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…

Descartes (c.1630)

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/

Turing (c. 1950)

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?

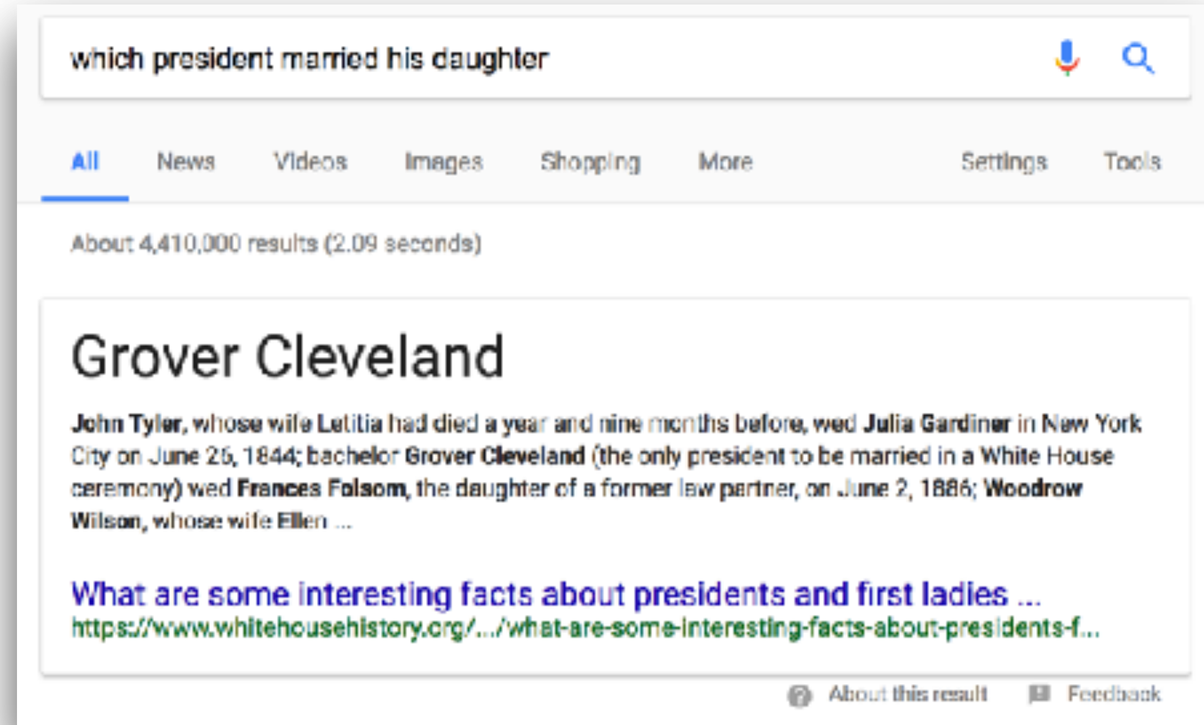
What is the Turing test? And are we all doomed now?

Computer Posing as Teenager Achieves Artificial-Intelligence Milestone

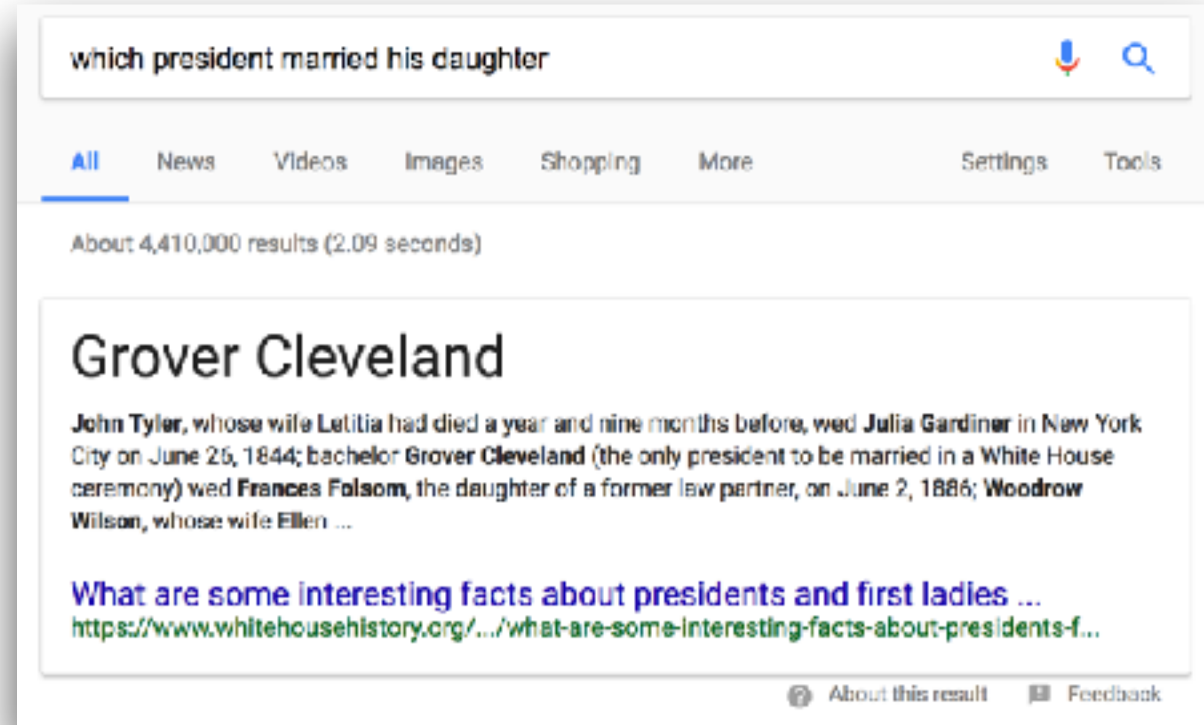
Computer AI passes Turing test in 'world first'

Passing the Turing Test…?

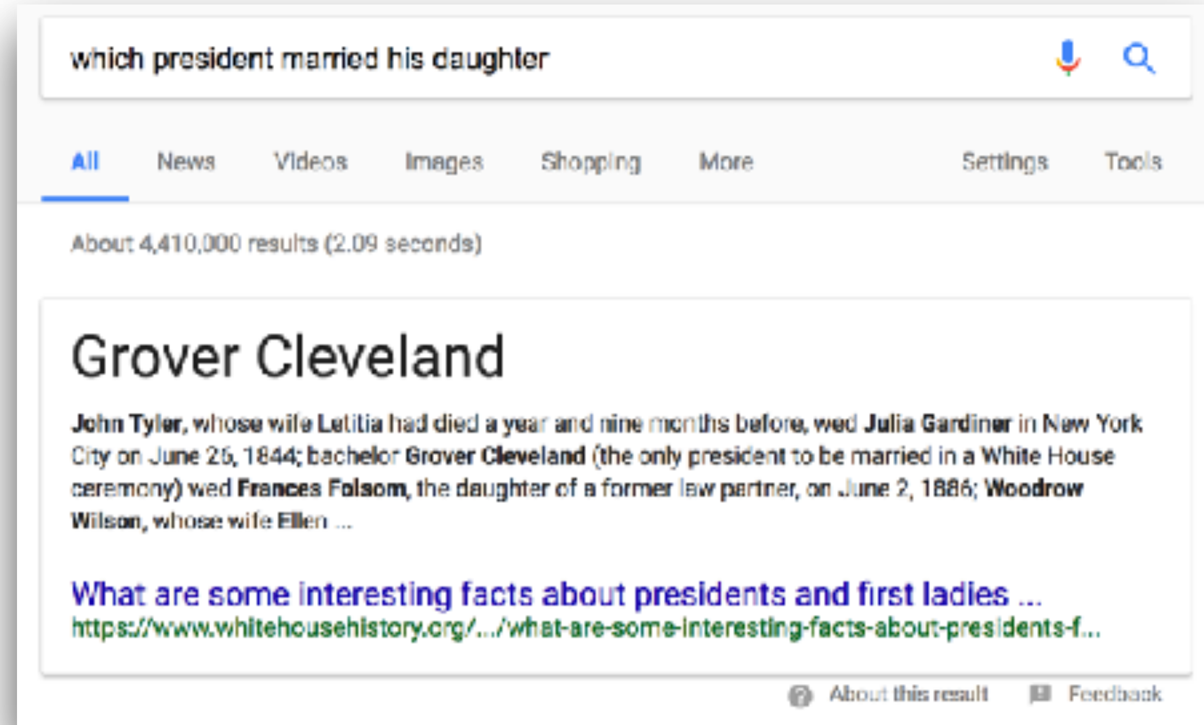
Passing the Turing Test…?



Passing the Turing Test…?

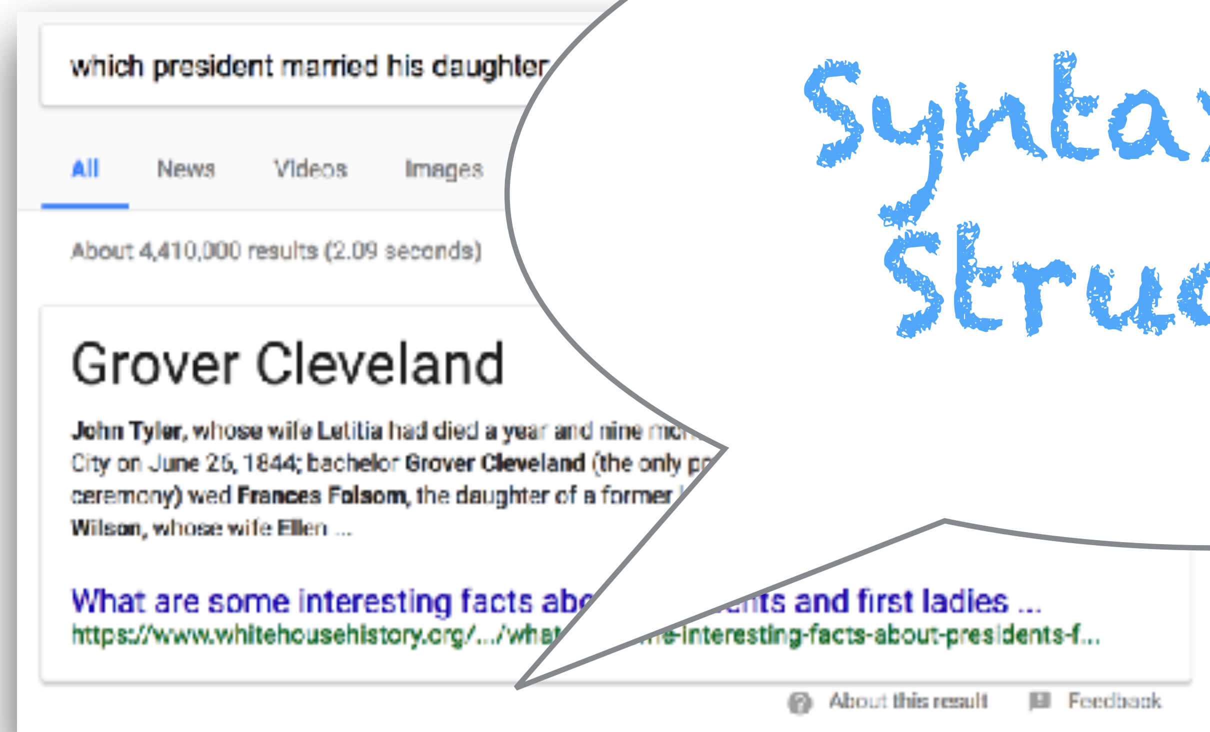


Passing the Turing Test…?



Passing the Turing Test…?

Syntax and Structure



Passing the Turing Test…?

Passing the Turing Test…?

Representing language is hard.

Crash Course: Non-Computational Linguistics

Crash Course: Non-~~Computational~~ Linguistics

Pencil-and-Paper NLP

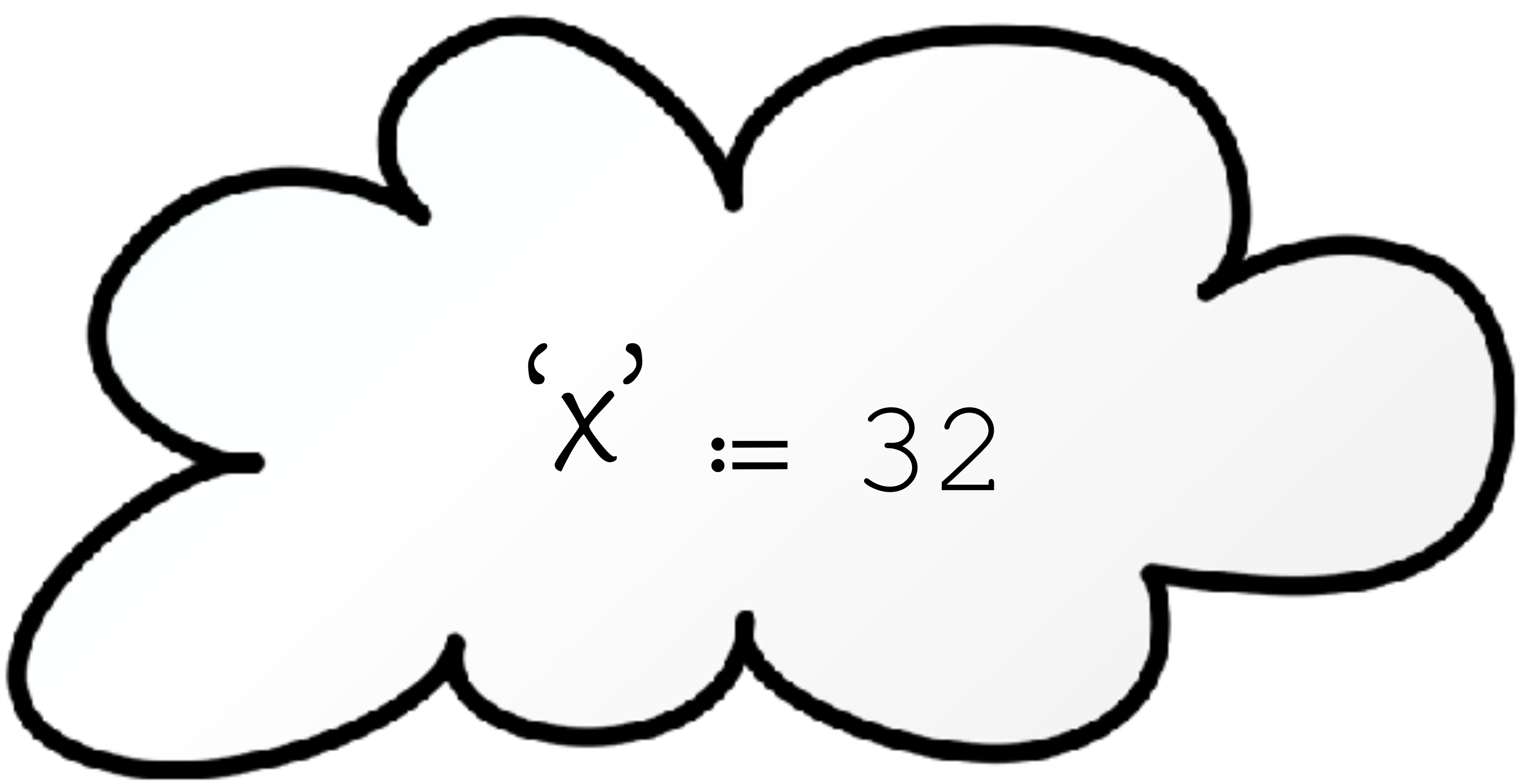
Model Theory

Model Theory

```
x > 17
```

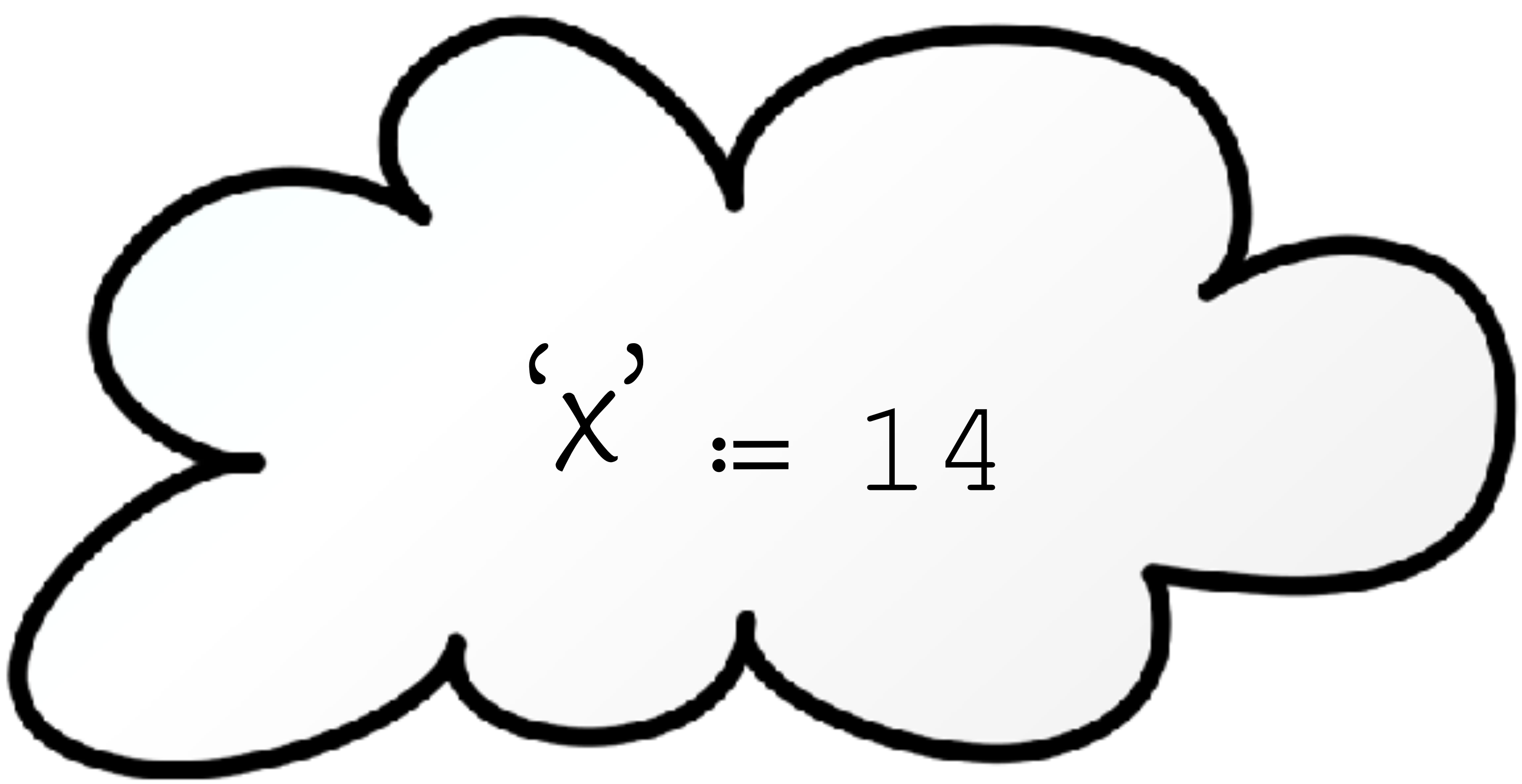
Model Theory

`x > 17` ✔



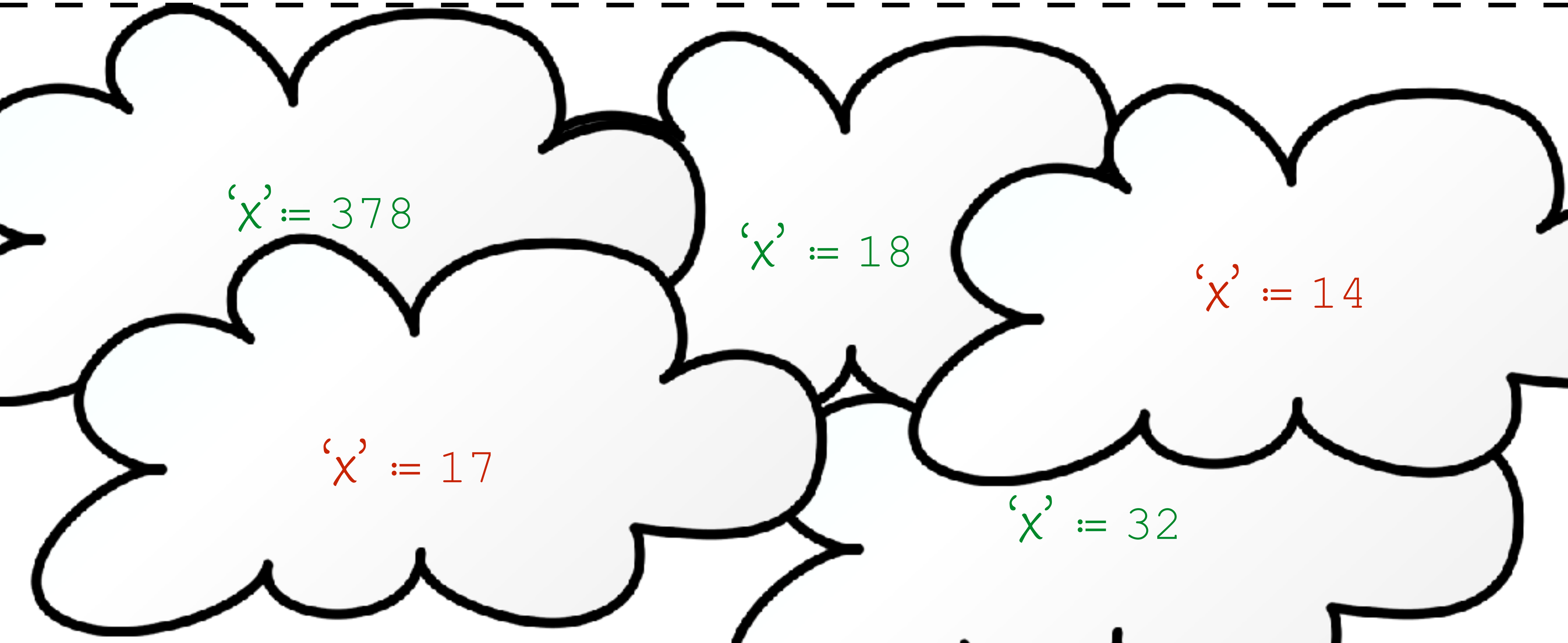
Model Theory

$x > 17$ ✘



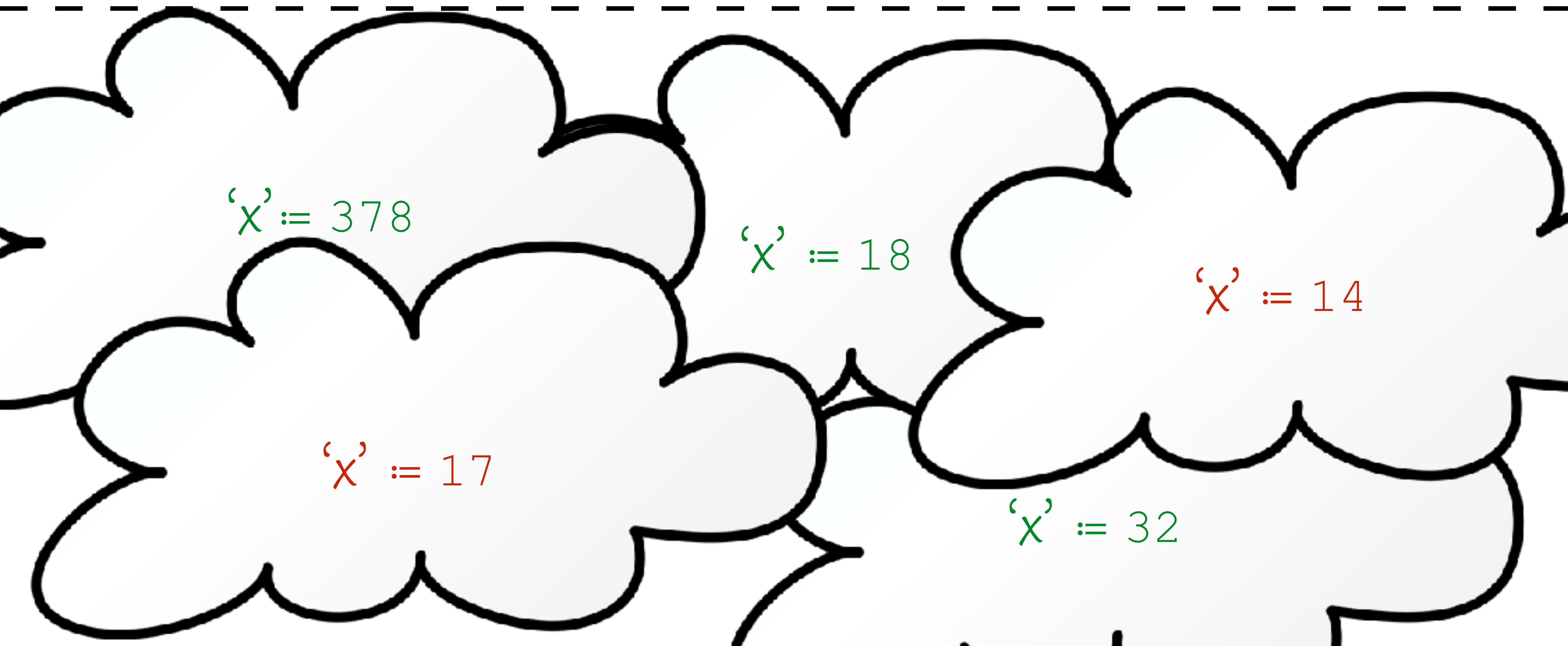
Model Theory

`x > 17`



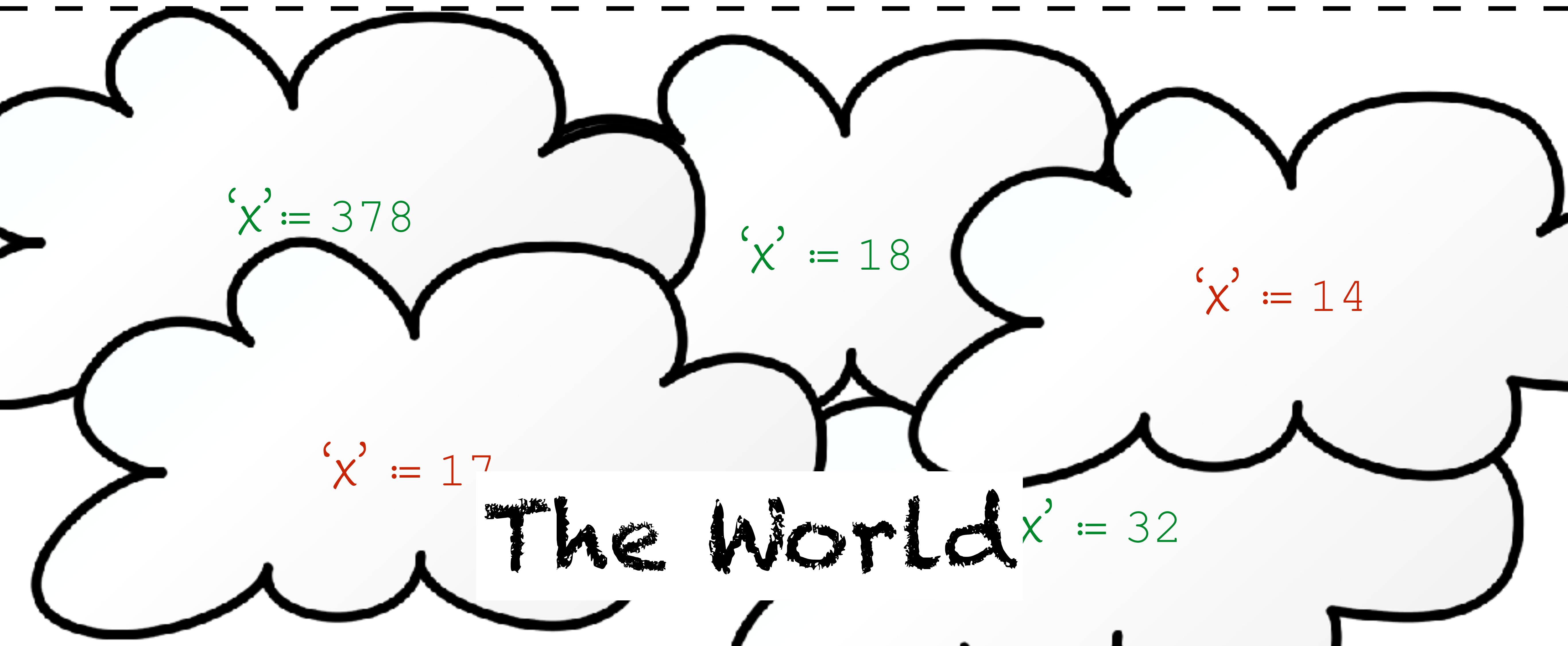
Model Theory

Language $x > 17$



Model Theory

Language   x > 17



Model Theory

Language

`x > 17`

The World (TBD)

