Should we care about linguistics?

Ellie Pavlick
Department of Computer Science
Brown University

Deep Learning is Taking Over NLP!

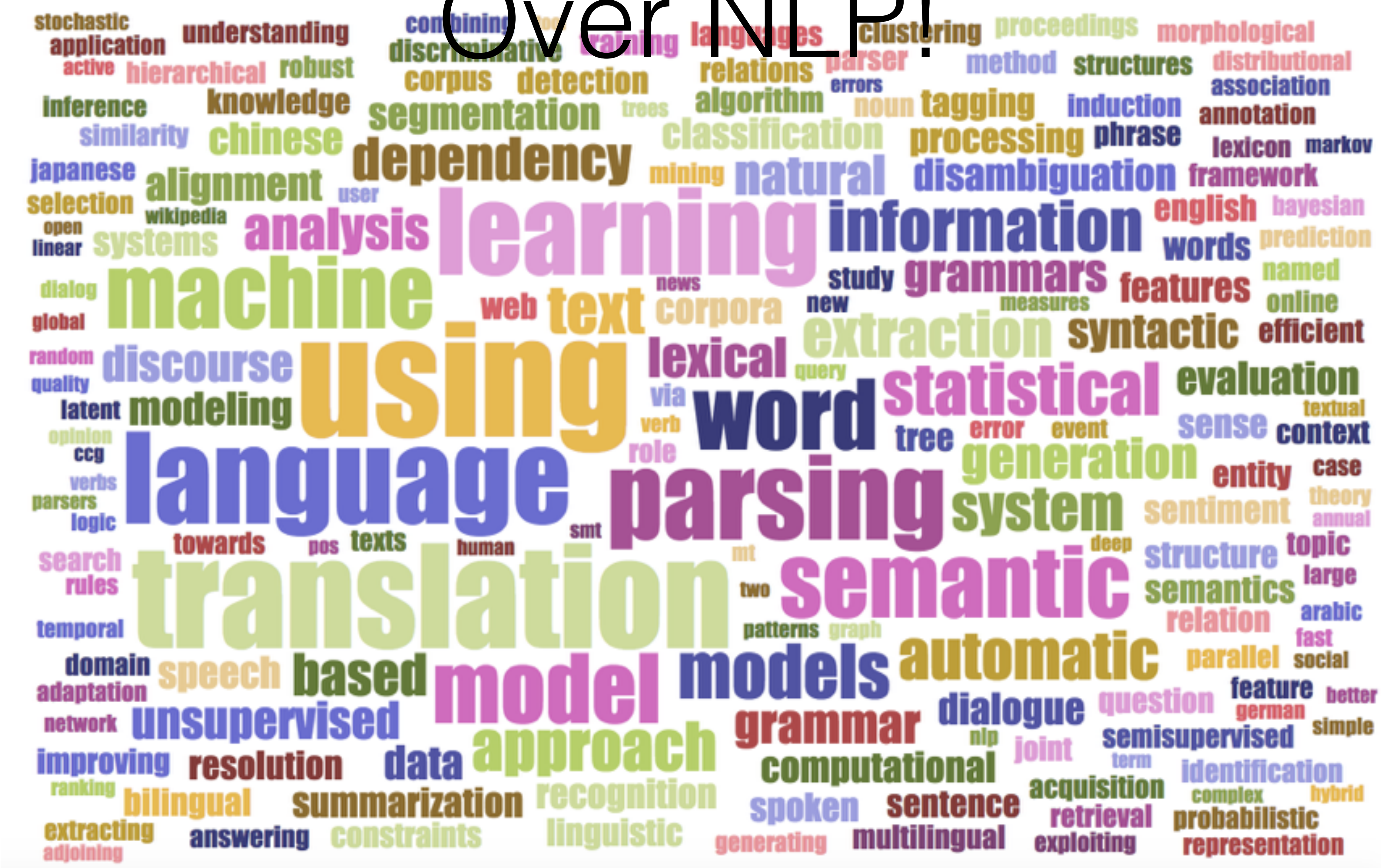
Deep Learning is Taking Over NLP!

Titles of ACL Papers, 1979

Deep Learning is Taking Over NLP!

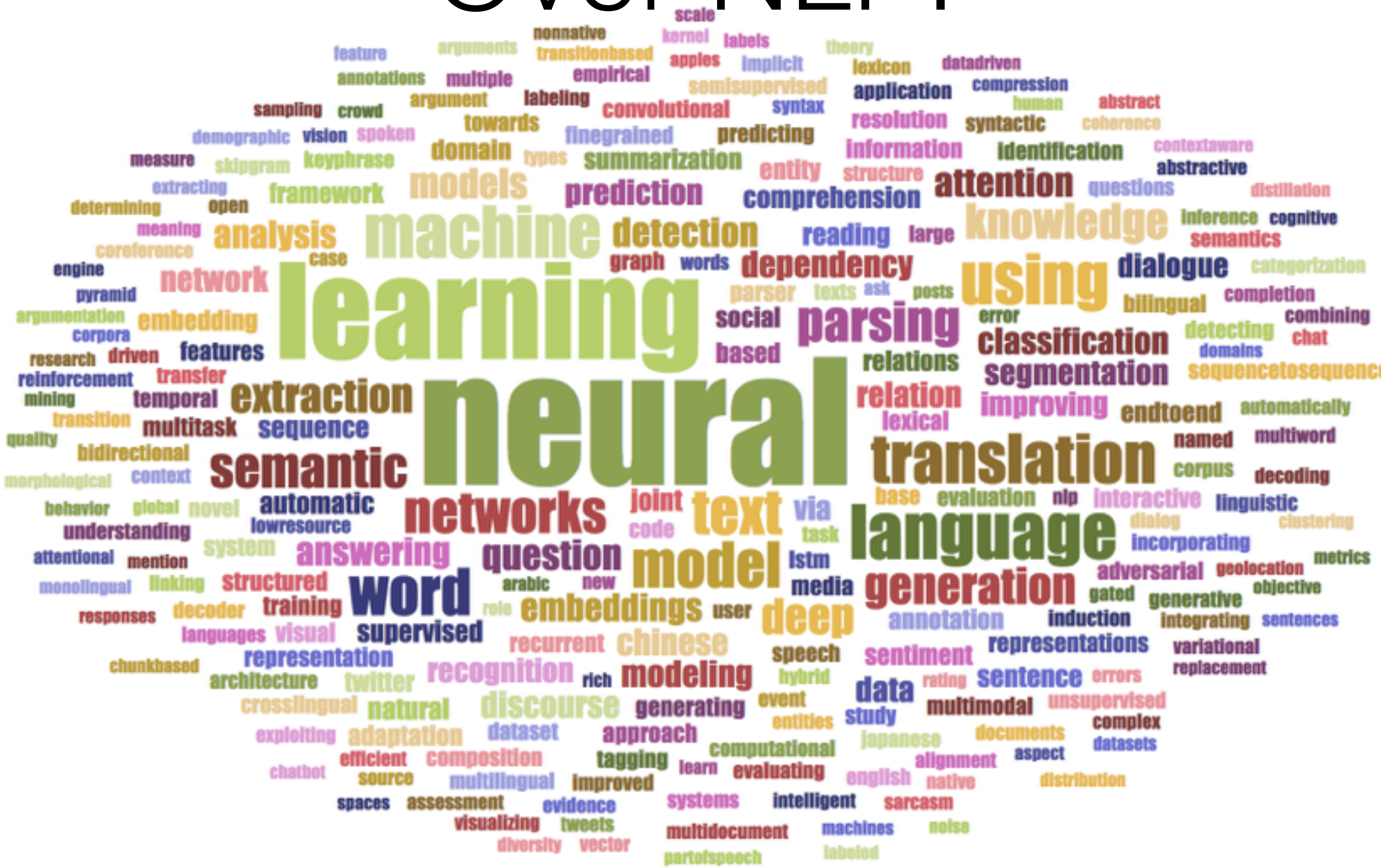
Titles of ACL Papers, 1987

Deep Learning is Taking Over NLP!



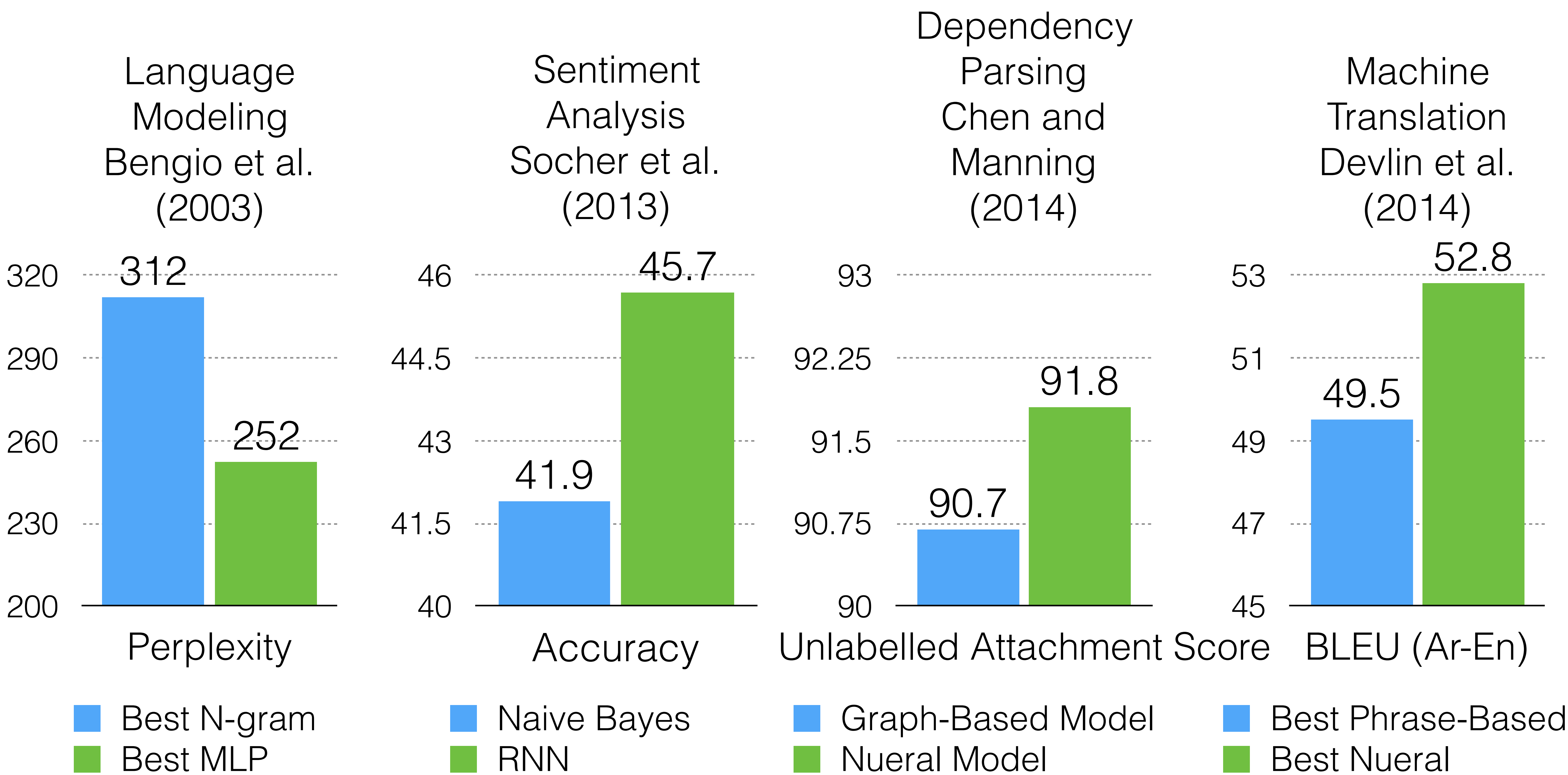
Titles of ACL Papers, everything pre-2015

Deep Learning is Taking Over NLP!



