: dependency parsing

The boy wants to go to New York City.

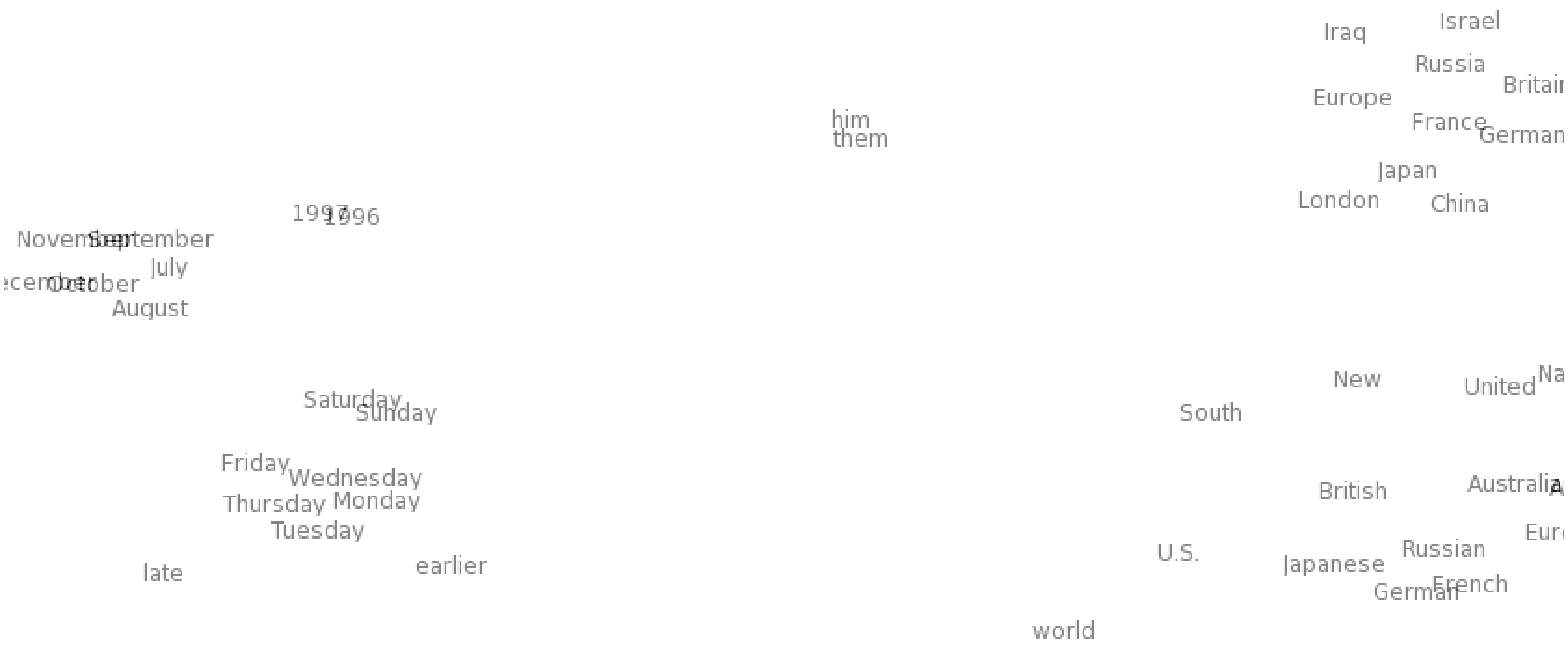


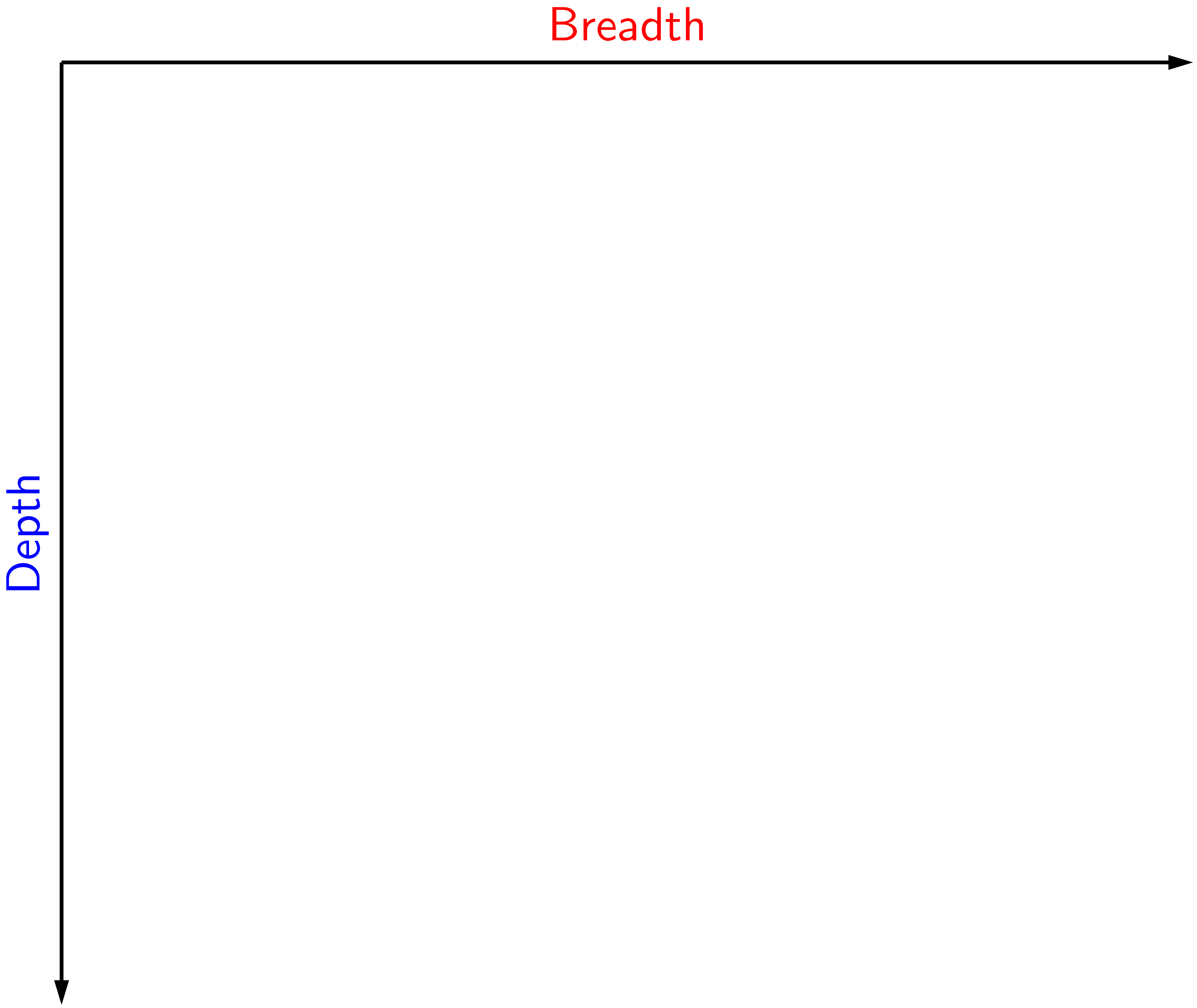
[Schütze, 1993; Bengio et al. 2003; Mikolov 2013; etc.]

Statistical NLP: word vectors

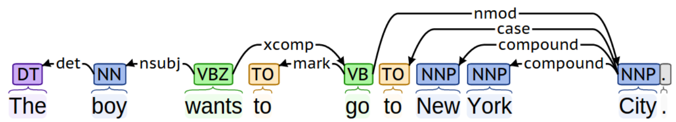
[Schütze, 1993; Bengio et al. 2003; Mikolov 2013; etc.]

Statistical NLP: word vectors




Breadth
Depth

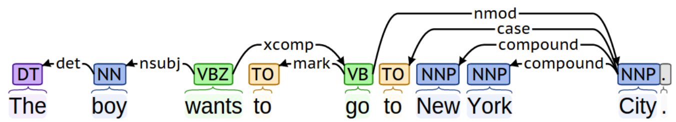
Breadth

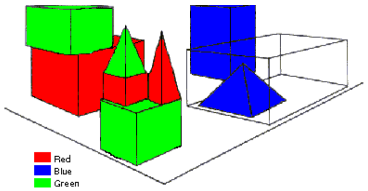


Depth

Breadth

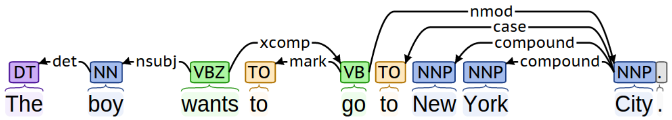
Depth

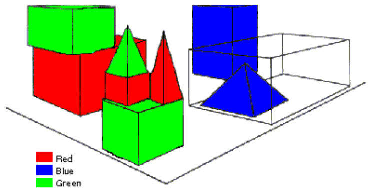


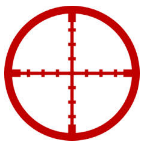


Breadth

Depth







Relevance for ML

Opportunity for transfer of ideas between ML and NLP

Relevance for ML

Opportunity for transfer of ideas between ML and NLP

- mid-1970s: **HMMs** for speech recognition $\Rightarrow$ probabilistic models

Relevance for ML

Opportunity for transfer of ideas between ML and NLP

- mid-1970s: **HMMs** for speech recognition $\Rightarrow$ probabilistic models
- early 2000s: **conditional random fields** for part-of-speech tagging $\Rightarrow$ structured prediction

Relevance for ML

Opportunity for transfer of ideas between ML and NLP

- mid-1970s: **HMMs** for speech recognition $\Rightarrow$ probabilistic models
- early 2000s: **conditional random fields** for part-of-speech tagging $\Rightarrow$ structured prediction
- early 2000s: **Latent Dirichlet Allocation** for modeling text documents $\Rightarrow$ topic modeling

Relevance for ML

Opportunity for transfer of ideas between ML and NLP

- mid-1970s: **HMMs** for speech recognition $\Rightarrow$ probabilistic models
- early 2000s: **conditional random fields** for part-of-speech tagging $\Rightarrow$ structured prediction
- early 2000s: **Latent Dirichlet Allocation** for modeling text documents $\Rightarrow$ topic modeling
- mid 2010s: **attention-based sequence-to-sequence models** for machine translation $\Rightarrow$ recurrent neural networks with memory

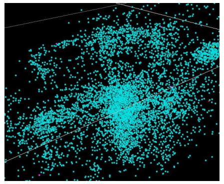
Relevance for ML

Opportunity for transfer of ideas between ML and NLP

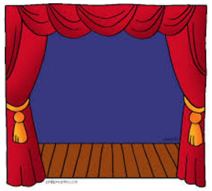
- mid-1970s: **HMMs** for speech recognition $\Rightarrow$ probabilistic models
- early 2000s: **conditional random fields** for part-of-speech tagging $\Rightarrow$ structured prediction
- early 2000s: **Latent Dirichlet Allocation** for modeling text documents $\Rightarrow$ topic modeling
- mid 2010s: **attention-based sequence-to-sequence models** for machine translation $\Rightarrow$ recurrent neural networks with memory
- now: **???** for natural language understanding

Outline

Properties of language



Distributional semantics



Frame semantics

Model-theoretic semantics

Interactive learning

Reflections

Levels of linguistic analyses

natural language utterance

Levels of linguistic analyses

$\boxed{\textbf{Syntax}\text{: what is grammatical?}}$

natural language utterance

Levels of linguistic analyses

$\boxed{\textbf{Semantics}\text{: what does it mean?}}$

$\boxed{\textbf{Syntax}\text{: what is grammatical?}}$

natural language utterance

Levels of linguistic analyses

Pragmatics: what does it do?

Semantics: what does it mean?

Syntax: what is grammatical?

natural language utterance

Analogy with programming languages

Syntax: no compiler errors

Semantics: no implementation bugs

Pragmatics: implemented the right algorithm

Analogy with programming languages

Syntax: no compiler errors

Semantics: no implementation bugs

Pragmatics: implemented the right algorithm

Different syntax, same semantics (5):

$$2 + 3 \Leftrightarrow 3 + 2$$

Analogy with programming languages

Syntax: no compiler errors

Semantics: no implementation bugs

Pragmatics: implemented the right algorithm

Different syntax, same semantics (5):

$$2 + 3 \Leftrightarrow 3 + 2$$

Same syntax, different semantics (1 and 1.5):

$$3 \;/\; 2 \text{ (Python 2.7)} \not\Leftrightarrow 3 \;/\; 2 \text{ (Python 3)}$$

Analogy with programming languages

Syntax: no compiler errors

Semantics: no implementation bugs

Pragmatics: implemented the right algorithm

Different syntax, same semantics (5):

$$2 + 3 \Leftrightarrow 3 + 2$$

Same syntax, different semantics (1 and 1.5):

$$3 \;/\; 2 \text{ (Python 2.7)} \not\Leftrightarrow 3 \;/\; 2 \text{ (Python 3)}$$

Good semantics, bad pragmatics:

correct implementation of deep neural network
for estimating coin flip prob.

What's a word?

light

What's a word?

light

Multi-word expressions: meaning unit beyond a word

light bulb

What's a word?

light

Multi-word expressions: meaning unit beyond a word

light bulb

Morphology: meaning unit within a word

light *lighten* *lightening* *relight*

What's a word?

light

Multi-word expressions: meaning unit beyond a word

light bulb

Morphology: meaning unit within a word

light *lighten* *lightening* *relight*

Polysemy: one word has multiple meanings (**word senses**)

- *The* ***light*** *was filtered through a soft glass window.*
- *He stepped into the* ***light****.*
- *This lamp* ***lights*** *up the room.*
- *The load is not* ***light****.*

Synonymy

Words:

confusing

Synonymy

Words:

confusing *unclear* *perplexing* *mystifying*

Synonymy

Words:

confusing *unclear* *perplexing* *mystifying*

Sentences:

I have fond memories of my childhood.

I reflect on my childhood with a certain fondness.

I enjoy thinking back to when I was a kid.

Synonymy

Words:

confusing *unclear* *perplexing* *mystifying*

Sentences:

I have fond memories of my childhood.

I reflect on my childhood with a certain fondness.

I enjoy thinking back to when I was a kid.

Beware: no true equivalence due to subtle diferences in meaning; think **distance metric**

Synonymy

Words:

confusing   *unclear*   *perplexing*   *mystifying*

Sentences:

I have fond memories of my childhood.

I reflect on my childhood with a certain fondness.

I enjoy thinking back to when I was a kid.

Beware: no true equivalence due to subtle diferences in meaning; think **distance metric**

But there's more to meaning than similarity...

Other lexical relations

Hyponymy (is-a):

a **cat** is a **mammal**

Other lexical relations

Hyponymy (is-a):

a **cat** is a **mammal**

Meronomy (has-a):

a **cat** has a **tail**

Other lexical relations

Hyponymy (is-a):

a **cat** is a **mammal**

Meronomy (has-a):

a **cat** has a **tail**

Useful for **entailment**:

I am giving a talk about natural language semantics.

$\Rightarrow$

I am speaking.

Compositional semantics

Two ideas: **model theory** and **compositionality**

Model theory: sentences refer to the world

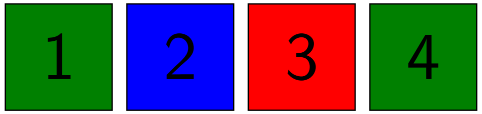
Block 2 is blue.

Compositional semantics

Two ideas: **model theory** and **compositionality**

Model theory: sentences refer to the world

Block 2 is blue.

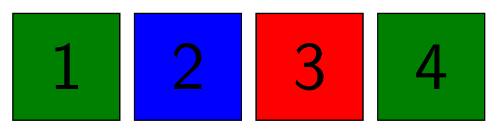


Compositional semantics

Two ideas: **model theory** and **compositionality**

Model theory: sentences refer to the world

Block 2 is blue.



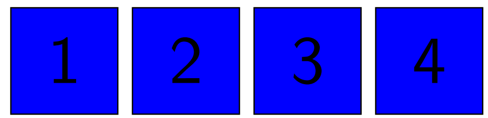
Compositionality: meaning of whole is meaning of parts

The [block left of the red block] is blue.