Model Theory

Language

$$\frac{\begin{array}{c} x > y \\ y > z \end{array}}{x > z}$$

The World (TBD)

Model Theory

Language

Variables
(to be grounded)

$$\frac{\begin{array}{c} x > y \\ y > z \end{array}}{x > z}$$

The World (TBD)

Model Theory

Language

$$\frac{\begin{array}{c} x > y \\ y > z \end{array}}{x > z}$$

Relations (defined)

- -

The World (TBD)

Model Theory

Entailment

$$\frac{\begin{array}{c} \mathtt{x} > \mathtt{y} \\ \mathtt{z} > \mathtt{w} \end{array}}{\mathtt{x} > \mathtt{w}}$$

The World (TBD)

Model Theory

Entailment

$$\frac{\begin{array}{c} x > y \\ z > w \end{array}}{x > w}$$

---

x = 10 y = 5 z = 11 w = 8

Model Theory

Entailment

$$\frac{\begin{array}{c} x > y \\ z > w \end{array}}{x > w}$$

$x = 10$ $y = 5$ $z = 11$ $w = 8$

Model Theory

Entailment

$$\frac{\begin{array}{c} \texttt{x > y} \\ \texttt{z > w} \end{array}}{\texttt{x > w}}$$

---

x = 10 y = 5 z = 11 w = 8

Model Theory

Entailment

$$\frac{\begin{array}{c} x > y \\ z > w \end{array}}{x > w}$$

$x = 10 \quad y = 5 \quad z = 11 \quad w = 8$

Model Theory

Entailment

$$\frac{\begin{array}{c} x > y \\ z > w \end{array}}{x > w}$$

✔

x = 10 y = 5 z = 11 w = 8

Model Theory

Entailment

$$\frac{\begin{array}{c} x > y \\ z > w \end{array}}{x > w}$$

✗

x = 10 y = 5 z = 12 w = 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

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

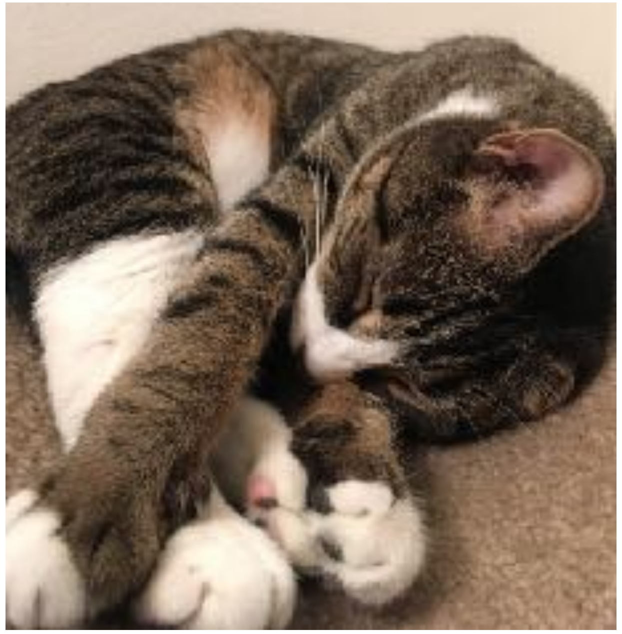
`Broca is a bird` ✔

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 a entailment

All **birds** are **gray**

Broca is a **bird**

Broca is **gray**

✔

Predicates

Formal Semantics

the notion of a entailment

All birds are gray
Broca is a bird

Broca is gray

✔

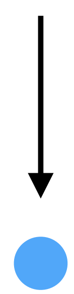
Higher-Order Relations

Formal Semantics

Formal Semantics

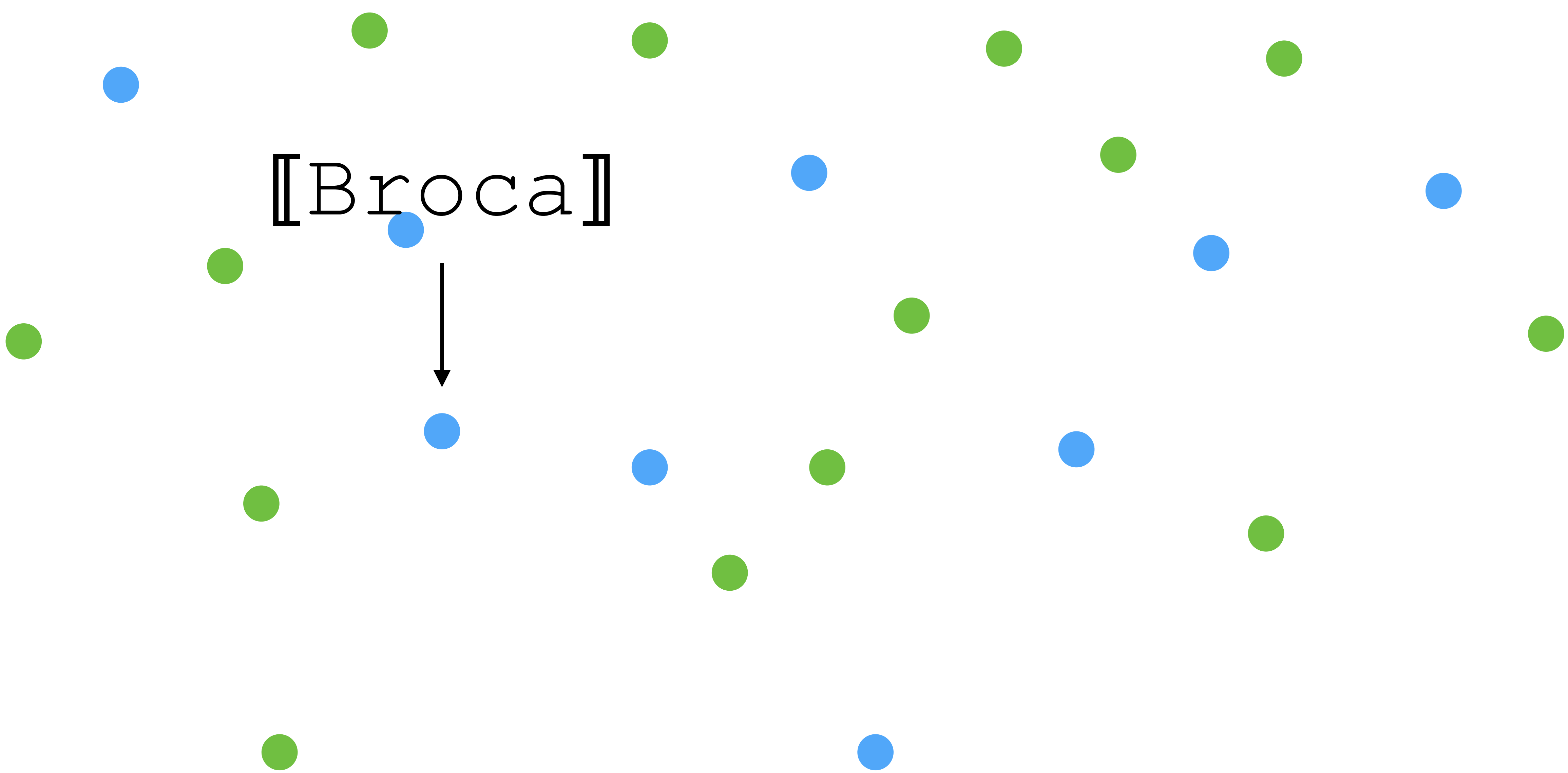
Entities

$[\![\texttt{Broca}]\!]$

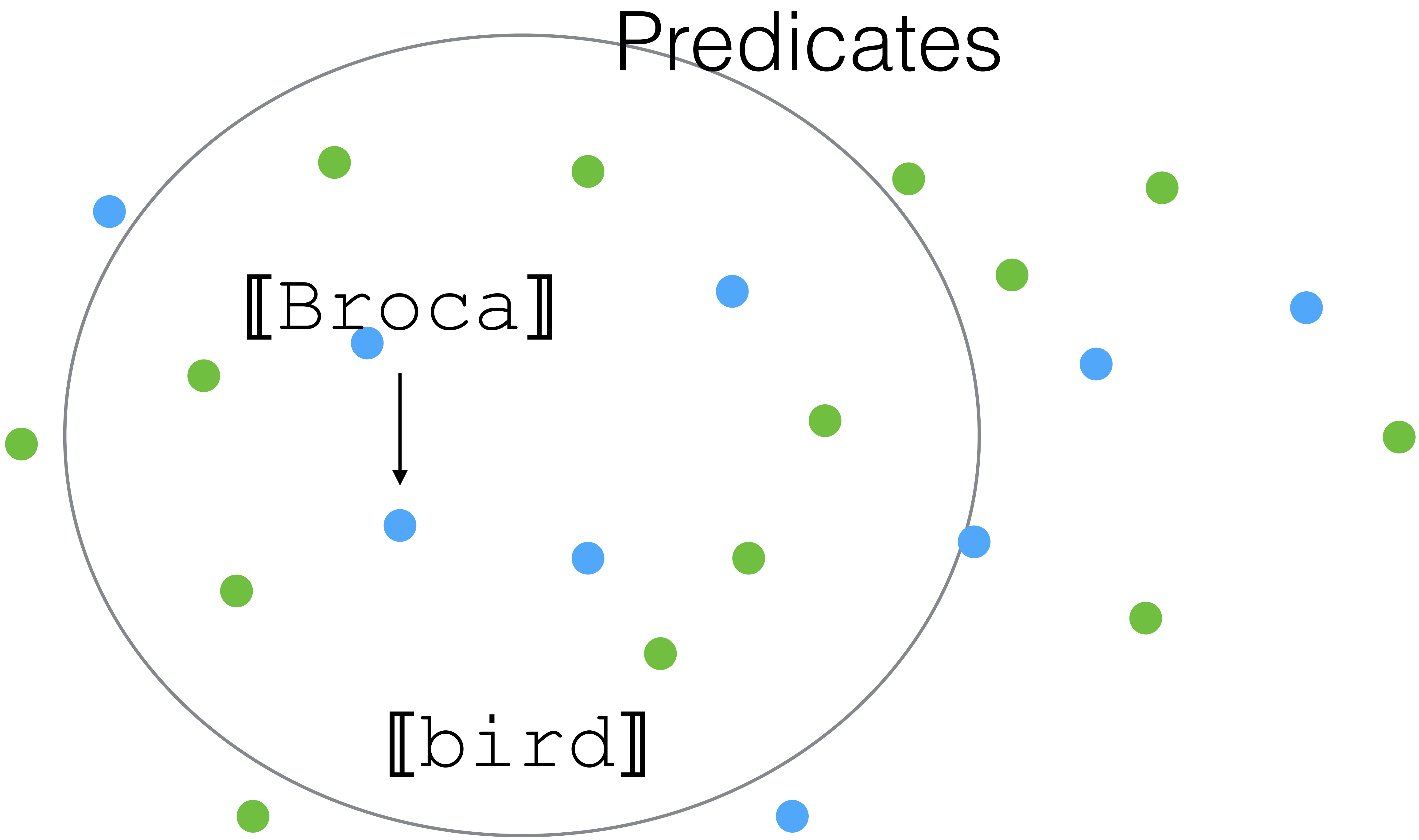


Formal Semantics

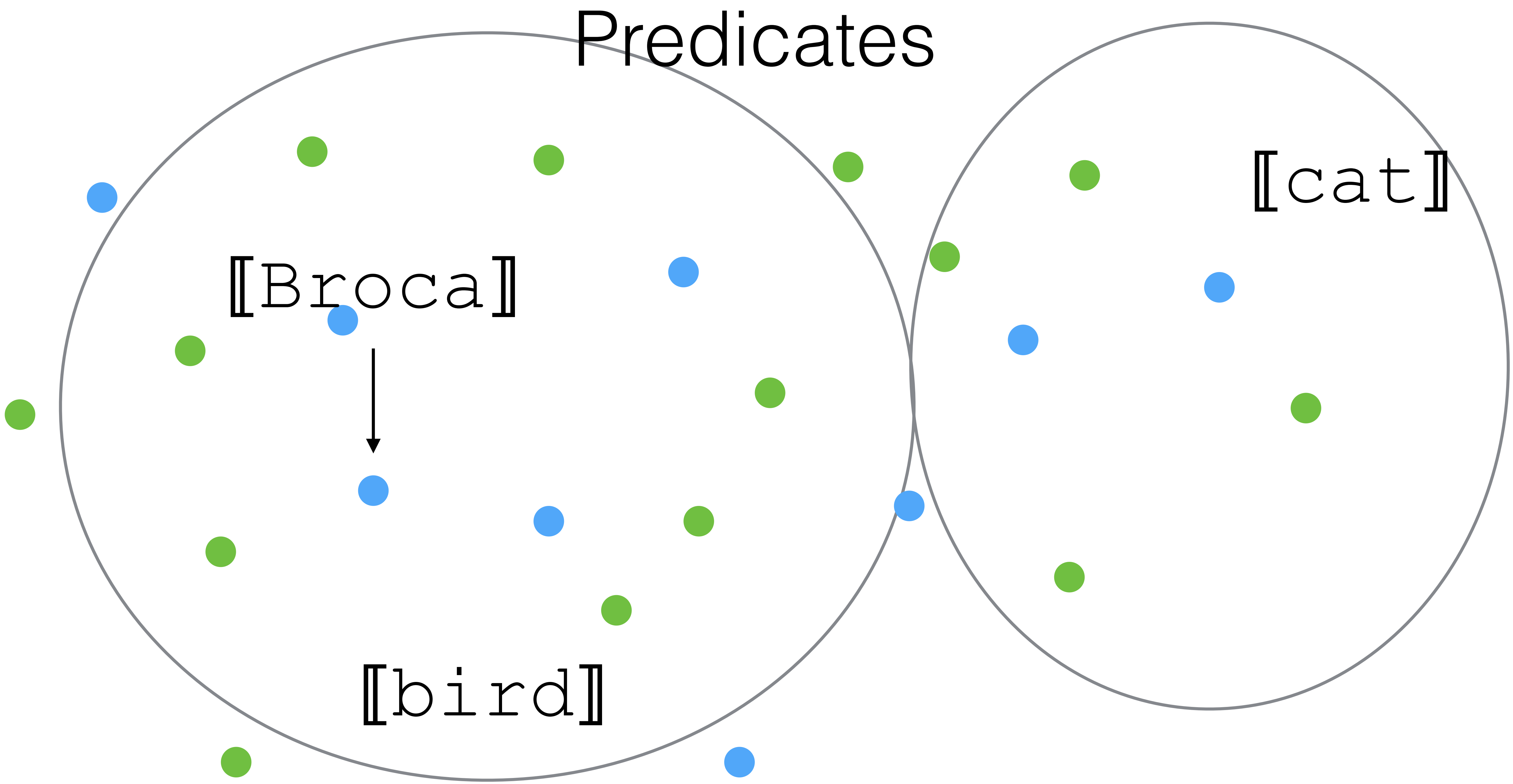
Entities



Formal Semantics



Formal Semantics



Formal Semantics

Predicates

Broca is a bird

$$\Longleftrightarrow$$

$$\forall x(\llbracket \mathrm{Broca} \rrbracket(x) \Rightarrow \llbracket \mathrm{bird} \rrbracket(x))$$

Formal Semantics

Predicates

Broca is a bird

$$\Longleftrightarrow$$

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

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

Formal Semantics

Predicates

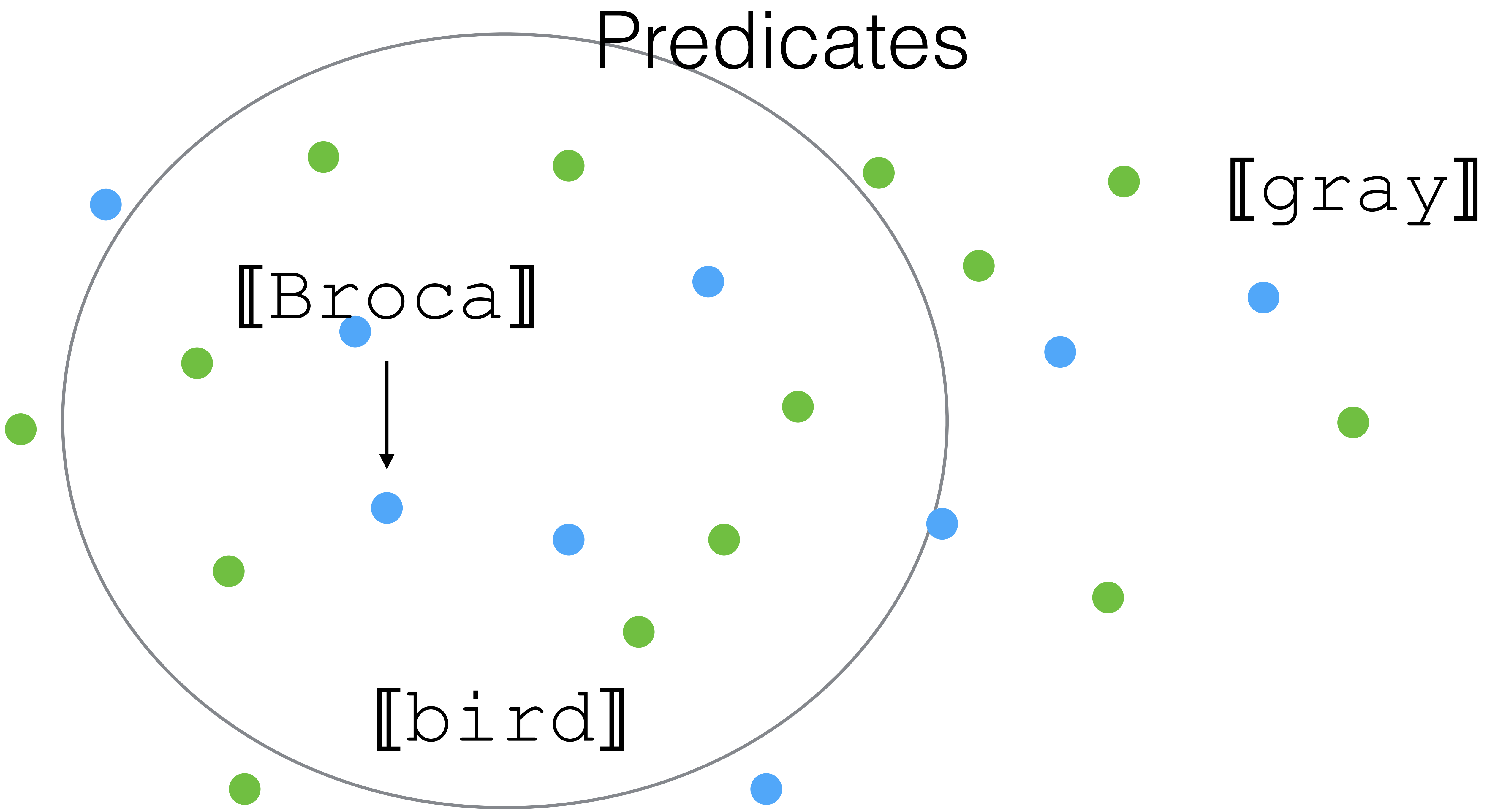
Broca is a bird

$$\Longleftrightarrow$$

$$\forall x ([\![\mathtt{Broca}]\!](x) \Rightarrow [\![\mathtt{bird}]\!](x))$$

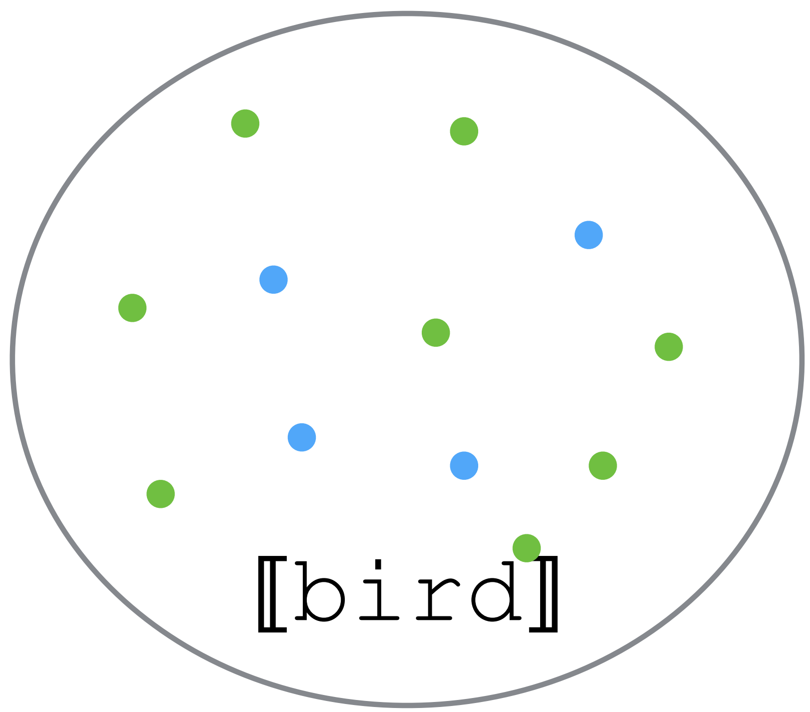
Takes entity as argument.
Returns true if x is an element of the set referred to by “bird”.

Formal Semantics

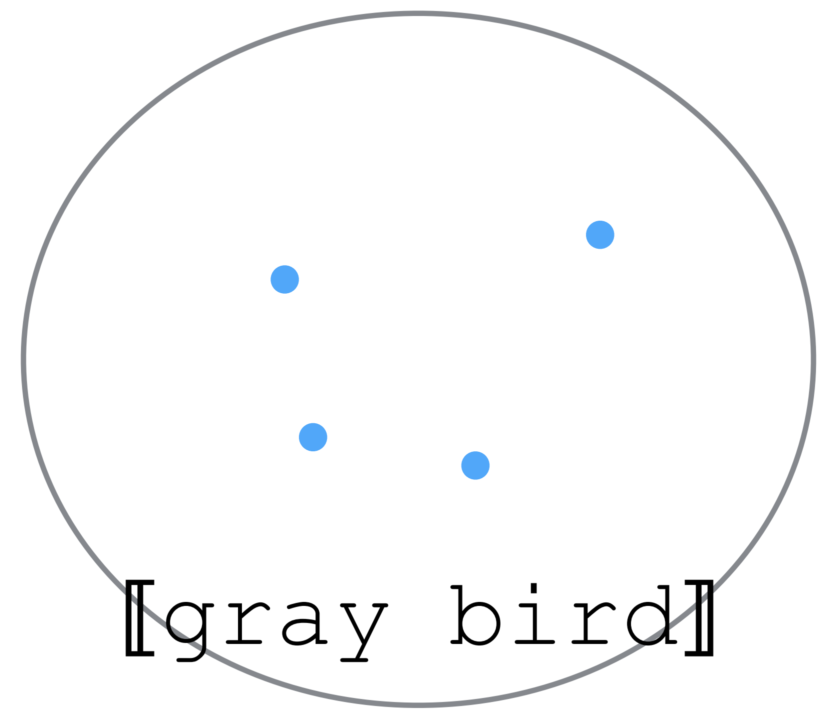


Formal Semantics

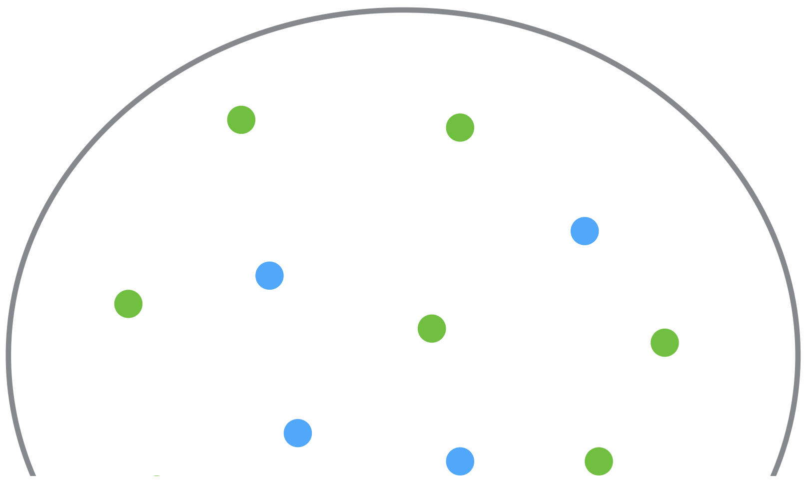
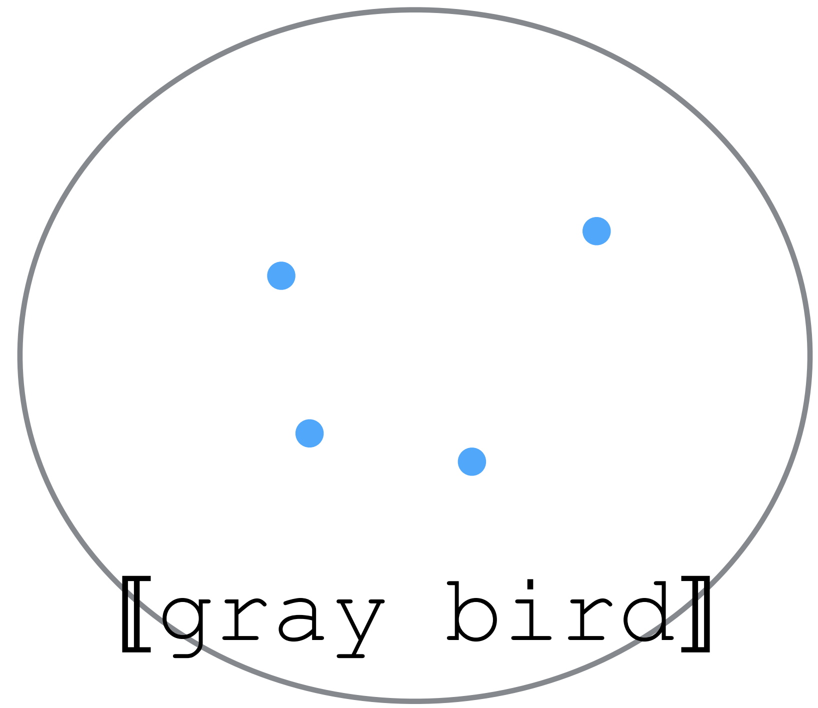
⟦gray⟧ (