Titles of ACL Papers, 2017

SOTA on Classic NLP Tasks



New Enthusiasm for End-to-End NLU Tasks

New Enthusiasm for End-to-End NLU Tasks

Recognizing Textual Entailment (RTE)

New Enthusiasm for End-to-End NLU Tasks

Recognizing Textual Entailment (RTE)

A man inspects the uniform of a
figure in some East Asian country.
+
The man is sleeping.

New Enthusiasm for End-to-End NLU Tasks

Recognizing Textual Entailment (RTE)

premise

A man inspects the uniform of a figure in some East Asian country.

+

The man is sleeping.

New Enthusiasm for End-to-End NLU Tasks

Recognizing Textual Entailment (RTE)

premise

A man inspects the uniform of a figure in some East Asian country.

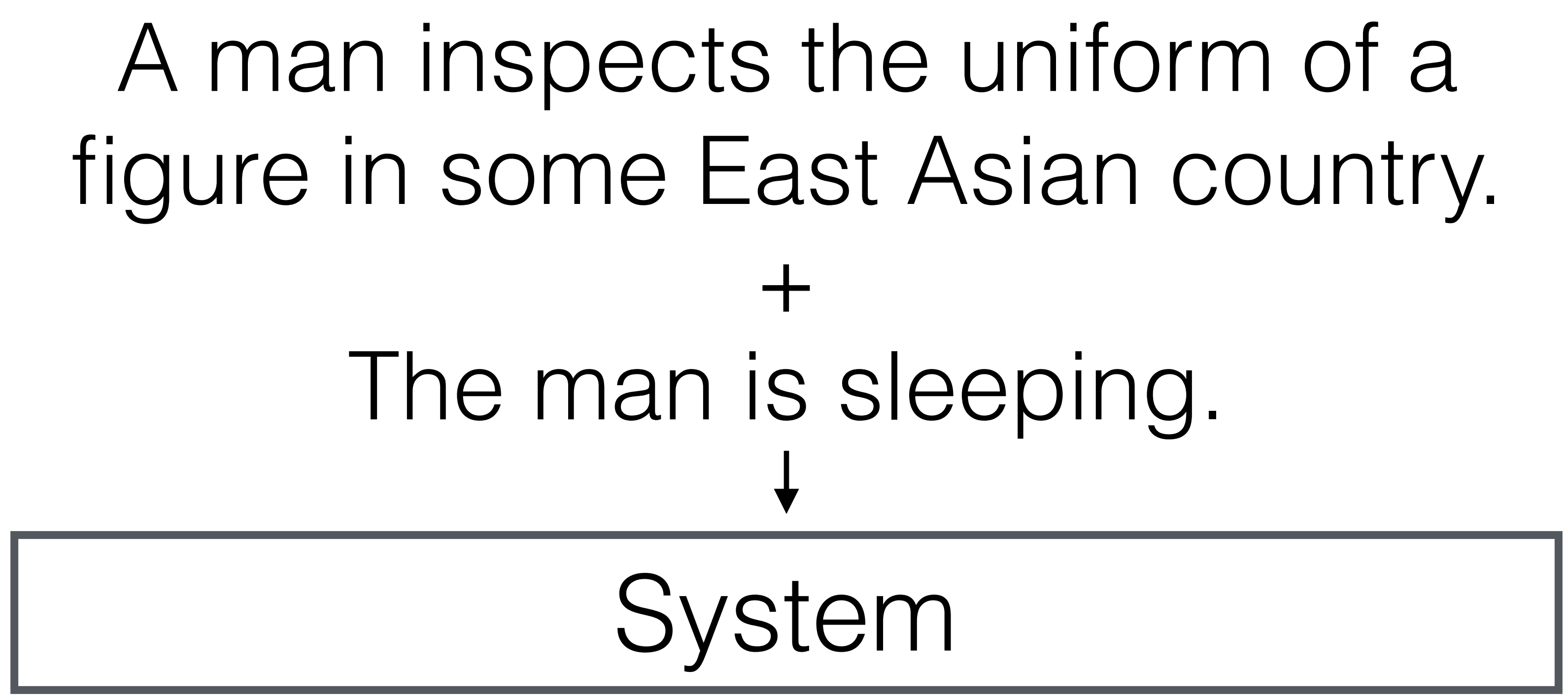
+

The man is sleeping.

hypothesis

New Enthusiasm for End-to-End NLU Tasks

Recognizing Textual Entailment (RTE)



New Enthusiasm for End-to-End NLU Tasks

Recognizing Textual Entailment (RTE)

A man inspects the uniform of a figure in some East Asian country.

+

The man is sleeping.

↓

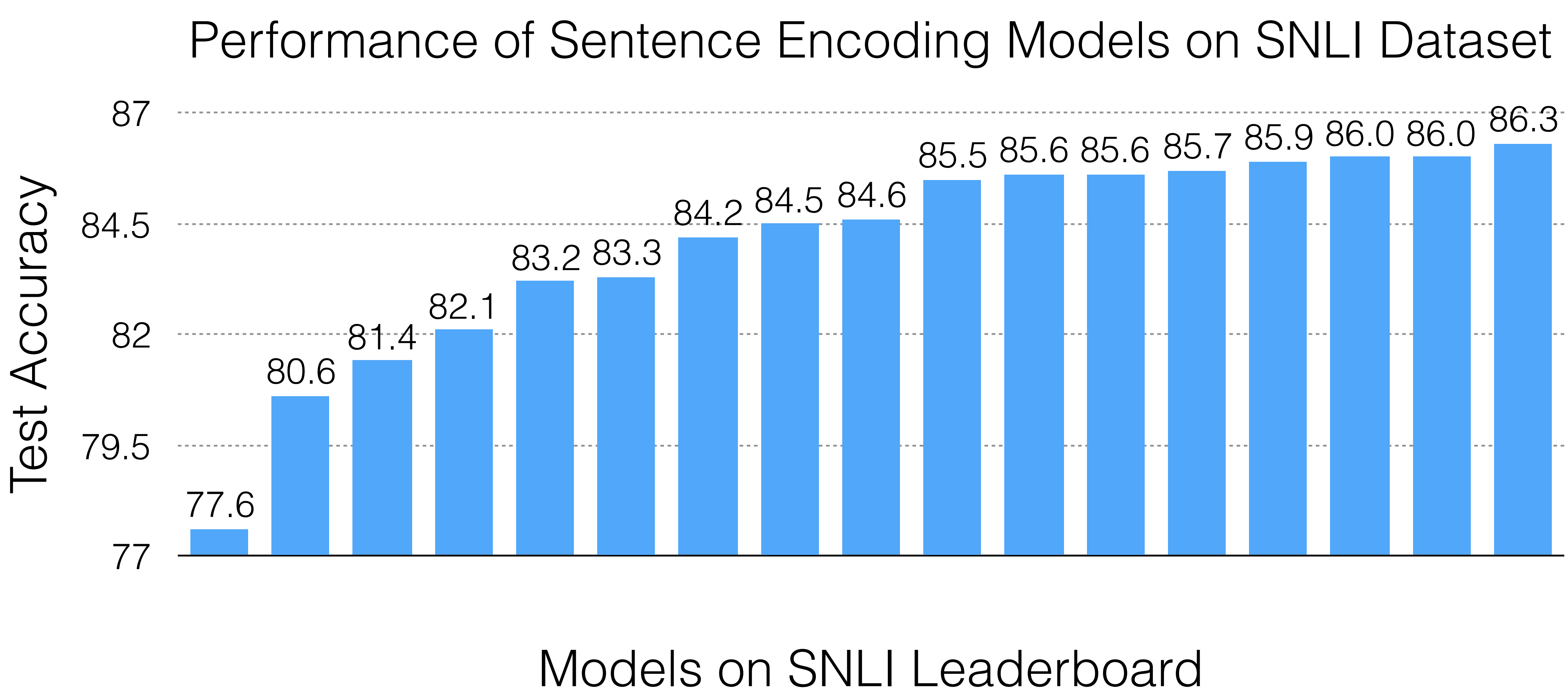
System

↓

False

New Enthusiasm for End-to-End NLU Tasks

Recognizing Textual Entailment (RTE)



A large annotated corpus for learning natural language inference.
Bowman et al. (2015)

New Enthusiasm for End-to-End NLU Tasks

Reading Comprehension

What is Southern California often abbreviated as?

SQuAD: 100,000+ Questions for Machine Comprehension of Text.
Rajpurkar et al. (2016)

New Enthusiasm for End-to-End NLU Tasks

Reading Comprehension

What is Southern California often abbreviated as?

+

Southern California, often abbreviated SoCal, is a geographic and cultural region that generally comprises California's southernmost 10 counties. The region is traditionally described as "eight counties", based on demographics and economic ties: Imperial, Los Angeles, Orange, Riverside, San Bernardino, San Diego, Santa Barbara, and Ventura. The more extensive 10-county definition, including Kern and San Luis Obispo counties, is also used based on historical political divisions. Southern California is a major economic center for the state of California and the United States.

SQuAD: 100,000+ Questions for Machine Comprehension of Text.
Rajpurkar et al. (2016)

New Enthusiasm for End-to-End NLU Tasks

Reading Comprehension

What is Southern California often abbreviated as?