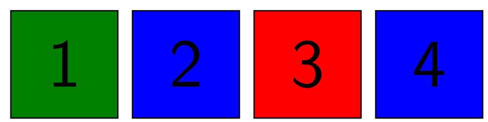
Quantifiers

Universal and existential quantification:

Every *block is blue.*



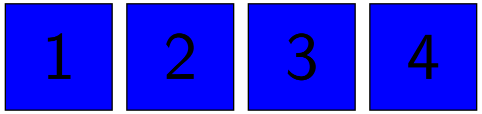
Some *block is blue.*



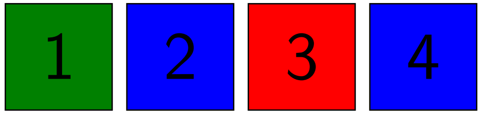
Quantifiers

Universal and existential quantification:

Every *block is blue.*



Some *block is blue.*



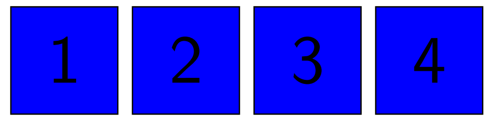
Quantifier scope ambiguity:

Every *non-blue block is next to* **some** *blue block.*

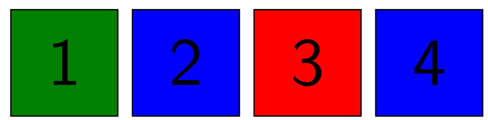
Quantifiers

Universal and existential quantification:

Every *block is blue.*



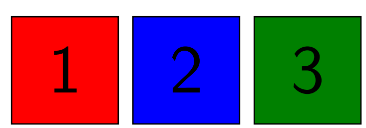
Some *block is blue.*



Quantifier scope ambiguity:

Every *non-blue block is next to* **some** *blue block.*

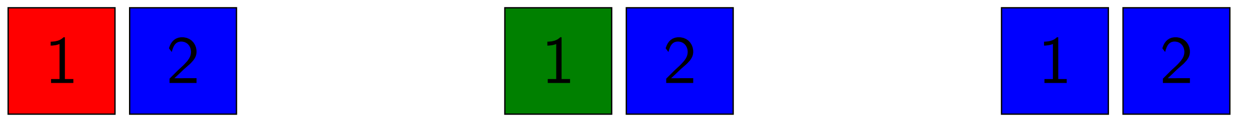
Every *non-blue block is next to* **some** *blue block.*



Multiple possible worlds

Modality:

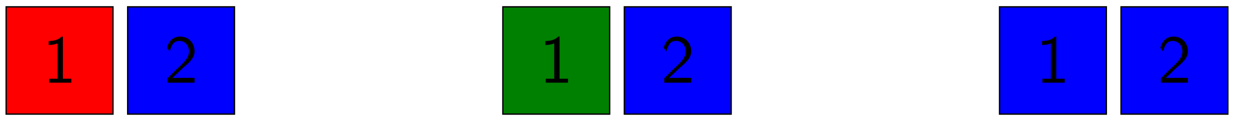
Block 2 **must** *be blue. Block 1* **could** *be red.*



Multiple possible worlds

Modality:

Block 2 **must** *be blue. Block 1* **could** *be red.*



Beliefs:

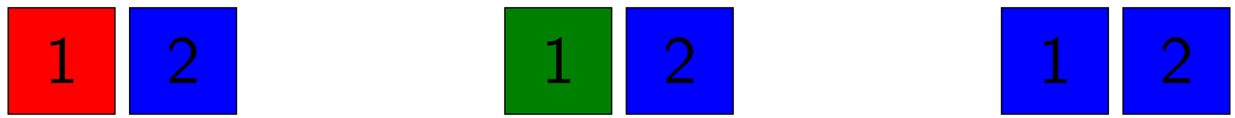
Clark Kent

Superman

Multiple possible worlds

Modality:

Block 2 **must** *be blue. Block 1* **could** *be red.*



Beliefs:

Clark Kent

Superman

Lois **believes** *Superman is a hero.*

$\neq$

Lois **believes** *Clark Kent is a hero.*

Pragmatics

Conversational implicature: new material **suggested** (not logically implied) by sentence

- *A: What on earth has happened to the roast beef?*

  B: The dog is looking very happy.

Pragmatics

Conversational implicature: new material **suggested** (not logically implied) by sentence

- *A: What on earth has happened to the roast beef?*

  B: The dog is looking very happy.

- Implicature: *The dog ate the roast beef.*

Pragmatics

Conversational implicature: new material **suggested** (not logically implied) by sentence

- *A: What on earth has happened to the roast beef?*

  B: The dog is looking very happy.

- Implicature: *The dog ate the roast beef.*

Presupposition: background **assumption** independent of truth of sentence

- *I have stopped eating meat.*

Pragmatics

Conversational implicature: new material **suggested** (not logically implied) by sentence

- *A: What on earth has happened to the roast beef?*

  B: The dog is looking very happy.
- Implicature: *The dog ate the roast beef.*

Presupposition: background **assumption** independent of truth of sentence

- *I have stopped eating meat.*
- Presupposition: *I once was eating meat.*

Pragmatics

Semantics: what does it mean **literally**?

Pragmatics: what is the speaker really conveying?

Pragmatics

Semantics: what does it mean **literally**?

Pragmatics: what is the speaker really conveying?

- Underlying principle (Grice, 1975): language is cooperative game between speaker and listener
- Implicatures and presuppositions depend on people and context and involves soft inference (machine learning opportunities here!)

Vagueness, ambiguity, uncertainty

Vagueness: does not specify full information

I had a **late** *lunch.*

Vagueness, ambiguity, uncertainty

Vagueness: does not specify full information

I had a ***late*** *lunch.*

Ambiguity: more than one possible (precise) interpretations

One morning I shot an elephant ***in*** *my pajamas.*

Vagueness, ambiguity, uncertainty

Vagueness: does not specify full information

I had a ***late*** *lunch.*

Ambiguity: more than one possible (precise) interpretations

One morning I shot an elephant ***in*** *my pajamas.*
How he got in my pajamas, I don't know. — Groucho Marx

Vagueness, ambiguity, uncertainty

Vagueness: does not specify full information

I had a ***late*** *lunch.*

Ambiguity: more than one possible (precise) interpretations

One morning I shot an elephant ***in*** *my pajamas.*
How he got in my pajamas, I don't know. — Groucho Marx

Uncertainty: due to an imperfect statistical model

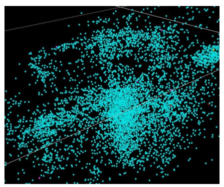
The witness was being ***contumacious****.*

Summary so far

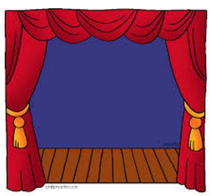
- Analyses: syntax, semantics, pragmatics
- Lexical semantics: synonymy, hyponymy/meronymy
- Compositional semantics: model theory, compositionality
- Challenges: polysemy, vagueness, ambiguity, uncertainty

Outline

Properties of language



Distributional semantics



Frame semantics

Model-theoretic semantics

Interactive learning

Reflections

Distributional semantics: warmup

The new design has _____ *lines.*

Let's try to keep the kitchen _____.

I forgot to _____ *out the cabinet.*

Distributional semantics: warmup

The new design has _____ *lines.*

Let's try to keep the kitchen _____.

I forgot to _____ *out the cabinet.*

What does _____ mean?

Distributional semantics

The new design has _____ lines.

Observation: **context** can tell us a lot about word meaning

Context: local window around a word occurrence (for now)

Distributional semantics

The new design has _____ lines.

Observation: **context** can tell us a lot about word meaning

Context: local window around a word occurrence (for now)

Roots in linguistics:

- **Distributional hypothesis**: Semantically similar words occur in similar contexts [Harris, 1954]
- "You shall know a word by the company it keeps." [Firth, 1957]

Distributional semantics

The new design has _____ lines.

Observation: **context** can tell us a lot about word meaning

Context: local window around a word occurrence (for now)

Roots in linguistics:

- **Distributional hypothesis**: Semantically similar words occur in similar contexts [Harris, 1954]
- "You shall know a word by the company it keeps." [Firth, 1957]
- Contrast: Chomsky's generative grammar (lots of hidden prior structure, no data)

Distributional semantics

The new design has _____ lines.

Observation: **context** can tell us a lot about word meaning

Context: local window around a word occurrence (for now)

Roots in linguistics:

- **Distributional hypothesis**: Semantically similar words occur in similar contexts [Harris, 1954]
- "You shall know a word by the company it keeps." [Firth, 1957]
- Contrast: Chomsky's generative grammar (lots of hidden prior structure, no data)

Upshot: **data-driven**!

General recipe

1. Form a **word-context matrix** of counts (data)

context c

word w N

General recipe

1. Form a **word-context matrix** of counts (data)

context c

word w N

2. Perform **dimensionality reduction** (generalize)

word w Θ $\Rightarrow$ word vectors $\theta_w \in \mathbb{R}^d$

[Deerwater/Dumais/Furnas/Landauer/Harshman, 1990]

Latent semantic analysis

Data:

Doc1: *Cats have tails.*

Doc2: *Dogs have tails.*

Latent semantic analysis

Data:

Doc1: *Cats have tails.*

Doc2: *Dogs have tails.*

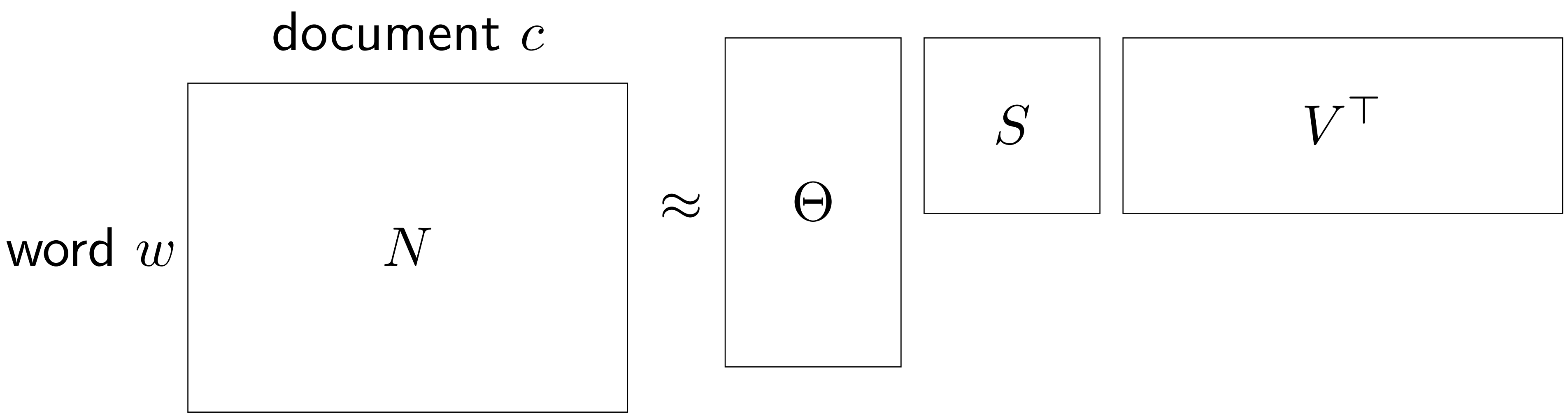
Matrix: contexts = **documents** that word appear in

	Doc1	Doc2
cats	1	0
dogs	0	1
have	1	1
tails	1	1

[Deerwater/Dumais/Furnas/Landauer/Harshman, 1990]

Latent semantic analysis

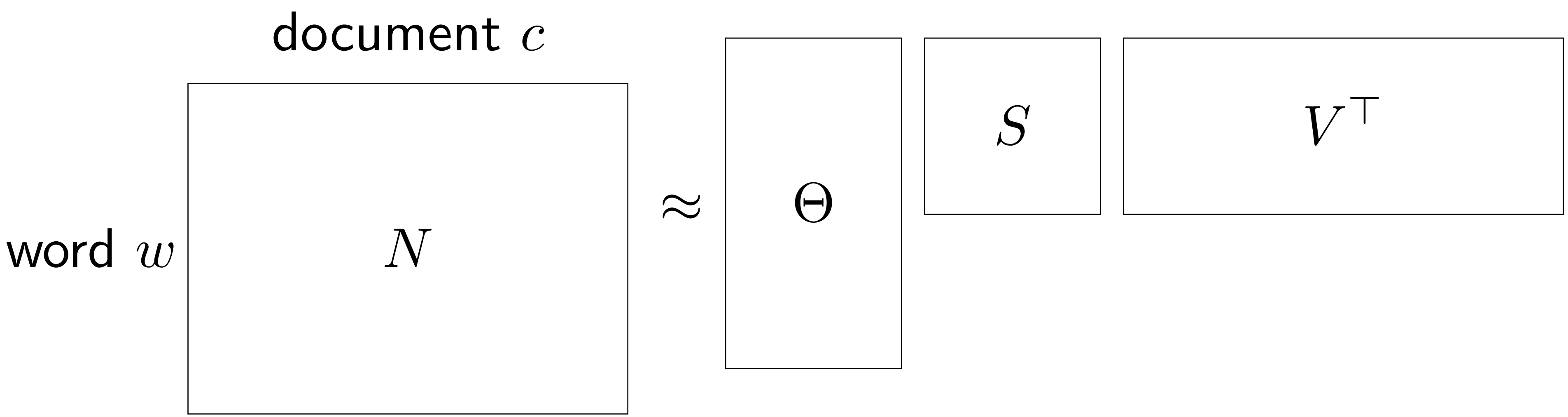
Dimensionality reduction: **SVD**



[Deerwater/Dumais/Furnas/Landauer/Harshman, 1990]

Latent semantic analysis

Dimensionality reduction: **SVD**



- Used for information retrieval
- Match query to documents in latent space rather than on keywords

Skip-gram model with negative sampling

Data:

Cats and dogs have tails.

[Mikolov/Sutskever/Chen/Corrado/Dean, 2013 (word2vec)]

Skip-gram model with negative sampling

Data:

Cats and dogs have tails.

Form matrix: contexts = words in a window

	cats	and	dogs	have	tails
cats	0	1	0	0	0
and	1	0	1	0	0
dogs	0	1	0	1	0
have	0	0	1	0	1
tails	0	0	0	1	0

Skip-gram model with negative sampling

Dimensionality reduction: **logistic regression with SGD**

[Mikolov/Sutskever/Chen/Corrado/Dean, 2013 (word2vec)]

Skip-gram model with negative sampling

Dimensionality reduction: **logistic regression with SGD**

Model: predict good (w, c) using logistic regression

$$p_\theta(g = 1 \mid w, c) = (1 + \exp(\theta_w \cdot \beta_c))^{-1}$$

Skip-gram model with negative sampling

Dimensionality reduction: **logistic regression with SGD**

Model: predict good (w, c) using logistic regression

$$p_\theta(g = 1 \mid w, c) = (1 + \exp(\theta_w \cdot \beta_c))^{-1}$$

Positives: (w, c) from data

Negatives: (w, c') for irrelevant c' (k times more)

+(cats, AI) −(cats, linguistics) −(cats, statistics)

Other models

Multinomial models:

- HMM word clustering [Brown et al., 1992]
- Latent Dirichlet Allocation [Blei et al., 2003]

Other models

Multinomial models:

- HMM word clustering [Brown et al., 1992]
- Latent Dirichlet Allocation [Blei et al., 2003]

Neural network models:

- Multi-tasking neural network [Weston/Collobert, 2008]

Other models

Multinomial models:

- HMM word clustering [Brown et al., 1992]
- Latent Dirichlet Allocation [Blei et al., 2003]

Neural network models:

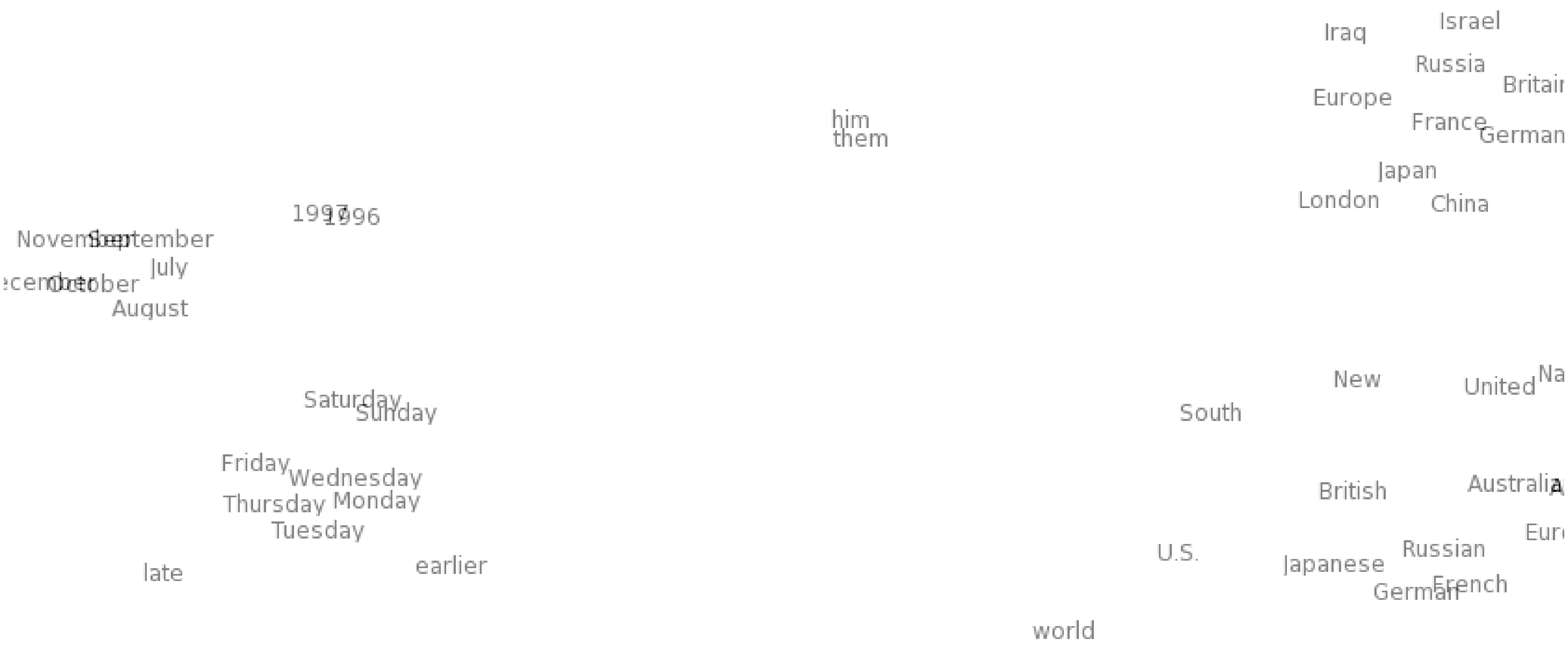
- Multi-tasking neural network [Weston/Collobert, 2008]