) =



Formal Semantics

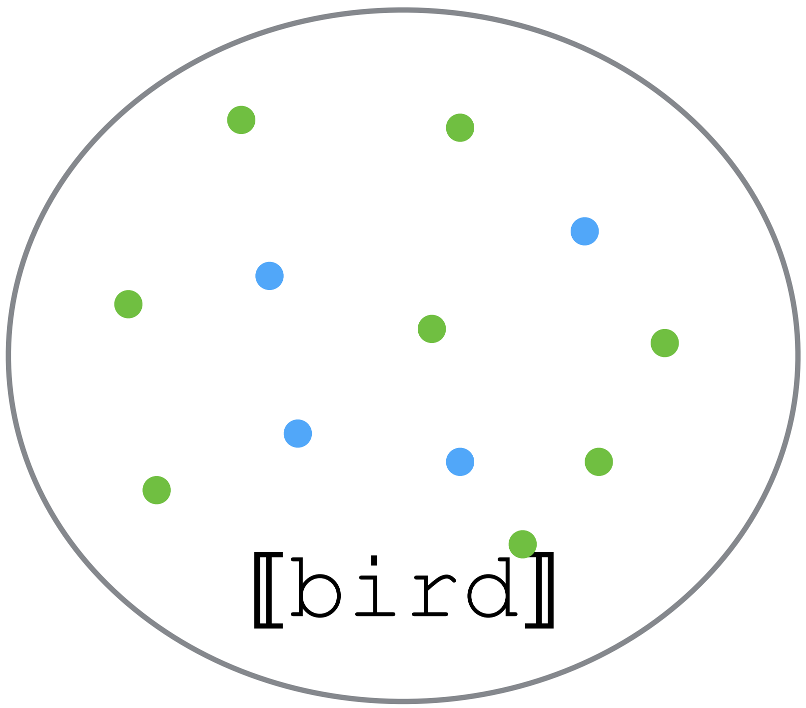
$[\![\texttt{gray}]\!]$ () = 

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

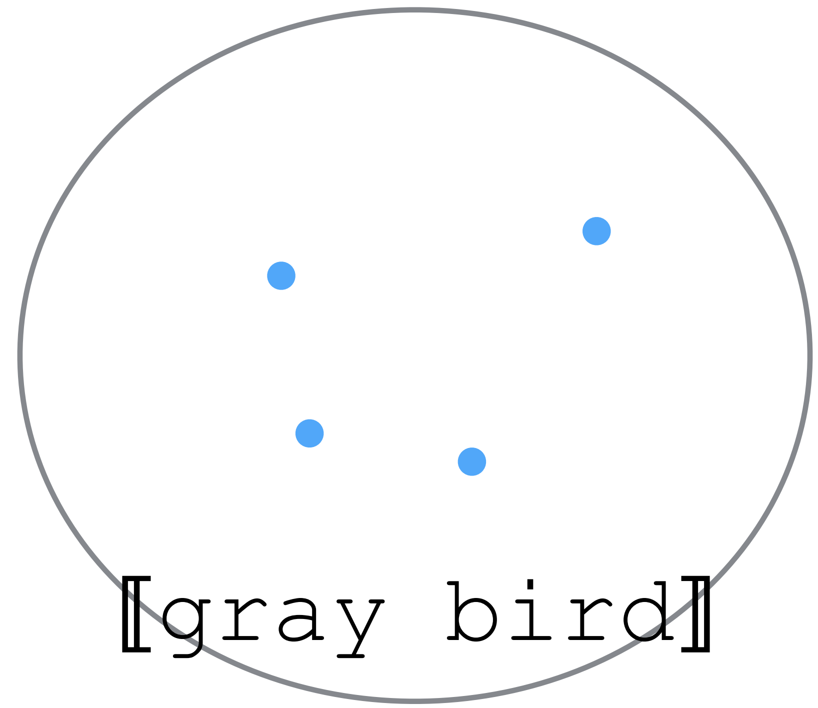
$$[\![\texttt{gray bird}]\!](x) \iff ([\![\texttt{bird}]\!](x) \wedge [\![\texttt{gray}]\!](x))$$

Formal Semantics

⟦gray⟧ (



) =



Broca is a gray bird

entails

Broca is a bird

Formal Semantics

Broca is a gray bird

entails

Broca is a bird

Formal Semantics

$$\forall x(\llbracket \texttt{Broca} \rrbracket(x) \Rightarrow \llbracket \texttt{gray bird} \rrbracket(x))$$

entails

Broca is a bird

Formal Semantics

$$\forall x\,(\llbracket \texttt{Broca} \rrbracket(x) \Rightarrow \llbracket \texttt{gray bird} \rrbracket(x))$$

$$\Rightarrow$$

Broca is a bird

Formal Semantics

$$\forall \mathtt{x}(\llbracket \mathtt{Broca} \rrbracket(\mathtt{x}) \Rightarrow \llbracket \mathtt{gray\ bird} \rrbracket(\mathtt{x}))$$

$$\Rightarrow$$

$$\forall \mathtt{x}(\llbracket \mathtt{Broca} \rrbracket(\mathtt{x}) \Rightarrow \llbracket \mathtt{bird} \rrbracket(\mathtt{x}))$$

Formal Semantics

$$\forall x(\llbracket \texttt{Broca} \rrbracket(x) \Rightarrow$$

$$(\llbracket \texttt{bird} \rrbracket(x) \;\wedge\; \llbracket \texttt{gray} \rrbracket(x))$$

$$\Rightarrow$$

$$\forall x(\llbracket \texttt{Broca} \rrbracket(x) \Rightarrow \llbracket \texttt{bird} \rrbracket(x))$$

Formal Semantics

$$\forall x (\llbracket \texttt{Broca} \rrbracket (x) \Rightarrow \llbracket \texttt{bird} \rrbracket (x))$$

$$\Rightarrow$$

$$\forall x (\llbracket \texttt{Broca} \rrbracket (x) \Rightarrow \llbracket \texttt{bird} \rrbracket (x))$$

✔

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

All gray birds are birds

$$[\![\mathtt{all}]\!](x) = \lambda P \lambda Q \forall x (P(x) \Rightarrow Q(x))$$

Formal Semantics

All gray birds are birds

$$[\![\texttt{all}]\!](x) = \lambda P \lambda Q \forall x (P(x) \Rightarrow Q(x))$$

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`

$$[\![\mathtt{all}]\!](\mathtt{x})(\mathtt{gray\_birds})(\mathtt{birds}) = \forall \mathtt{x}(\mathtt{gray\_bid}(\mathtt{x}) \Rightarrow \mathtt{bird}(\mathtt{x}))$$

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

Formal Semantics: Takeaways

Formal Semantics: Takeaways

- Language is like logic—we can <u>ground</u> symbols to the world, but we can also <u>reason abstractly</u> 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 <u>ground</u> symbols to the world, but we can also <u>reason abstractly</u> 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 <u>ground</u> symbols to the world, but we can also <u>reason abstractly</u> 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.

"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.

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.

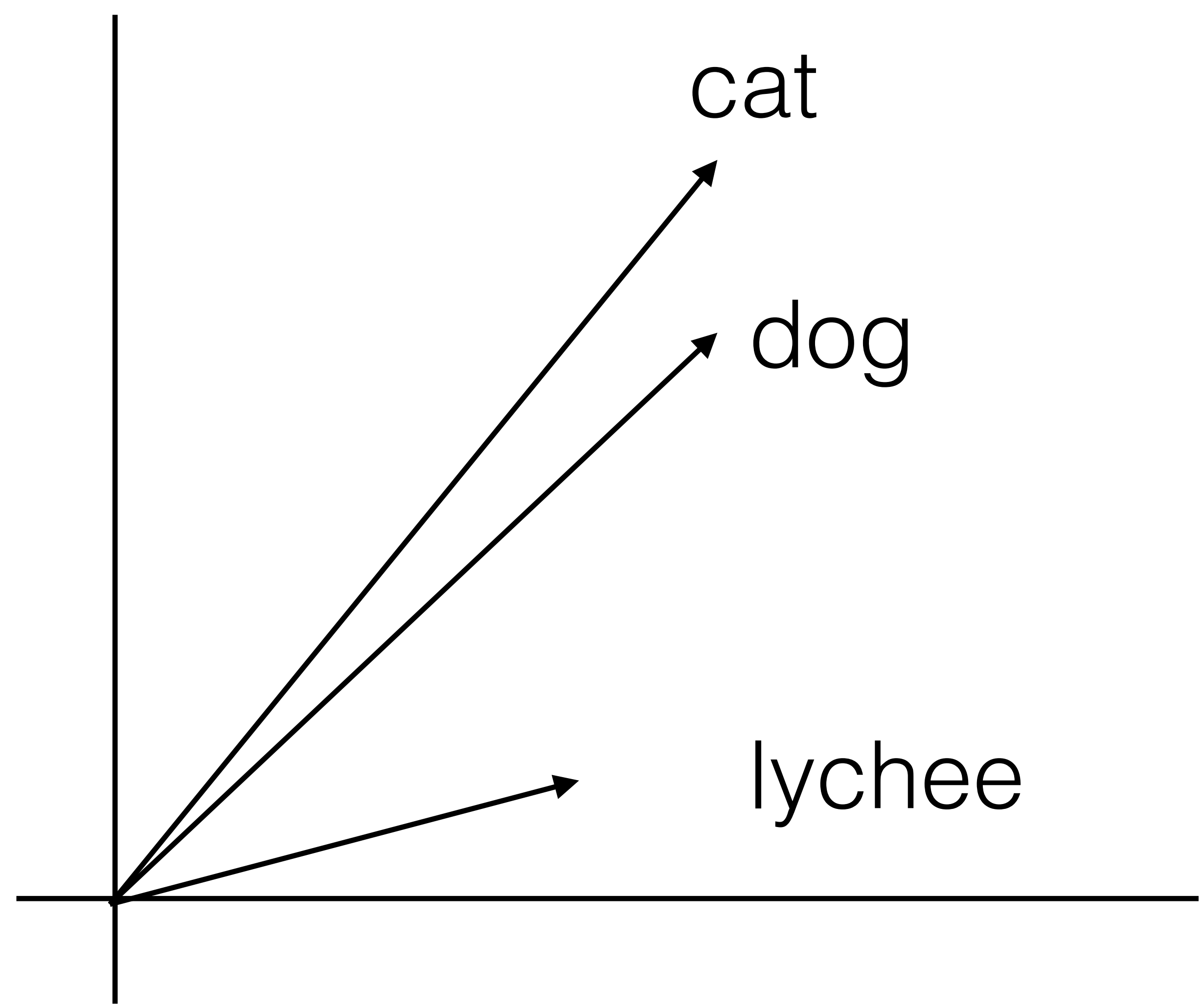
"The **cat** does not offer services," William Burroughs wrote. "The **cat** offers itself." But it does so with unapologetic ambivalence.

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.

	the	domes-tic	is	a	your	online	owners	breed	informa-tion	selec-tion
cat	1000	40	500	700	400	3	80	100	15	6
dog	1050	50	400	950	500	1	105	160	4	2
lychee	2000	2	500	1000	25	50	2	3	45	700

BOW Vector Space Models



Skip-Gram Model (word2vec)

Skip-Gram Model (word2vec)

	the	domes-tic	is	a	your	online	owners	breed	informa-tion	selec-tion
cat	1000	40	500	700	400	3	80	100	15	6

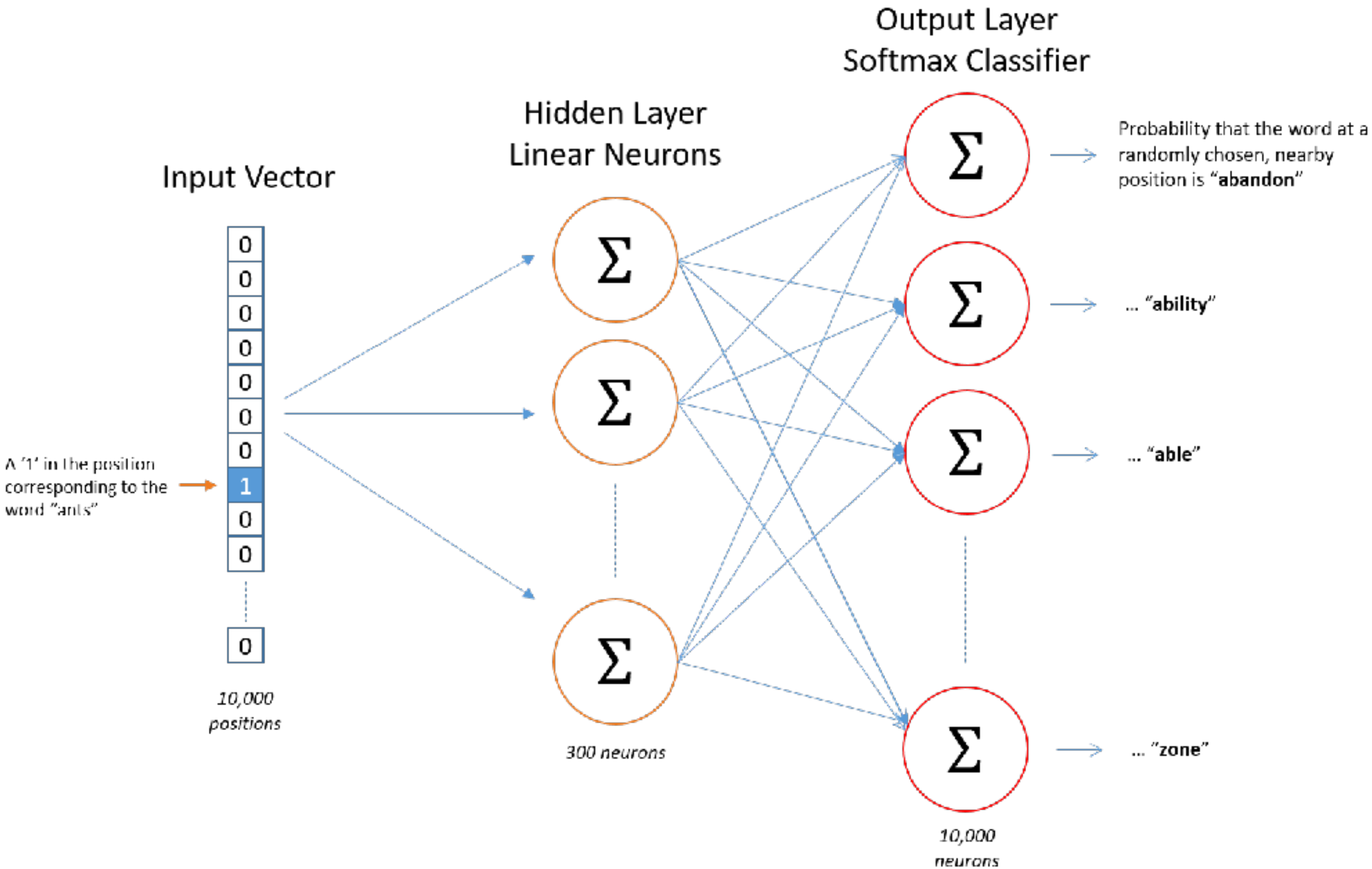
Skip-Gram Model (word2vec)

cat

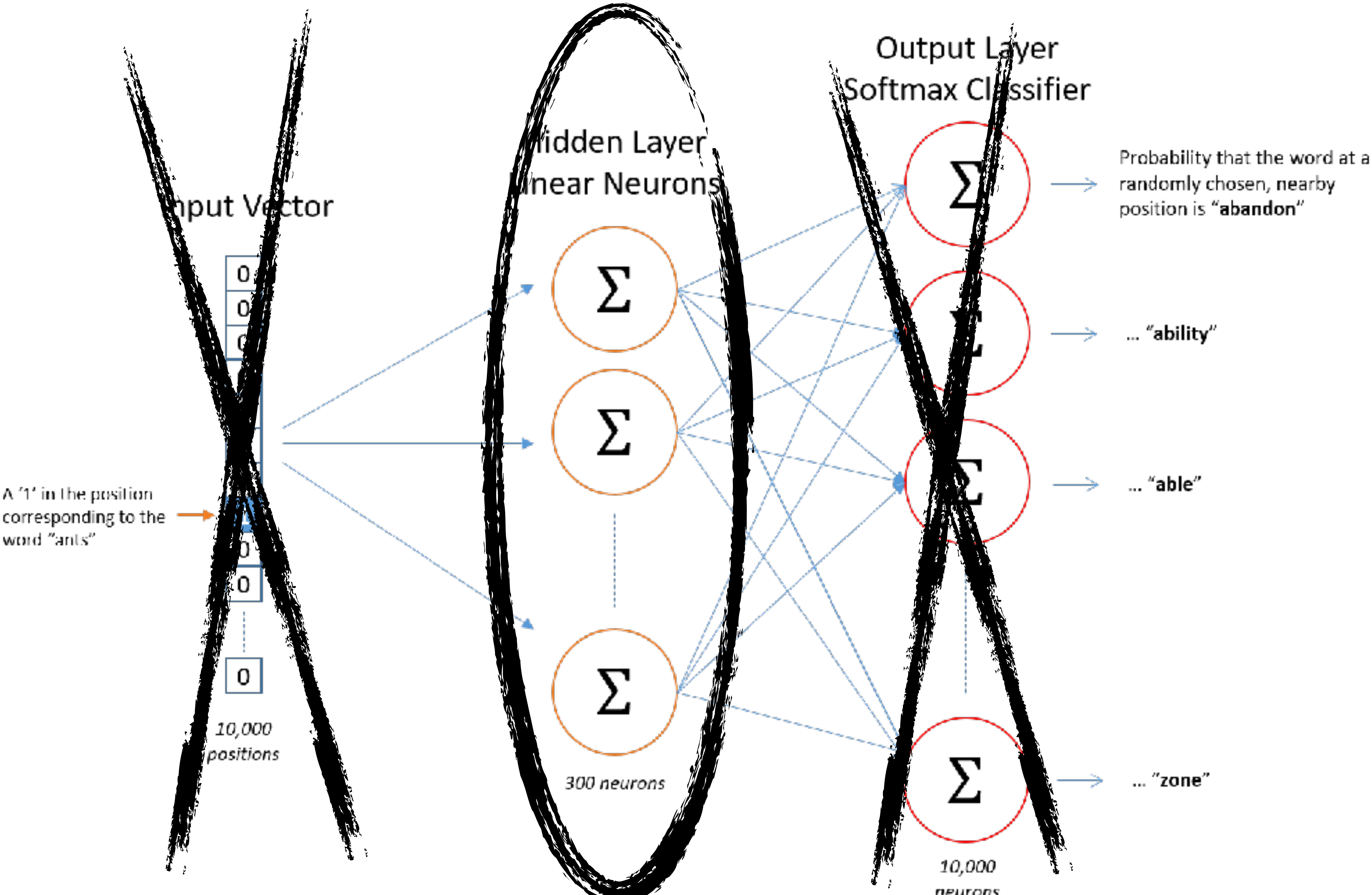
0
0
1
0
0
0
0
0
0
0

→

1000	the
40	domes-tic
500	is
700	a
400	your
3	online
80	owners
100	breed
15	informa-tion
6	selec-tion

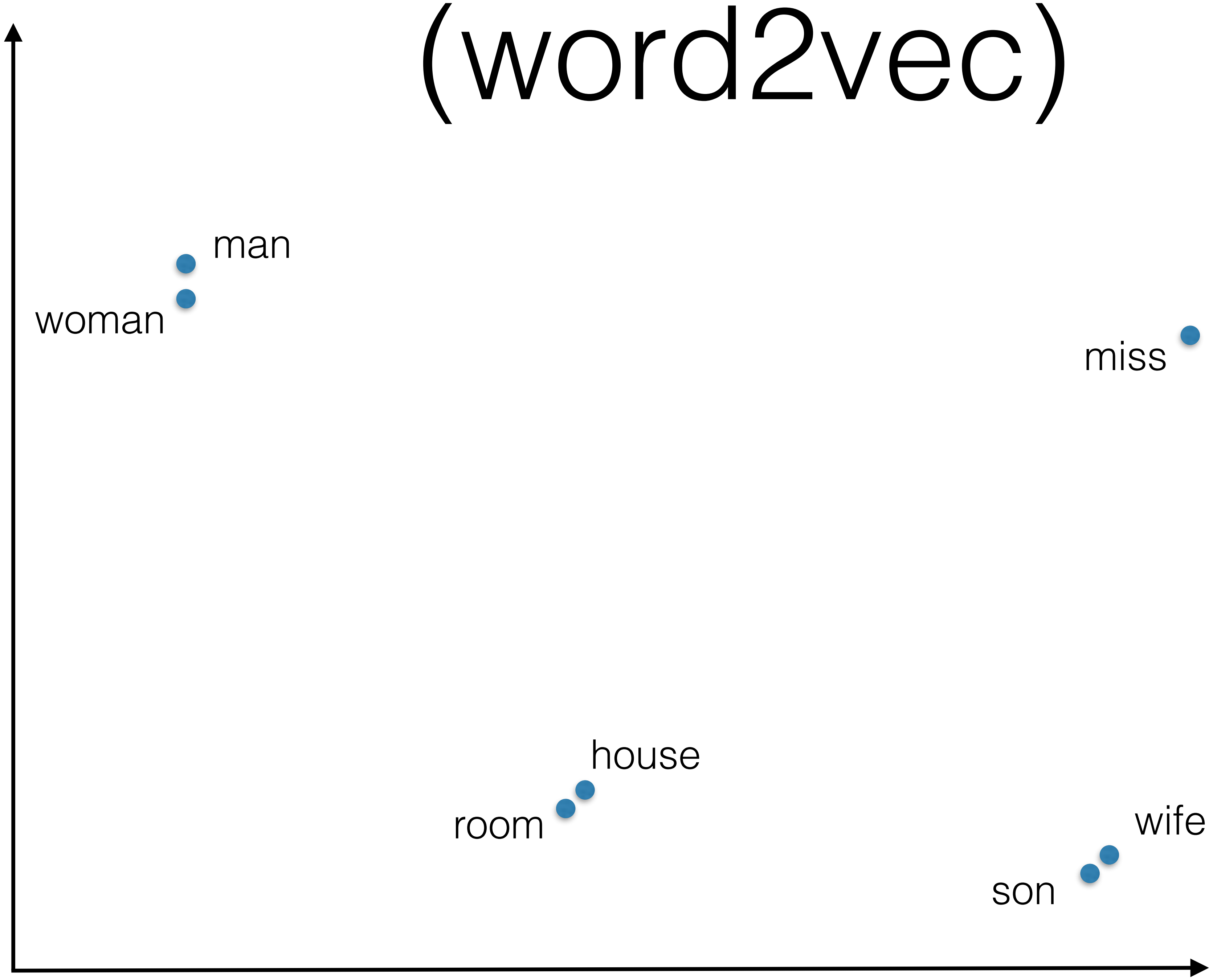


https://towardsdatascience.com/word2vec-skip-gram-model-part-1-intuition-78614e4d6e0b



https://towardsdatascience.com/word2vec-skip-gram-model-part-1-intuition-78614e4d6e0b

Skip-Gram Model (word2vec)



2-D projection of word vectors learned from *Pride and Prejudice* (http://www.ghostweather.com/files/word2vecpride/)

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.

"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.