+

Southern California, often abbreviated SoCal, is a geographic and cultural region that generally comprises California's southernmost 10 counties. The region is traditionally described as "eight counties", based on demographics and economic ties: Imperial, Los Angeles, Orange, Riverside, San Bernardino, San Diego, Santa Barbara, and Ventura. The more extensive 10-county definition, including Kern and San Luis Obispo counties, is also used based on historical political divisions. Southern California is a major economic center for the state of California and the United States.

↓

System

SQuAD: 100,000+ Questions for Machine Comprehension of Text.
Rajpurkar et al. (2016)

New Enthusiasm for End-to-End NLU Tasks

Reading Comprehension

What is Southern California often abbreviated as?

+

Southern California, often abbreviated SoCal, is a geographic and cultural region that generally comprises California's southernmost 10 counties. The region is traditionally described as "eight counties", based on demographics and economic ties: Imperial, Los Angeles, Orange, Riverside, San Bernardino, San Diego, Santa Barbara, and Ventura. The more extensive 10-county definition, including Kern and San Luis Obispo counties, is also used based on historical political divisions. Southern California is a major economic center for the state of California and the United States.

↓

System

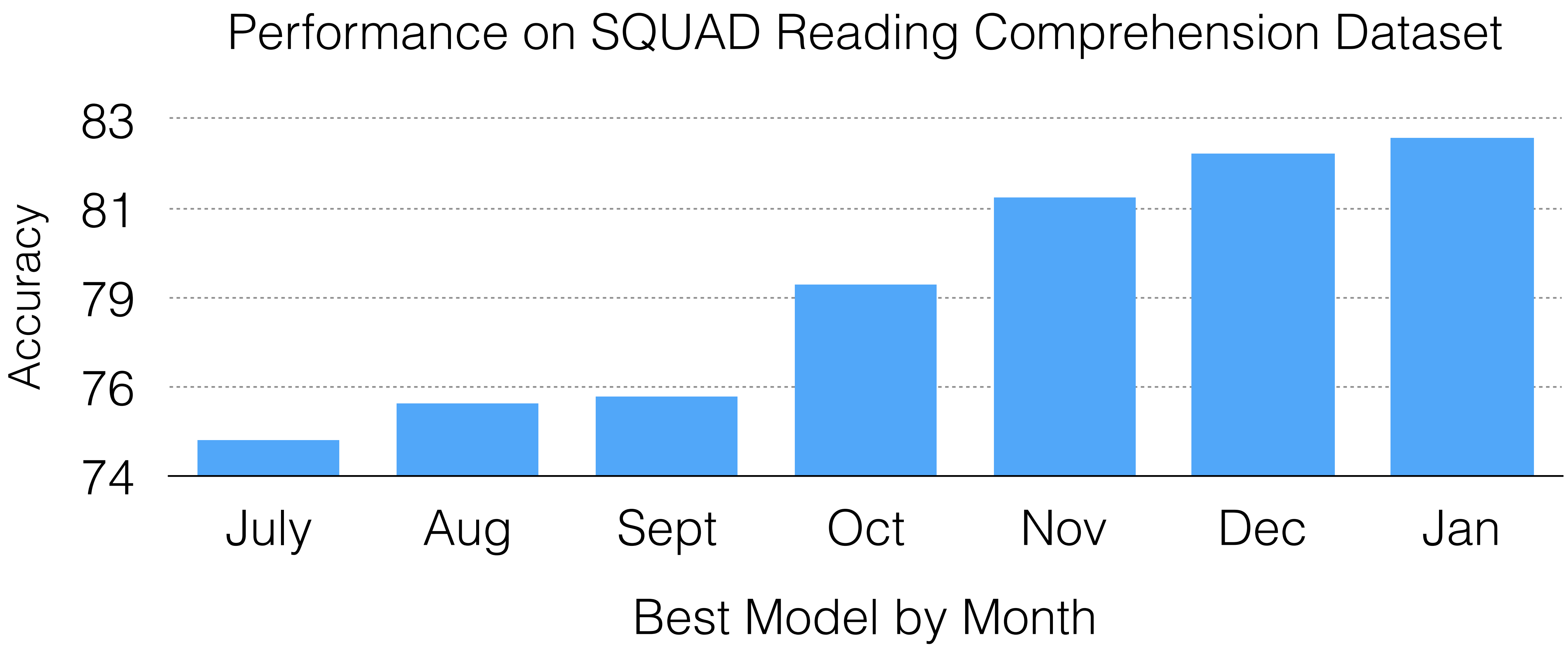
↓

SoCal

SQ ions for Machine Comprehension of Text.
Rajpurkar et al. (2016)

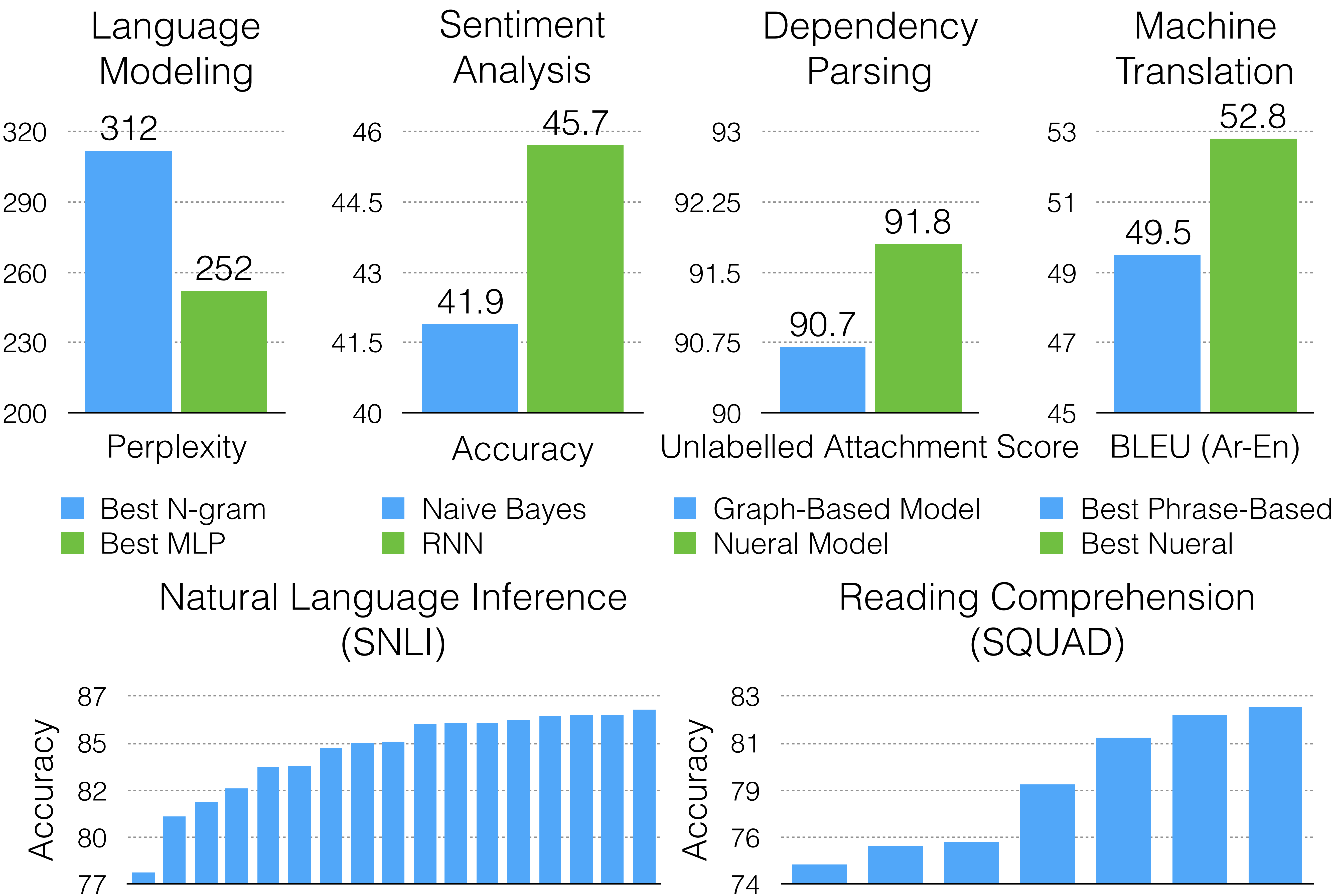
New Enthusiasm for End-to-End NLU Tasks

Reading Comprehension

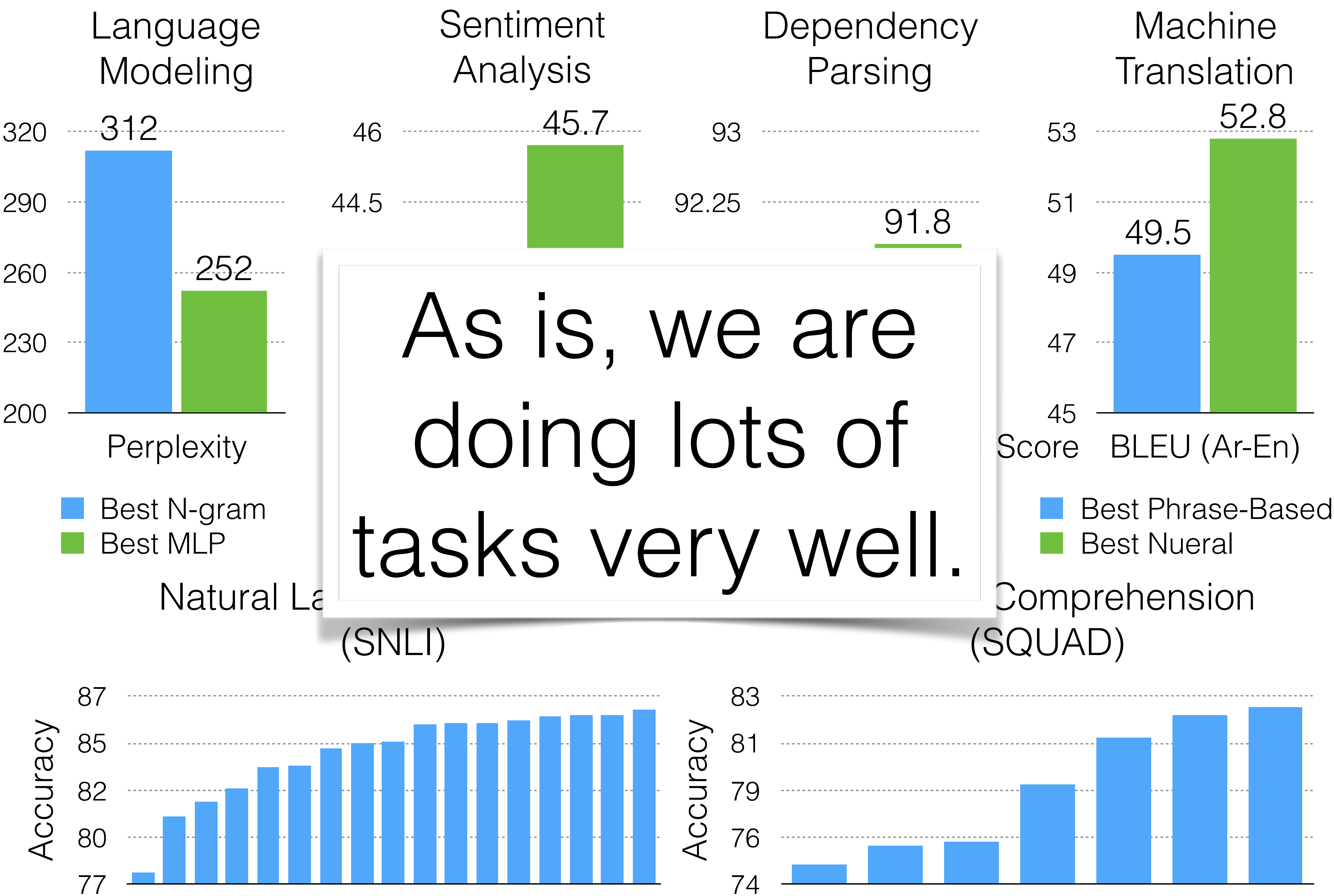


SQuAD: 100,000+ Questions for Machine Comprehension of Text.
Rajpurkar et al. (2016)

Should we care about linguistics?