Recurrent/recursive models: (can embed phrases too)

- Neural language models [Bengio et al., 2003]
- Neural machine translation [Sutskever/Vinyals/Le, 2014, Cho/Merrienboer/Bahdanau/Bengio, 2014]
- Recursive neural networks [Socher/Lin/Ng/Manning, 2011]

2D visualization of word vectors

2D visualization of word vectors



Nearest neighbors

cherish

(words)

adore
love
admire
embrace
rejoice

(contexts)

cherish
both
love
pride
thy

quasi-synonyms

Nearest neighbors

cherish	**tiger**
(words)	(words)
adore	*leopard*
love	*dhole*
admire	*warthog*
embrace	*rhinoceros*
rejoice	*lion*
(contexts)	(contexts)
cherish	*tiger*
both	*leopard*
love	*panthera*
pride	*woods*
thy	*puma*
quasi-synonyms	co-hyponyms

Nearest neighbors

cherish (words)	**tiger** (words)	**good** (words)
adore	*leopard*	*bad*
love	*dhole*	*decent*
admire	*warthog*	*excellent*
embrace	*rhinoceros*	*lousy*
rejoice	*lion*	*nice*

(contexts)	(contexts)	(contexts)
cherish	*tiger*	*faith*
both	*leopard*	*natured*
love	*panthera*	*luck*
pride	*woods*	*riddance*
thy	*puma*	*both*

quasi-synonyms | co-hyponyms | includes antonyms

Nearest neighbors

cherish	**tiger**	**good**
(words)	(words)	(words)
adore	*leopard*	*bad*
love	*dhole*	*decent*
admire	*warthog*	*excellent*
embrace	*rhinoceros*	*lousy*
rejoice	*lion*	*nice*
(contexts)	(contexts)	(contexts)
cherish	*tiger*	*faith*
both	*leopard*	*natured*
love	*panthera*	*luck*
pride	*woods*	*riddance*
thy	*puma*	*both*
quasi-synonyms	co-hyponyms	includes antonyms

Many things under **semantic similarity**!

Effect of context

Suppose *Barack Obama* always appear together (a **collocation**).

Effect of context

Suppose *Barack Obama* always appear together (a **collocation**).

Global context (document):

- same context $\Rightarrow \theta_{\text{Barack}}$ close to θ_{Obama}
- so-called more "semantic"

Effect of context

Suppose *Barack Obama* always appear together (a **collocation**).

Global context (document):

- same context $\Rightarrow \theta_{\text{Barack}}$ close to θ_{Obama}
- so-called more "semantic"

Local context (neighbors):

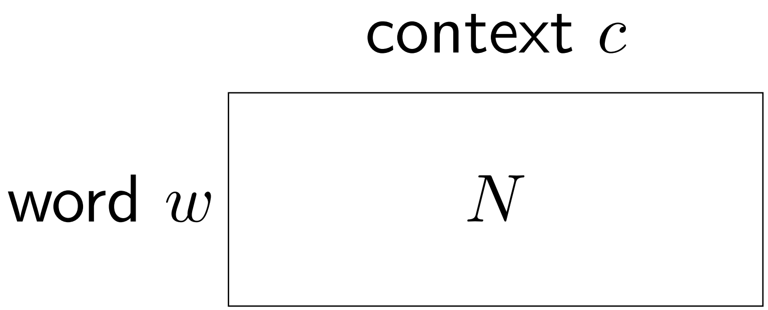
- different context $\Rightarrow \theta_{\text{Barack}}$ far from θ_{Obama}
- so-called more "syntactic"

Summary so far

- Premise: semantics = context of word/phrase

Summary so far

- Premise: semantics = context of word/phrase
- Recipe: form word-context matrix + dimensionality reduction



Summary so far

- Premise: semantics = context of word/phrase
- Recipe: form word-context matrix + dimensionality reduction

context c

word w $\boxed{N}$

Pros:

- Simple models, leverage tons of raw text
- Context captures nuanced information about usage
- Word vectors useful in downstream tasks

Food for thought

What **contexts**?

- No such thing as pure unsupervised learning, representation depends on choice of context (e.g., global/local/task-specific)

- Language is not just text in isolation, context should include world/environment

Food for thought

What **contexts**?

- No such thing as pure unsupervised learning, representation depends on choice of context (e.g., global/local/task-specific)

- Language is not just text in isolation, context should include world/environment

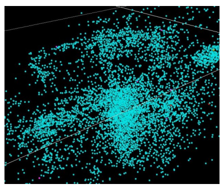
Examples to ponder:

Cynthia sold the bike for \$200.

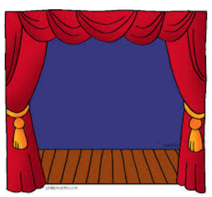
The bike sold for \$200.

Outline

Properties of language



Distributional semantics



Frame semantics

Model-theoretic semantics

Interactive learning

Reflections

Word meaning revisited

sold

Word meaning revisited

sold

Distributional semantics: all the contexts in which *sold* occurs

...was sold by... *...sold me that piece of...*

- Can find similar words/contexts and generalize (dimensionality reduction), but monolithic (no internal structure on word vectors)

Word meaning revisited

sold

Distributional semantics: all the contexts in which *sold* occurs

...was sold by... *...sold me that piece of...*

- Can find similar words/contexts and generalize (dimensionality reduction), but monolithic (no internal structure on word vectors)

Frame semantics: meaning given by a **frame**, a stereotypical situation

Commercial transaction

SELLER : ?

BUYER : ?

GOODS : ?

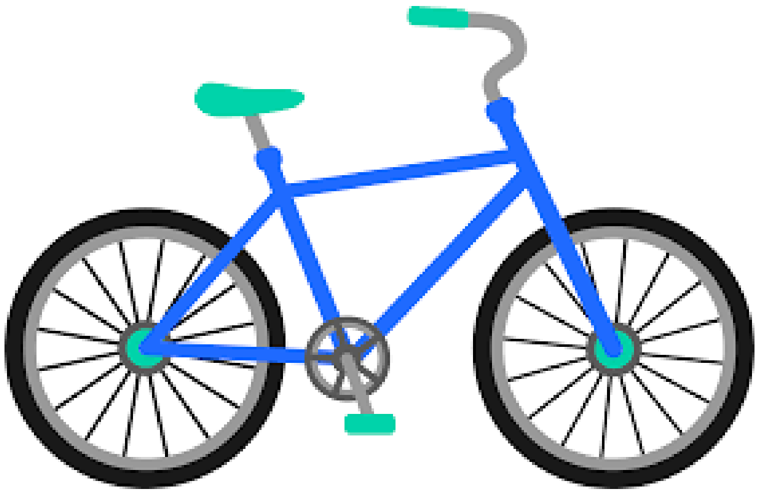
PRICE : ?

An example

Cynthia sold the bike for \$200.

An example

Cynthia sold the bike for $200.



Commercial transaction

SELLER : *Cynthia*

GOODS : *the bike*

PRICE : *$200*

Two properties of frames

Prototypical: don't need to handle all the cases

widow

[Fillmore, 1977; Langacker, 1987]

Two properties of frames

Prototypical: don't need to handle all the cases

widow

- Frame: woman marries one man, man dies

Two properties of frames

Prototypical: don't need to handle all the cases

widow

- Frame: woman marries one man, man dies
- What if a woman has 3 husbands, 2 of which died?

Two properties of frames

Prototypical: don't need to handle all the cases

widow

- Frame: woman marries one man, man dies
- What if a woman has 3 husbands, 2 of which died?

Profiling: highlight one aspect

- *sell* is seller-centric, *buy* is buyer-centric

Cynthia sold the bike (to Bob).
Bob bought the bike (from Cynthia).

Two properties of frames

Prototypical: don't need to handle all the cases

widow

- Frame: woman marries one man, man dies
- What if a woman has 3 husbands, 2 of which died?

Profiling: highlight one aspect

- *sell* is seller-centric, *buy* is buyer-centric

Cynthia sold the bike (to Bob).

Bob bought the bike (from Cynthia).

- *rob* highlights person, *steal* highlights goods

Cynthia robbed Bob (of the bike).

Cynthia stole the bike (from Bob).

Historical developments

Linguistics:

- Case grammar [Fillmore, 1968]: introduced idea of deep semantic roles (agents, themes, patients) which are tied to surface syntax (subjects, objects)

Historical developments

Linguistics:

- Case grammar [Fillmore, 1968]: introduced idea of deep semantic roles (agents, themes, patients) which are tied to surface syntax (subjects, objects)

AI / cognitive science:

- Frames [Minsky, 1975]: "a data-structure for representing a stereotyped situation, like...a child's birthday party"

Historical developments

Linguistics:

- Case grammar [Fillmore, 1968]: introduced idea of deep semantic roles (agents, themes, patients) which are tied to surface syntax (subjects, objects)

AI / cognitive science:

- Frames [Minsky, 1975]: "a data-structure for representing a stereotyped situation, like...a child's birthday party"
- Scripts [Schank & Abelson, 1977]: represent procedural knowledge (going to a restaurant)

Historical developments

Linguistics:

- Case grammar [Fillmore, 1968]: introduced idea of deep semantic roles (agents, themes, patients) which are tied to surface syntax (subjects, objects)

AI / cognitive science:

- Frames [Minsky, 1975]: "a data-structure for representing a stereotyped situation, like...a child's birthday party"
- Scripts [Schank & Abelson, 1977]: represent procedural knowledge (going to a restaurant)
- Frames [Fillmore, 1977]: coherent individuatable perception, memory, experience, action, or object

Historical developments

Linguistics:

- Case grammar [Fillmore, 1968]: introduced idea of deep semantic roles (agents, themes, patients) which are tied to surface syntax (subjects, objects)

AI / cognitive science:

- Frames [Minsky, 1975]: "a data-structure for representing a stereotyped situation, like...a child's birthday party"
- Scripts [Schank & Abelson, 1977]: represent procedural knowledge (going to a restaurant)
- Frames [Fillmore, 1977]: coherent individuatable perception, memory, experience, action, or object

NLP:

- FrameNet (1998) and PropBank (2002)

From syntax to semantics

Commercial transaction

SELLER : *Cynthia*
BUYER : *Bob*
GOODS : *the bike*
PRICE : *$200*

From syntax to semantics

Commercial transaction

SELLER : *Cynthia*
BUYER : *Bob*
GOODS : *the bike*
PRICE : *\$200*

Many **syntactic alternations** with different arguments/verbs:

Cynthia sold the bike to Bob for \$200.

The bike sold for \$200.

From syntax to semantics

Commercial transaction

SELLER : *Cynthia*
BUYER : *Bob*
GOODS : *the bike*
PRICE : *$200*

Many **syntactic alternations** with different arguments/verbs:

Cynthia sold the bike to Bob for $200.

The bike sold for $200.

Bob bought the bike from Cynthia.

The bike was bought by Bob.

The bike was bought for $200.

The bike was bought for $200 by Bob.

From syntax to semantics

Commercial transaction

SELLER : *Cynthia*

BUYER : *Bob*

GOODS : *the bike*

PRICE : *$200*

Many **syntactic alternations** with different arguments/verbs:

Cynthia sold the bike to Bob for $200.

The bike sold for $200.

Bob bought the bike from Cynthia.

The bike was bought by Bob.

The bike was bought for $200.

The bike was bought for $200 by Bob.

Goal: **syntactic positions** $\Rightarrow$ **semantic roles**

Semantic role labeling

Task:

Input: *Cynthia sold the bike to Bob for* $\$$*200*

Semantic role labeling

Task:

Input:	*Cynthia*	*sold*	*the bike*	*to*	*Bob*	*for*	*$200*
Output:	SELLER	**PREDICATE**	GOODS		BUYER		PRICE

Semantic role labeling

Task:

Input:	*Cynthia*	*sold*	*the bike*	*to*	*Bob*	*for*	*$200*
Output:	SELLER	**PREDICATE**	GOODS		BUYER		PRICE

Subtasks:

1. Frame identification (PREDICATE)
2. Argument identification (SELLER, GOODS, etc.)

A brief history

- First system (on FrameNet) [Gildea/Jurafsky, 2002]
- CoNLL shared tasks [2004, 2005]
- Use ILP to enforce constraints on arguments [Punyakanok/Roth/Yih, 2008]
- No feature engineering or parse trees [Collobert/Weston, 2008]
- Semi-supervised frame identification [Das/Smith, 2011]
- Embeddings for frame identification [Hermann/Das/Weston/Ganchev, 2014]
- Dynamic programming for some argument constraints [Tackstrom/Ganchev/Das, 2015]

Abstract meaning representation (AMR)

Semantic role labeling:

- predicate + semantic roles

[Banarescu et al., 2013]

Abstract meaning representation (AMR)

Semantic role labeling:

- predicate + semantic roles

Named-entity recognition:

$\overbrace{\text{Cynthia}}^{\text{Person}}$ went back to $\overbrace{\text{Lille}}^{\text{Loc}}$ because she liked it.

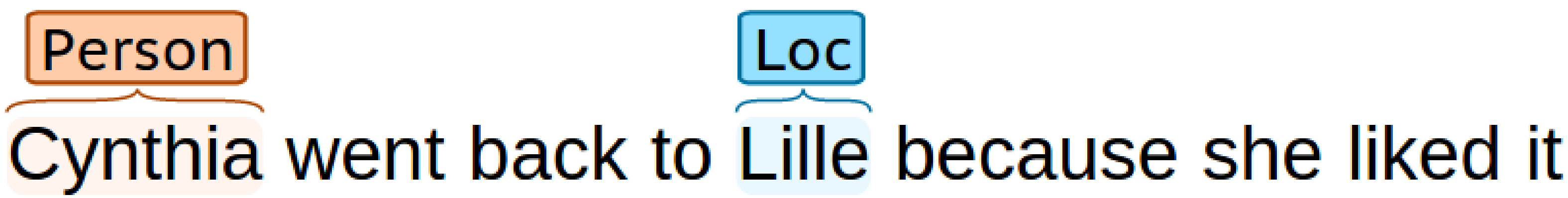
[Banaresku et al., 2013]

Abstract meaning representation (AMR)

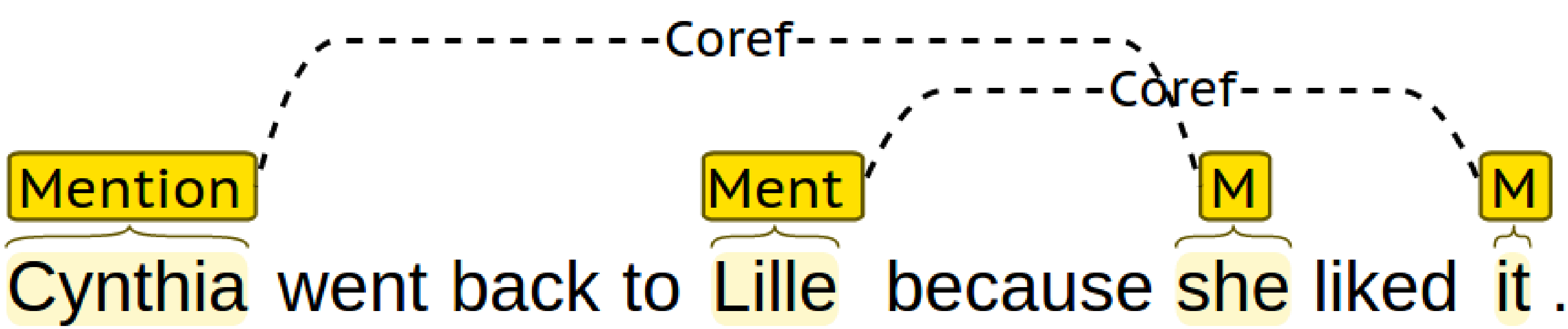
Semantic role labeling:

- predicate + semantic roles

Named-entity recognition:



Coreference resolution:

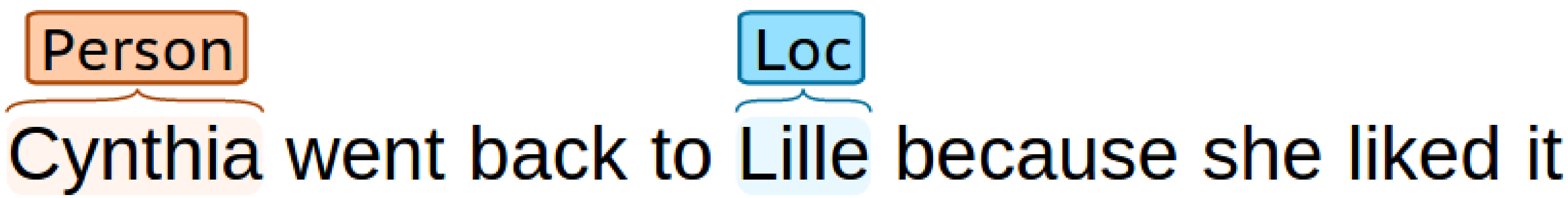


Abstract meaning representation (AMR)

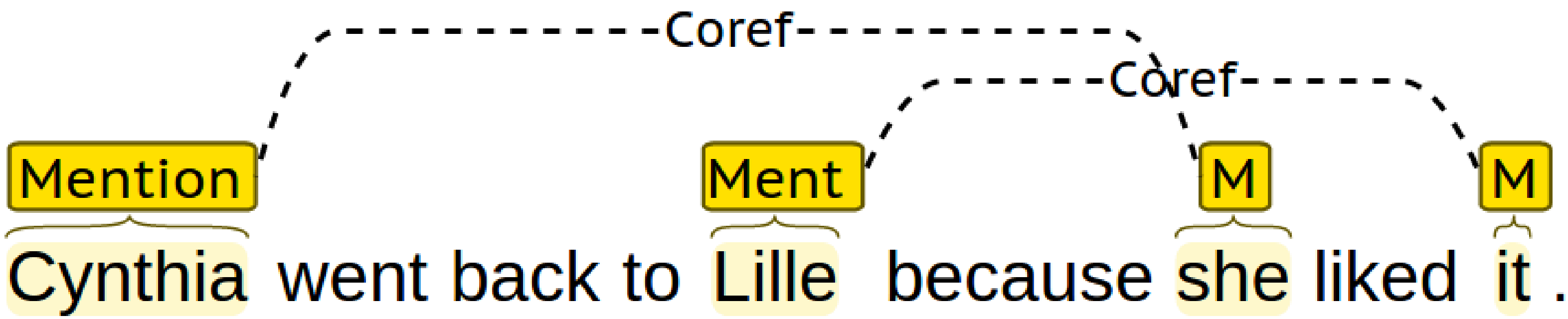
Semantic role labeling:

- predicate + semantic roles

Named-entity recognition:



Coreference resolution:



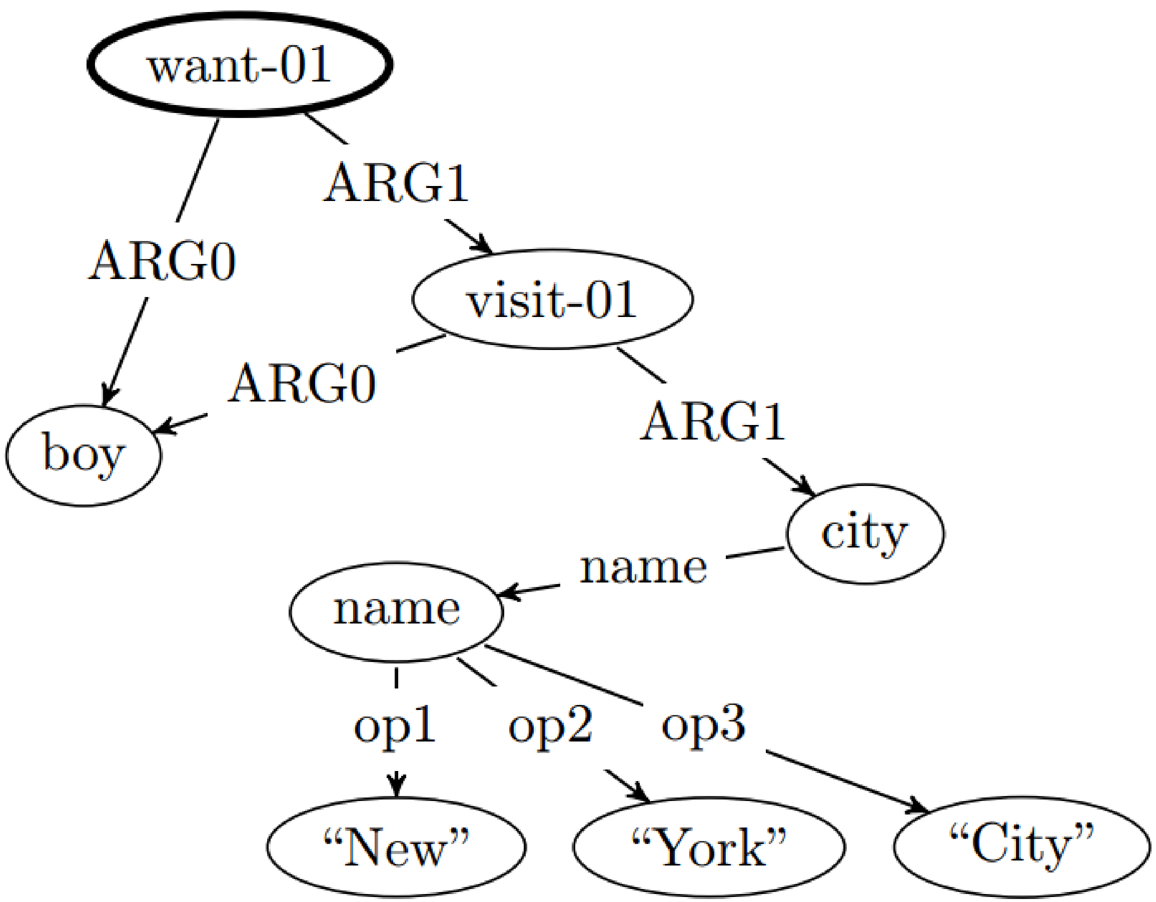
Motivation of AMR: **unify all semantic annotation**

AMR parsing task

Input: sentence

The boy wants to go to New York City.

Output: graph



Summary so far

- Frames: stereotypical situations that provide rich structure for understanding

Summary so far

- Frames: stereotypical situations that provide rich structure for understanding

- Semantic role labeling (FrameNet, PropBank): resource and task that operationalize frames

- AMR graphs: unified broad-coverage semantic annotation

Summary so far

- Frames: stereotypical situations that provide rich structure for understanding
- Semantic role labeling (FrameNet, PropBank): resource and task that operationalize frames
- AMR graphs: unified broad-coverage semantic annotation
- Methods: classification (featurize a structured object), structured prediction (not a tractable structure)

Food for thought

- Both distributional semantics (DS) and frame semantics (FS) involve compression/abstraction

- Frame semantics exposes more structure, more tied to an external world, but requires more supervision

Food for thought

- Both distributional semantics (DS) and frame semantics (FS) involve compression/abstraction

- Frame semantics exposes more structure, more tied to an external world, but requires more supervision

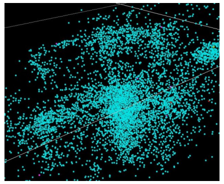
Examples to ponder:

Cynthia went to the bike shop ***yesterday****.*

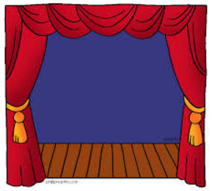
Cynthia bought the ***cheapest*** *bike.*

Outline

Properties of language



Distributional semantics



Frame semantics

Model-theoretic semantics

Interactive learning

Reflections

Types of semantics

Every *non-blue block is next to* **some** *blue block.*

Types of semantics

Every *non-blue block is next to* **some** *blue block.*

Distributional semantics: *block* is like *brick*, *some* is like *every*

Types of semantics

Every *non-blue block is next to* **some** *blue block.*

Distributional semantics: *block* is like *brick*, *some* is like *every*

Frame semantics: *is next to* has two arguments, *block* and *block*

Types of semantics

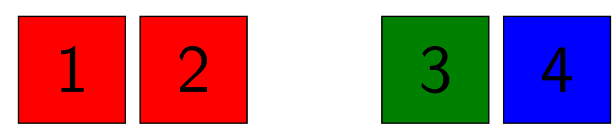
***Every** non-blue block is next to **some** blue block.*

Distributional semantics: *block* is like *brick*, *some* is like *every*

Frame semantics: *is next to* has two arguments, *block* and *block*

Model-theoretic semantics: tell the difference between

and




[database]

Executable semantic parsing

What is the largest city in Europe by population?

[database]

Executable semantic parsing

What is the largest city in Europe by population?

semantic parsing

Cities

[database]

Executable semantic parsing

What is the largest city in Europe by population?

semantic parsing

Cities Europe

[database]

Executable semantic parsing

What is the largest city in Europe by population?

semantic parsing

Cities ContainedBy(Europe)


[database]

Executable semantic parsing

What is the largest city in Europe by population?

semantic parsing

Cities ∩ ContainedBy(Europe)

[database]

Executable semantic parsing

What is the largest city in Europe by population?

$\Downarrow$ semantic parsing

$\text{Cities} \cap \text{ContainedBy(Europe)} \qquad \text{Population}$

[database]

Executable semantic parsing

What is the largest city in Europe by population?

semantic parsing

$$\mathrm{argmax}(\mathrm{Cities} \cap \mathrm{ContainedBy}(\mathrm{Europe}), \mathrm{Population})$$

[database]

Executable semantic parsing

What is the largest city in Europe by population?

$\big\downarrow$ semantic parsing

$\mathrm{argmax}(\mathrm{Cities} \cap \mathrm{ContainedBy}(\mathrm{Europe}), \mathrm{Population})$

$\big\downarrow$ execute

`Istanbul`

Executable semantic parsing

Remind me to buy milk after my last meeting on Monday.


[calendar]

Executable semantic parsing

Remind me to buy milk after my last meeting on Monday.

semantic parsing

$$\mathrm{Add}(\mathrm{Buy}(\mathrm{Milk}), \mathrm{argmax}(\mathrm{Meetings} \cap \mathrm{HasDate}(2016\text{-}07\text{-}18), \mathrm{EndTime}))$$

[calendar]

Executable semantic parsing

Remind me to buy milk after my last meeting on Monday.

semantic parsing

$\mathrm{Add}(\mathrm{Buy}(\mathrm{Milk}), \mathrm{argmax}(\mathrm{Meetings} \cap \mathrm{HasDate}(2016\text{-}07\text{-}18), \mathrm{EndTime}))$

execute

`[reminder added]`


[context]

Executable semantic parsing

[sentence]

semantic parsing

[program]

execute

`[behavior]`

A brief history of semantic parsing

GeoQuery [Zelle & Mooney 1996]

Inductive logic programming [Tang & Mooney 2001]

CCG [Zettlemoyer & Collins 2005]

String kernels [Kate & Mooney 2006]

Synchronous grammars [Wong & Mooney 2007]

Relaxed CCG [Zettlemoyer & Collins 2007]

Learning from world [Clarke et al. 2010]

Higher-order unification [Kwiatkowski et al. 2011]

Learning from answers [Liang et al. 2011]

Language + vision [Matsusek et al. 2012]

Regular expressions [Kushman et al. 2013]

Large-scale KBs [Berant et al.; Kwiatkowski et al. 2013]

Instruction following [Artzi & Zettlemoyer 2013]

Reduction to paraphrasing [Berant & Liang 2014]

Compositionality on tables [Pasupat & Liang, 2015]

Dataset from logical forms [Wang et al. 2015]

Richard Montague

Compositional semantics

cities in Europe

$\text{Cities} \cap \text{ContainedBy}(\text{Europe})$

Richard Montague

Compositional semantics

cities in Europe

Cities

cities

Cities ∩ ContainedBy(Europe)

Richard Montague

Compositional semantics

cities in Europe

Cities

cities

ContainedBy

in

Cities ∩ ContainedBy(Europe)

Richard Montague

Compositional semantics

cities in Europe

Cities

cities

ContainedBy

in

Europe

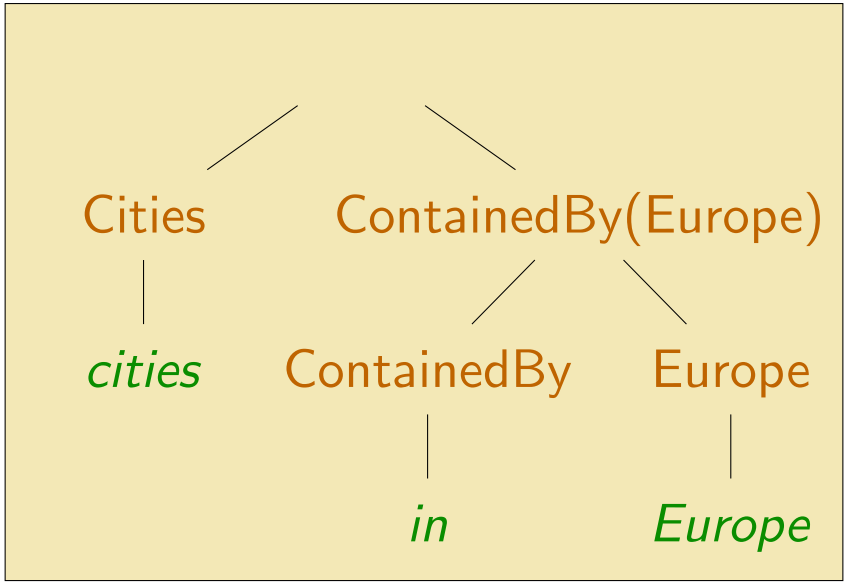
Europe

$$\text{Cities} \cap \text{ContainedBy}(\text{Europe})$$

Richard Montague

Compositional semantics

cities in Europe

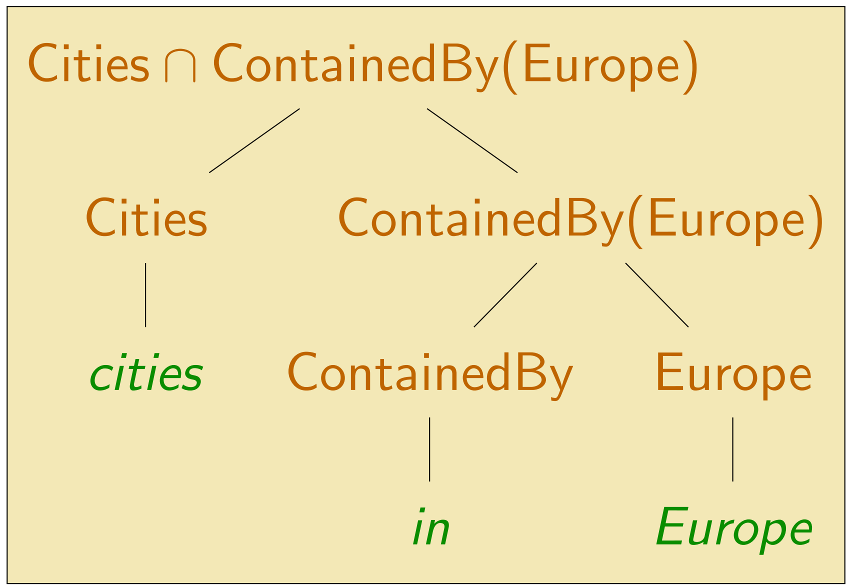


$$\text{Cities} \cap \text{ContainedBy}(\text{Europe})$$

Richard Montague

Compositional semantics

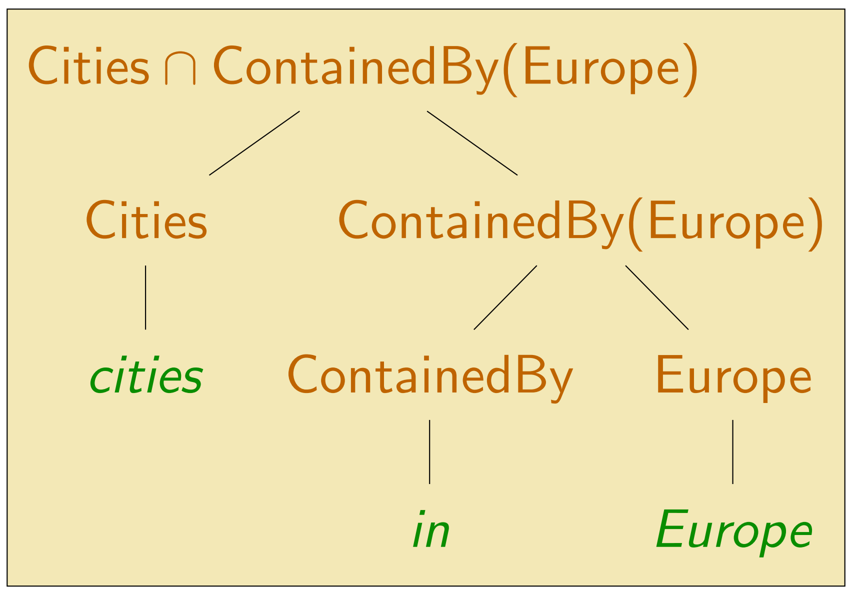
cities in Europe



Cities ∩ ContainedBy(Europe)