Representing Context

Australian scientist discovers star with telescope

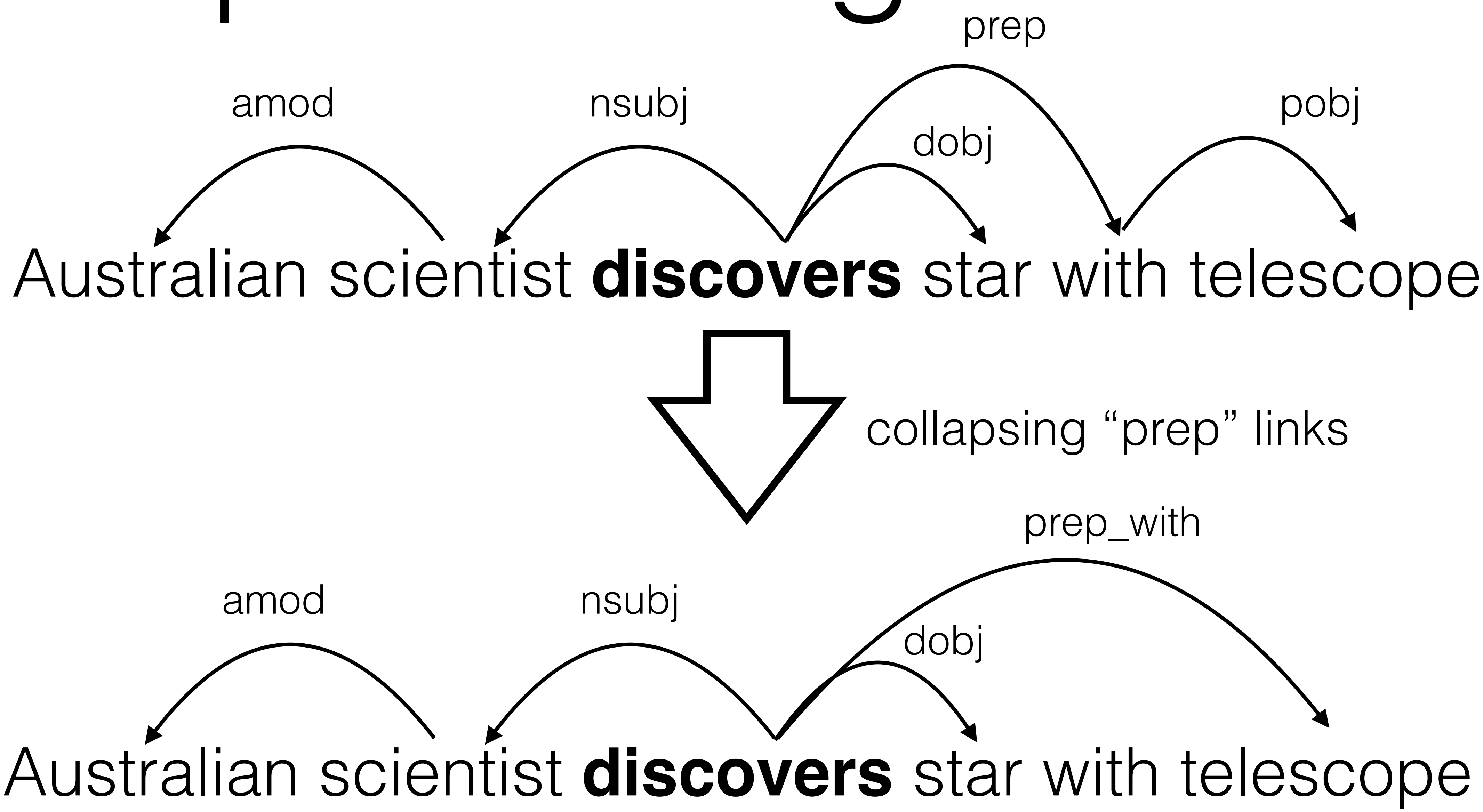
Representing Context

Australian scientist **discovers** star with telescope

Skip-Gram contexts with ***n=2***

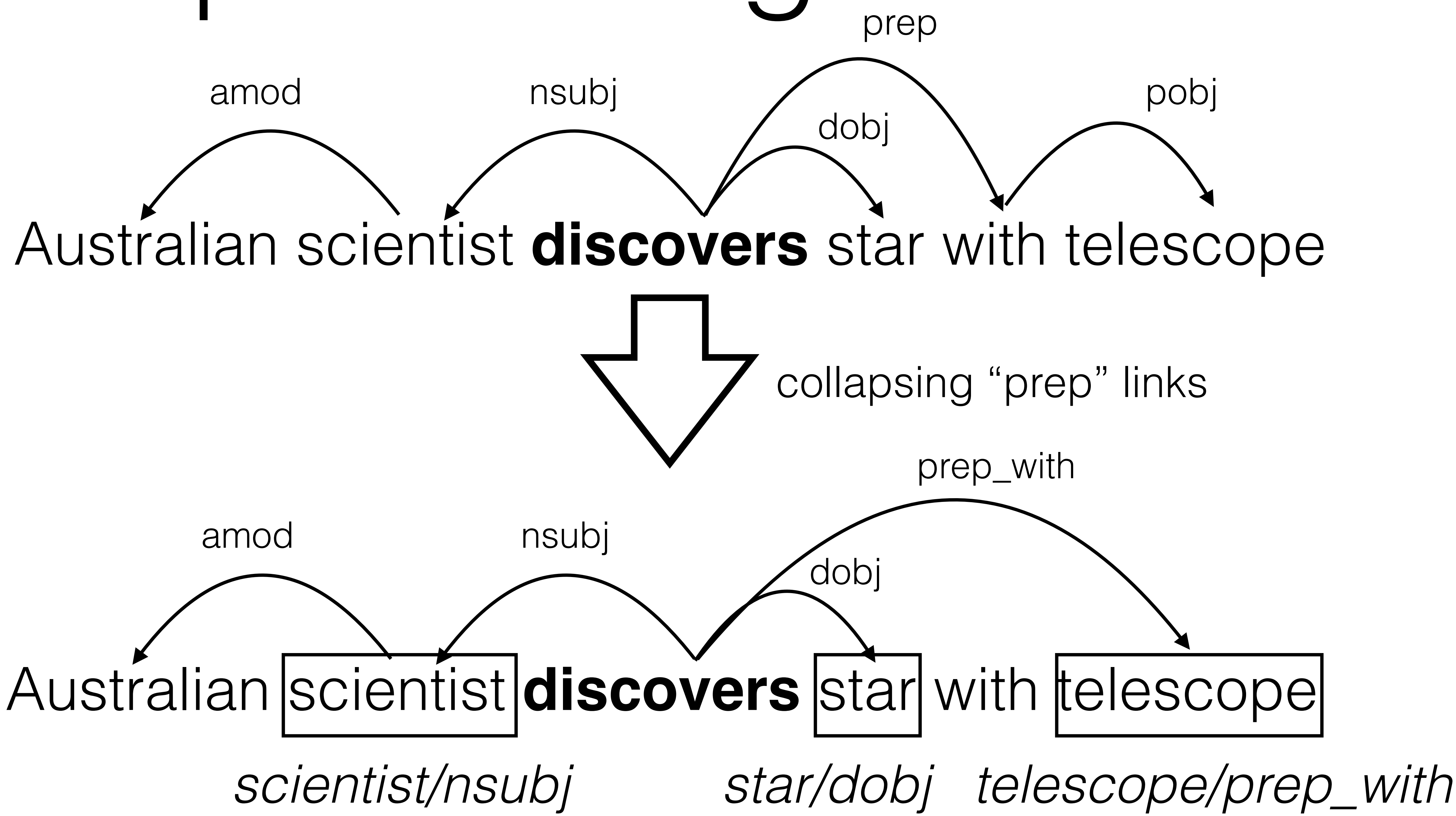
Levy and Goldberg, 2014

Representing Context



Levy and Goldberg, 2014

Representing Context



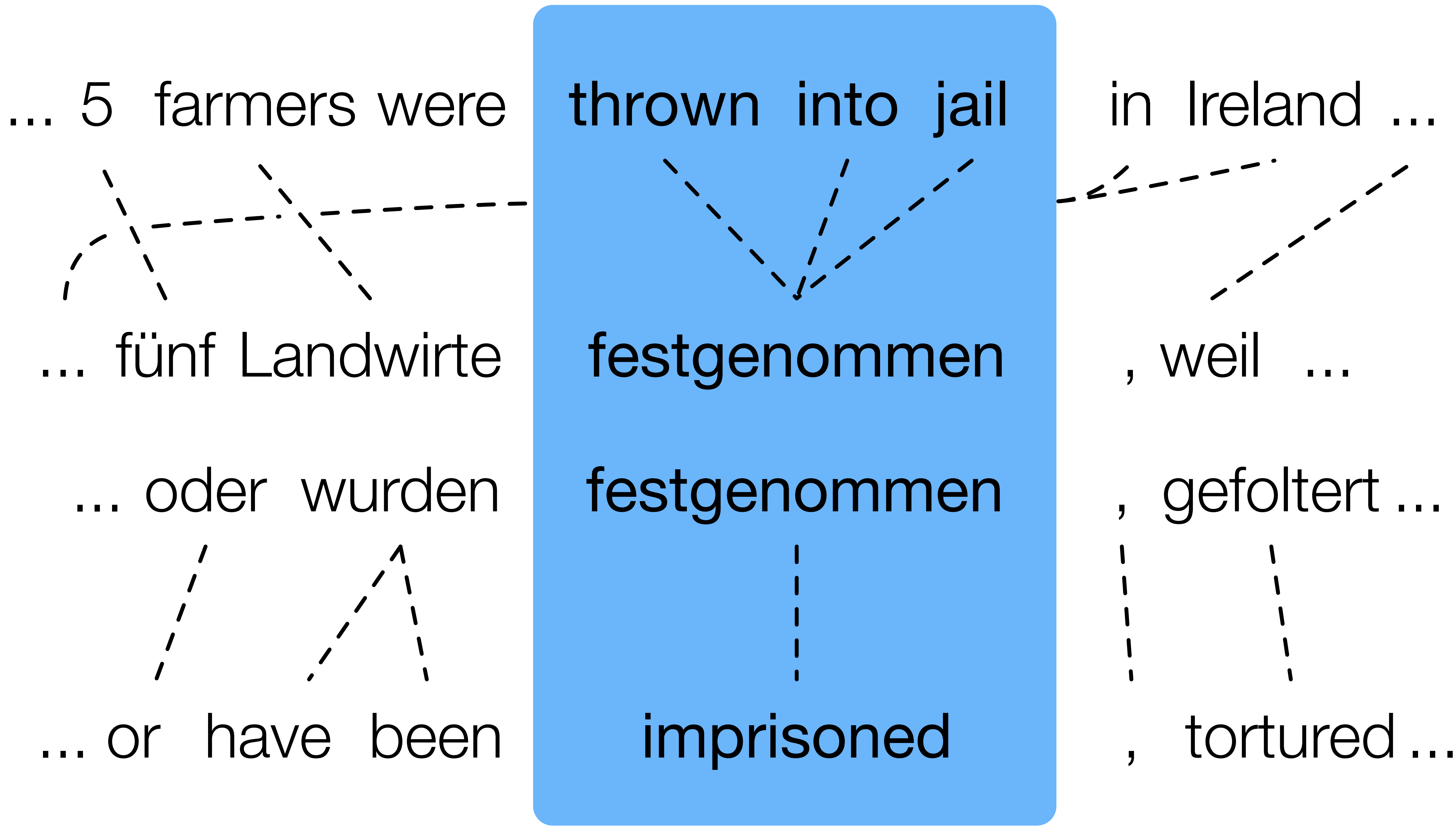
Levy and Goldberg, 2014

Representing Context

Target Word	BoW5	Deps
batman	nightwing aquaman catwoman superman manhunter	superman superboy supergirl catwoman aquaman
hogwarts	dumbledore hallows half-blood malfoy snape	sunnydale collinwood calarts greendale millfield
turing	nondeterministic non-deterministic computability deterministic finite-state	pauling hotelling heting lessing hamming
florida	gainesville fla jacksonville tampa lauderdale	texas louisiana georgia california carolina

Levy and Goldberg, 2014

Representing Context



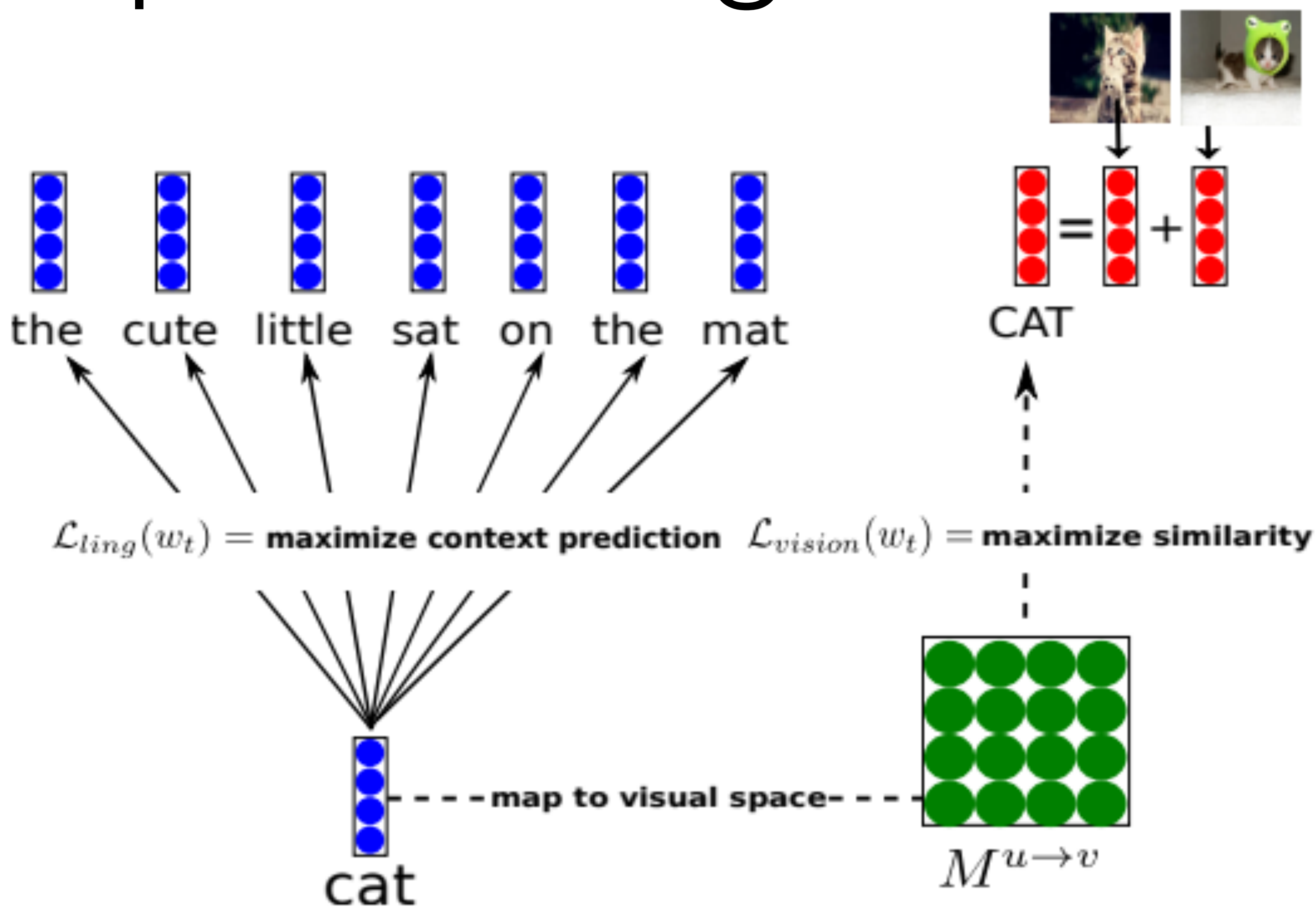
Bannard and Callison-Burch, 2005

Representing Context

Cosine Similarity		Monolingual (symmetric)		Bilingual	
$\sqsupset$	shades/the shade	$\neg$	large/small	$\equiv$	dad/father
$\sqsupset$	yard/backyard	$\equiv$	few/several	$\sqsupset$	some kid/child
#	each other/man	$\neg$	different/same	$\equiv$	a lot of/many
$\sqsupset$	picture/drawing	$\neg$	other/same	$\equiv$	female/woman
$\sim$	practice/target	$\neg$	put/take	$\equiv$	male/man

Pavlick et al, 2015

Representing Context



Lazaridou et al. (2015)

Representing Context

Target	SKIP-GRAM	MMSKIP-GRAM-B
donut	fridge, diner, candy	pizza, sushi, sandwich
owl	pheasant, woodpecker, squirrel	eagle, falcon, hawk
mural	sculpture, painting, portrait	painting, portrait, sculpture
tobacco	coffee, cigarette, corn	cigarette, cigar, smoking
depth	size, bottom, meter	sea, size, underwater
chaos	anarchy, despair, demon	demon, anarchy, shadow

Lazaridou et al. (2015)

Distributional Semantics: Takeaways

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

Natural Language Inference	Logical Forms

Sentence-Level Semantics

Natural Language Inference	Logical Forms
given a premise *p* and a hypothesis *h*, predict whether *p* entails *h*	given a sentence *s* return an executable representation (e.g. mathematical formula, SQL query…)

Sentence-Level Semantics

Natural Language Inference	Logical Forms
ungrounded—relate text to other text	grounded—relate text to tables in a database, or actions on a robot

Sentence-Level Semantics

Natural Language Inference	Logical Forms
ungrounded—relate text to other text	grounded—relate text to tables in a database, or actions on a robot

Semantic Parsing

Semantic Parsing

What is the largest state?

NAME	Alaska
Abbrev.	AK
SIZE	663,268
CAPITAL	Juneau

Alaska

Semantic Parsing

What is the largest state?

NAME	Alaska
Abbrev.	AK
SIZE	663,268
CAPITAL	Juneau

Alaska

Semantic Parsing

What is the largest state? →

...
Prolog
SQL
FunQL
First-order logic
λ-calculus
...

→

NAME	Alaska
Abbrev.	AK
SIZE	663,268
CAPITAL	Juneau

→ Alaska

Supervised Semantic Parsing

What is the largest state? → λ-calculus →

NAME	Alaska
Abbrev.	AK
SIZE	663,268
CAPITAL	Juneau

→ Alaska

Zettlemoyer and Collins (2012)

Supervised Semantic Parsing

What is the largest state?

`argmax(λx.state(x),λx.size(x))`

Zettlemoyer and Collins (2012)

Supervised Semantic Parsing

What is the largest state?

`argmax(λx.state(x),λx.size(x))`

Zettlemoyer and Collins (2012)

Combinatory Categorial Grammar

state := `NP : state`

Zettlemoyer and Collins (2012)

Combinatory Categorial Grammar

state := `NP : state`

Syntax

Zettlemoyer and Collins (2012)

Combinatory Categorial Grammar

Semantics

state := NP : state

Syntax

Zettlemoyer and Collins (2012)

Combinatory Categorial Grammar

borders := `(S\NP)/NP : λx.λy.borders(y,x)`

Zettlemoyer and Collins (2012)

Combinatory Categorial Grammar

borders := **(S\NP)/NP** : λx.λy.borders(y,x)

utah	borders	idaho
NP	(S**NP**)/**NP**	**NP**

Zettlemoyer and Collins (2012)

Combinatory Categorial Grammar

borders := (S\NP)/NP : **λx.λy.borders(y,x)**

borders(utah,idaho)

Zettlemoyer and Collins (2012)

Combinatory Categorial Grammar

utah	borders	idaho
NP	(S\NP)/NP	NP
utah	λx.λy.borders(y,x)	idaho

Zettlemoyer and Collins (2012)

Combinatory Categorial Grammar

utah	borders	idaho
NP	(S\NP)/**NP**	**NP**
utah	**λx**.λy.borders(y,**x**)	**idaho**

Zettlemoyer and Collins (2012)

Combinatory Categorial Grammar

utah	borders idaho
NP	(S\NP)
utah	λy.borders(y,idaho)

Zettlemoyer and Collins (2012)

Combinatory Categorial Grammar

utah	borders idaho
NP	(S**NP**)
utah	**λy**.borders(**y**,idaho)

Zettlemoyer and Collins (2012)

Combinatory Categorial Grammar

utah borders idaho

S

`borders(utah,idaho)`

Zettlemoyer and Collins (2012)

Supervised Semantic Parsing

Does Utah border Idaho? → `borders(utah,idaho)` →

(utah, idaho)	TRUE
(utah, michigan)	FALSE
(idaho, michigan)	FALSE
(idaho, montatana)	TRUE

→ Yes

Zettlemoyer and Collins (2012)

Supervised Semantic Parsing

Does Utah border Idaho? → `borders(utah,idaho)` →

(utah, idaho)	TRUE
(utah, michigan)	FALSE
(idaho, michigan)	FALSE
(idaho, montatana)	TRUE

→ Yes

What is the largest state?	S:argmax(λx.state(x),λx.size(x))
Utah borders Idaho	S:borders(utah,idaho)
highest point in the US	NP:argmax(λx.point(x),λx.elevation(x))
city in California	NP:λx.(city(x)∧location(x,CA))

training data

Zettlemoyer and Collins (2012)

Weakly-supervised Semantic Parsing

Does Utah border Idaho?

`borders(utah,idaho)`

(utah, idaho)	TRUE
(utah, michigan)	FALSE
(idaho, michigan)	FALSE
(idaho, montatana)	TRUE

Yes

Liang et al. (2012)

Weakly-supervised Semantic Parsing

Does Utah border Idaho? → `borders(utah,idaho)` →

(utah, idaho)	TRUE
(utah, michigan)	FALSE
(idaho, michigan)	FALSE
(idaho, montatana)	TRUE

→ Yes

Liang et al. (2012)

Weakly-supervised Semantic Parsing

Does Utah border Idaho? → `borders(utah,idaho)` →

(utah, idaho)	TRUE
(utah, michigan)	FALSE
(idaho, michigan)	FALSE
(idaho, montatana)	TRUE

→ Yes

Liang et al. (2012)

Weakly-supervised Semantic Parsing

Liang et al. (2012)

Weakly-supervised Semantic Parsing

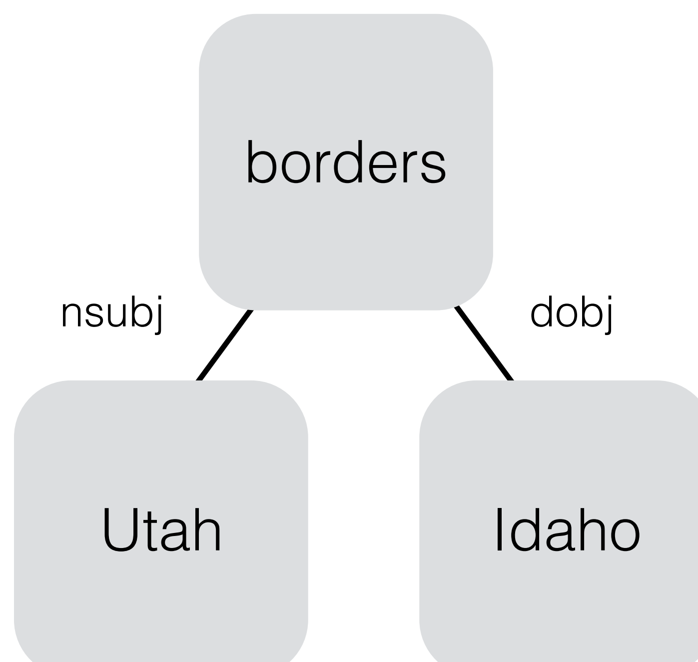
What is the largest state?	Alaska
Utah borders Idaho	TRUE
highest point in the US	Mt. McKinley
city in California	Los Angeles, San Fransisco...

training data

Liang et al. (2012)

Dependency-based compositional semantics

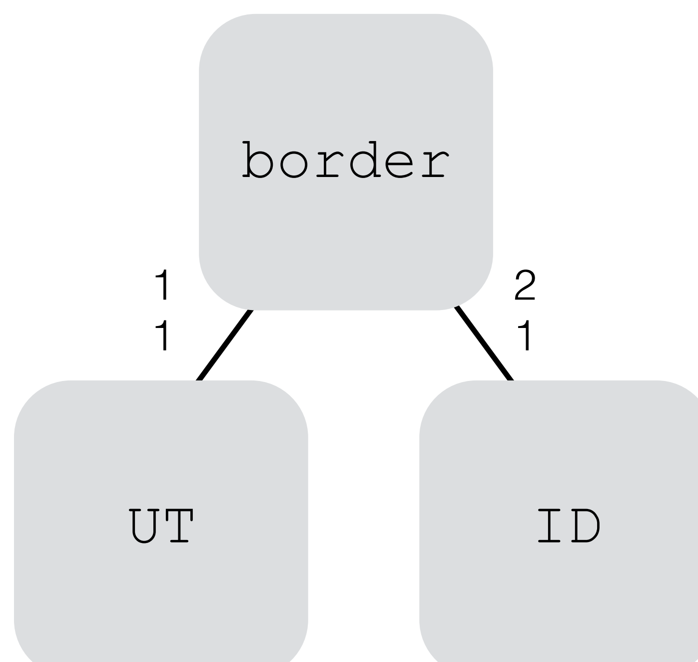
Utah borders Idaho



Liang et al. (2012)

Dependency-based compositional semantics

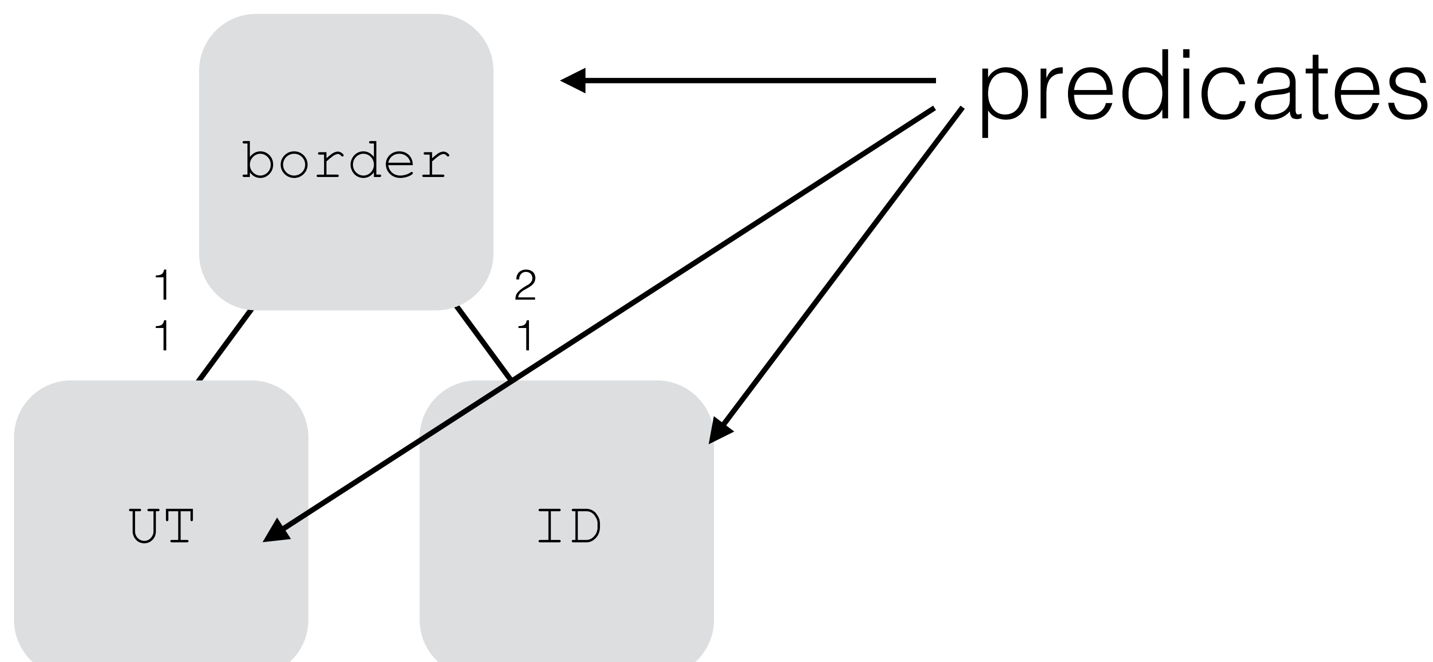
Utah borders Idaho



Liang et al. (2012)

Dependency-based compositional semantics

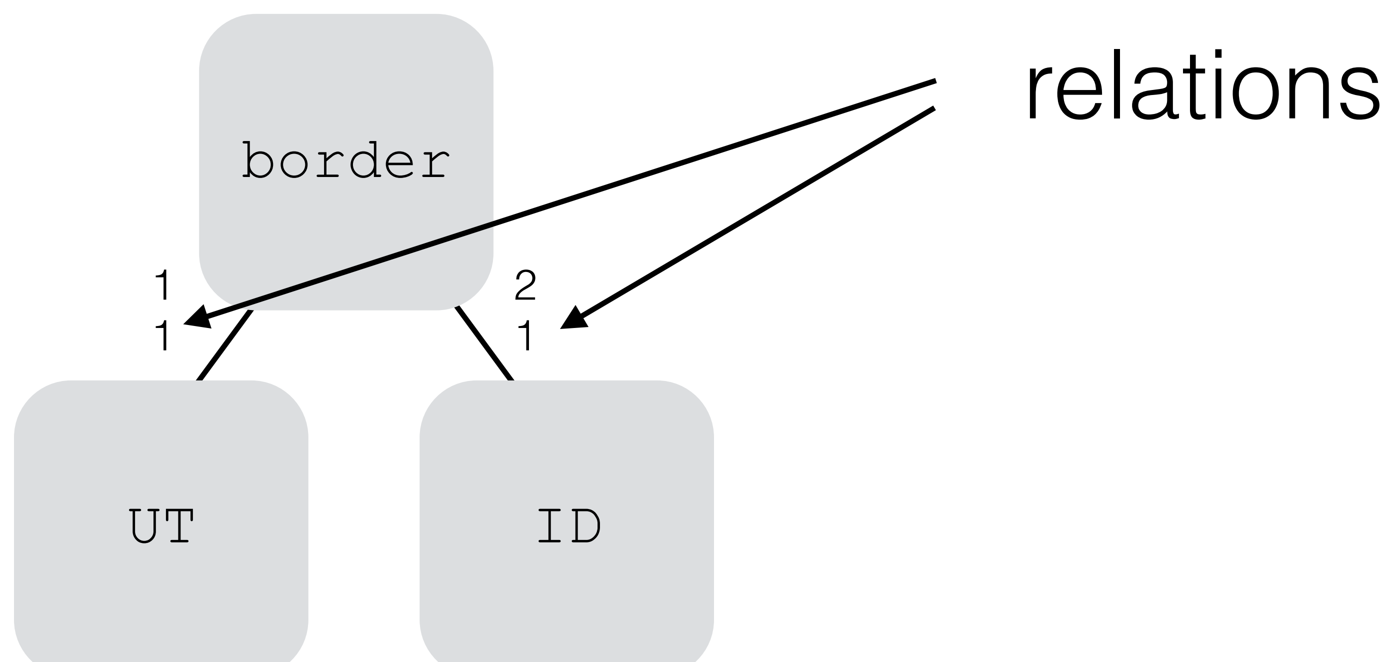
Utah borders Idaho



Liang et al. (2012)

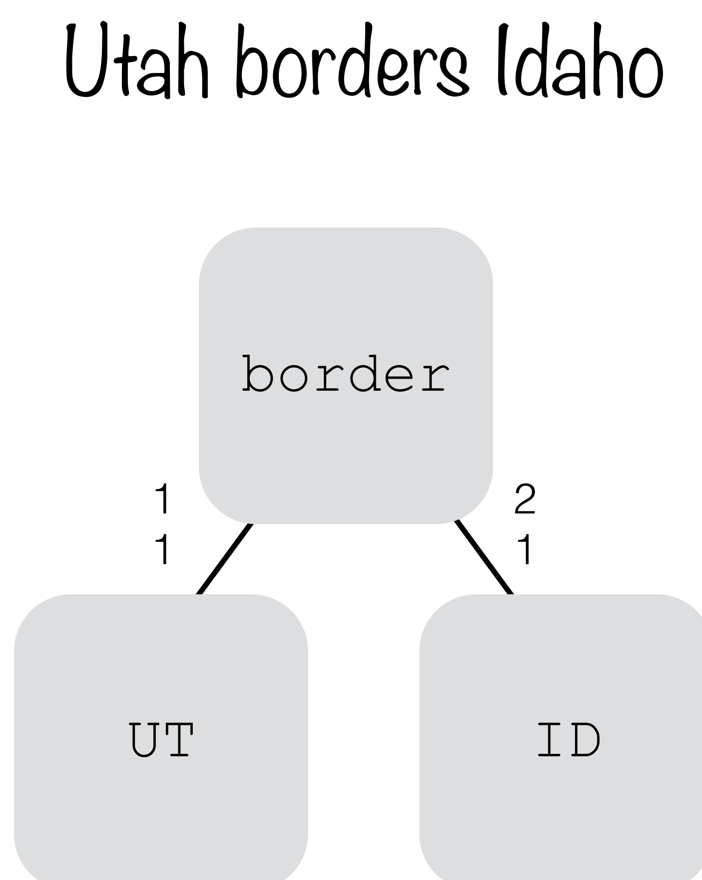
Dependency-based compositional semantics

Utah borders Idaho



Liang et al. (2012)