Language Modeling
320
290
260
230
200
312
252
Perplexity
Best N-gram
Best MLP
Sentiment Analysis
46
44.5
43
41.5
40
41.9
45.7
Accuracy
Naive Bayes
RNN
Dependency Parsing
93
92.25
91.5
90.75
90
90.7
91.8
Unlabelled Attachment Score
Graph-Based Model
Nueral Model
Machine Translation
53
51
49
47
45
49.5
52.8
BLEU (Ar-En)
Best Phrase-Based
Best Nueral
Natural Language Inference (SNLI)
Accuracy
87
85
82
80
77
Reading Comprehension (SQUAD)
Accuracy
83
81
79
76
74



Language Modeling
320
290
260
230
200
312
252
Perplexity
Best N-gram
Best MLP
Sentiment Analysis
46
44.5
45.7
Dependency Parsing
93
92.25
91.8
Score
Machine Translation
53
51
49
47
45
52.8
49.5
BLEU (Ar-En)
Best Phrase-Based
Best Nueral
Natural La
(SNLI)
Accuracy
87
85
82
80
77
Comprehension
(SQUAD)
Accuracy
83
81
79
76
74
As is, we are doing lots of tasks very well.

What are our systems learning?

A man inspects the uniform of a figure in some East Asian country.

↓

The man is sleeping.

SNLI dataset (Bowman, 2015)

What are our systems learning?

A man **inspects** the uniform of a figure in some East Asian country.

The man is **sleeping**.

SNLI dataset (Bowman, 2015)

What are our systems learning?

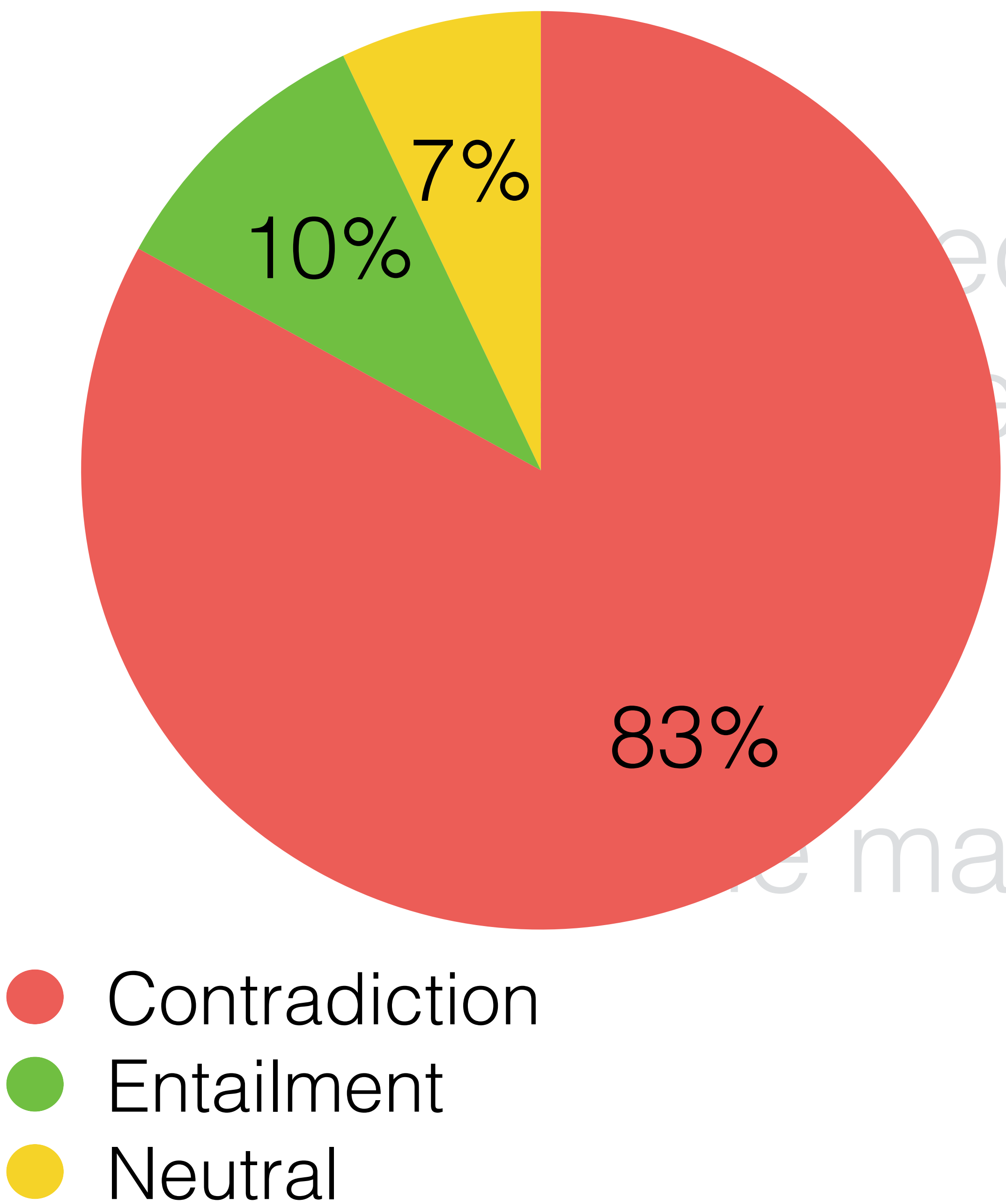
A man inspects the uniform of a figure in some East Asian country.

↓

The man is **sleeping**.

SNLI dataset (Bowman, 2015)

What are our systems learning?

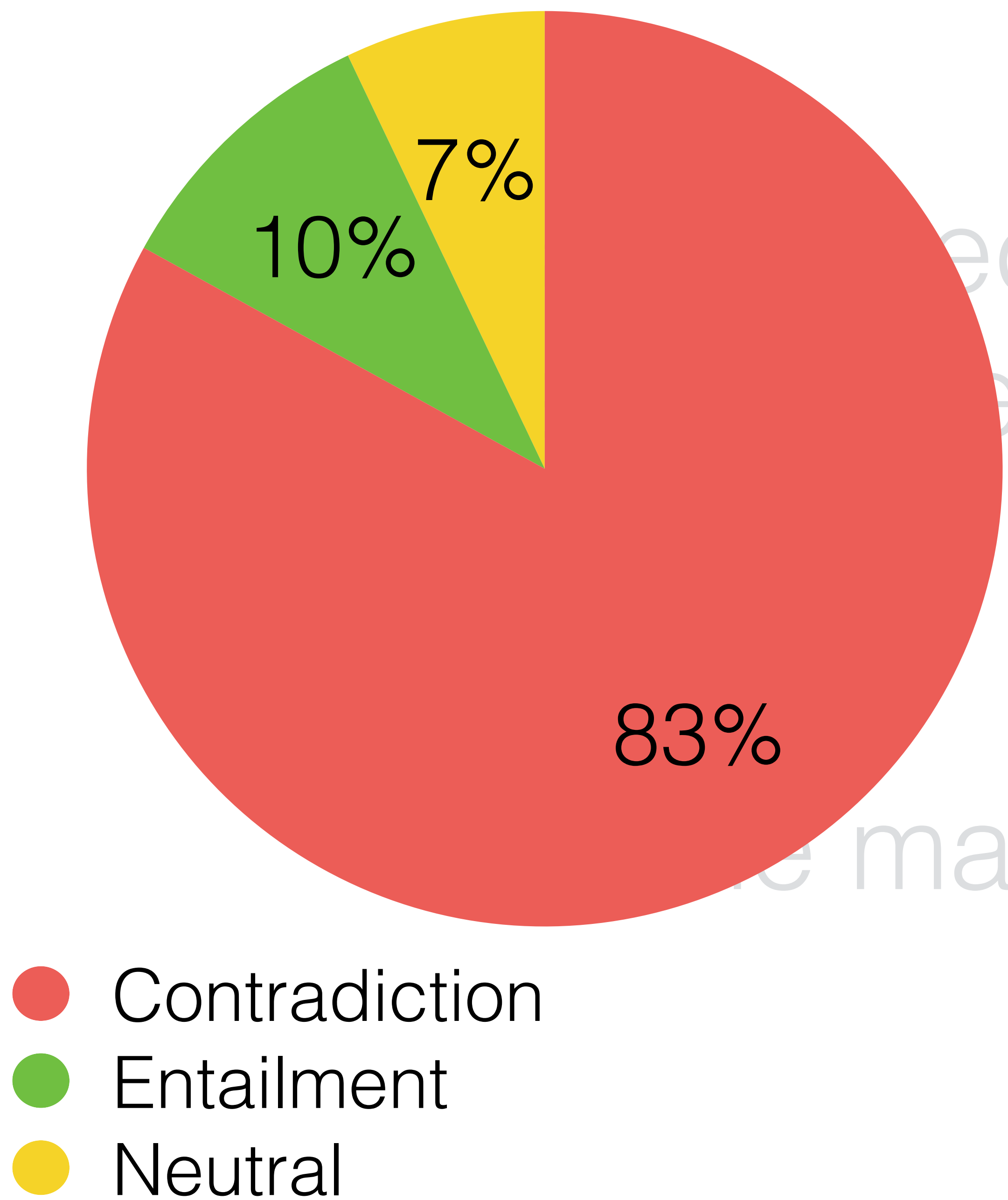


ects the uniform of a
e East Asian country.

e man is sleeping.