Language variation

cities in Europe



Cities ∩ ContainedBy(Europe)

Language variation

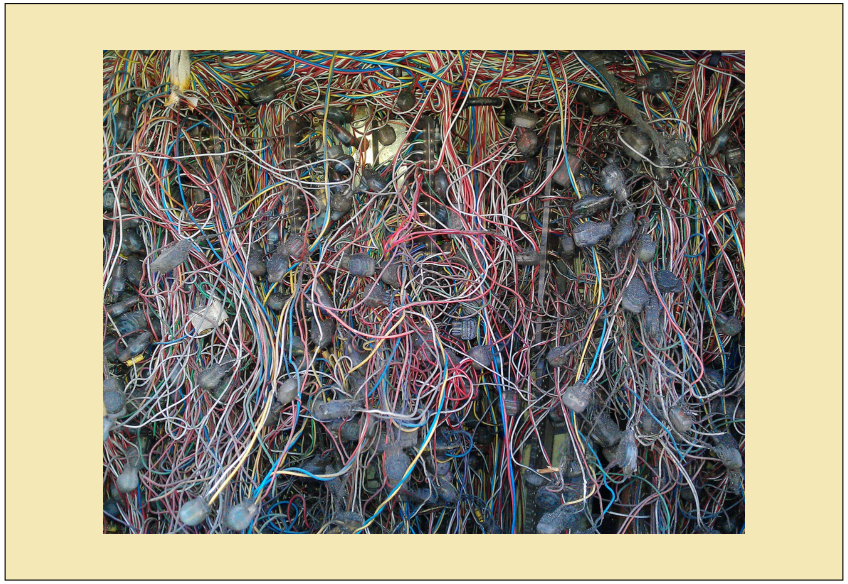
cities in Europe

European cities

cities that are in Europe

cities located in Europe

cities on the European continent



$\text{Cities} \cap \text{ContainedBy}(\text{Europe})$

Language variation

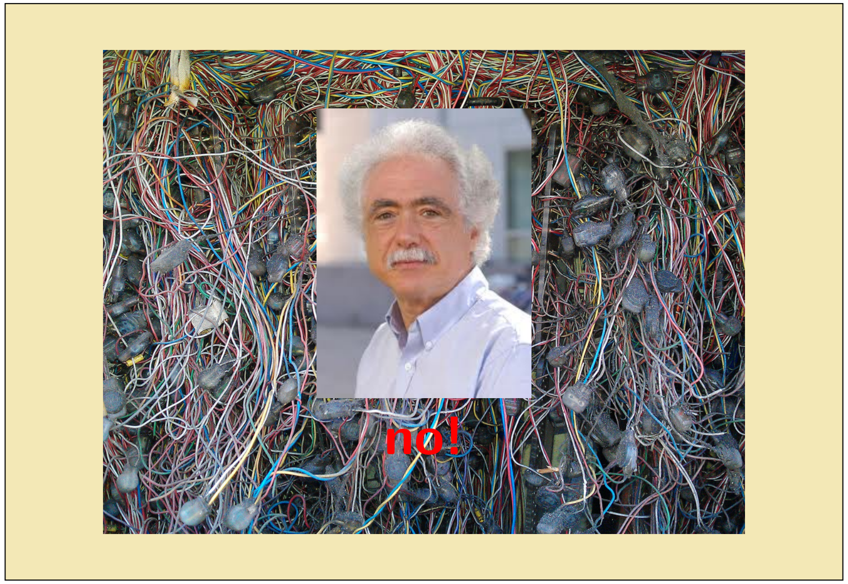
cities in Europe

European cities

cities that are in Europe

cities located in Europe

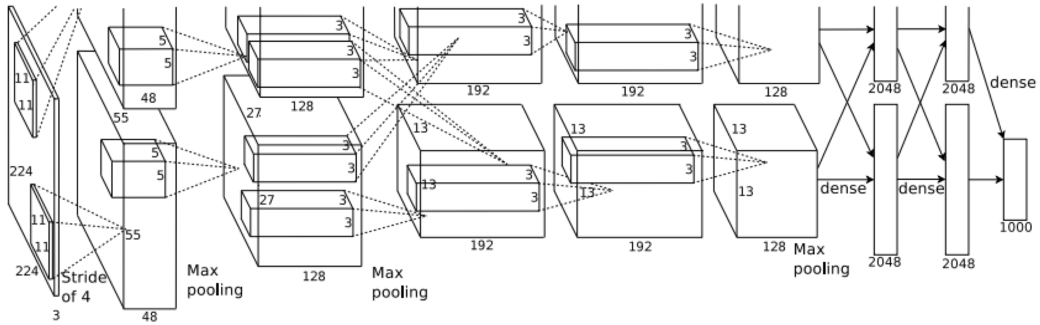
cities on the European continent



Cities $\cap$ ContainedBy(Europe)

Deep learning

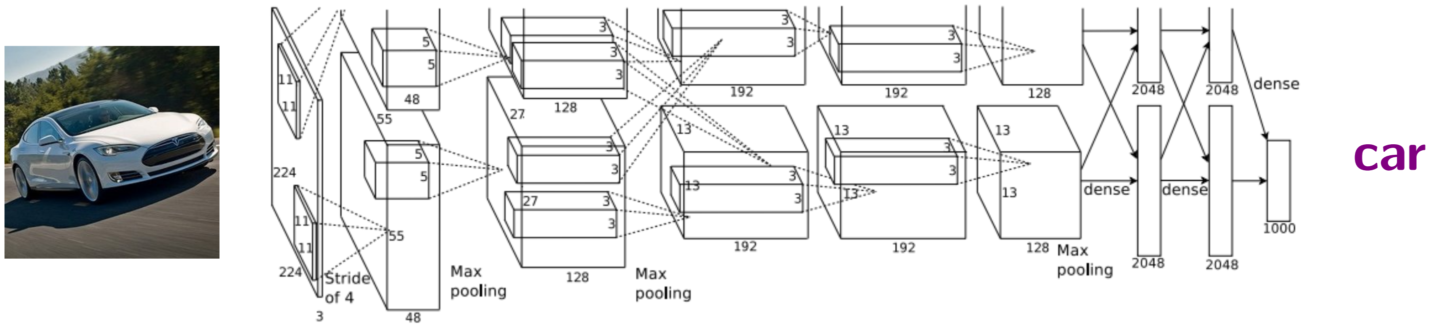
Object recognition: Krizhevsky/Sutskever/Hinton (2012)



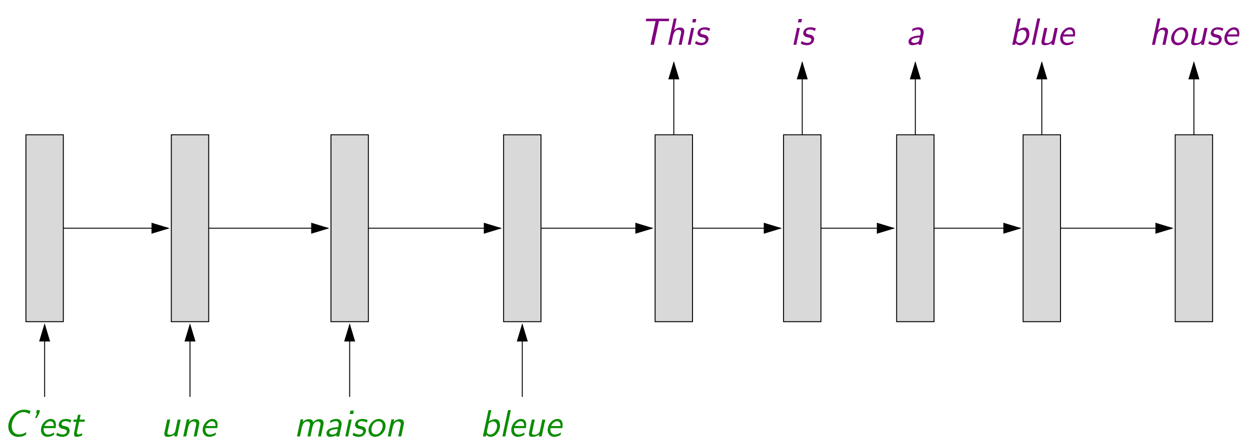
car

Deep learning

Object recognition: Krizhevsky/Sutskever/Hinton (2012)

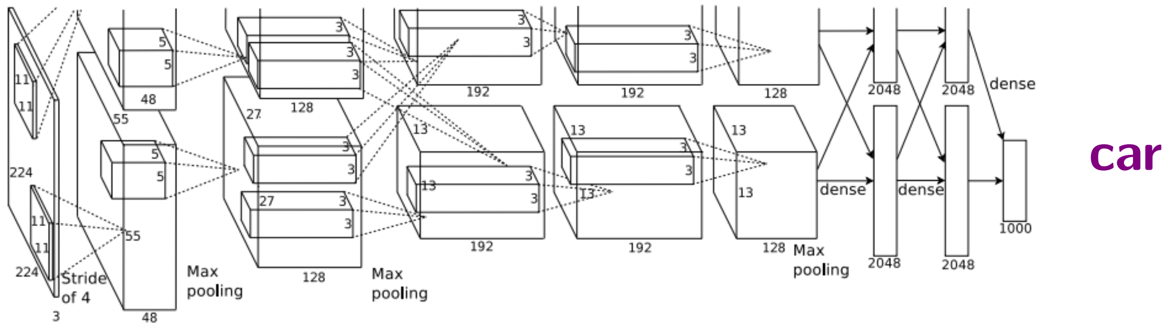


Machine translation: Sutskever/Vinyals/Le (2014)

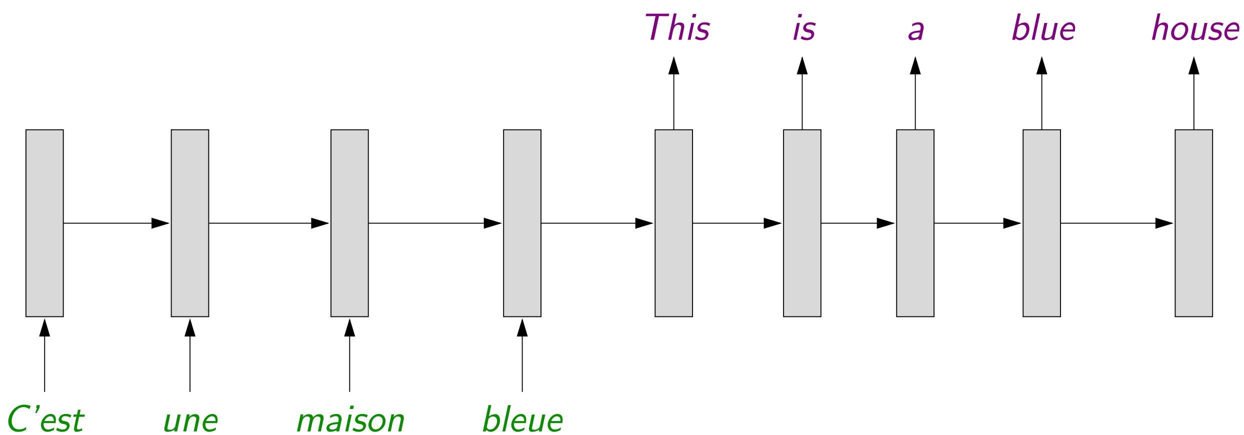


Deep learning

Object recognition: Krizhevsky/Sutskever/Hinton (2012)



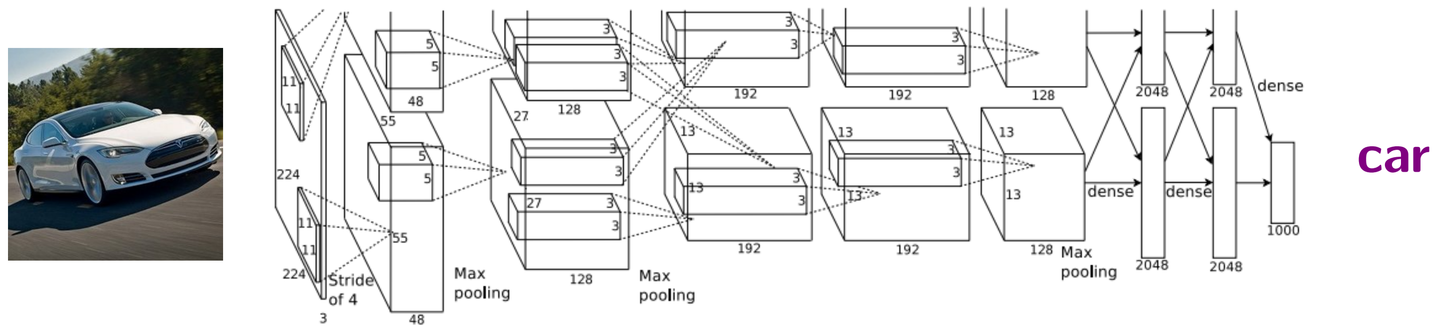
Machine translation: Sutskever/Vinyals/Le (2014)



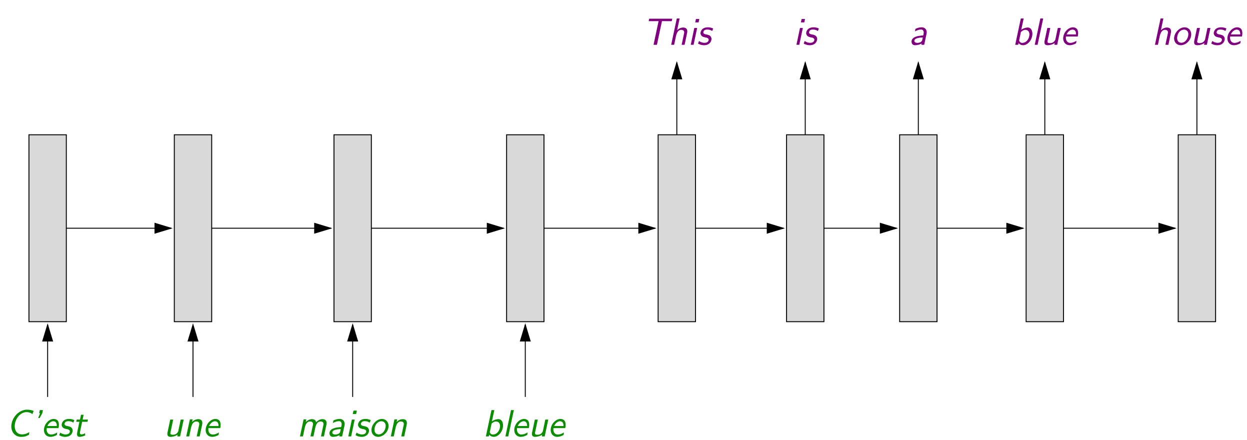
accuracy

Deep learning

Object recognition: Krizhevsky/Sutskever/Hinton (2012)



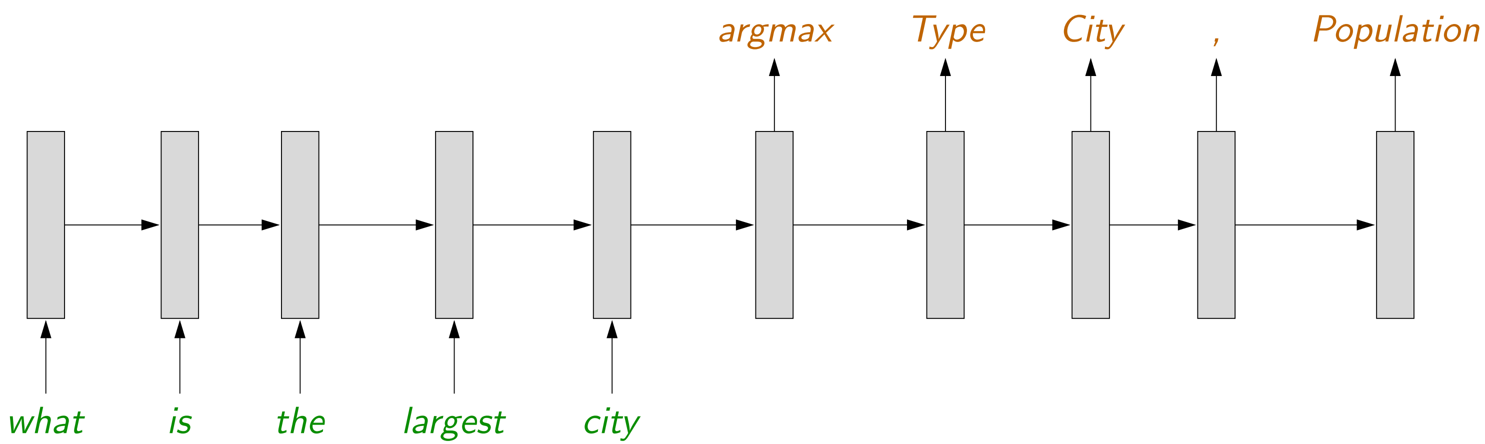
Machine translation: Sutskever/Vinyals/Le (2014)