Dependency-based compositional semantics

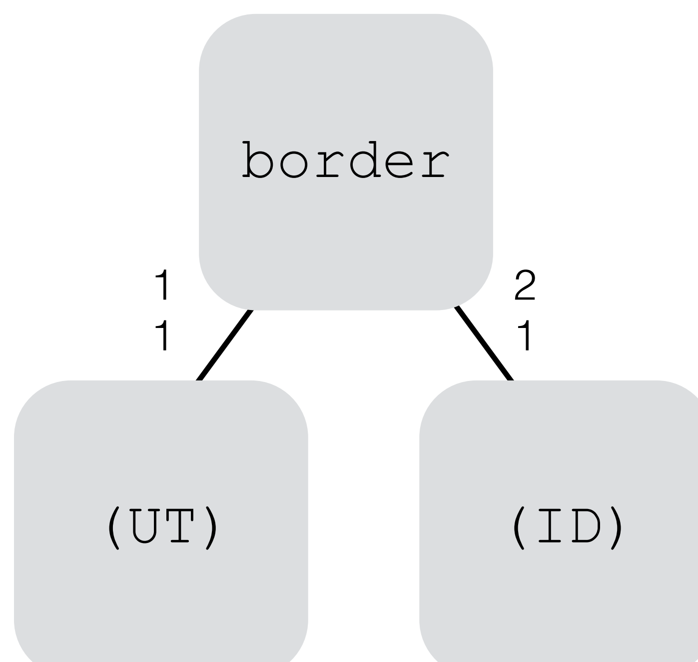


(UT,ID)	TRUE
(UT,MI)	FALSE
(ID,MI)	FALSE
(ID,MT)	TRUE

Liang et al. (2012)

Dependency-based compositional semantics

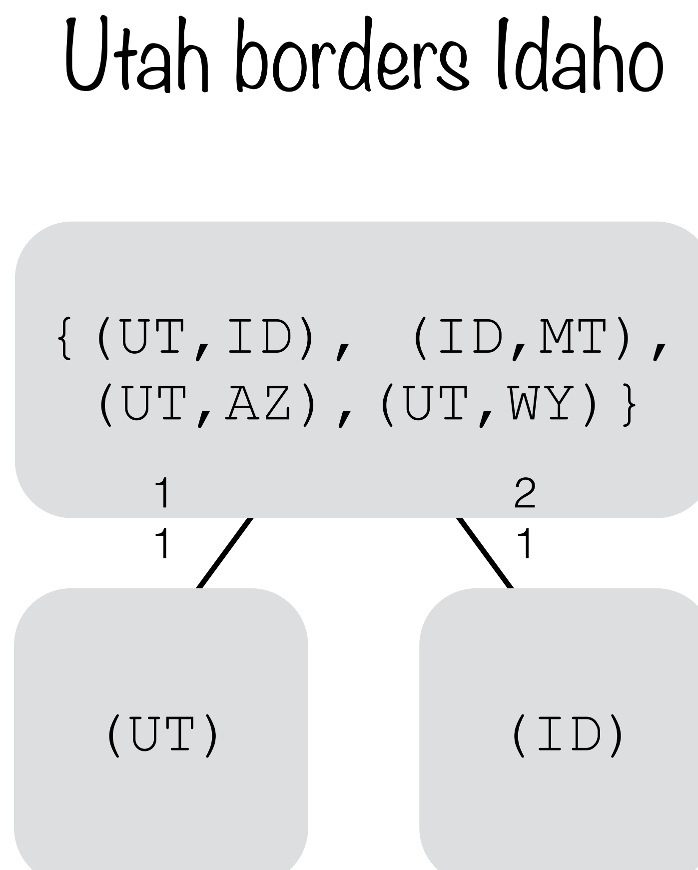
Utah borders Idaho



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

Liang et al. (2012)

Dependency-based compositional semantics

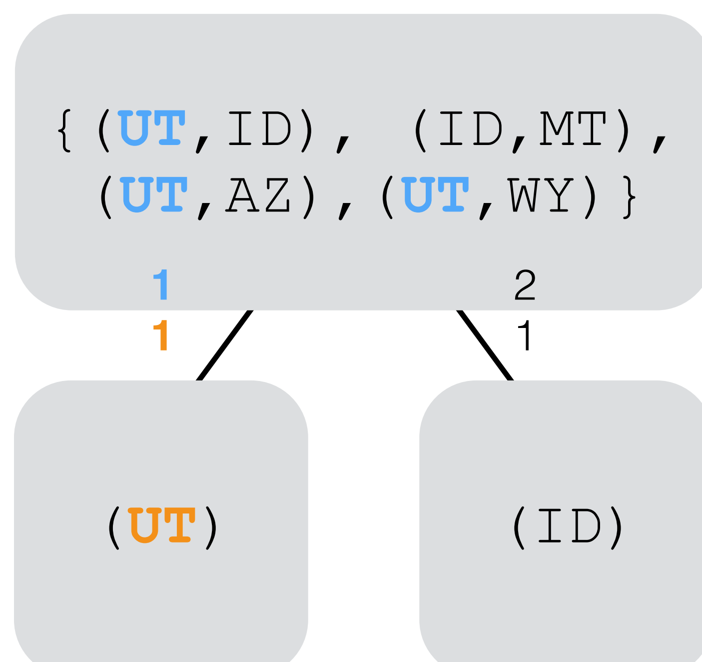


(UT,ID)	TRUE
(UT,MI)	FALSE
(ID,MI)	FALSE
(ID,MT)	TRUE

Liang et al. (2012)

Dependency-based compositional semantics

Utah borders Idaho

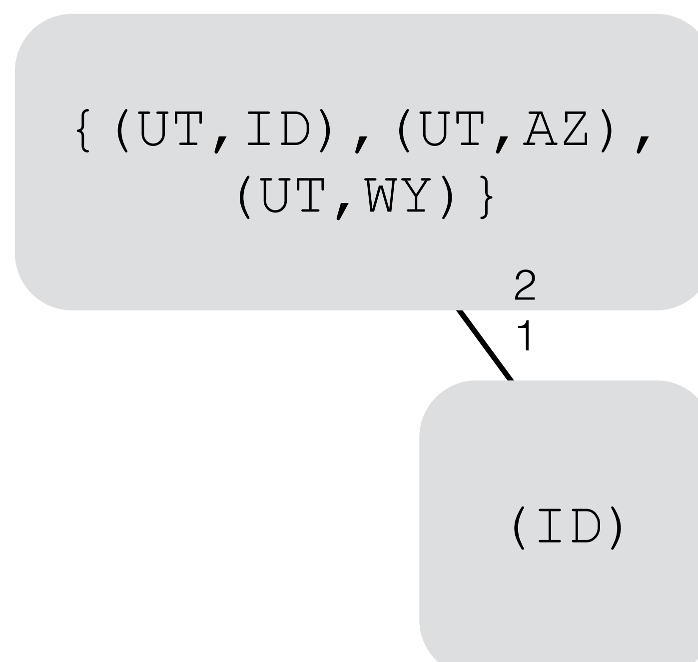


(UT,ID)	TRUE
(UT,MI)	FALSE
(ID,MI)	FALSE
(ID,MT)	TRUE

Liang et al. (2012)

Dependency-based compositional semantics

Utah borders Idaho

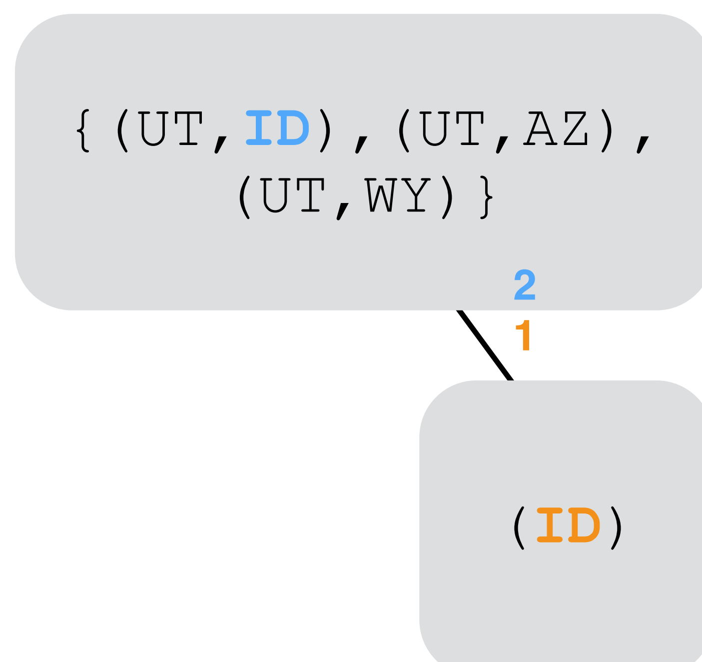


(UT,ID)	TRUE
(UT,MI)	FALSE
(ID,MI)	FALSE
(ID,MT)	TRUE

Liang et al. (2012)

Dependency-based compositional semantics

Utah borders Idaho



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

Liang et al. (2012)

Dependency-based compositional semantics

Utah borders Idaho

{(UT,ID)}

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

Liang et al. (2012)

Dependency-based compositional semantics

Utah borders Idaho

TRUE!

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

Liang et al. (2012)

Dependency-based compositional semantics

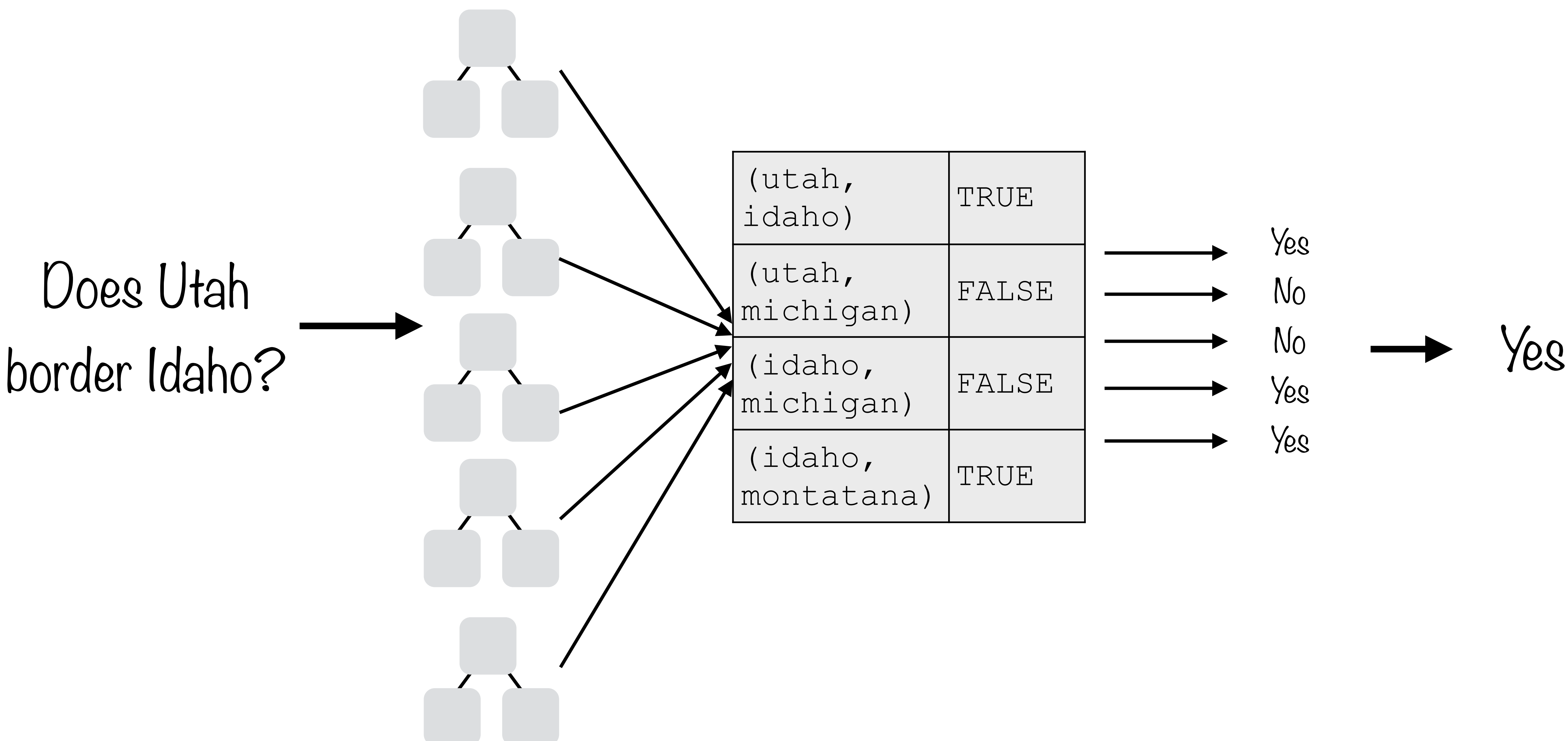
Does Utah border Idaho? → `borders(utah,idaho)` →

(utah, idaho)	TRUE
(utah, michigan)	FALSE
(idaho, michigan)	FALSE
(idaho, montatana)	TRUE

→ Yes

Liang et al. (2012)

Dependency-based compositional semantics



Liang et al. (2012)

Sentence-Level Semantics

Natural Language Inference	Logical Forms
ungrounded—relate text to other text	grounded—relate text to tables in a database, or actions on a robot

Compositional Distributional Semantics

Compositional Distributional Semantics

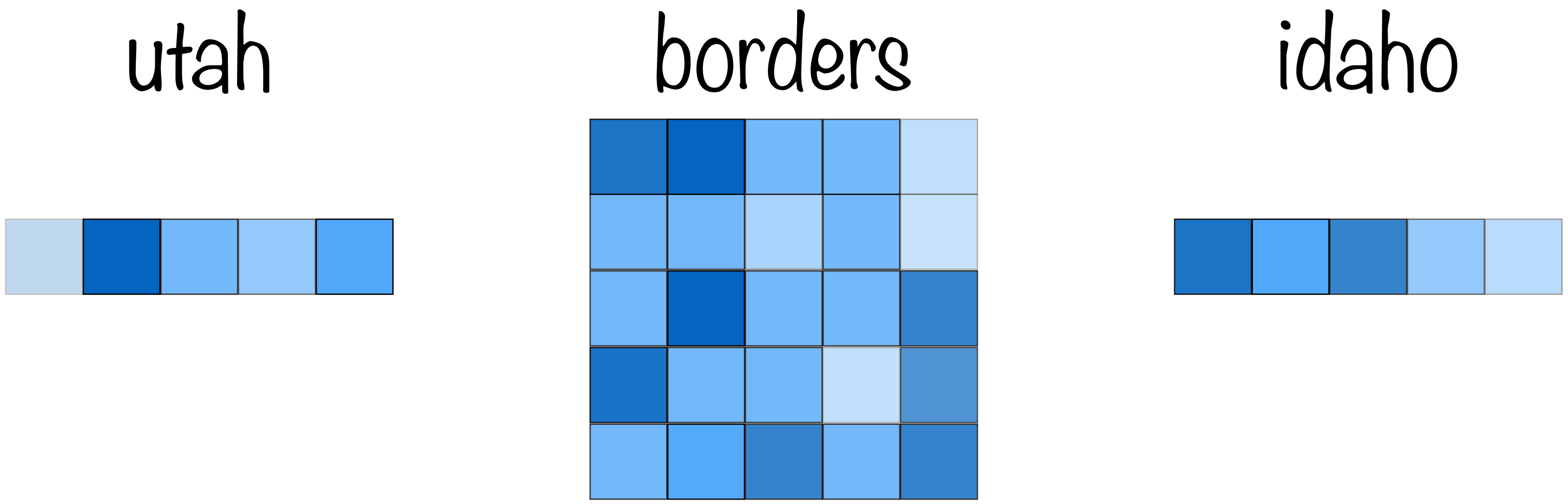
utah	borders	idaho
NP	(S\NP)/NP	NP

Grefenstette and Sadrzadeh (2011)

Compositional Distributional Semantics

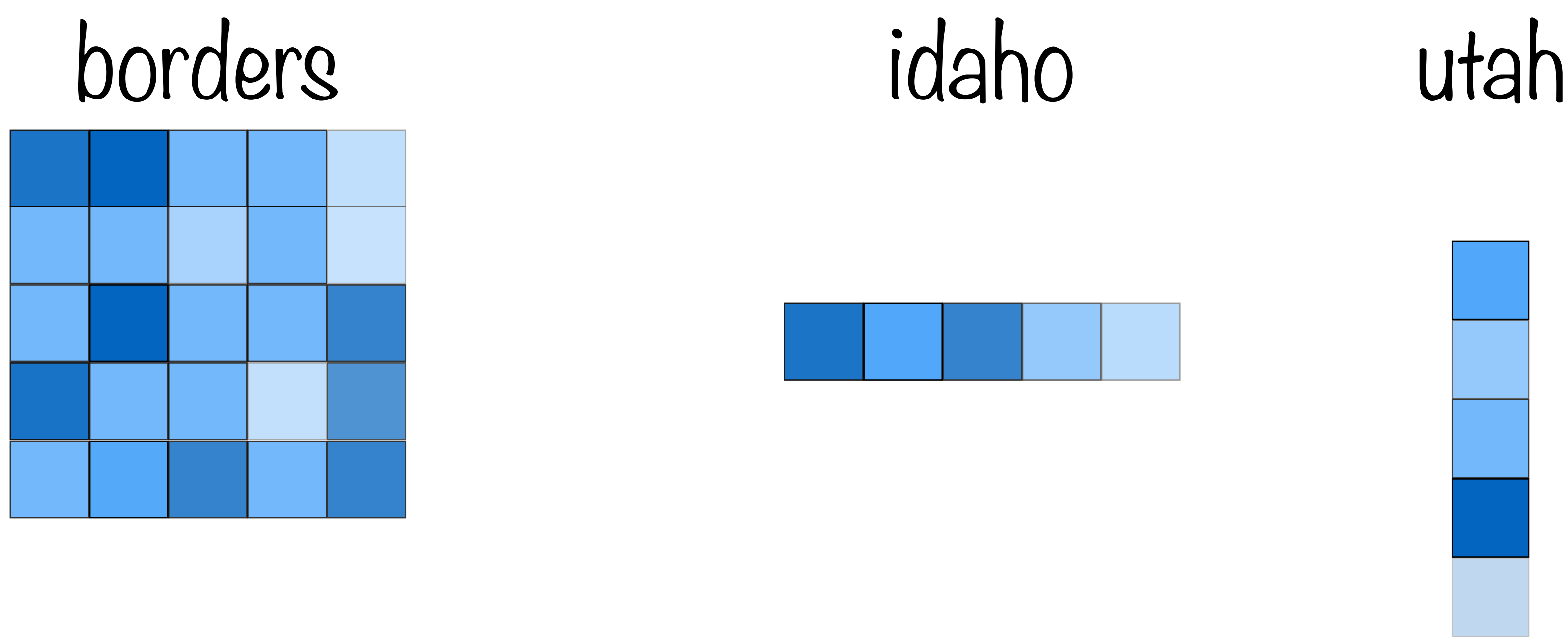
Grefenstette and Sadrzadeh (2011)

Compositional Distributional Semantics



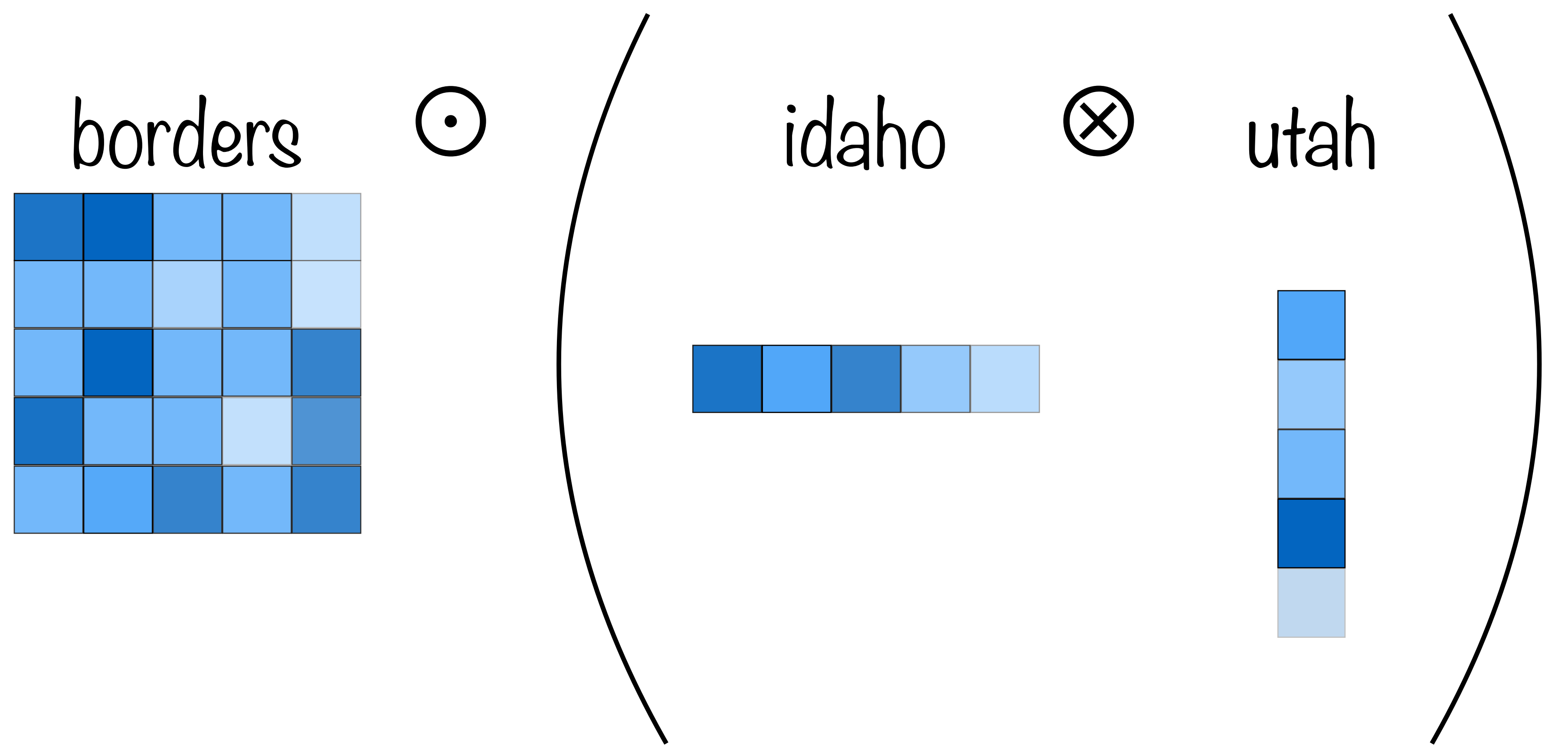
Grefenstette and Sadrzadeh (2011)

Compositional Distributional Semantics



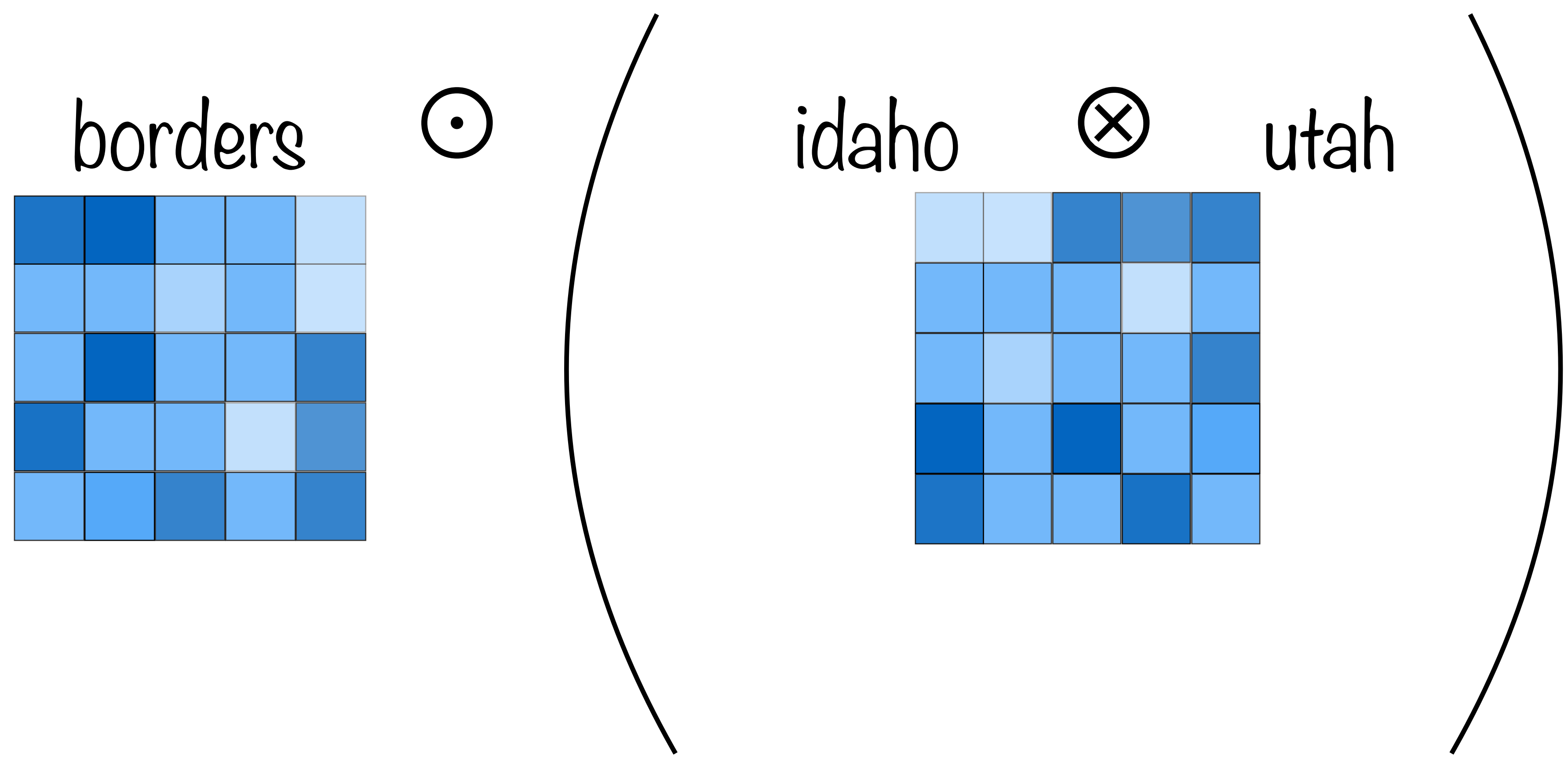
Grefenstette and Sadrzadeh (2011)

Compositional Distributional Semantics



Grefenstette and Sadrzadeh (2011)

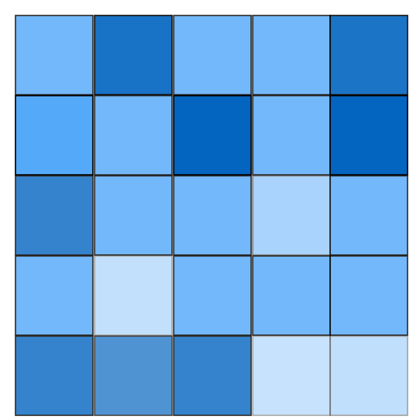
Compositional Distributional Semantics



Grefenstette and Sadrzadeh (2011)

Compositional Distributional Semantics

borders $\odot$ idaho $\otimes$ utah

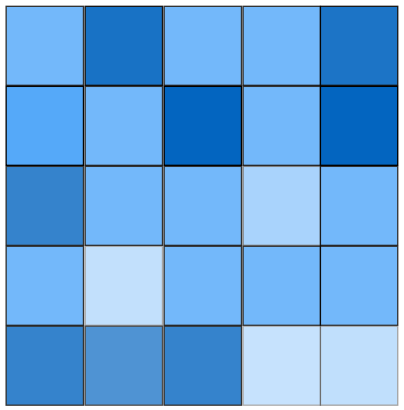


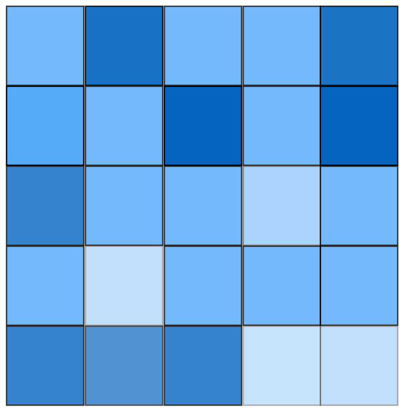
Grefenstette and Sadrzadeh (2011)

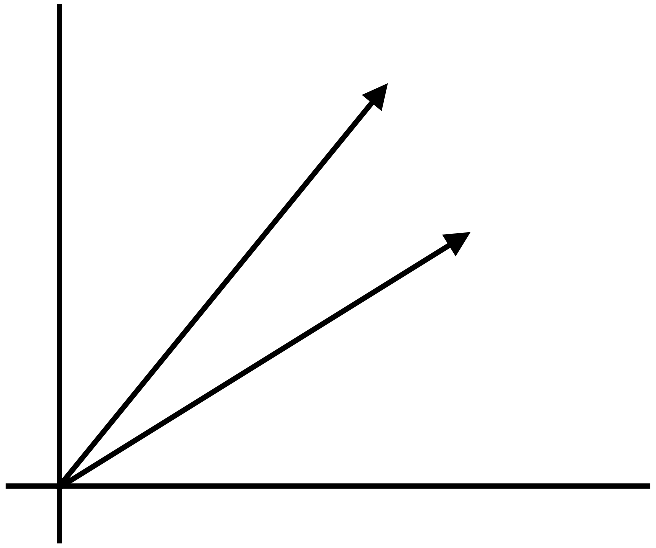
Compositional Distributional Semantics

borders $\odot$ idaho $\otimes$ utah

is near $\odot$ boise $\otimes$ ogden







Grefenstette and Sadrzadeh (2011)

Compositional Distributional Semantics

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

Table 2: Nearest 3 neighbors of specific ANs.