SNLI dataset (Bowman, 2015)

What are our systems learning?



Current SOTA is 86%

Tao Shen et al. 2018
300D Reinforced Self-Attention Network

man is sleeping.

SNLI dataset (Bowman, 2015)

What are our systems learning?

What is Southern California often abbreviated as?

+

Southern California, often abbreviated SoCal, is a geographic and cultural region that generally comprises California's southernmost 10 counties. The region is traditionally described as "eight counties", based on demographics and economic ties: Imperial, Los Angeles, Orange, Riverside, San Bernardino, San Diego, Santa Barbara, and Ventura. The more extensive 10-county definition, including Kern and San Luis Obispo counties, is also used based on historical political divisions. Southern California is a major economic center for the state of California and the United States.

↓

SoCal

Adversarial Examples for Evaluating Reading Comprehension Systems
Jia and Liang (2017)

What are our systems learning?

What is Southern California often abbreviated as?

+

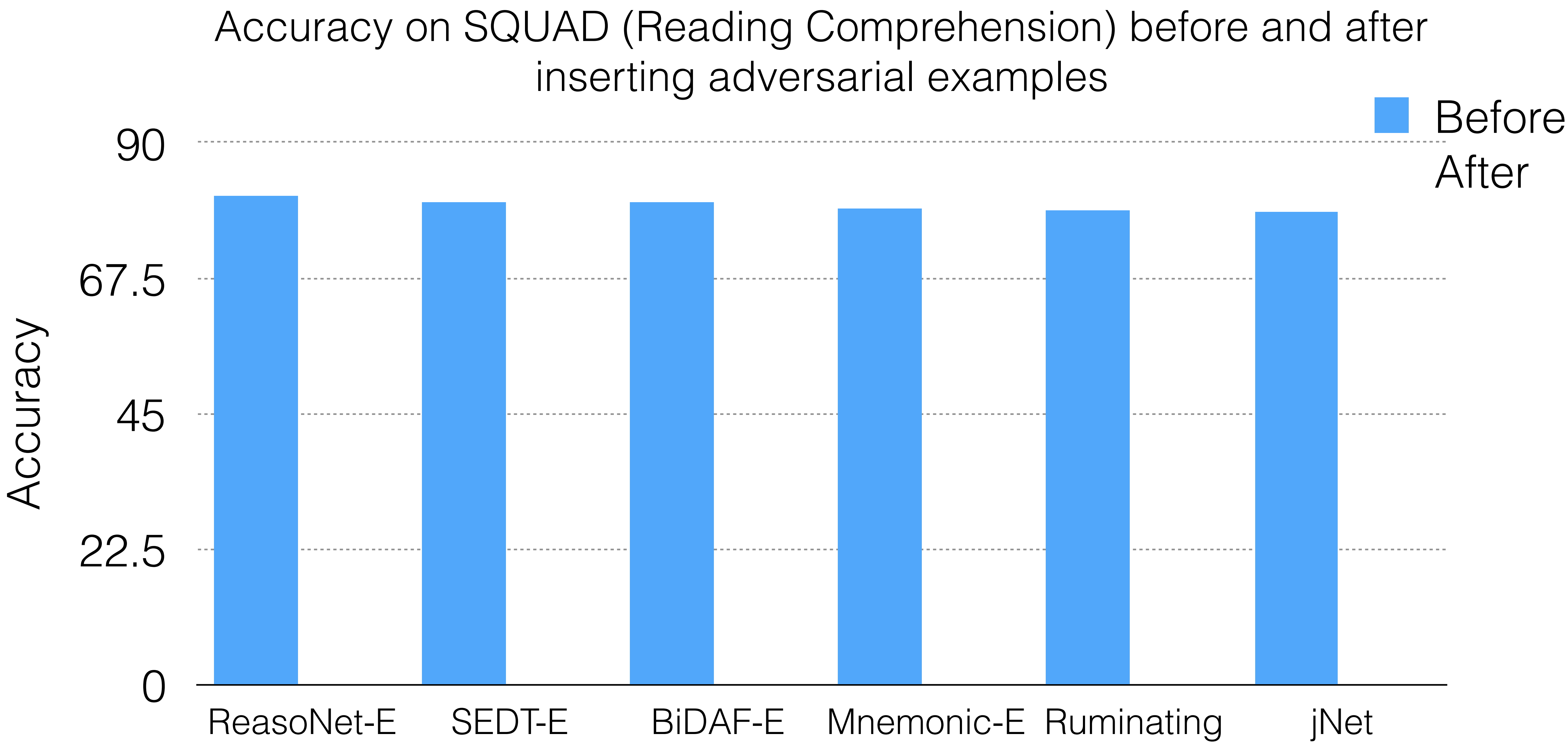
Southern California, often abbreviated SoCal, is a geographic and cultural region that generally comprises California's southernmost 10 counties. The region is traditionally described as "eight counties", based on demographics and economic ties: Imperial, Los Angeles, Orange, Riverside, San Bernardino, San Diego, Santa Barbara, and Ventura. The more extensive 10-county definition, including Kern and San Luis Obispo counties, is also used based on historical political divisions. Southern California is a major economic center for the state of California and the United States. Norther California is often abbreviated NorCal.

↓

SoCal

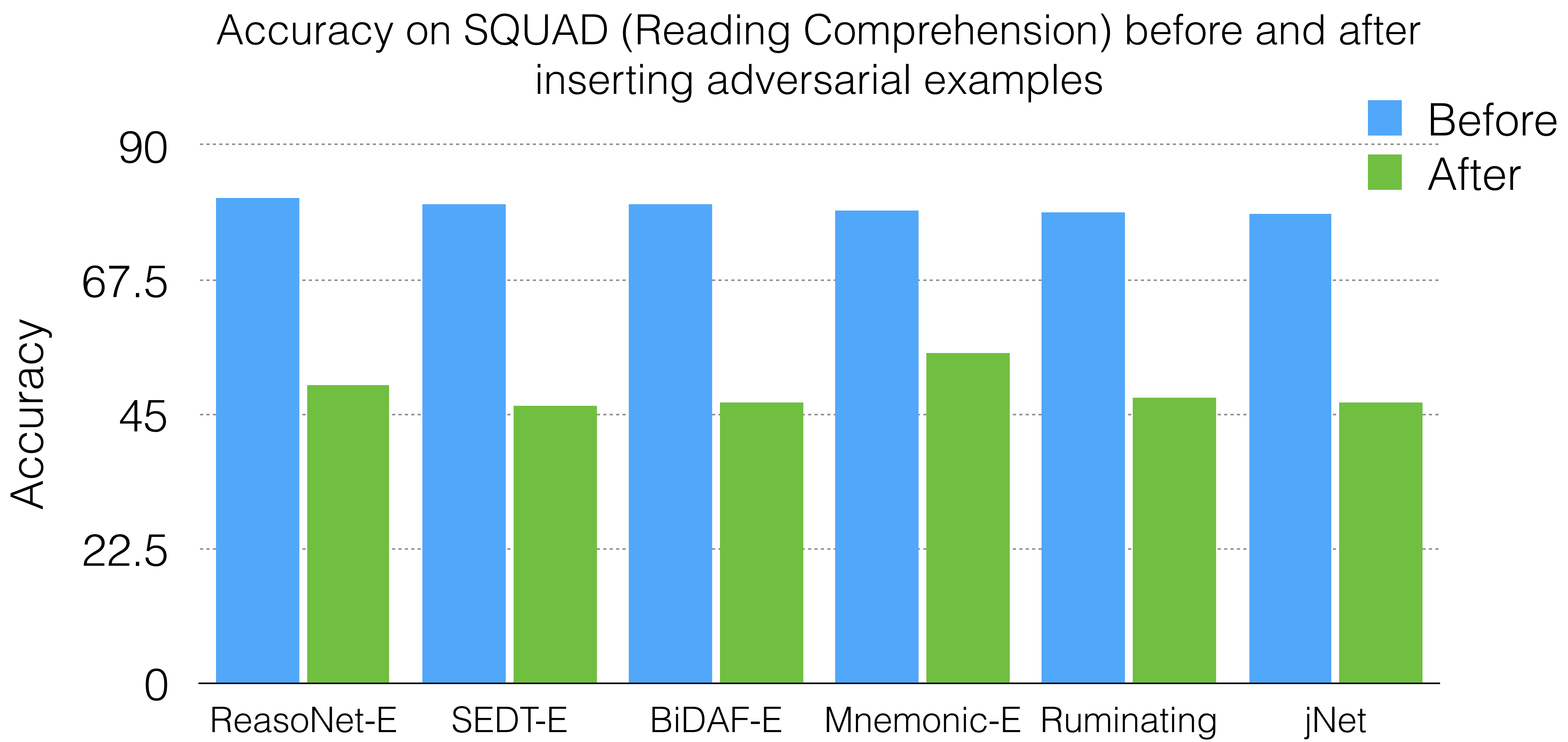
Adversarial Examples for Evaluating Reading Comprehension Systems
Jia and Liang (2017)

What are our systems learning?



Adversarial Examples for Evaluating Reading Comprehension Systems
Jia and Liang (2017)

What are our systems learning?



Adversarial Examples for Evaluating Reading Comprehension Systems
Jia and Liang (2017)

What do we want our systems to learn?

What do we want our systems to learn?

Word Embedding-based Antonym Detection using Thesauri and Distributional Information (Ono et al. 2015)

Retrofitting Word Vectors to Semantic Lexicons (Faruqui et al 2015)

Pre-Trained Word Embeddings

Counter-fitting Word Vectors to Linguistic Constraints (Mrkšić et al. 2016)

Low-Dimensional Embeddings of Logic (Rocktaschel et al. 2014)

Integrating Distributional Lexical Contrast into Word Embeddings for Antonym-Synonym Distinction (Nguyen et al. 2016)

Learning Semantic Word Embeddings based on Ordinal Knowledge Constraints (Liu et al 2015)

Identifying and Exploiting Hearst Patterns in Distributional Vectors for Lexical Entailment (Roller and Erk 2016)

SENSEMBED: Learning Sense Embeddings… (Iacobacci et al 2015)

AutoExtend: Extending Word Embeddings…(Rothe and Schutze 2015)

What do we want our systems to learn?