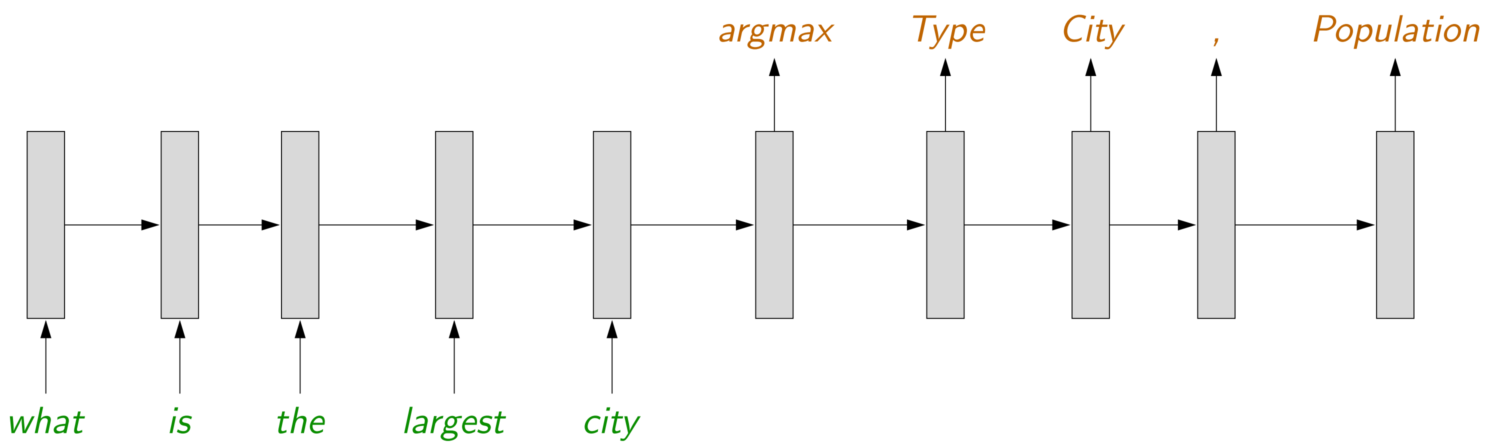
accuracy

simplicity

Neural semantic parsing



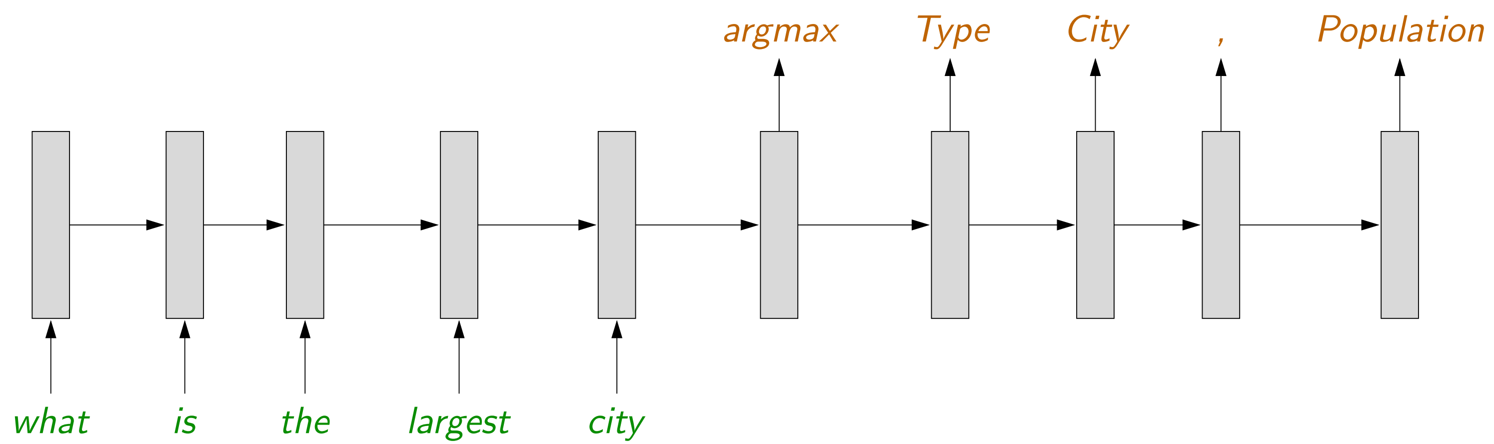
Neural semantic parsing



- Learn semantic composition without predefined grammar

[Jia & Liang, 2016]

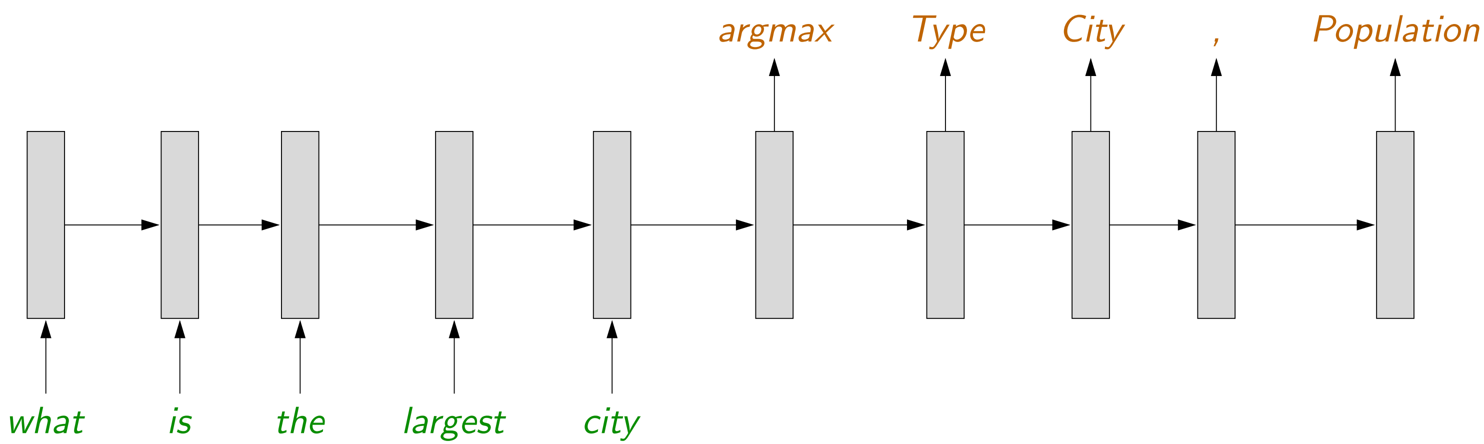
Neural semantic parsing



- Learn semantic composition without predefined grammar
- Encode compositionality through **data recombination**

what's the capital of Germany?	CapitalOf(Germany)
what countries border France?	Borders(France)

Neural semantic parsing



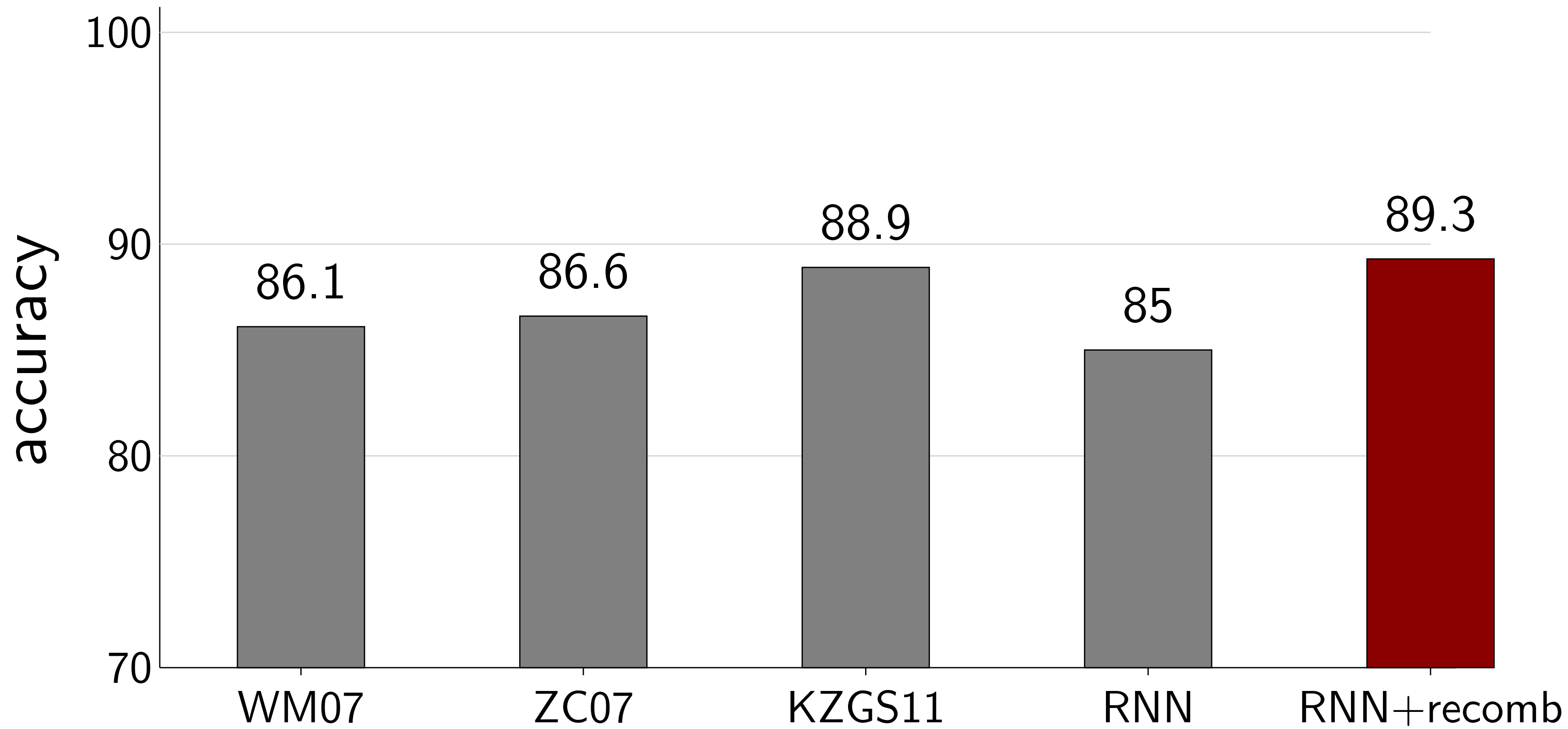
- Learn semantic composition without predefined grammar
- Encode compositionality through **data recombination**

what's the capital of Germany?	CapitalOf(Germany)
what countries border France?	Borders(France)
what's the capital of France?	CapitalOf(France)
what countries border Germany?	Borders(Germany)

Neural semantic parsing

Dataset: US Geography dataset (Zelle & Mooney, 1996)

What is the highest point in Florida?



state-of-art, simpler

[context]

Summary so far

What is the largest city in Europe by population?

semantic parsing

argmax(Cities $\cap$ ContainedBy(Europe), Population)

execute

Istanbul

- Language encodes computation
- Semantic parsing represents language as programs
- Recurrent neural networks + semantic parsing

[Zelle & Mooney, 1996; Zettlemoyer & Collins, 2005; Clarke et al. 2010; Liang et al., 2011]

Training data for semantic parsing

Heavy supervision

What's Bulgaria's capital?

CapitalOf(Bulgaria)

When was Walmart started?

DateFounded(Walmart)

What movies has Tom Cruise been in?

Movies $\cap$ Starring(TomCruise)

...

[Zelle & Mooney, 1996; Zettlemoyer & Collins, 2005; Clarke et al. 2010; Liang et al., 2011]

Training data for semantic parsing

Heavy supervision

What's Bulgaria's capital?

CapitalOf(Bulgaria)

When was Walmart started?

DateFounded(Walmart)

What movies has Tom Cruise been in?

Movies ∩ Starring(TomCruise)

...

[Zelle & Mooney, 1996; Zettlemoyer & Collins, 2005; Clarke et al. 2010; Liang et al., 2011]

Training data for semantic parsing

Heavy supervision	**Light supervision**
What's Bulgaria's capital?	*What's Bulgaria's capital?*
CapitalOf(Bulgaria)	`Sofia`
When was Walmart started?	*When was Walmart started?*
DateFounded(Walmart)	`1962`
What movies has Tom Cruise been in?	*What movies has Tom Cruise been in?*
Movies ∩ Starring(TomCruise)	`TopGun,VanillaSky,...`
...	...

[Zelle & Mooney, 1996; Zettlemoyer & Collins, 2005; Clarke et al. 2010; Liang et al., 2011]

Training data for semantic parsing

Heavy supervision	**Light supervision**
What's Bulgaria's capital?	*What's Bulgaria's capital?*
CapitalOf(Bulgaria)	`Sofia`
When was Walmart started?	*When was Walmart started?*
DateFounded(Walmart)	`1962`
What movies has Tom Cruise been in?	*What movies has Tom Cruise been in?*
Movies $\cap$ Starring(TomCruise)	`TopGun,VanillaSky,...`
...	...

[Zelle & Mooney, 1996; Zettlemoyer & Collins, 2005; Clarke et al. 2010; Liang et al., 2011]

Training data for semantic parsing

Heavy supervision

What's Bulgaria's capital?

CapitalOf(Bulgaria)

When was Walmart started?

DateFounded(Walmart)

What movies has Tom Cruise been in?

Movies $\cap$ Starring(TomCruise)

...

Light supervision

What's Bulgaria's capital?

`Sofia`

When was Walmart started?

`1962`

What movies has Tom Cruise been in?

`TopGun,VanillaSky,...`

...

Training intuition

Where did Mozart tupress?

Vienna

Training intuition

Where did Mozart tupress?

`PlaceOfBirth(WolfgangMozart)`

`PlaceOfDeath(WolfgangMozart)`

`PlaceOfMarriage(WolfgangMozart)`

Vienna

Training intuition

Where did Mozart tupress?

`PlaceOfBirth(WolfgangMozart)`	$\Rightarrow$ Salzburg
`PlaceOfDeath(WolfgangMozart)`	$\Rightarrow$ Vienna
`PlaceOfMarriage(WolfgangMozart)`	$\Rightarrow$ Vienna

Vienna

Training intuition

Where did Mozart tupress?

~~`PlaceOfBirth(WolfgangMozart)`~~	~~$\Rightarrow$ Salzburg~~
`PlaceOfDeath(WolfgangMozart)`	$\Rightarrow$ Vienna
`PlaceOfMarriage(WolfgangMozart)`	$\Rightarrow$ Vienna

Vienna

Training intuition

Where did Mozart tupress?

~~$\texttt{PlaceOfBirth(WolfgangMozart)} \Rightarrow \text{Salzburg}$~~

$\texttt{PlaceOfDeath(WolfgangMozart)} \Rightarrow \text{Vienna}$

$\texttt{PlaceOfMarriage(WolfgangMozart)} \Rightarrow \text{Vienna}$

Vienna

Where did Hogarth tupress?

Training intuition

Where did Mozart tupress?

~~PlaceOfBirth(WolfgangMozart) ⇒ Salzburg~~

PlaceOfDeath(WolfgangMozart) ⇒ Vienna

PlaceOfMarriage(WolfgangMozart) ⇒ Vienna

Vienna

Where did Hogarth tupress?

PlaceOfBirth(WilliamHogarth)

PlaceOfDeath(WilliamHogarth)

PlaceOfMarriage(WilliamHogarth)

London

Training intuition

Where did Mozart tupress?

~~`PlaceOfBirth(WolfgangMozart)` ⇒ Salzburg~~

`PlaceOfDeath(WolfgangMozart)` ⇒ Vienna

`PlaceOfMarriage(WolfgangMozart)` ⇒ Vienna

Vienna

Where did Hogarth tupress?

`PlaceOfBirth(WilliamHogarth)` ⇒ London

`PlaceOfDeath(WilliamHogarth)` ⇒ London

`PlaceOfMarriage(WilliamHogarth)` ⇒ Paddington

London

Training intuition

Where did Mozart tupress?

~~`PlaceOfBirth(WolfgangMozart)` ⇒ Salzburg~~

`PlaceOfDeath(WolfgangMozart)` ⇒ Vienna

`PlaceOfMarriage(WolfgangMozart)` ⇒ Vienna

Vienna

Where did Hogarth tupress?

`PlaceOfBirth(WilliamHogarth)` ⇒ London

`PlaceOfDeath(WilliamHogarth)` ⇒ London

~~`PlaceOfMarriage(WilliamHogarth)` ⇒ Paddington~~

London

Training intuition

Where did Mozart tupress?

~~PlaceOfBirth(WolfgangMozart) ⇒ Salzburg~~

$\boxed{\texttt{PlaceOfDeath(WolfgangMozart)} \Rightarrow \text{Vienna}}$

PlaceOfMarriage(WolfgangMozart) ⇒ Vienna

Vienna

Where did Hogarth tupress?