Grefenstette and Sadrzadeh (2011)

Compositional Distributional Semantics

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

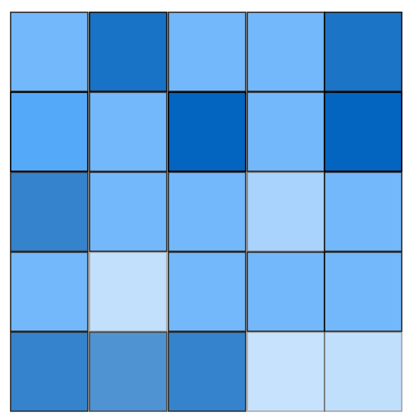
Table 2: Nearest 3 neighbors of specific ANs.

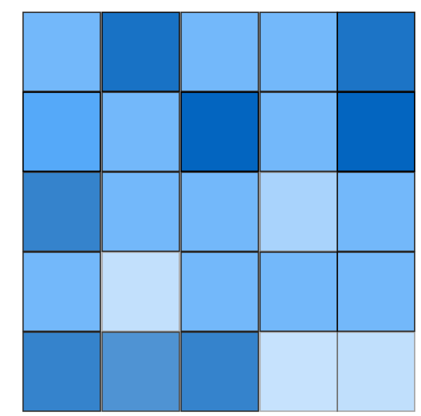
Grefenstette and Sadrzadeh (2011)

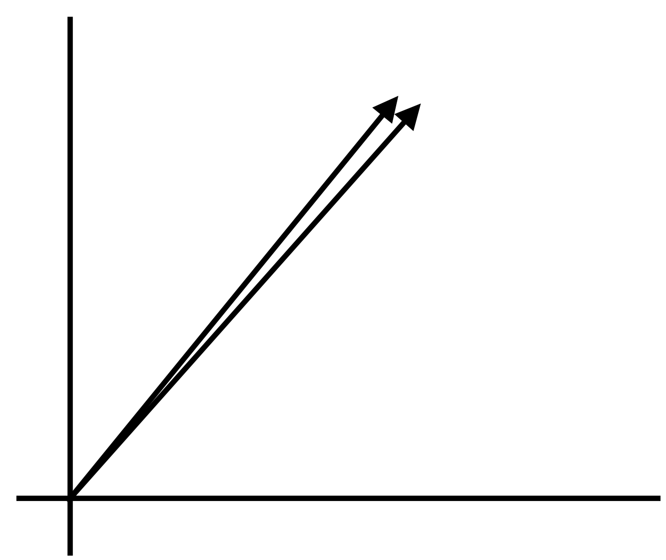
Compositional Distributional Semantics

borders $\odot$ idaho $\otimes$ utah

borders $\odot$ idaho $\otimes$ texas







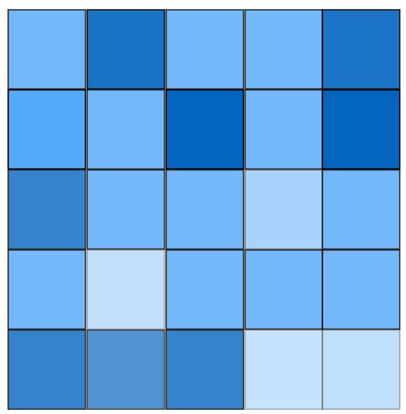
???

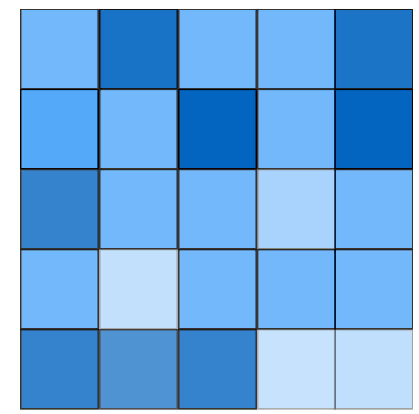
Grefenstette and Sadrzadeh (2011)

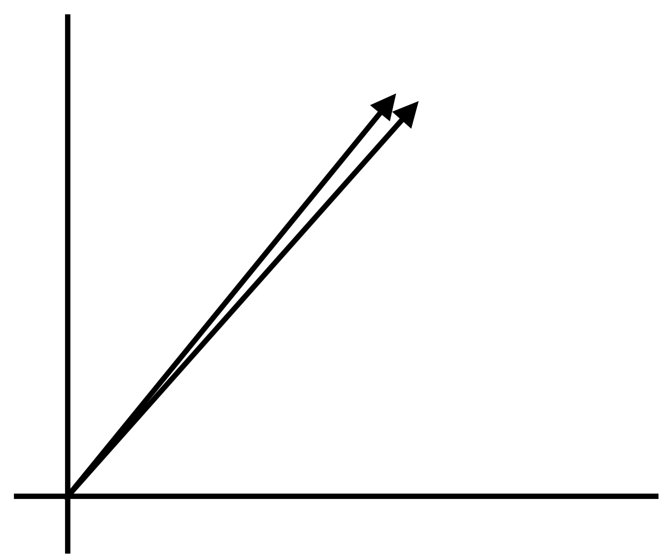
Compositional Distributional Semantics

borders $\odot$ idaho $\otimes$ utah

doesn't border $\odot$ idaho $\otimes$ utah



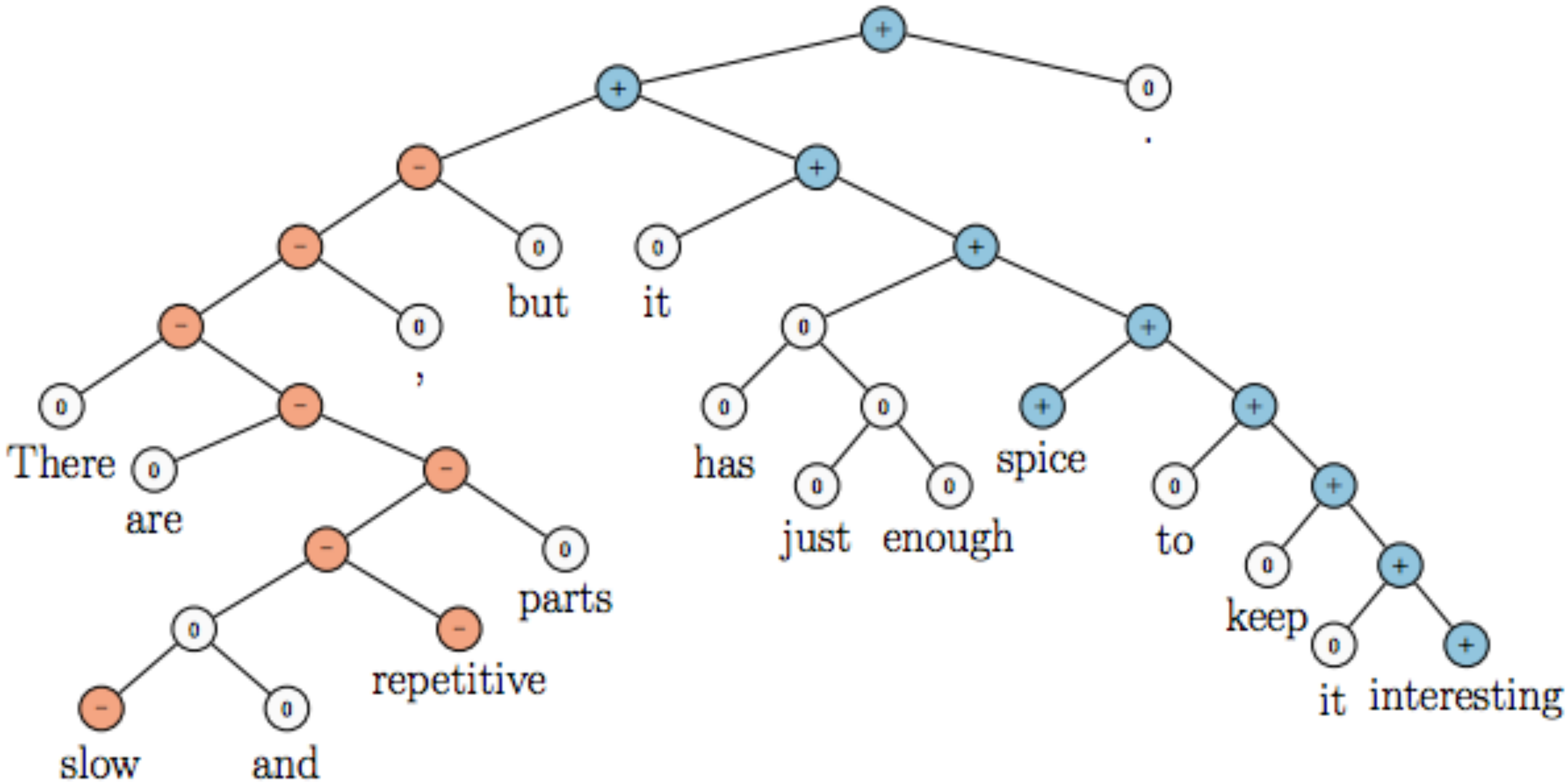




???

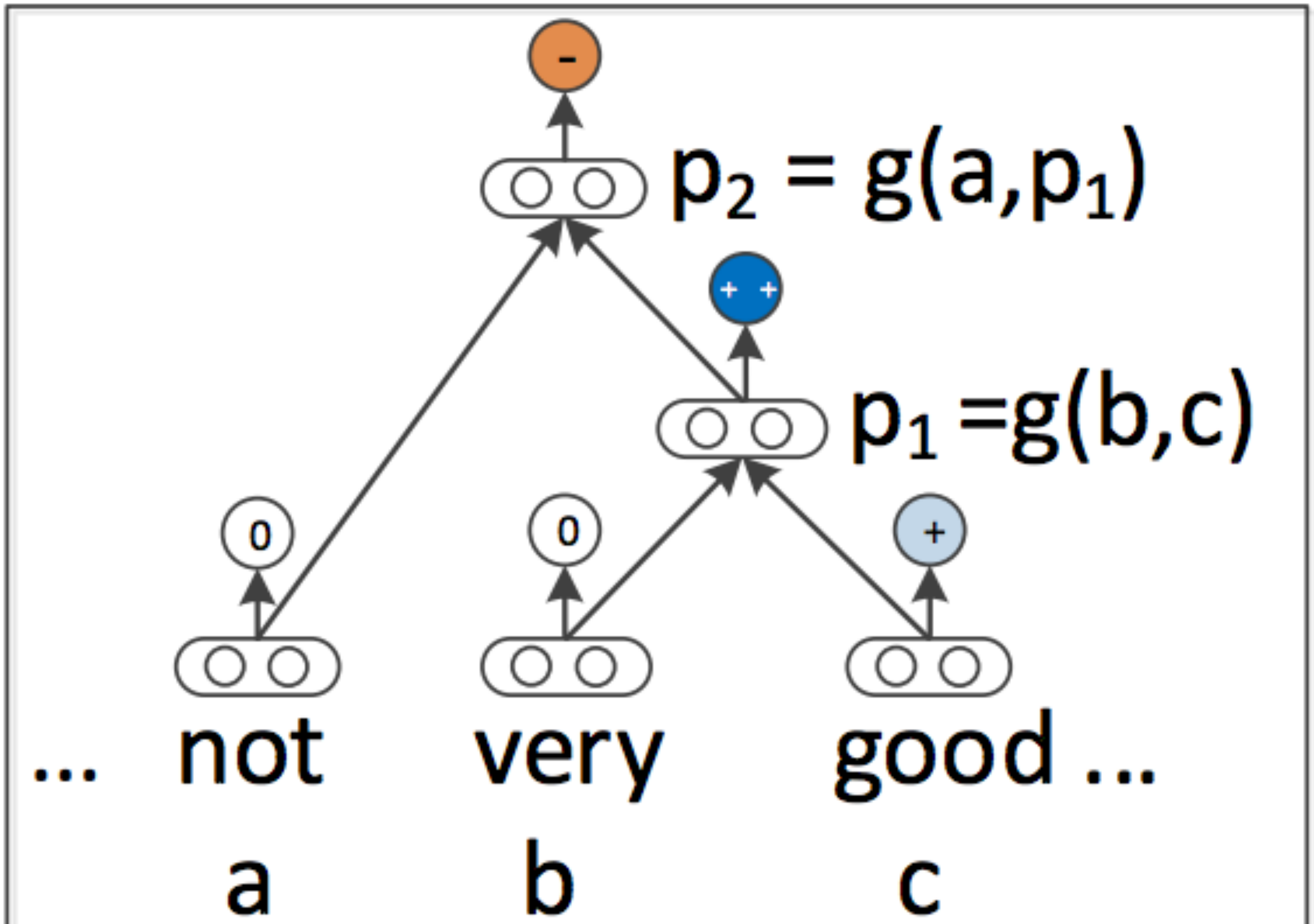
Grefenstette and Sadrzadeh (2011)

Compositional Distributional Semantics



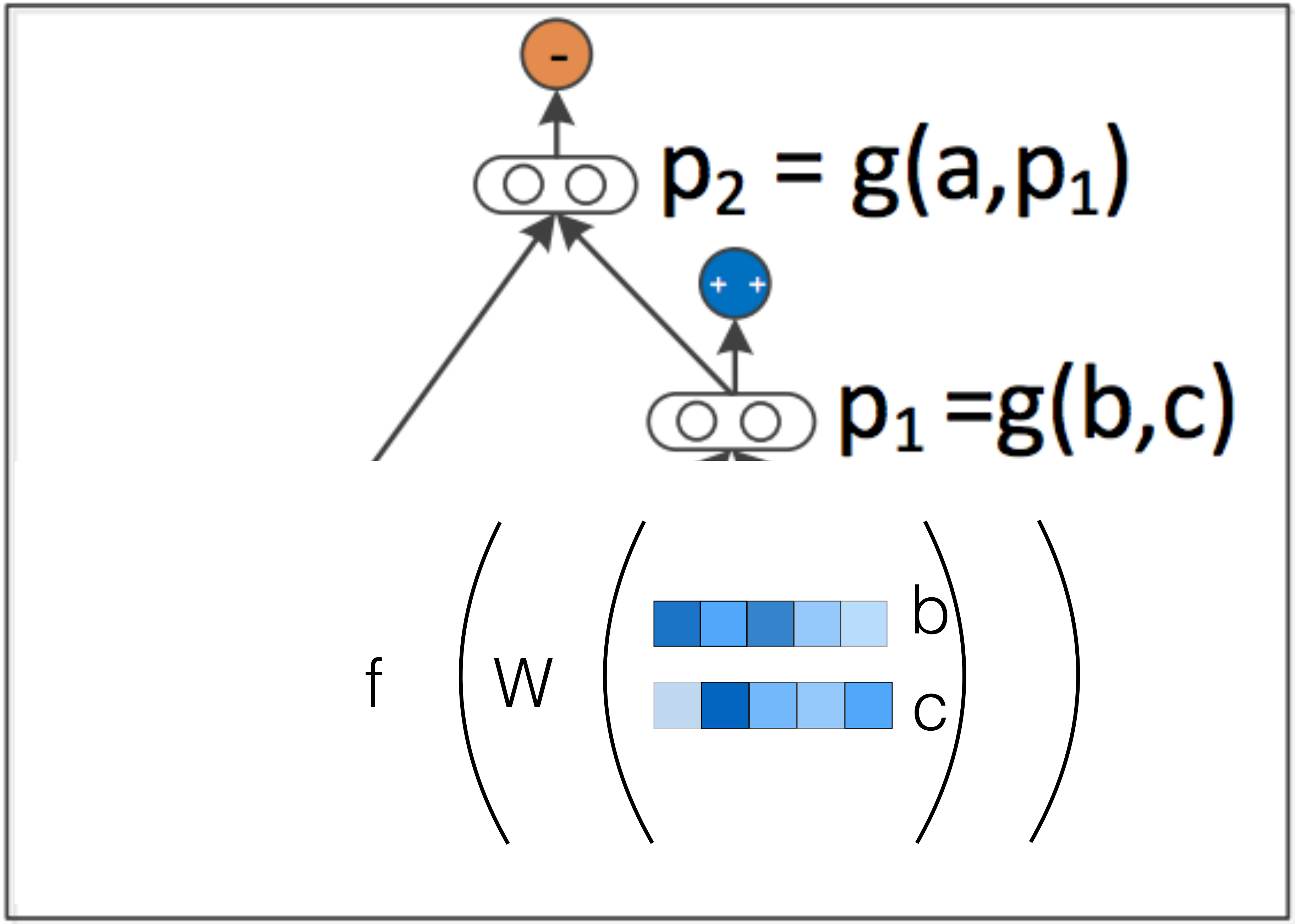
Socher et al (2013)

Compositional Distributional Semantics



Socher et al (2013)

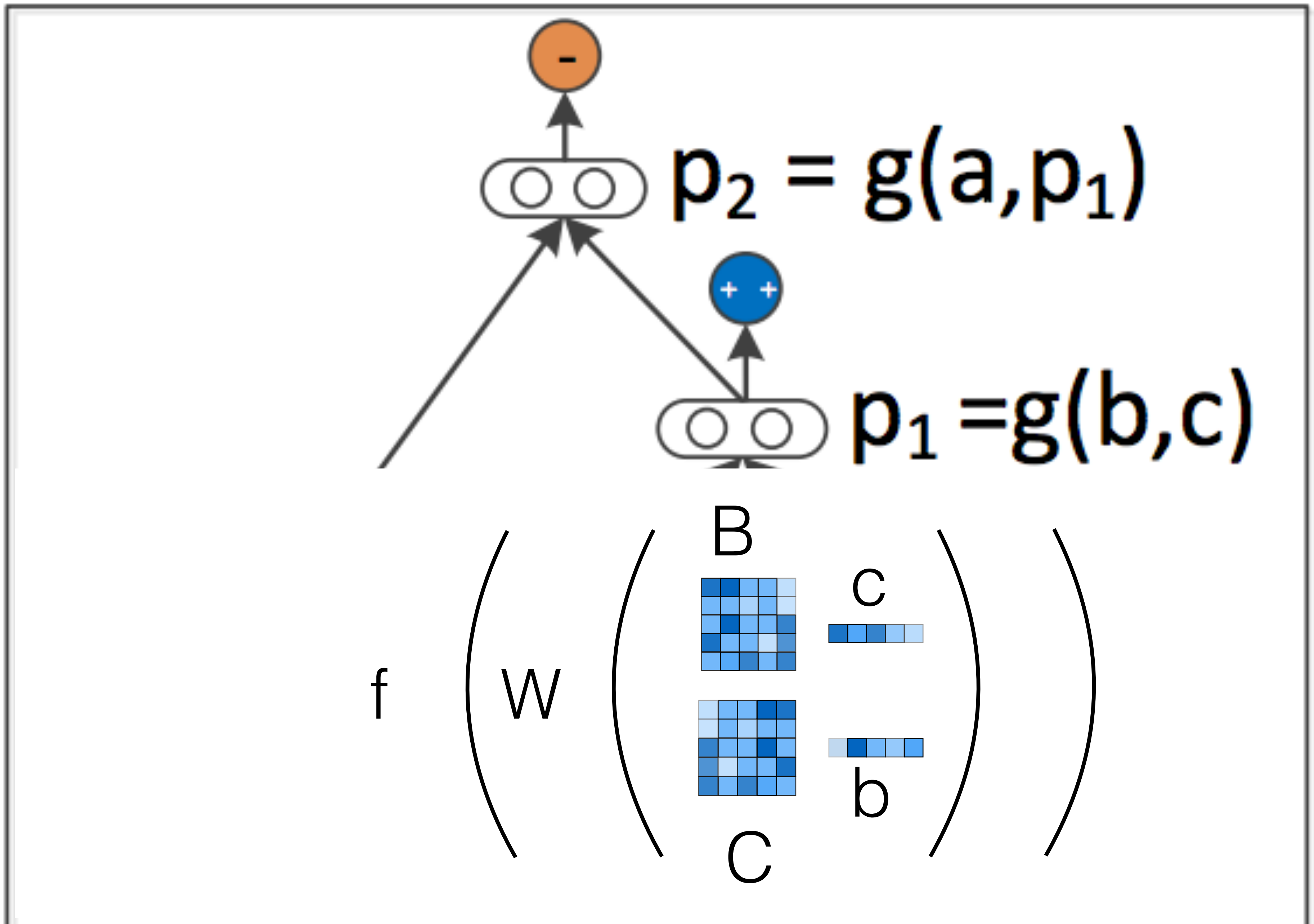
Compositional Distributional Semantics



Basic RNN

Socher et al (2013)

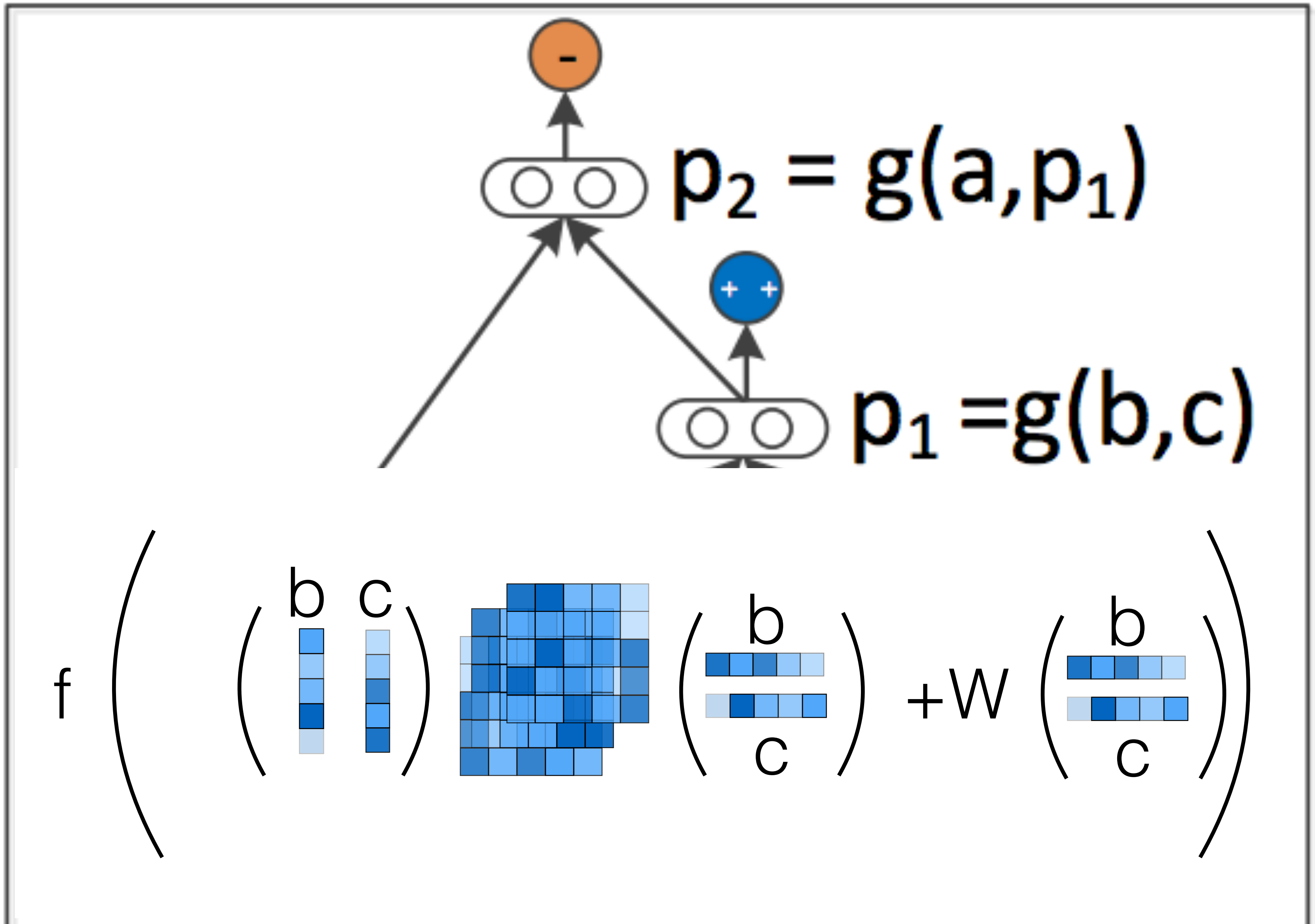
Compositional Distributional Semantics



Matrix-Vector RNN

Socher et al (2013)

Compositional Distributional Semantics



Recursive Neural Tensor Network (2013)

Compositional Distributional Semantics

Model	Accuracy	
	Negated Positive	Negated Negative
biNB	19.0	27.3
RNN	33.3	45.5
MV-RNN	52.4	54.6
RNTN	**71.4**	**81.8**

Socher et al (2013)

Compositional Distributional Semantics

Model	Accuracy	
	Negated Positive	Negated Negative
biNB	19.0	27.3
RNN	33.3	45.5
MV-RNN	52.4	54.6
RNTN	**71.4**	**81.8**

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

Socher et al (2013)

Compositional Distributional Semantics

Model	Accuracy	
	Negated Positive	Negated Negative
biNB	19.0	27.3
RNN	33.3	45.5
MV-RNN	52.4	54.6
RNTN	**71.4**	**81.8**

The movie was [not] terrible.

Socher et al (2013)

Natural Logic

Natural Logic

Every person danced.

Every young woman danced.

Natural Logic

Every person danced.

$\forall x(\mathrm{person}(x) \rightarrow \mathrm{danced}(x))$

$\forall x((\mathrm{woman}(x) \wedge \mathrm{young}(x)) \rightarrow \mathrm{danced}(x))$

Every young woman danced.

Natural Logic

Every person danced.