Word Embedding-based Antonym Detection using Thesauri and Distributional Information (Ono et al. 2015)

Retrofitting Word Vectors to Semantic Lexicons (Faruqui et al 2015)

Pre-Trained Word Embeddings

Counter-fitting Word Vectors to Linguistic Constraints (Mrkšić et al. 2016)

Low-Dimensional Embeddings of Logic (Rocktaschel et al. 2014)

Integrating Distributional Lexical Contrast into Word Embeddings for Antonym-Synonym Distinction (Nguyen et al. 2016)

Learning Semantic Word Embeddings based on Ordinal Knowledge Constraints (Liu et al 2015)

Identifying and Exploiting Hearst Patterns in Distributional Vectors for Lexical Entailment (Roller and Erk 2016)

SENSEMBED: Learning Sense Embeddings… (Iacobacci et al 2015)

AutoExtend: Extending Word Embeddings…(Rothe and Schutze 2015)

Skip-Thought Vectors (Kiros et al. 2015)

Distributed Representations of Sentences and Documents (Quoc et al. 2014)

A Structured Self-attentive Sentence Embedding (Lin et al. 2017)

Task-Independent Sentence Embeddings

Supervised Learning of Universal Sentence Representations from Natural Language Inference Data (Conneau et al. 2017)

Siamese CBOW: Optimizing Word Embeddings for Sentence Representations (Kenter et al. 2016)

Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection (Socher et al. 2011)

Towards Universal Paraphrastic Sentence Embeddings (Wieting et al. 2016)

What do we want our systems to learn?

Word Embedding-based Antonym Detection using Thesauri and Distributional Information (Ono et al. 2015)

Retrofitting Word Vectors to Semantic Lexicons (Faruqui et al 2015)

Pre-Trained Word Embeddings

Counter-fitting Word Vectors to Linguistic Constraints (Mrkšić et al. 2016)

Low-Dimensional Embeddings of Logic (Rocktaschel et al. 2014)

Integrating Distributional Lexical Contrast into Word Embeddings for Antonym-Synonym Distinction (Nguyen et al. 2016)

Learning Semantic Word Embeddings based on Ordinal Knowledge Constraints (Liu et al 2015)

Identifying and Exploiting Hearst Patterns in Distributional Vectors for Lexical Entailment (Roller and Erk 2016)

SENSEMBED: Learning Sense Embeddings… (Iacobacci et al 2015)

AutoExtend: Extending Word Embeddings…(Rothe and Schutze 2015)

Skip-Thought Vectors (Kiros et al. 2015)

Distributed Representations of Sentences and Documents (Quoc et al. 2014)

A Structured Self-attentive Sentence Embedding (Lin et al. 2017)

Task-Independent Sentence Embeddings

Supervised Learning of Universal Sentence Representations from Natural Language Inference Data (Conneau et al. 2017)

Siamese CBOW: Optimizing Word Embeddings for Sentence Representations (Kenter et al. 2016)

Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection (Socher et al. 2011)

Towards Universal Paraphrastic Sentence Embeddings (Wieting et al. 2016)

Knowledge Representation and Grounding

Learning Structured Embeddings of Knowledge Bases (Bordes et al. 2011)

Embedding Multimodal Relational Data (Open review 2018)

DeViSE: A Deep Visual-Semantic Embedding Model (Frome et al 2013)

Multimodal Neural Language Models (Kiros et al 2014)

Learning Semantic Hierarchies via Word Embeddings (Fu et al. 2014)

Deep Visual-Semantic Alignments for Generating Image Descriptions (Karpathy and Fei Fei 2015)

Improved Representation Learning for Predicting Commonsense Ontologies (Li et al 2017)

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

Language → $\lambda g.(\lambda x.g\ (x\ x))(\lambda x.g\ (x\ x))$

Language $\rightarrow$ $$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$
$$h_t = o_t \circ \sigma_h(c_t)$$

Language → 0101101101010010101010

Language → 01011011010100101010

The Semantics-Pragmatics Interface

The Semantics-Pragmatics Interface

The Semantics-Pragmatics Interface

I went to the beach over vacation.

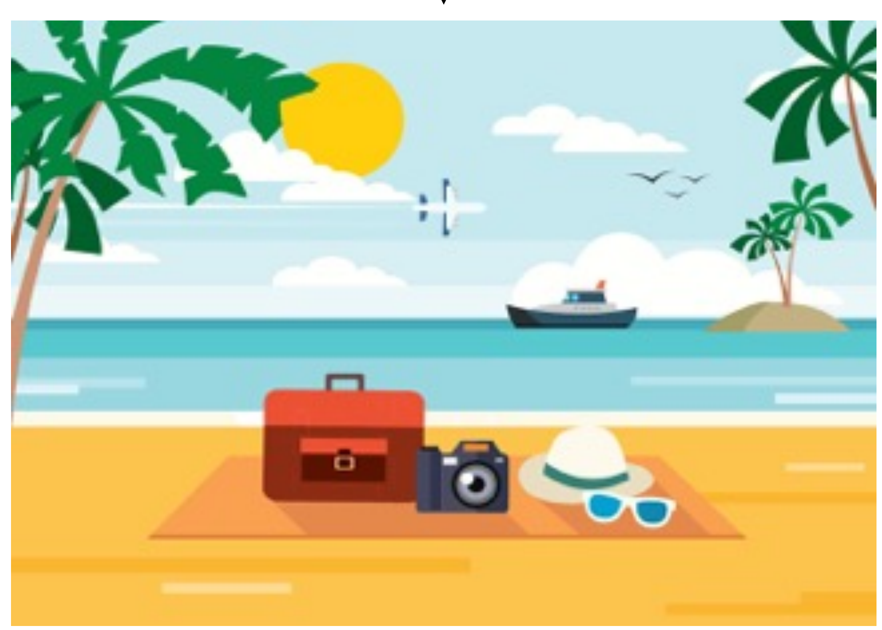
The Semantics-Pragmatics Interface

I went to the beach over vacation.

The Semantics-Pragmatics Interface

Semantics

I went to the beach over vacation.

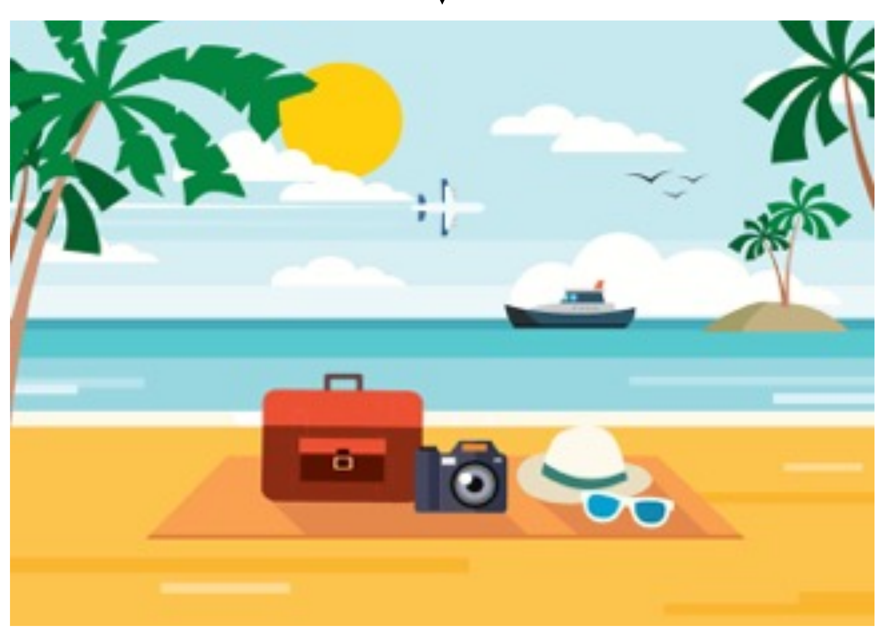


The Semantics-Pragmatics Interface

Semantics

I went to the beach over vacation.

Pragmatics



The Semantics-Pragmatics Interface

I went to the beach over vacation.

Semantics

agmatics

Context TBD

The Semantics-Pragmatics Interface

Semantics

I went to the beach over vacation.

↕

I laid out in the sun.

- -

agmatics

Context TBD

The Semantics-Pragmatics Interface

Semantics

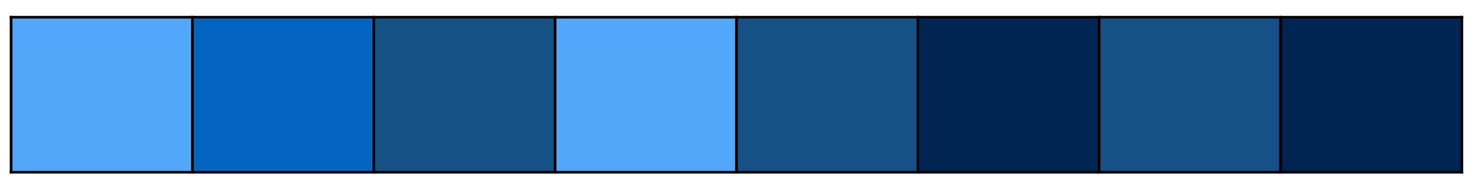
I went to the beach over vacation.

I laid out in the sun.

Pragmatics

What “belongs” in the representation of a word?

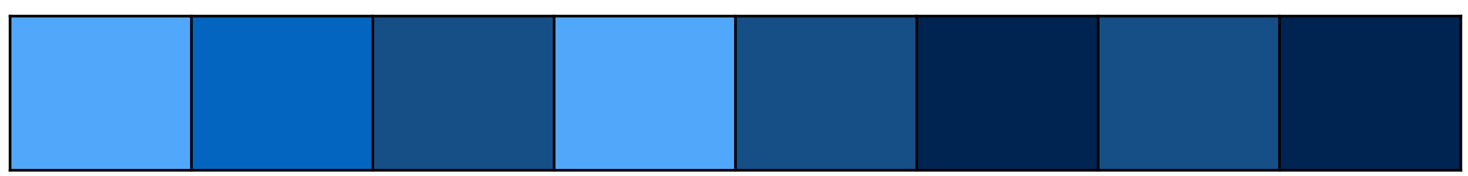
beach



What “belongs” in the representation of a word?

location near
the water

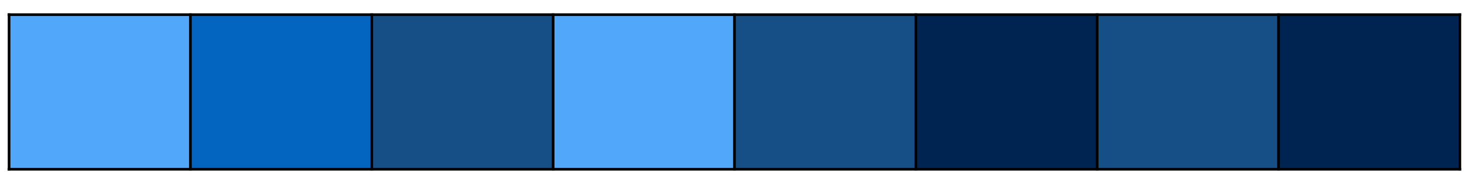
beach



What “belongs” in the representation of a word?