PlaceOfBirth(WilliamHogarth) ⇒ London

$\boxed{\texttt{PlaceOfDeath(WilliamHogarth)} \Rightarrow \text{London}}$

~~PlaceOfMarriage(WilliamHogarth) ⇒ Paddington~~

London

Searching...

Greece held its last Summer Olympics in which year?

?

2004

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...	...	...	...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

Searching...

Greece held its last Summer Olympics in which year?

$\mathbb{R}$[Index].Country.Greece

2004

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...	...	...	...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

Searching...

Greece held its last Summer Olympics in which year?

$\mathbb{R}$[Nations].Country.Greece

2004

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...	...	...	...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

Searching...

Greece held its last Summer Olympics in which year?

argmax(Country.Greece, Nations)

2004

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...	...	...	...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

Searching...

Greece held its last Summer Olympics in which year?

$$\text{argmax}(\text{Country.Greece}, \text{Index})$$

2004

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...	...	...	...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

Searching...

Greece held its last Summer Olympics in which year?

... (thousands of logical forms later) ...

2004

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...	...	...	...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

Searching...

Greece held its last Summer Olympics in which year?

$\boxed{\mathbb{R}[\text{Date}].\mathbb{R}[\text{Year}].\text{argmax}(\text{Country.Greece}, \text{Index})}$

2004

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...	...	...	...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

WikiTableQuestions

Year	Competition	Venue	Position	Event	Notes
Representing Poland					
2001	World Youth Championships	Debrecen, Hungary	2nd	400 m	47.12
			1st	Medley relay	1:50.46
	European Junior Championships	Grosseto, Italy	1st	4x400 m relay	3:06.12
2002	World Junior Championships	Kingston, Jamaica	4th	4×400m relay	3:06.25
2003	European Junior Championships	Tampere, Finland	3rd	400 m	46.69
			2nd	4x400 m relay	3:08.62
2005	European U23 Championships	Erfurt, Germany	11th (sf)	400 m	46.62
			1st	4x400 m relay	3:04.41
	Universiade	Izmir, Turkey	7th	400 m	46.89
			1st	4x400 m relay	3:02.57
2006	World Indoor Championships	Moscow, Russia	2nd (h)	4x400 m relay	3:06.10
	European Championships	Gothenburg, Sweden	3rd	4x400 m relay	3:01.73
2007	European Indoor Championships	Birmingham, United Kingdom	3rd	4x400 m relay	3:08.14
	Universiade	Bangkok, Thailand	7th	400 m	46.85
			1st	4x400 m relay	3:02.05
2008	World Indoor Championships	Valencia, Spain	4th	4x400 m relay	3:08.76
	Olympic Games	Beijing, China	7th	4x400 m relay	3:00.32
2009	Universiade	Belgrade, Serbia	2nd	4x400 m relay	3:05.69

In what city did Piotr's last 1st place finish occur?

Language & world

Talks | Spring 2017

Natural Language Understanding: Foundations and State-of-the-Art

Friday, January 27th, 2017 2:00 pm – 3:30 pm

Event: Foundations of Machine Learning Boot Camp

Add to Calendar

Speaker: Percy Liang, Stanford University

How long is Percy's talk?

Language & world

Talks | Spring 2017

Natural Language Understanding: Foundations and State-of-the-Art

Friday, January 27th, 2017 2:00 pm – 3:30 pm

Event: Foundations of Machine Learning Boot Camp

Add to Calendar

Speaker: Percy Liang, Stanford University

How long is Percy's talk?

LengthOf(PercyTalk)

Language & world

Talks | Spring 2017

Natural Language Understanding: Foundations and State-of-the-Art

Friday, January 27th, 2017 2:00 pm – 3:30 pm

Event: Foundations of Machine Learning Boot Camp

Add to Calendar

Speaker: Percy Liang, Stanford University

How long is Percy's talk?

Header(PercyTalk)[1] – Header(PercyTalk)[0]

Language & world

Talks | Spring 2017

Natural Language Understanding: Foundations and State-of-the-Art

Friday, January 27th, 2017 2:00 pm – 3:30 pm

Event: Foundations of Machine Learning Boot Camp

Add to Calendar

Speaker: Percy Liang, Stanford University

How long is Percy's talk?

Header(PercyTalk)[1] – Header(PercyTalk)[0]

natural language understanding

language	world

[context]

Summary so far

What is the largest city in Europe by population?

$\Big\downarrow$ semantic parsing

?

$\Big\downarrow$ execute

Istanbul

- Two ideas: model theory and compositionality, both about factorization / **generalization**
- Applications: question answering, natural language interfaces to robots, programming by natural language
- Search is hard, needed even to get training signal

Food for thought

- Learning from denotations is hard; implicity moving from easy to harder examples; don't have good formalism yet

Food for thought

- Learning from denotations is hard; implicity moving from easy to harder examples; don't have good formalism yet

- Semantic parsing works on short sentences (user to computer); distributional/frame semantics has broader coverage; how to bridge the gap?

Food for thought

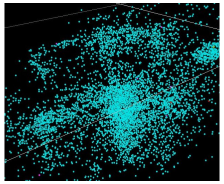
- Learning from denotations is hard; implicity moving from easy to harder examples; don't have good formalism yet

- Semantic parsing works on short sentences (user to computer); distributional/frame semantics has broader coverage; how to bridge the gap?

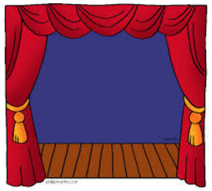
- What is the best way to produce answer from question? Logical forms are means to an end. Fully neural [Neelakantan et al. 2016]?

Outline

Properties of language



Distributional semantics



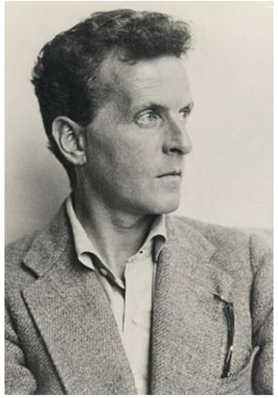
Frame semantics

Model-theoretic semantics

Interactive learning

Reflections

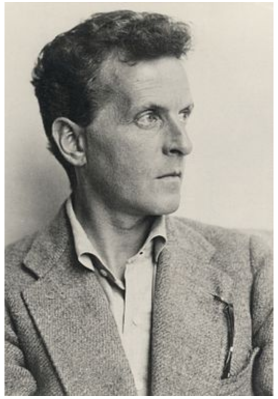
Language game



Wittgenstein (1953):

Language derives its meaning from use.

Language game

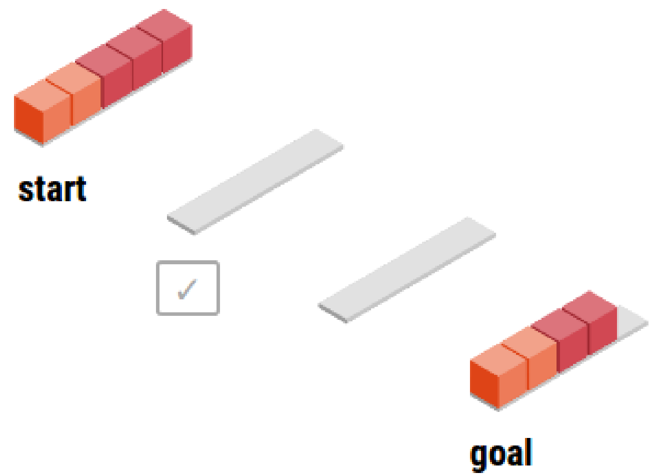


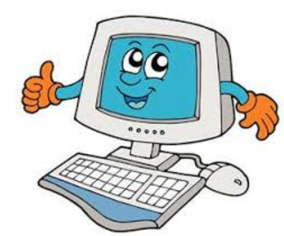
Wittgenstein (1953):

Language derives its meaning from use.

[Wang et al. 2016]

SHRDLURN





sees goal

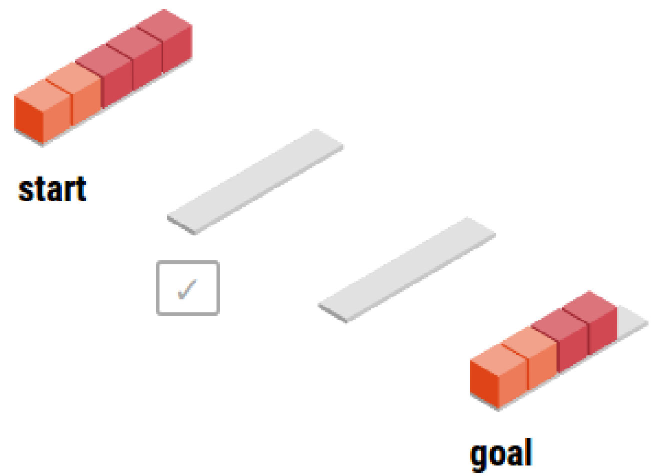
has language

perform actions

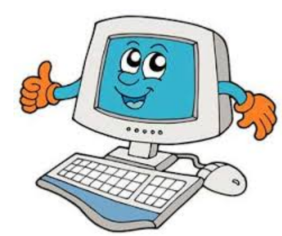
no language

[Wang et al. 2016]

SHRDLURN



remove red



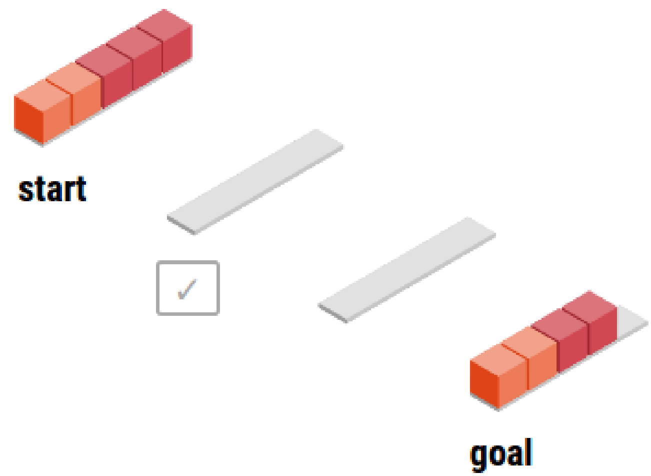
sees goal

has language

perform actions

no language

SHRDLURN



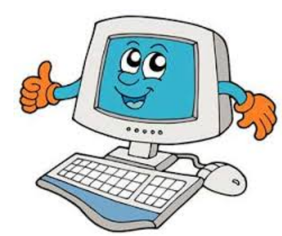
remove red

add(hascolor(red))

add(hascolor(brown))

remove(hascolor(red))

remove(hascolor(brown))



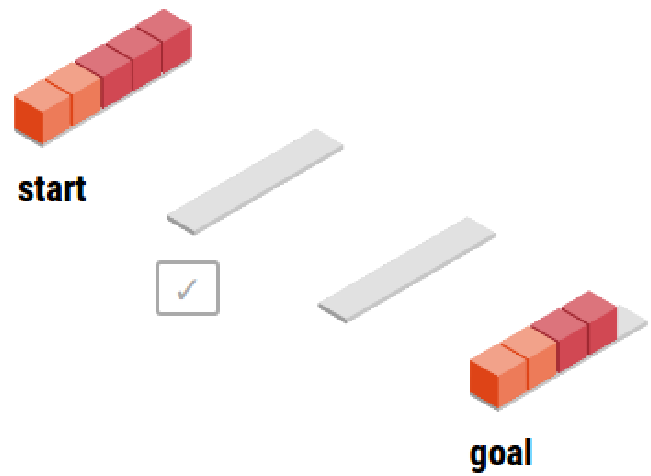
sees goal

has language

perform actions

no language

SHRDLURN



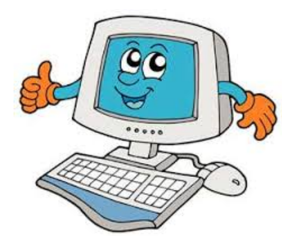
remove red

add(hascolor(red))

add(hascolor(brown))

remove(hascolor(red))

remove(hascolor(brown))



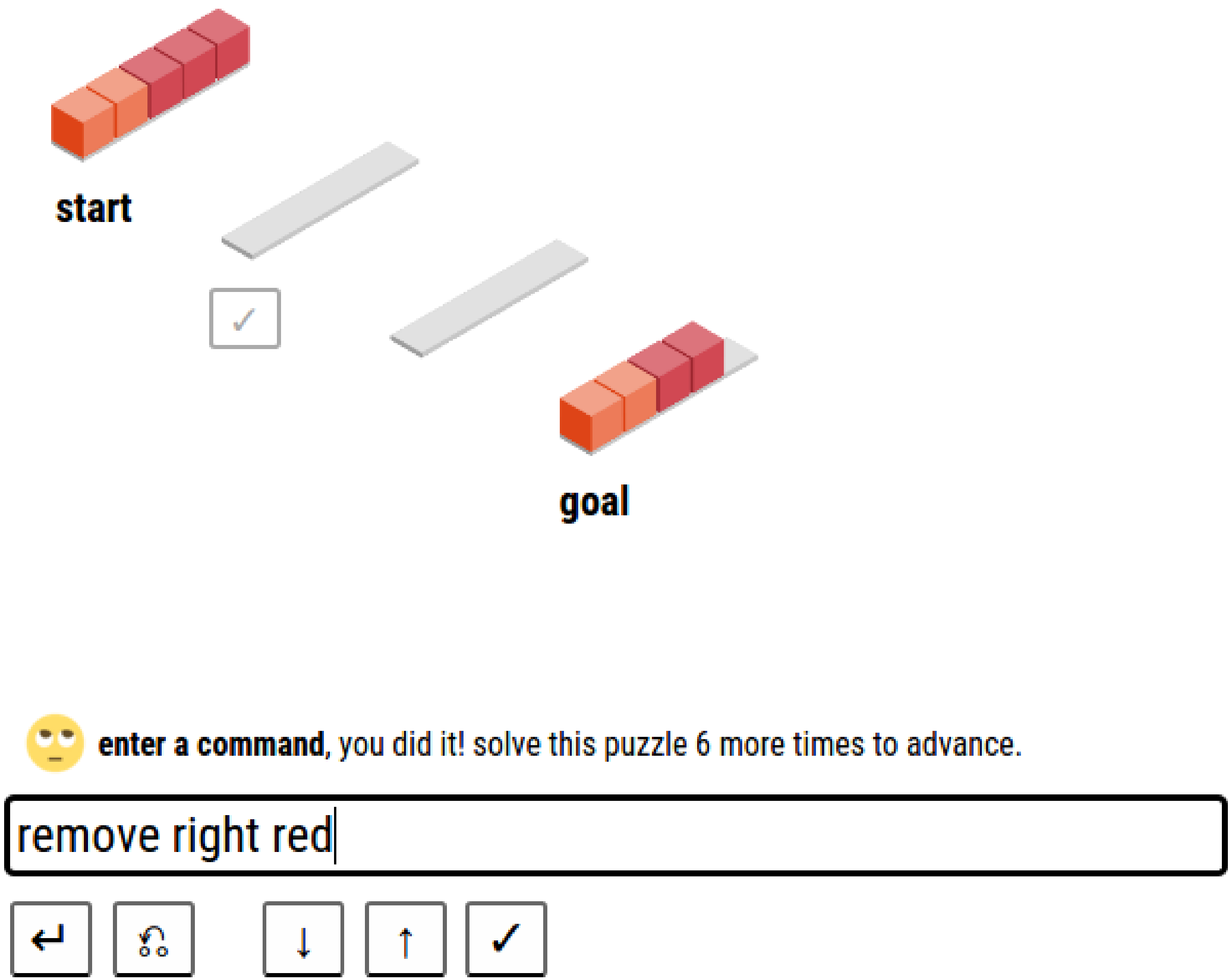
sees goal

has language

perform actions

no language

SHRDLURN



shrdlurn.sidaw.xyz/acl16

Experiments

100 players from Amazon Mechanical Turk

6 hours $\Rightarrow$ 10K utterances

Results: top players (rank 1-20)

(3.01)

rem cy pos 1
stack or blk pos 4
rem blk pos 2 thru 5
rem blk pos 2 thru 4
stack bn blk pos 1 thru 2
fill bn blk
stack or blk pos 2 thru 6
rem cy blk pos 2 fill rd blk

(2.72)

Remove the center block
Remove the red block
Remove all red blocks
Remove the first orange block
Put a brown block on the first brown block
Add blue block on first blue block

(2.78)

remove the brown block
remove all orange blocks
put brown block on orange blocks
put orange blocks on all blocks
put blue block on leftmost blue block in top row

Results: average players (rank 21-50)

(9.17)

reinsert pink
take brown
put in pink
remove two pink from second layer
Add two red to second layer in odd intervals
Add five pink to second layer
Remove one blue and one brown from bottom layer

(7.18)

move second cube
double red with blue
double first red with red
triple second and fourth with orange
add red
remove orange on row two
add blue to column two
add brown on first and third