$\forall x(\text{person}(x) \rightarrow \text{danced}(x))$

$\forall x(\text{woman}(x) \rightarrow \text{person}(x))$

$\forall x(\forall P((P(x) \wedge \text{young}(x)) \rightarrow P(x)))$

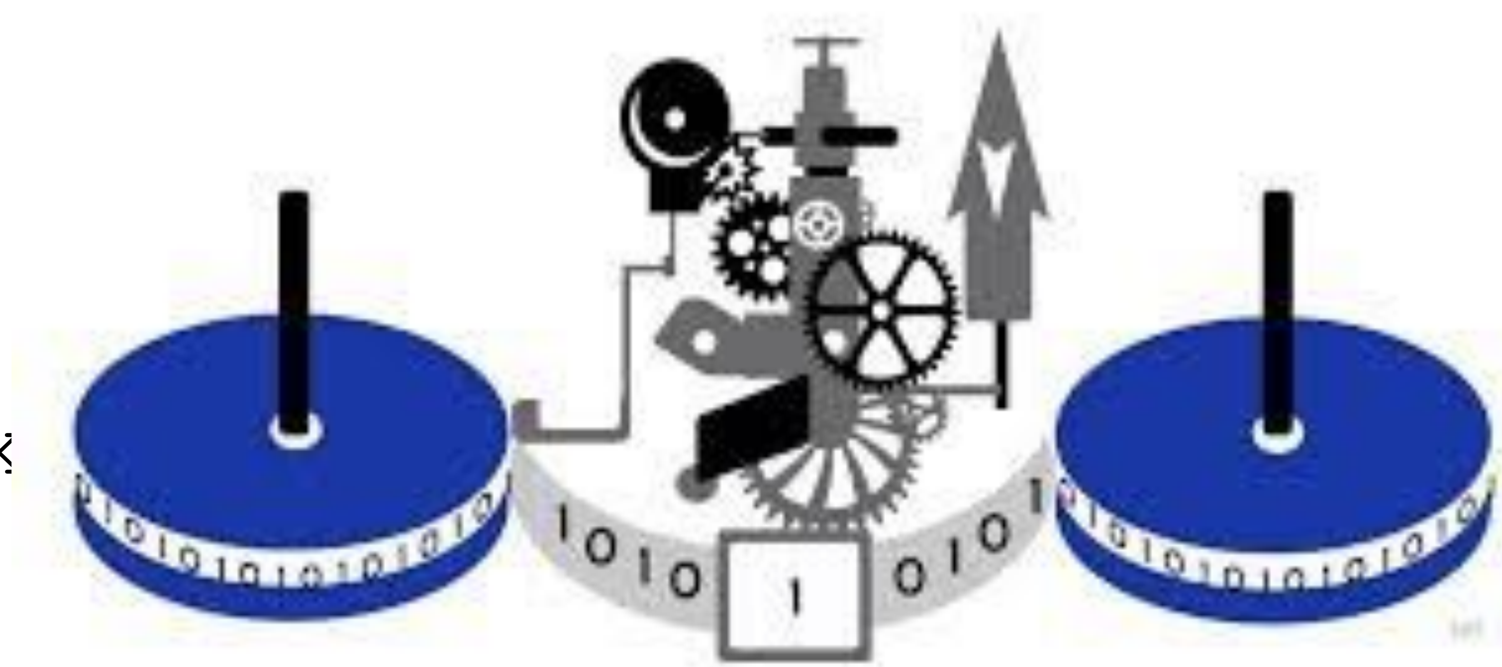
$\forall x((\text{woman}(x) \wedge \text{young}(x)) \rightarrow \text{danced}(x))$

Every young woman danced.

Natural Logic

Every person danced.

∀x))



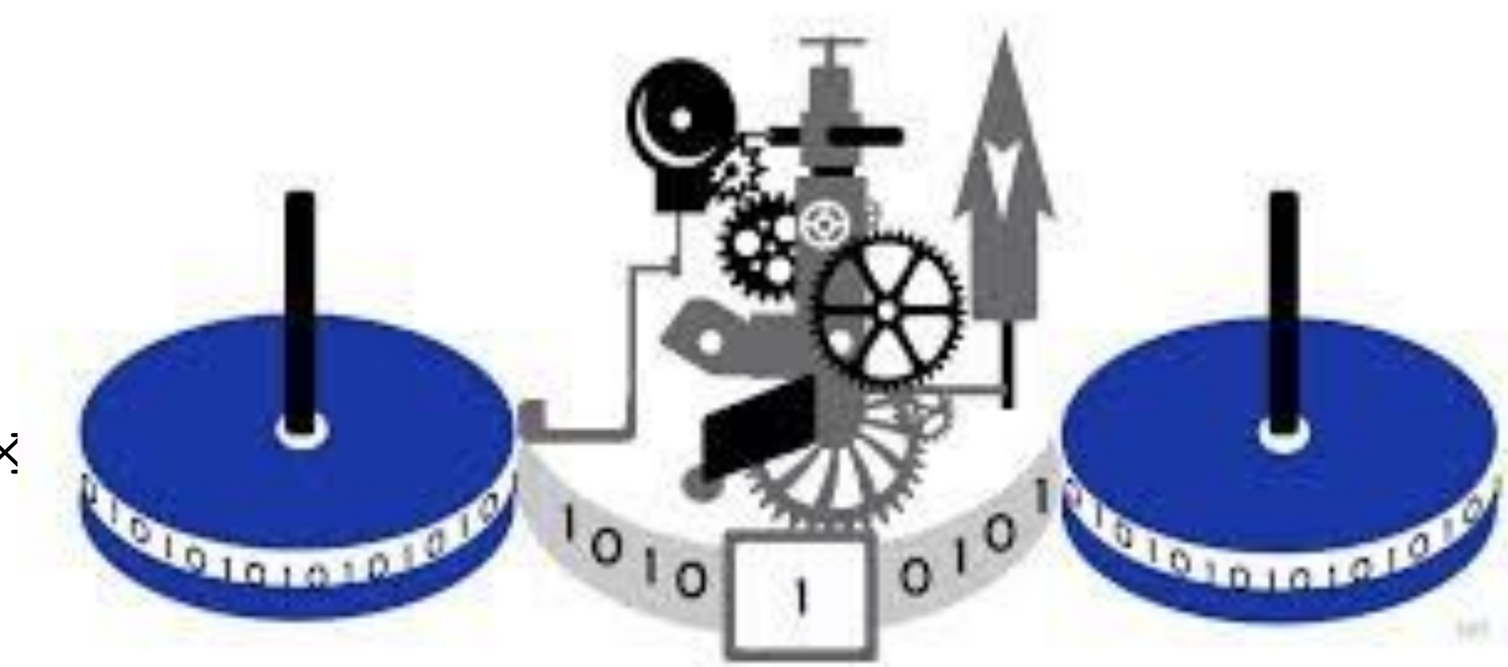
∀x (x))

Every young woman danced.

Natural Logic

Every person danced.

$\forall x$ $))$



$\forall x\ ($ $x))$

Every young woman danced.

✔

Natural Logic

Every person danced.

$$\forall x(\text{person}(x) \rightarrow \text{danced}(x))$$

$$\forall x(\text{woman}(x) \rightarrow \text{person}(x))$$

$$\forall x(\forall P((P(x) \wedge \text{young}(x)) \rightarrow P(x)))$$

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

Every young woman danced.

Natural Logic

Every person danced.

$$\forall x(\mathrm{person}(x) \rightarrow \mathrm{danced}(x))$$

$$\forall x(\mathrm{woman}(x) \rightarrow \mathrm{person}(x))$$

$$\forall x(\forall P((P(x) \wedge \mathrm{young}(x)) \rightarrow P(x)))$$

$$\forall x((\mathrm{woman}(x) \wedge \mathrm{young}(x)) \rightarrow \mathrm{danced}(x))$$

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.

person $\sqsupset$ young woman

every person $\sqsubset$ every young woman

Every **young woman** danced.

downward monotone

Natural Logic

Every **person** danced.

$\sqsubset$ ✔

Every **young woman** danced.

downward monotone

The NatLog System

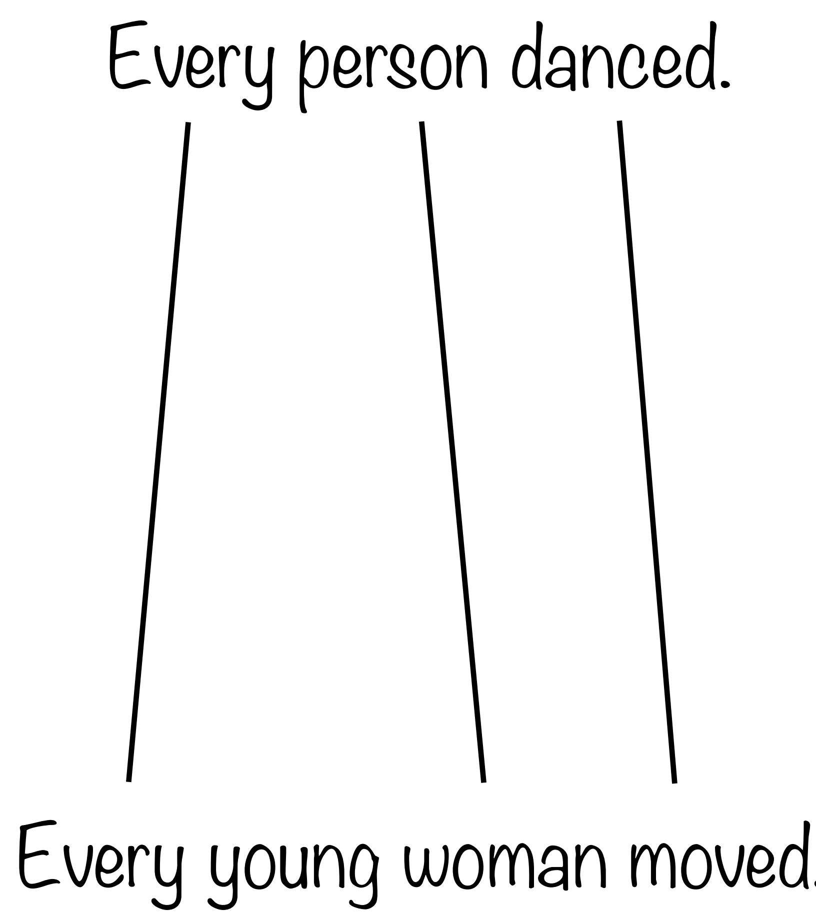
Every person danced.

Every young woman moved.

McCartney (2009)

The NatLog System

Alignment



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.

McCartney (2009)

The NatLog System

Entailment Classification

Every person danced.

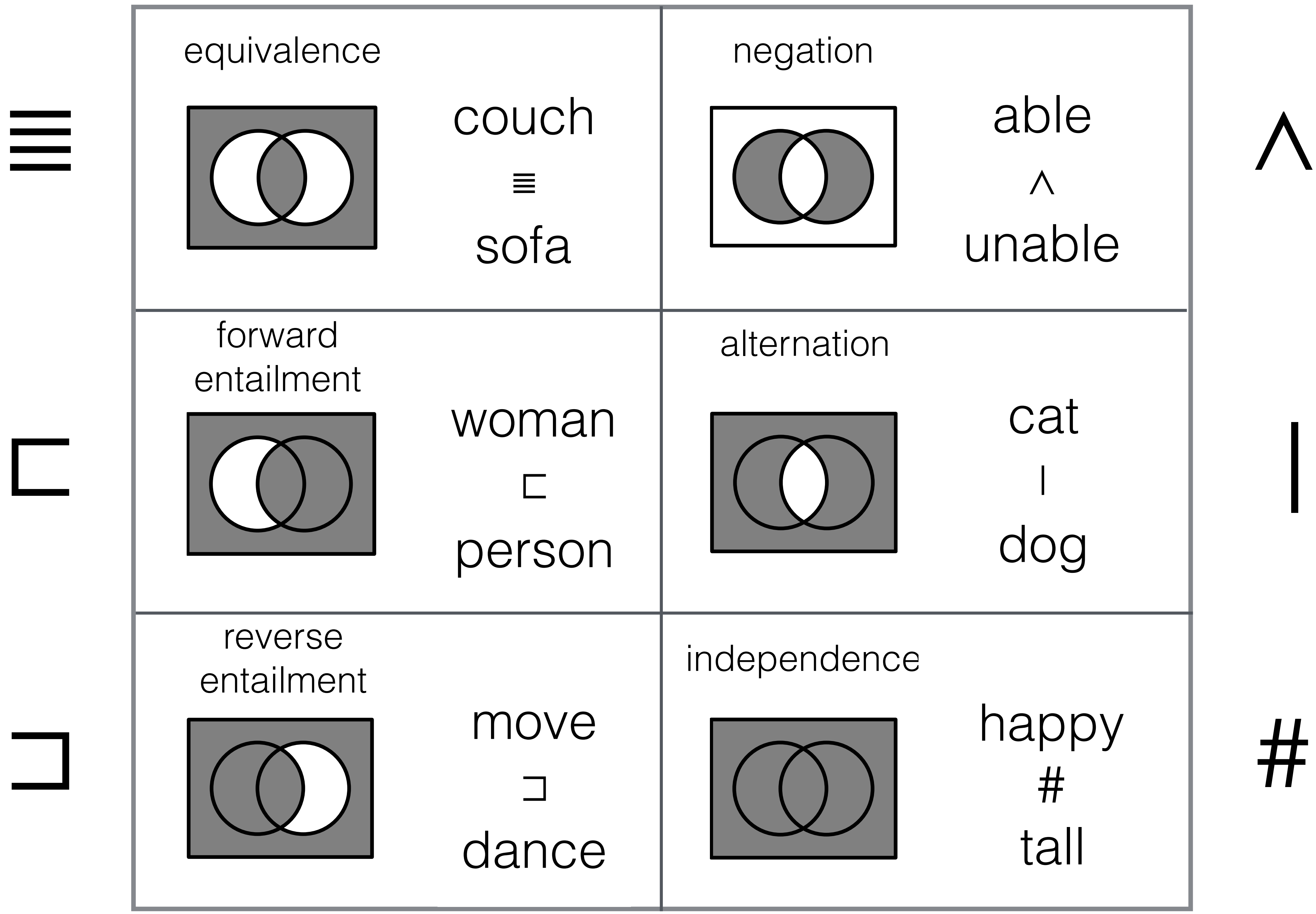
`SUB(person,woman)` Every woman danced.

`INS(young)` Every young woman danced.

`SUB(danced,moved)` Every young woman moved.

McCartney (2009)

The NatLog System



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.

McCartney (2009)

The NatLog System

Entailment Classification

Every person danced.

reverse entailment

`SUB(person,woman)` Every woman danced. $\sqsupset$

`INS(young)` Every young woman danced.

`SUB(danced,moved)` Every young woman moved.

McCartney (2009)

The NatLog System

Entailment Classification

Every person danced.

SUB(person,woman) Every woman danced. ⊏

reverse entailment

INS(young) Every young woman danced. ⊏

SUB(danced,moved) Every young woman moved.

McCartney (2009)

The NatLog System

Entailment Classification

Every person danced.

SUB(person,woman) Every woman danced. $\sqsubset$

INS(young) Every young woman danced. $\sqsubset$

forward entailment

SUB(danced,moved) Every young woman moved. $\sqsupset$

McCartney (2009)

The NatLog System

Entailment Classification

Every person danced.

`SUB(person,woman)` Every woman danced. $\sqsubset$

`INS(young)` Every young woman danced. $\sqsubset$

`SUB(danced,moved)` Every young woman moved. $\sqsupset$

McCartney (2009)

The NatLog System

Projectivity Marking

Every person danced.

McCartney (2009)

The NatLog System

Projectivity Marking

Every person danced.

upward
monotone

McCartney (2009)

The NatLog System

Projectivity Marking

Every person moved.

upward
monotone

$\sqsupset \longrightarrow \sqsupset$

danced $\sqsubset$ moved
every person danced $\sqsubset$ 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

$\sqsupset \longrightarrow \sqsubset$

person $\sqsupset$ young woman
every person $\sqsubset$ every young woman

McCartney (2009)

The NatLog System

Projectivity Marking

Every person danced.

`SUB(person,woman)`	Every woman danced.	$\sqsubset$
`INS(young)`	Every young woman danced.	$\sqsubset$
`SUB(danced,moved)`	Every young woman moved.	$\sqsupset$

McCartney (2009)

The NatLog System

Projectivity Marking

Every person danced.

`SUB(person,woman)` Every woman danced. $\sqsubset$

`INS(young)` Every young woman danced. $\sqsubset$

`SUB(danced,moved)` Every young woman moved. $\sqsupset$

McCartney (2009)

The NatLog System

Projectivity Marking

Every person danced.

`SUB(person,woman)`	Every woman danced.	$\sqsupset$
`INS(young)`	Every young woman danced.	$\sqsupset$
`SUB(danced,moved)`	Every young woman moved.	$\sqsupset$

McCartney (2009)

The NatLog System

Joining Entailment Relations

Every person danced.

`SUB(person,woman)`	Every woman danced.	$\sqsupset$
`INS(young)`	Every young woman danced.	$\sqsupset$
`SUB(danced,moved)`	Every young woman moved.	$\sqsupset$

McCartney (2009)

The NatLog System

Joining Entailment Relations

Every person danced.

`SUB(person,woman)`	Every woman danced.	$\sqsupset$
`INS(young)`	Every young woman danced.	$\sqsupset$
`SUB(danced,moved)`	Every young woman moved.	$\sqsupset$

$\sqsupset$

McCartney (2009)

The NatLog System

Joining Entailment Relations

Every person danced.

`SUB(person,woman)`	Every woman danced.	$\sqsupset$
`INS(young)`	Every young woman danced.	$\sqsupset$
`SUB(danced,moved)`	Every young woman moved.	$\sqsupset$

$\sqsupset$ ✔

McCartney (2009)

The NatLog System

Joining Entailment Relations

No person danced.

`SUB(person,woman)`	No woman danced.	$\sqsupset$
`INS(young)`	No young woman danced.	$\sqsupset$
`SUB(danced,moved)`	No young woman moved.	$\sqsubset$

$\#$

McCartney (2009)

Sentence-Level Semantics

Natural Language Inference	Logical Forms
ungrounded—relate text to other text	grounded—relate text to tables in a database, or actions on a robot

Denotational semantics

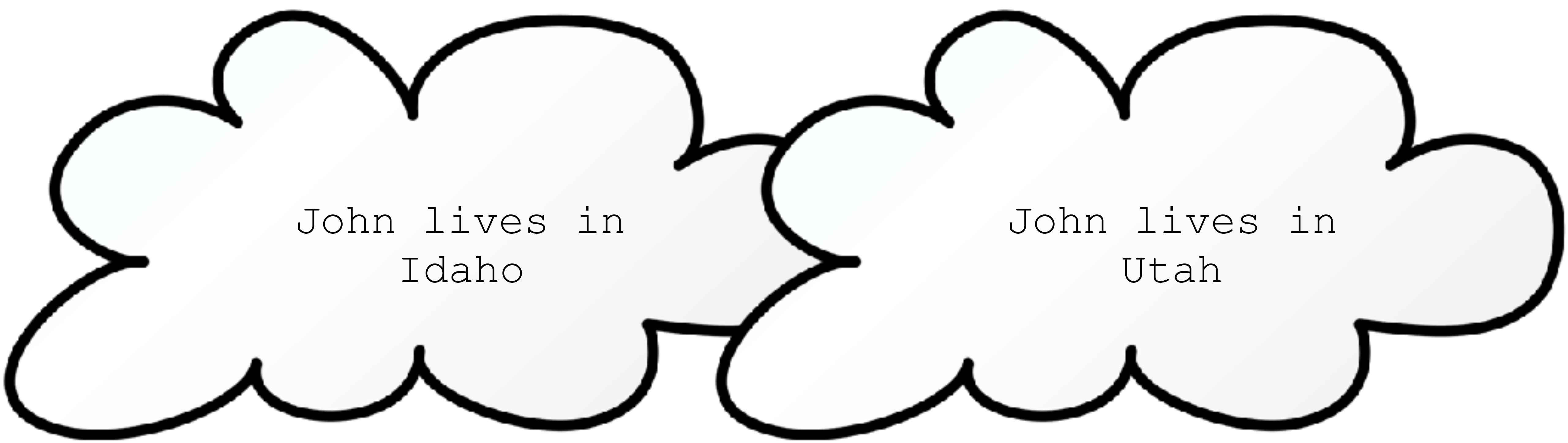
Idaho $\approx$ Utah

Denotational semantics

Idaho $\neq$ Utah

Denotational semantics

Idaho ≠ Utah



Denotation Graph

Young et al. (2014)

Denotation Graph

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)

Denotation Graph

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

Young et al. (2014)

Denotation Graph

A man in a tie

Young et al. (2014)

Denotation Graph

A man

Young et al. (2014)

Denotation Graph

A man

$\sqsupset$

A businessman in a yellow tie

Young et al. (2014)

Denotation Graph

play football

Distributional Similarity	Denotational Similarity
play game	tackle person
play rugby	hold football
play soccer	run down field
play on field	wear white jersey
play ball	avoid

Denotation Graph

yellow tie

red tie

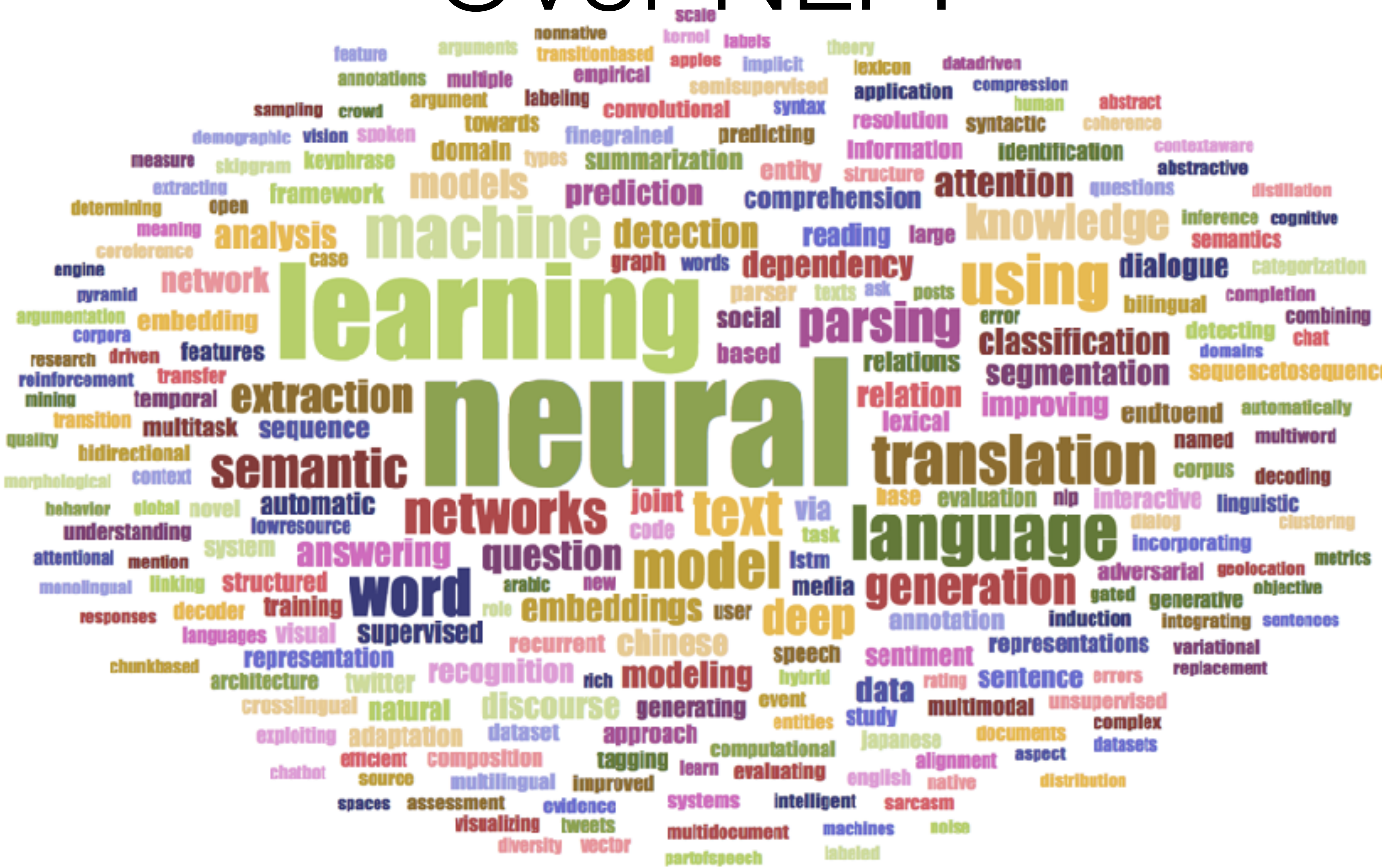
yellow dress

???

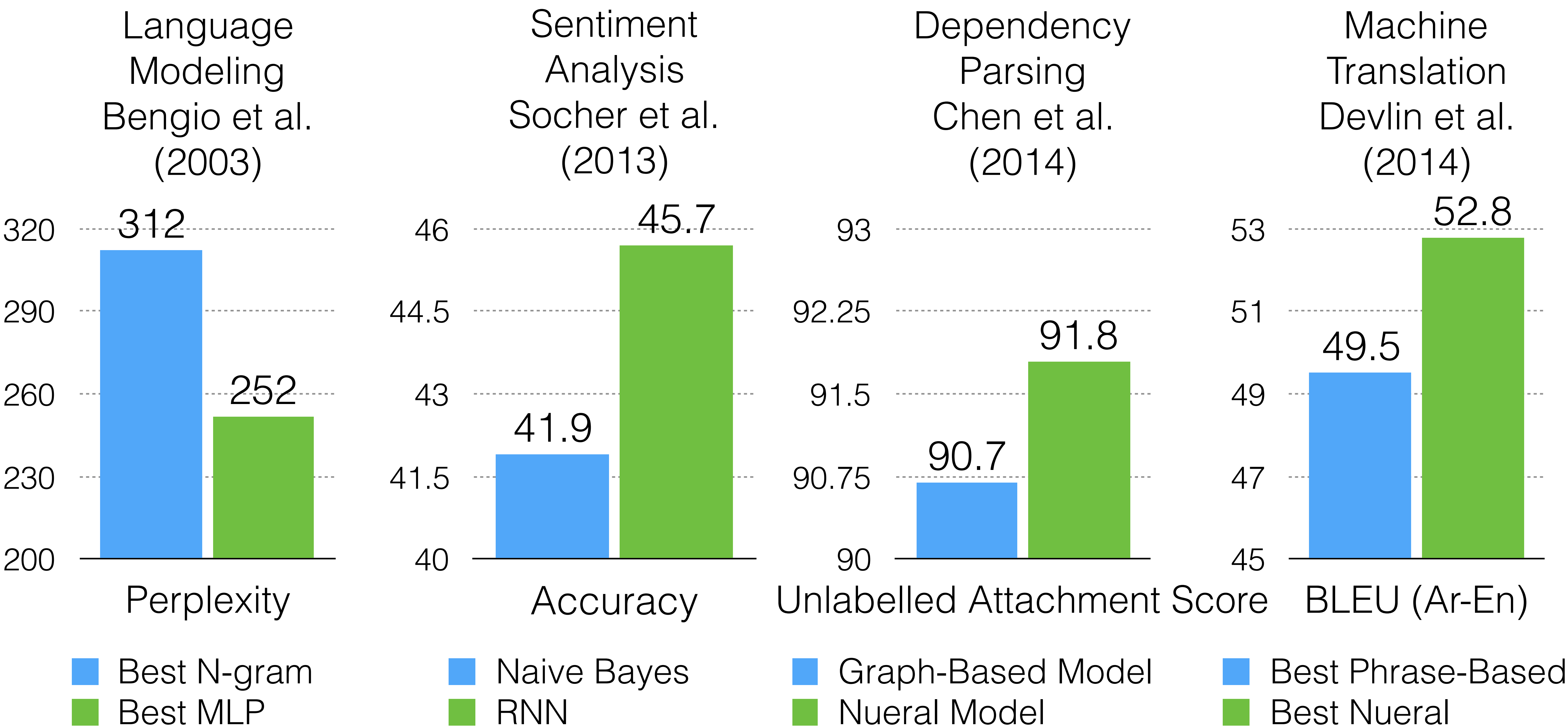
Pause: Questions!

So….uh…does this really matter?

Deep Learning is Taking Over NLP!



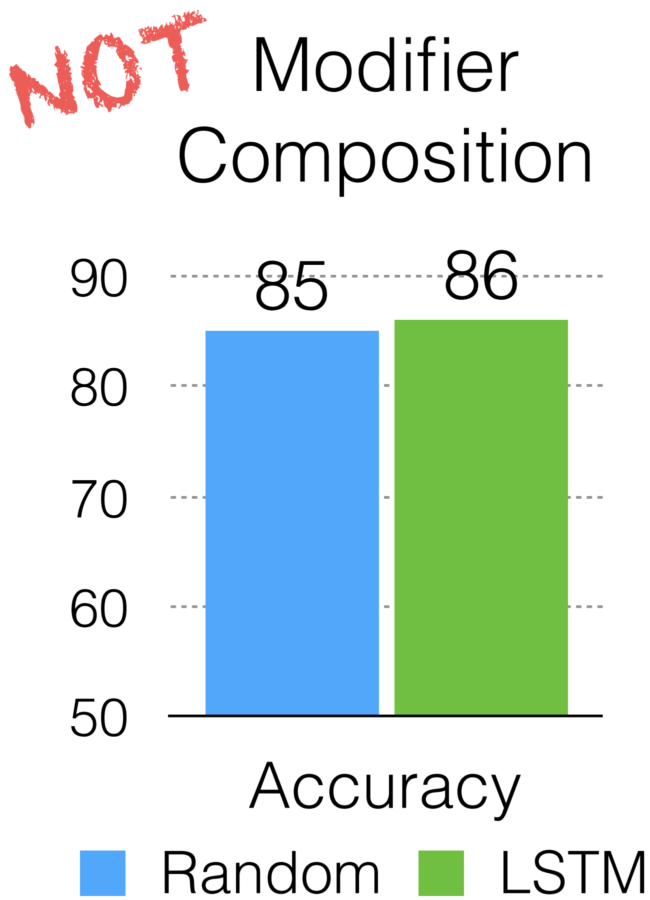
Titles of ACL Papers, 2017



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.