location near the water

beach

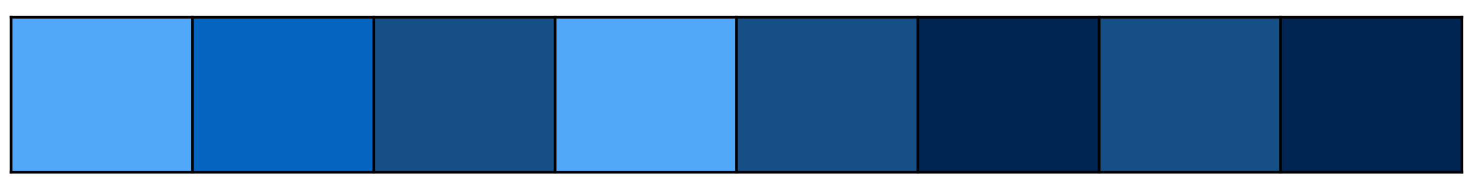


is a place

What “belongs” in the representation of a word?

location near the water

beach



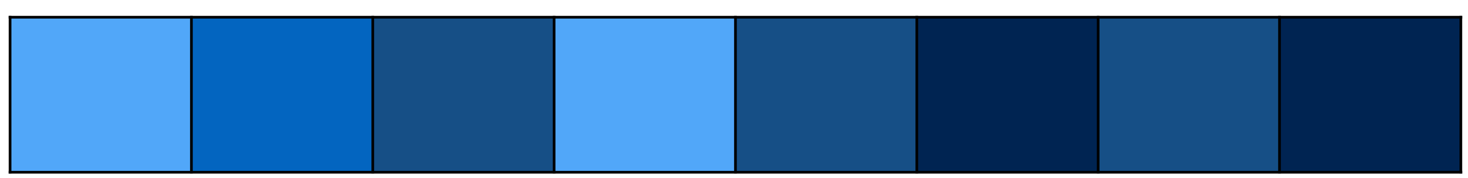
is a place

is a popular vacation place

What “belongs” in the representation of a word?

location near the water

beach



is a place

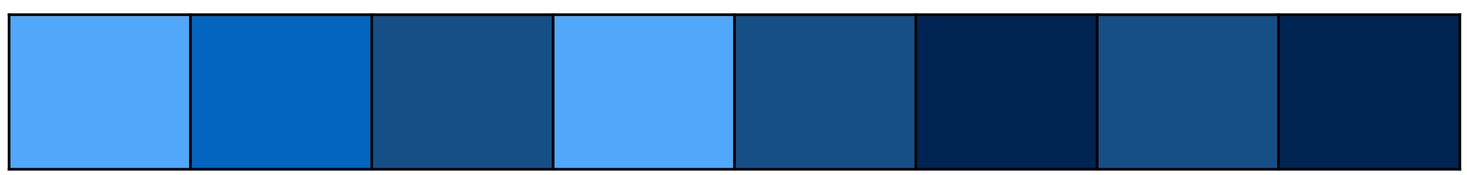
not indoors

is a popular vacation place

What “belongs” in the representation of a word?

location near the water

beach



is a place

not indoors

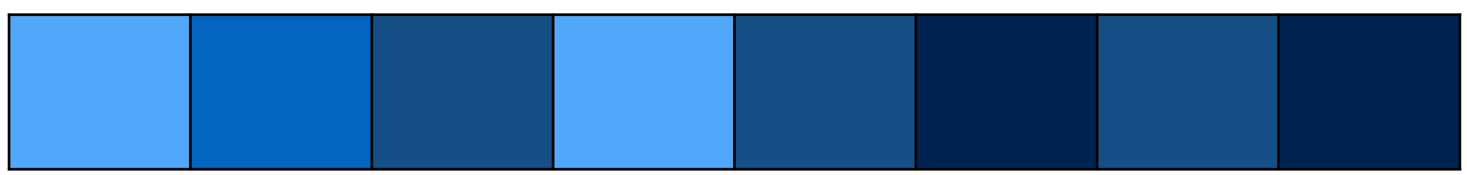
is a popular vacation place

may have palm trees

What “belongs” in the representation of a word?

location near the water

beach



is a place

P(palm trees)

not indoors

is a popular vacation place

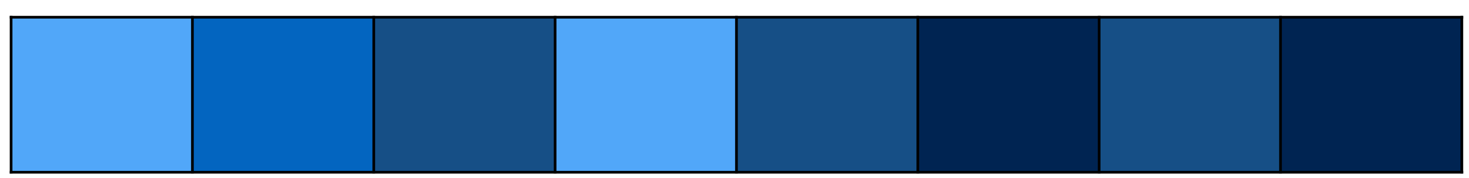
may have palm trees

What “belongs” in the representation of a word?

P(it is okay for a listener to imagine that it has palm trees)

location near the water

beach



is a place

P(palm trees)

not indoors

is a popular vacation place

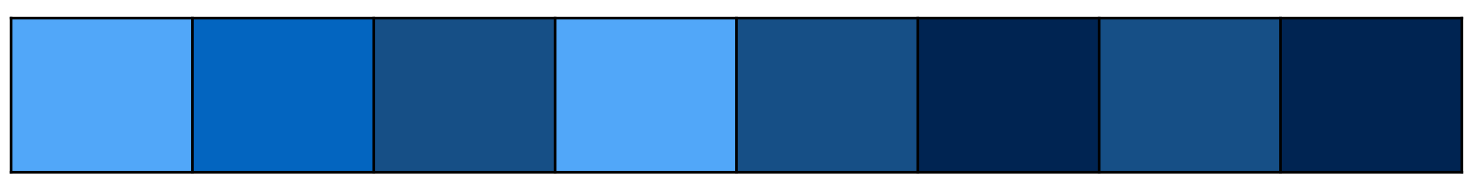
may have palm trees

What "belongs" in the representation of a word?

P(it is okay for a listener to imagine that it has palm trees)

location near the water

beach



is a place

P(palm trees)

not indoors

"Dictionary" Representation

is a popular vacation place

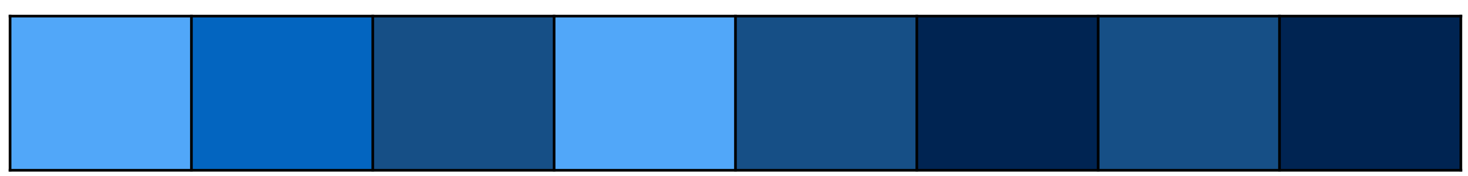
may have palm trees

What “belongs” in the representation of a word?

P(it is okay for a listener to imagine that it has palm trees)

location near the water

beach



is a place

P(palm trees)

not indoors

Taxonomic Representation

is a popular vacation place

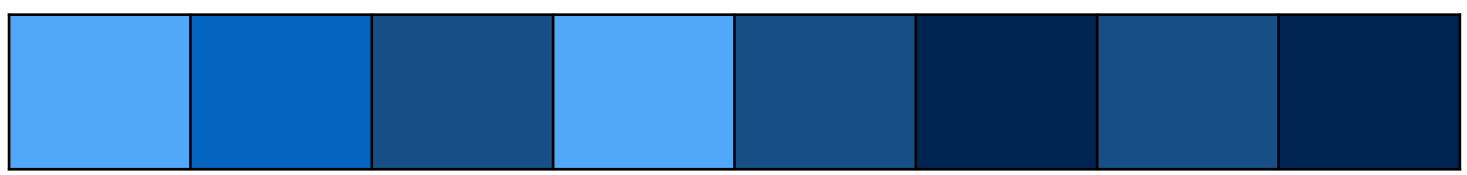
may have palm trees

What “belongs” in the representation of a word?

P(it is okay for a listener to imagine that it has palm trees)

location near the water

beach



is a place

Knowledge Base Representation

not indoors

P(palm trees)

is a popular vacation place

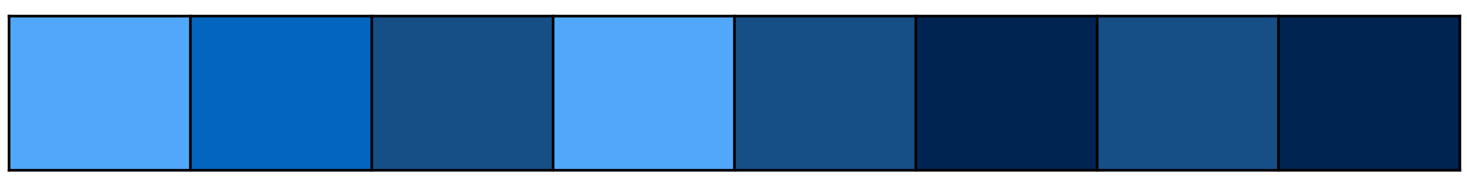
may have palm trees

What “belongs” in the representation of a word?

P(it is okay for a listener to imagine that it has palm trees)

location near the water

beach



is a place

P(palm trees)

not indoors

Prototype Representation

is a popular vacation place

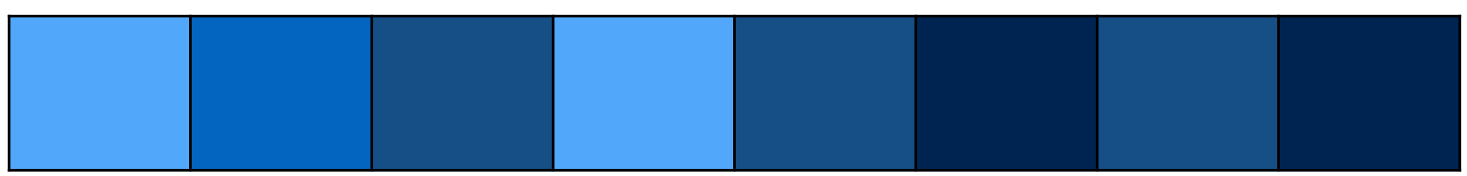
may have palm trees

What “belongs” in the representation of a word?