(8.37)

remove red
remove 1 red
remove 2 4 orange
add 2 red
add 1 2 3 4 blue
emove 1 3 5 orange
add 2 4 orange
add 2 orange
remove 2 3 brown
add 1 2 3 4 5 red
remove 2 3 4 5 6
remove 2
add 1 2 3 4 6 red

Results: worst players (rank 51-100)

(12.6)

'add red cubes on center left
center right
far left and far right'
'remove blue blocks on row two column two
row two column four'
remove red blocks in center left and center right on second row

(14.32)

laugh with me
red blocks with one aqua
aqua red alternate
brown red red orange aqua orange
red brown red brown red brown
space red orange red
second level red space red space red space

(14.15)

holdleftmost
holdbrown
holdleftmost
blueonblue
brownonblue1
blueonorange
holdblue
holdorange2
blueonred2
holdends1
holdrightend
hold2
orangeonorangerightmost

Results: interesting players

 (Polish)

usuń brazowe klocki

postaw pomarańczowy klocek na pierwszym klocku

postaw czerwone klocki na pomarańczowych

usuń pomarańczowe klocki w górnym rzedzie

Results: interesting players

(Polish)

usuń brazowe klocki

postaw pomarańczowy klocek na pierwszym klocku

postaw czerwone klocki na pomarańczowych

usuń pomarańczowe klocki w górnym rzedzie

(Polish notation)

rm scat + 1 c

+ 1 c

rm sh

+ 1 2 4 sh

+ 1 c

- 4 o

rm 1 r

+ 1 3 o

full fill c

rm o

full fill sh

- 1 3

full fill sh

rm sh

rm r

+ 2 3 r

rm o

+ 3 sh

+ 2 3 sh

Pragmatics: motivation

remove red

```
remove(hascolor(red))
```

Pragmatics: motivation

remove red

```
remove(hascolor(red))
```

remove cyan

Pragmatics: motivation

remove red

```
remove(hascolor(red))
```

remove cyan

```
remove(hascolor(red))

remove(hascolor(cyan))

remove(hascolor(brown))

remove(hascolor(orange))
```

Pragmatics: motivation

remove red

```
remove(hascolor(red))
```

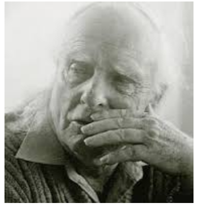
remove cyan

```
remove(hascolor(red))
remove(hascolor(cyan))
remove(hascolor(brown))
remove(hascolor(orange))
```

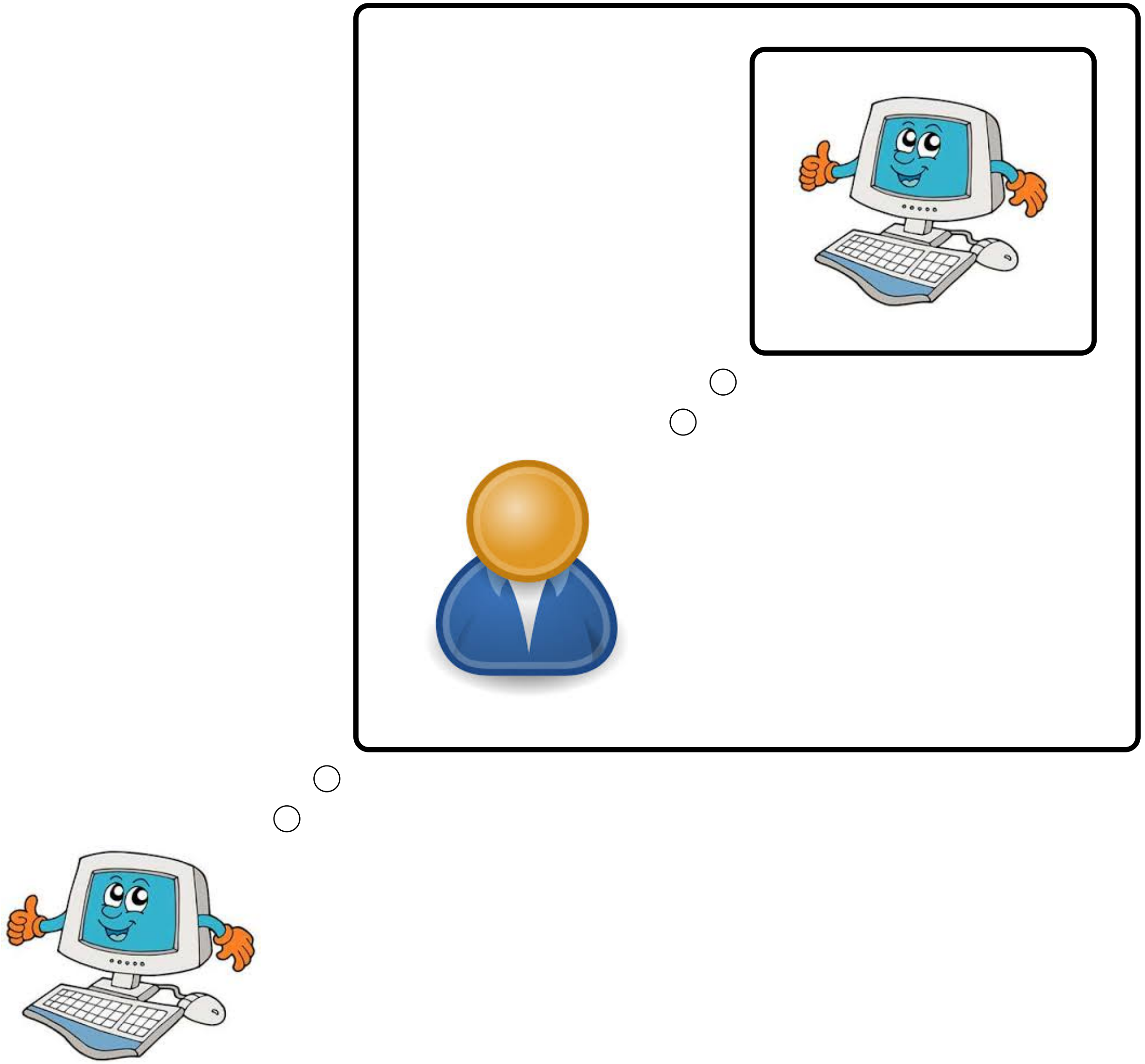
Key intuition: mutual exclusivity

[Golland et al. 2010; Frank/Goodman, 2012]

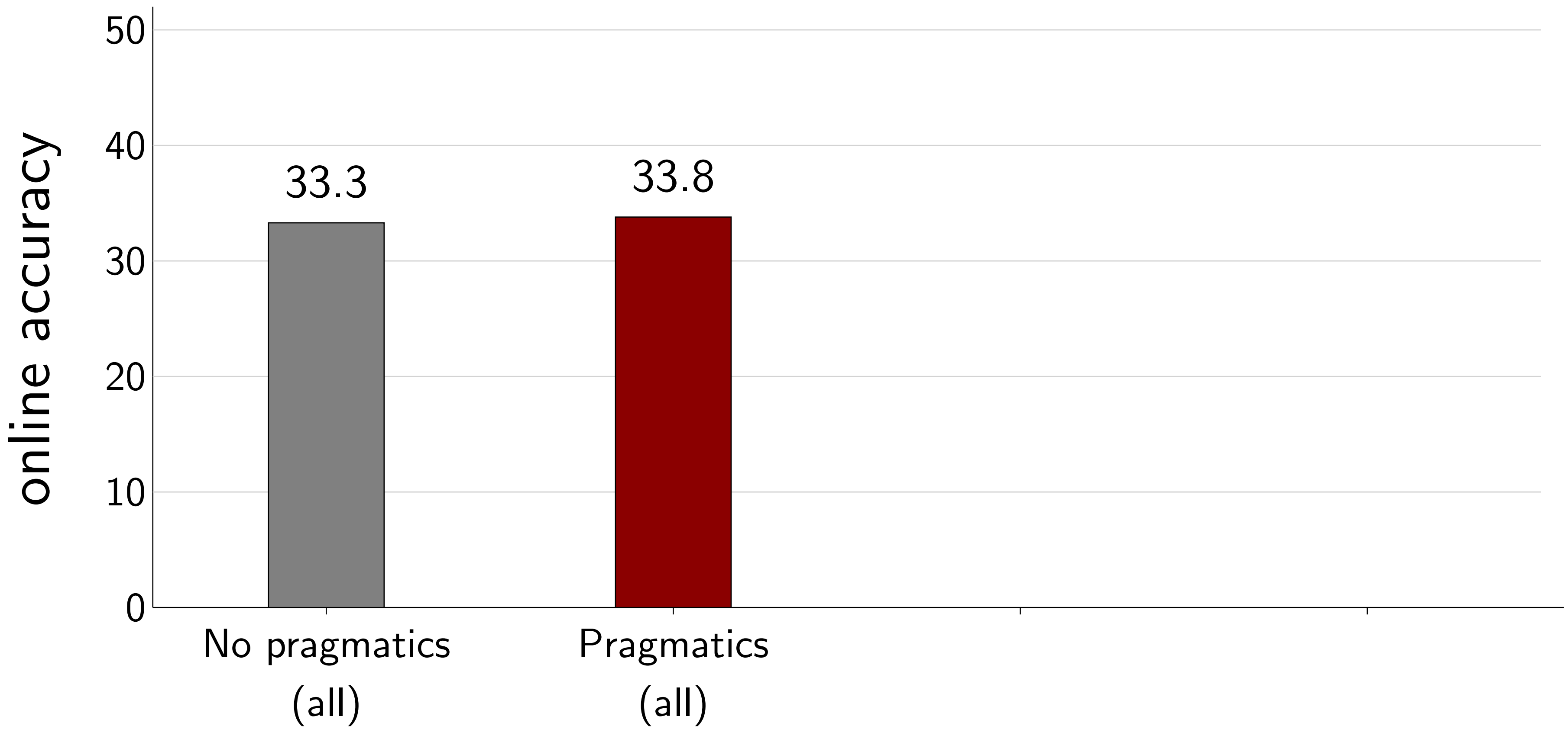
Pragmatics: model



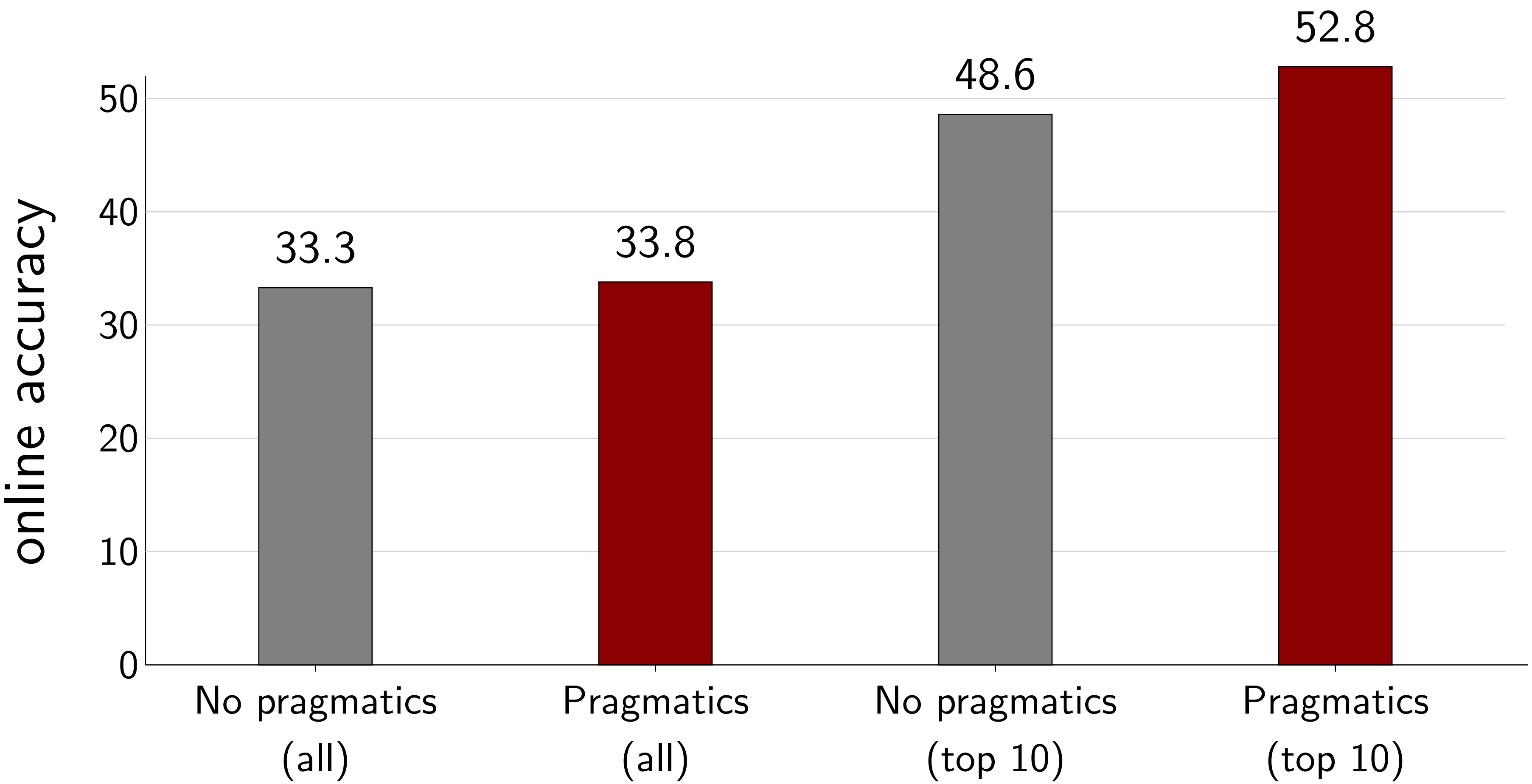
Paul Grice



Pragmatics: results



Pragmatics: results



pragmatics helps top (cooperative, rational) players

Summary so far

- Downstream goal drives language learning

Summary so far

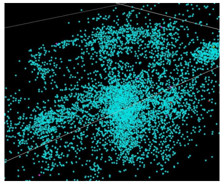
- Downstream goal drives language learning
- Both human and computer learn and adapt

Summary so far

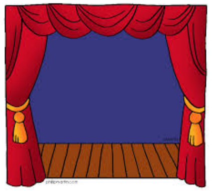
- Downstream goal drives language learning
- Both human and computer learn and adapt
- Require online learning with instance-level precision

Outline

Properties of language



Distributional semantics



Frame semantics

Model-theoretic semantics

Interactive learning

Reflections

Three types of semantics

1. Distributional semantics:
   - Pro: Most broadly applicable, ML-friendly
   - Con: Monolithic representations

Three types of semantics

1. Distributional semantics:
   - Pro: Most broadly applicable, ML-friendly
   - Con: Monolithic representations
2. Frame semantics:
   - Pro: More structured representations
   - Con: Not full representation of world

Three types of semantics

1. Distributional semantics:
   - Pro: Most broadly applicable, ML-friendly
   - Con: Monolithic representations
2. Frame semantics:
   - Pro: More structured representations
   - Con: Not full representation of world
3. Model-theoretic semantics:
   - Pro: Full world representation, rich semantics, end-to-end
   - Con: Narrower in scope

many opportunities for synthesis

[Rajpurkar et al. 2016]

Reading comprehension

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

Reading comprehension

[Rajpurkar et al. 2016]

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

SQuAD (100K examples)

`stanford-qa.com`

[Rajpurkar et al. 2016]

Reading comprehension

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

SQuAD (100K examples)

stanford-qa.com

Team	**F1**
MSR-A	82.2%
AI2	81.1%
Salesforce	80.4%
...	...
Log. regression	51.0%

[Rajpurkar et al. 2016]

Reading comprehension

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

SQuAD (100K examples)

`stanford-qa.com`

Team	**F1**
MSR-A	82.2%
AI2	81.1%
Salesforce	80.4%
...	...
Log. regression	51.0%

- Solve end-to-end task without explicit model of world

Dialogue

Time	User	Utterance
[12:21]	dell	well, can I move the drives?
[12:21]	cucho	dell: ah not like that
[12:21]	RC	dell: you can't move the drives
[12:21]	RC	dell: definitely not
[12:21]	dell	ok
[12:21]	dell	lol
[12:21]	RC	this is the problem with RAID:)
[12:21]	dell	RC haha yeah
[12:22]	dell	cucho, I guess I could just get an enclosure and copy via USB...
[12:22]	cucho	dell: i would advise you to get the disk

Ubuntu Dialogue Corpus [Lowe et al. 2015]

- Involves both understanding and generation
- Grounding? task-oriented versus chatbots
- Challenge: how to evaluate?

Takeaway 1/2

most of language understanding is about the world

Takeaway 2/2

language is about communication, interactive

Open questions

How to combine **logical** and **distributional**?

Open questions

How to combine **logical** and **distributional**?

How to represent **knowledge, context, memory**?

Open questions

How to combine **logical** and **distributional**?

How to represent **knowledge, context, memory**?

How to create **interactive** learning environment?

Questions?