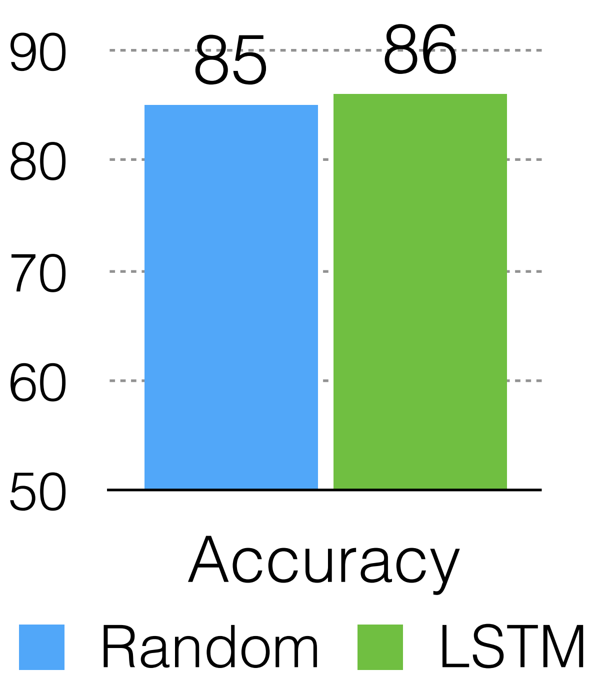
⇓

The **deadly attack** killed at least 12 civilians.

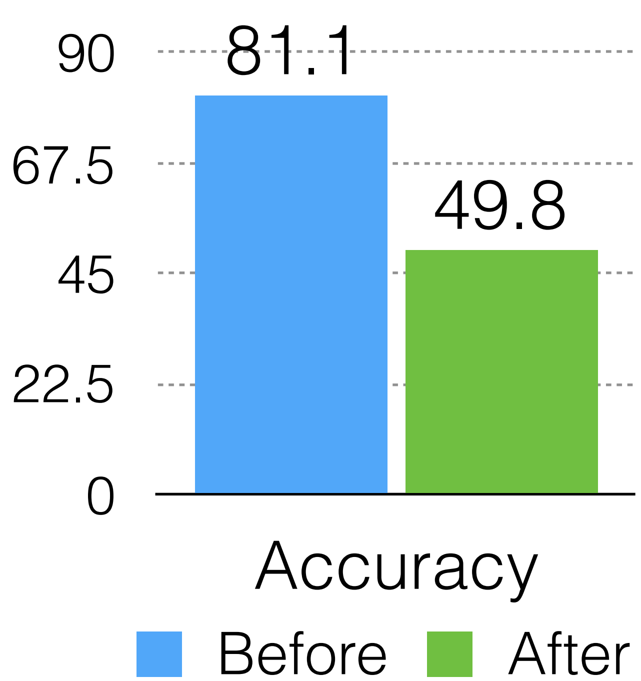
Pavlick and Callison-Burch (2016)

But what, exactly, are our systems learning?

NOT Modifier Composition



NOT Invariance to Distractors



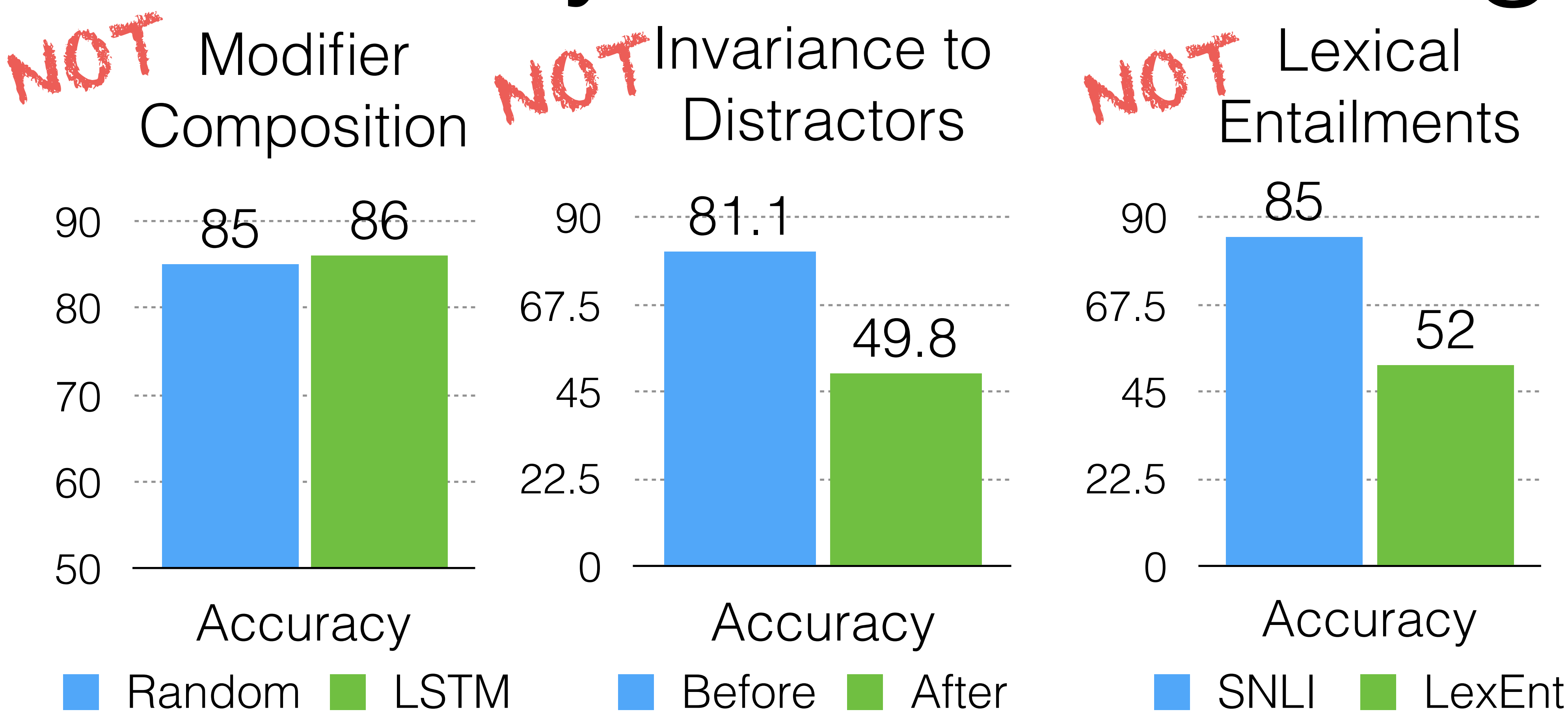
The past record was held by John Elway…
Jeff Dean had jersey number 37…

$\Downarrow$

John Elway

Jia and Liang (2017)

But what, exactly, are our systems learning?



The man is holding a **saxophone.**

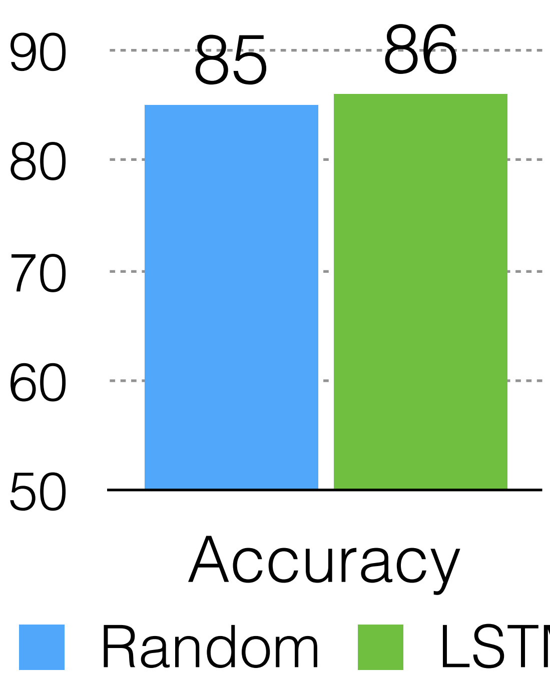
$\Downarrow$

The man is holding an **electric guitar**.

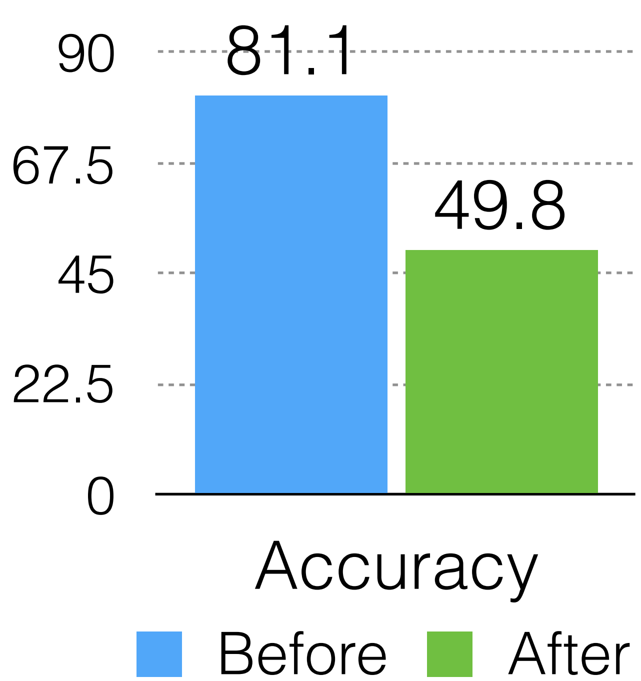
Glockner et al (ACL 2018)

But what, exactly, are our systems learning?

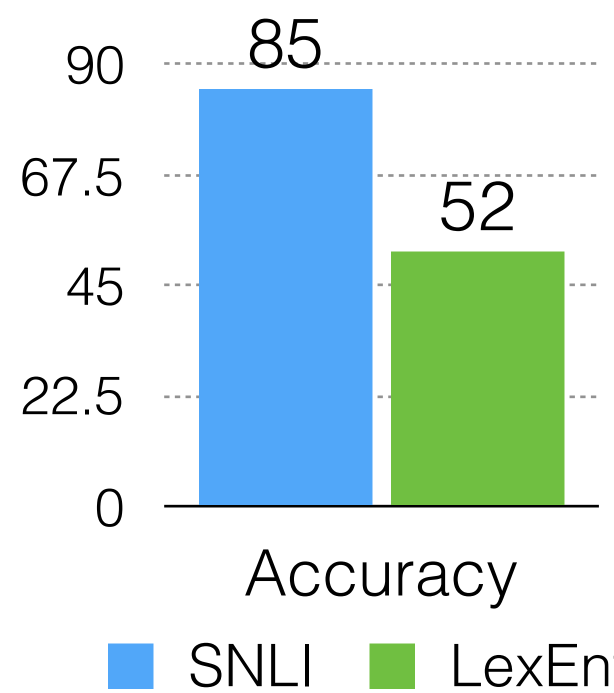
NOT Modifier Composition



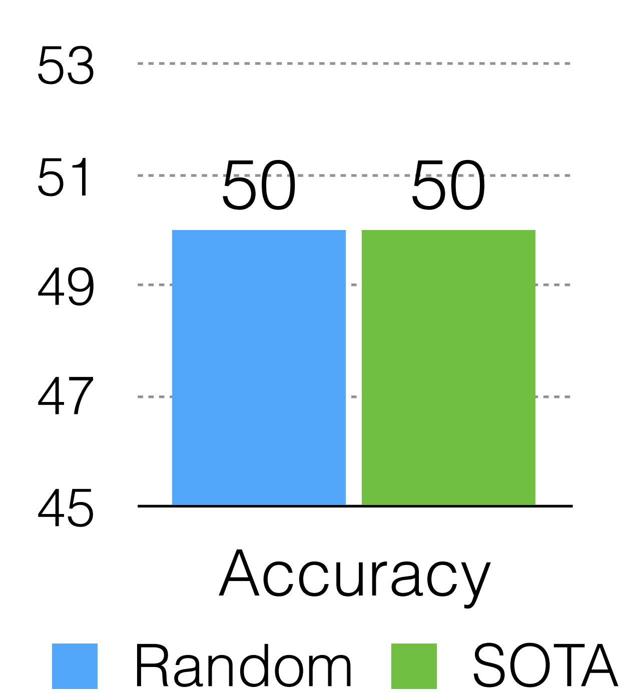
NOT Invariance to Distractors



NOT Lexical Entailments



NOT Negations



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

$\Downarrow$

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

Dasgupta et al (2018)

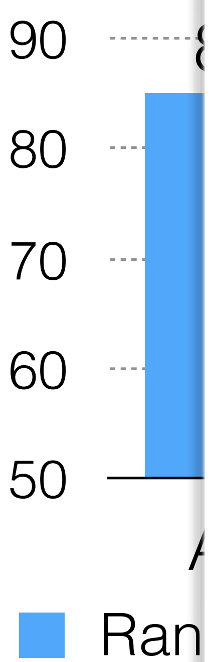
But what, exactly, are our systems learning?

NOT Modifier Composition

NOT Invariance to Distractors

NOT Lexical Entailments

NOT Negations



Probing Tasks Galore!

White et al (IJCNLP 2017)

Mahler et al (2017)

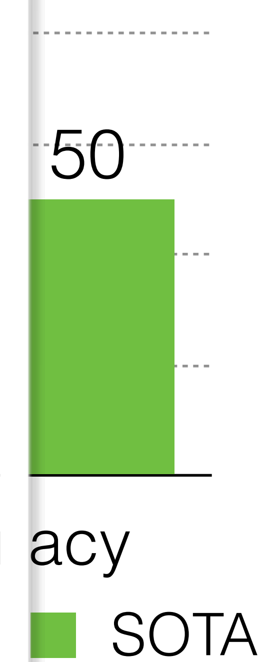
Ettinger et al. (EMNLP 2017)

Adi et al. (ICLR 2017)

Poliak et al. (*SEM 2018)

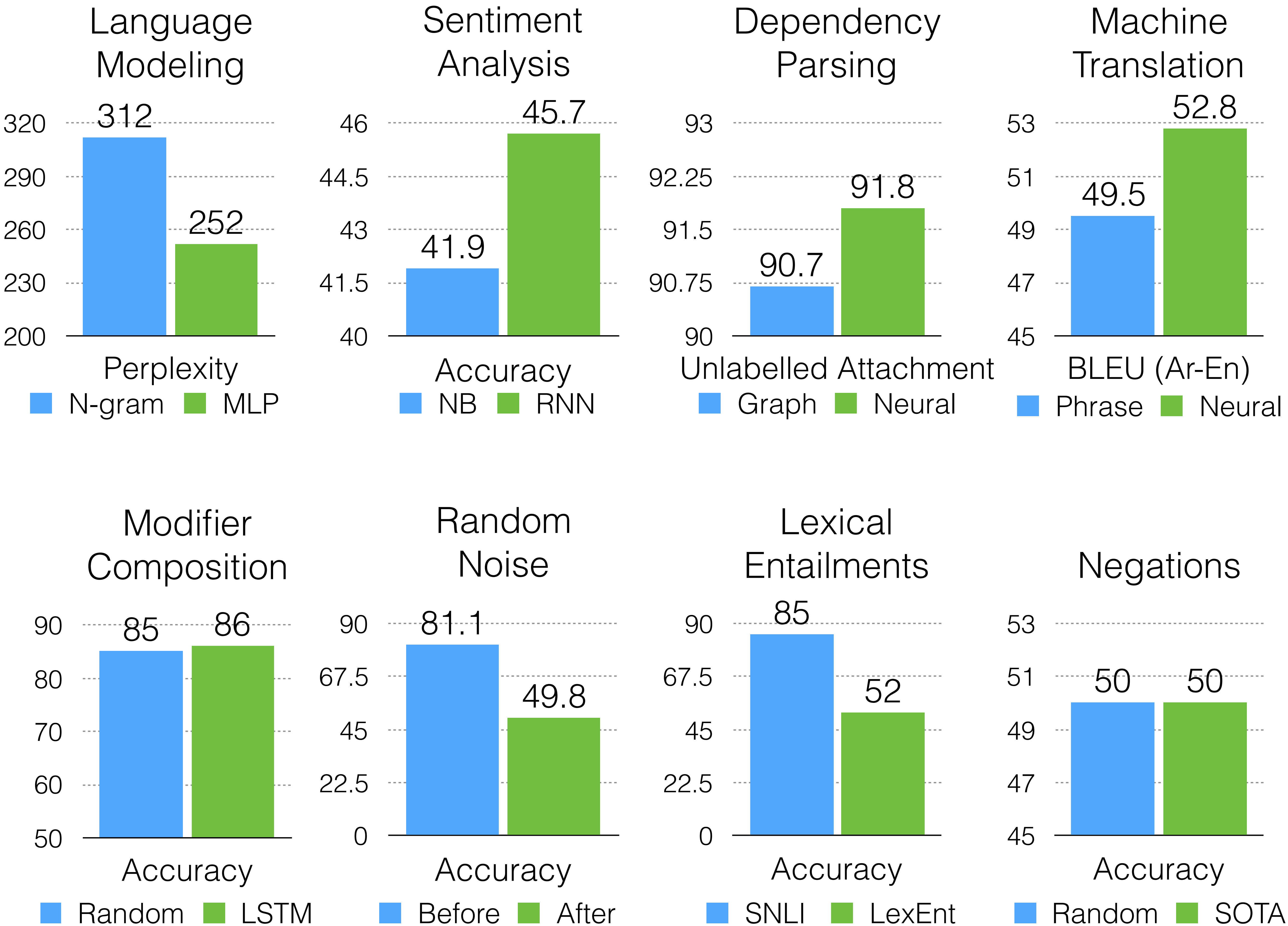
Conneau et al. (ACL 2018)

Zhu et al. (ACL 2018)

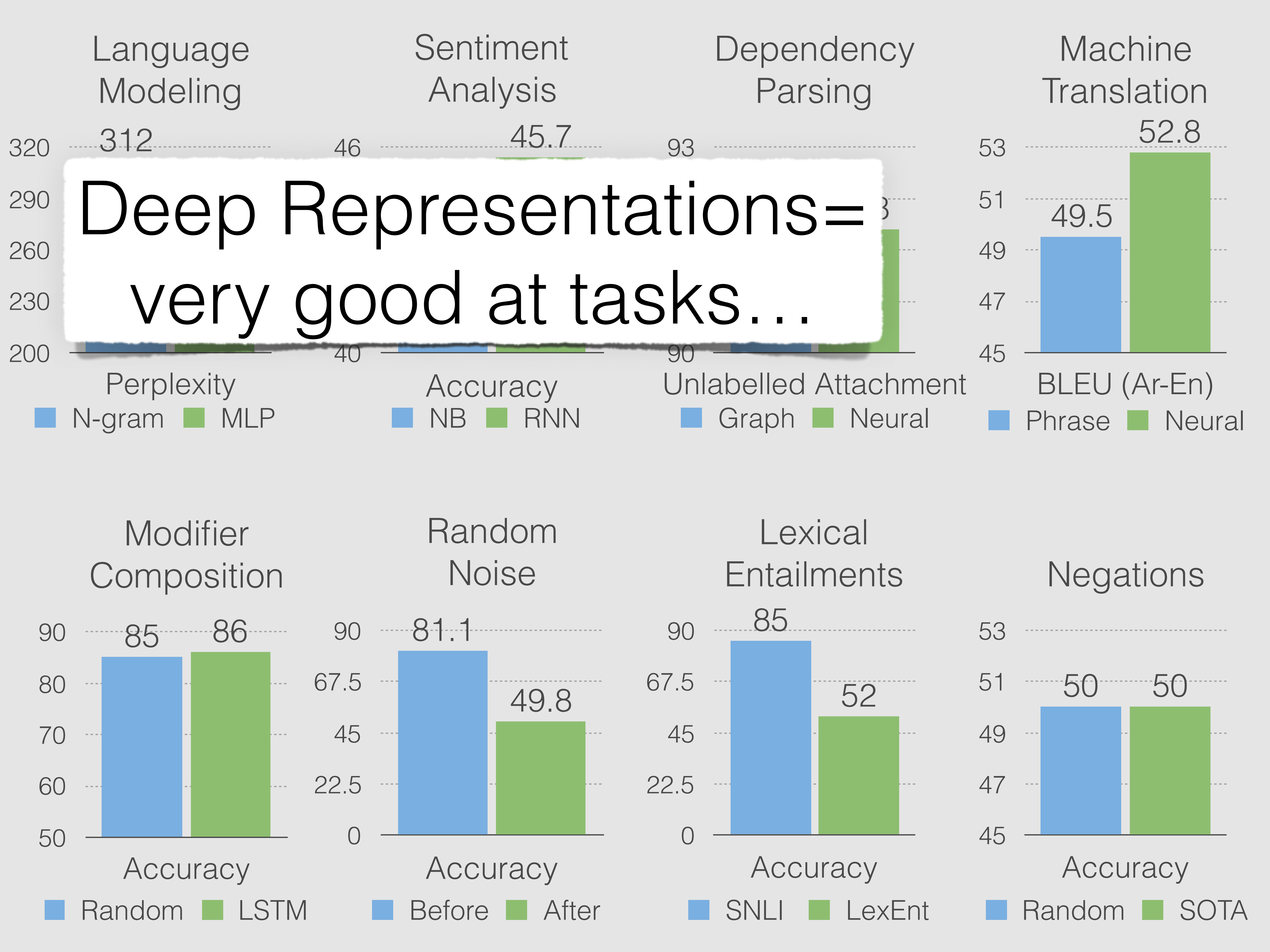


The woman is **less** cheerful than the man.

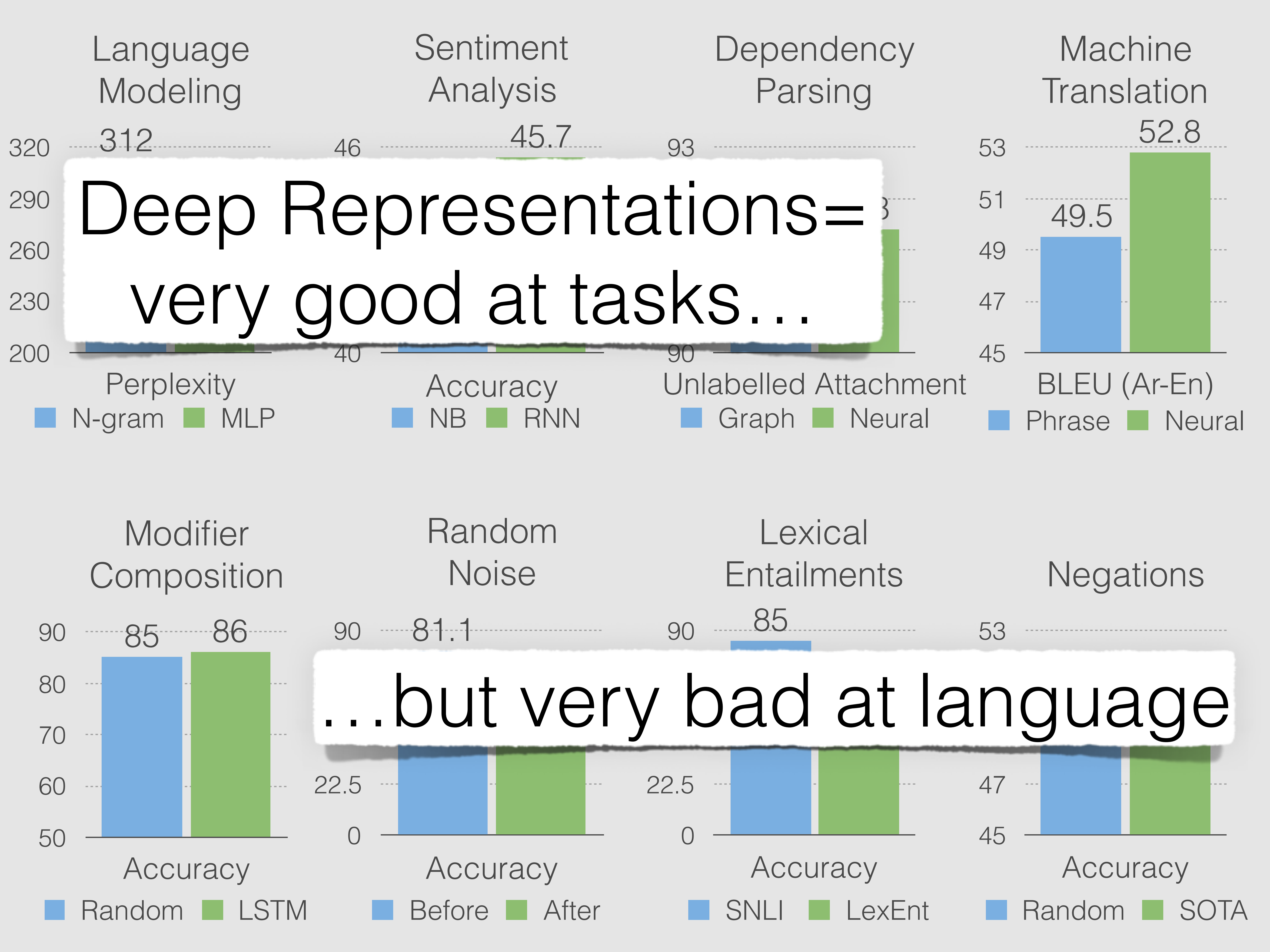
Dasgupta et al (2018)



Language Modeling
320
290
260
230
200
312
252
Perplexity
N-gram
MLP
Sentiment Analysis
46
44.5
43
41.5
40
41.9
45.7
Accuracy
NB
RNN
Dependency Parsing
93
92.25
91.5
90.75
90
90.7
91.8
Unlabelled Attachment
Graph
Neural
Machine Translation
53
51
49
47
45
49.5
52.8
BLEU (Ar-En)
Phrase
Neural
Modifier Composition
90
80
70
60
50
85
86
Accuracy
Random
LSTM
Random Noise
90
67.5
45
22.5
0
81.1
49.8
Accuracy
Before
After
Lexical Entailments
90
67.5
45
22.5
0
85
52
Accuracy
SNLI
LexEnt
Negations
53
51
49
47
45
50
50
Accuracy
Random
SOTA



Deep Representations=
very good at tasks…
Language Modeling
312
Perplexity
N-gram
MLP
Sentiment Analysis
45.7
Accuracy
NB
RNN
Dependency Parsing
Unlabelled Attachment
Graph
Neural
Machine Translation
52.8
49.5
BLEU (Ar-En)
Phrase
Neural
Modifier Composition
85
86
Accuracy
Random
LSTM
Random Noise
81.1
49.8
Accuracy
Before
After
Lexical Entailments
85
52
Accuracy
SNLI
LexEnt
Negations
50
50
Accuracy
Random
SOTA


Language Modeling
312
Perplexity
N-gram
MLP
Sentiment Analysis
45.7
Accuracy
NB
RNN
Dependency Parsing
Unlabelled Attachment
Graph
Neural
Machine Translation
52.8
49.5
BLEU (Ar-En)
Phrase
Neural
Deep Representations=
very good at tasks…
Modifier Composition
85
86
Accuracy
Random
LSTM
Random Noise
81.1
Accuracy
Before
After
Lexical Entailments
85
Accuracy
SNLI
LexEnt
Negations
Accuracy
Random
SOTA
…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

↓

Math

Language

$$\downarrow$$

$$\forall x \forall y (P(f(x)) \to \neg(Q(f(y), x))$$

Language

$$\downarrow$$

$$\lambda x.f(y, g(x)) \wedge h(y)$$

Language

↓

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$$

Language

↓

010110110101001010101
010100011010111011001
010101101010010101100
000110000001011111010

Language

↓

010110110101001010101

010100011010111011001

010101101010010101100

000110000001011111010

Questions!