P(it is okay for a listener to imagine that it has palm trees)

location near the water

beach



is a place

“Encyclopedic” Representation

not indoors

P(palm trees)

is a popular vacation place

may have palm trees

Is SkipGram enough?

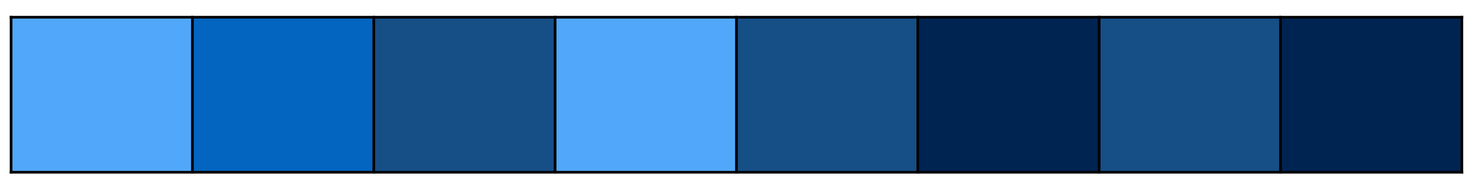
P(occurs after “sandy”)

P(occurs after “on”)

P(occurs before “vacation”)

P(occurs after “the”)

beach



P(occurs after “clandestine”)

P(occurs before “grinning”)

Distributional Contextual Representation

Should we care about linguistics?

Should we care about linguistics?

Yes.

Should we care about linguistics?

Yes.
Because we have to form and test hypotheses about what our word representations should capture.

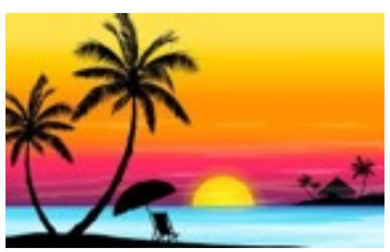
What “belongs” in the representation of a word?

beach

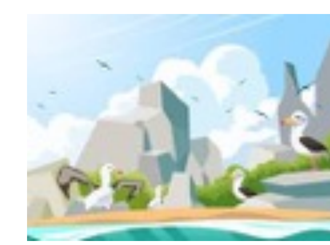
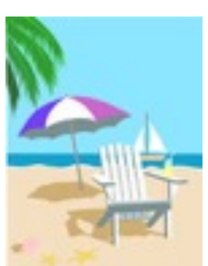
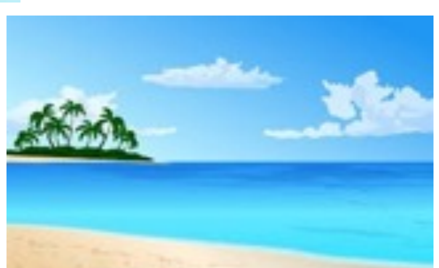
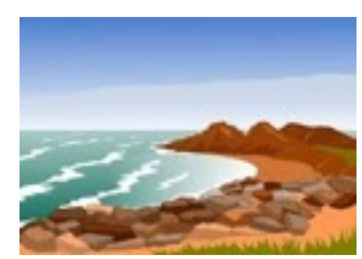
Model Theoretic Representation

What “belongs” in the representation of a word?

beach



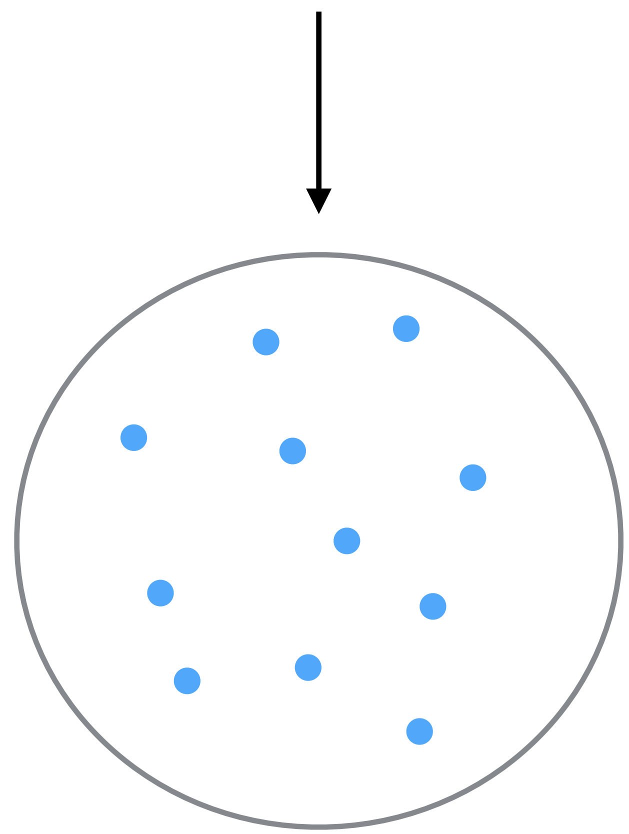


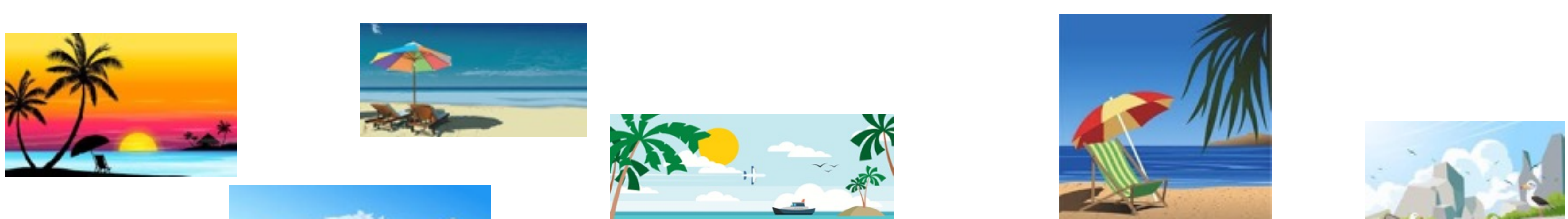





Model Theoretic Representation

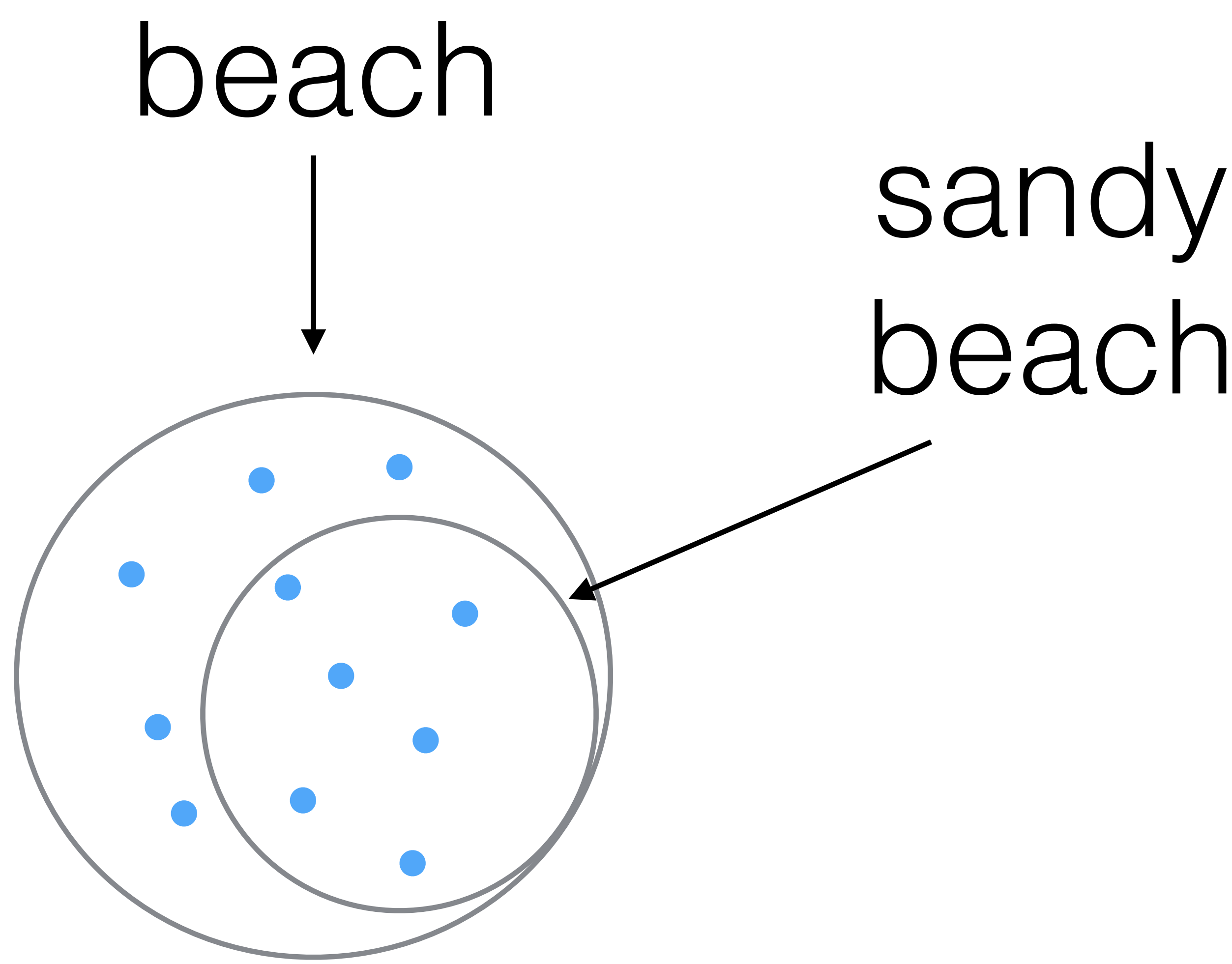
What "belongs" in the representation of a word?

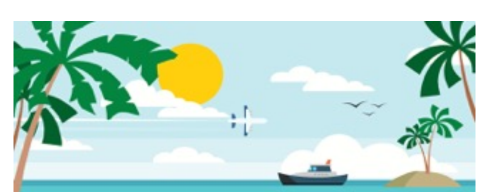
beach



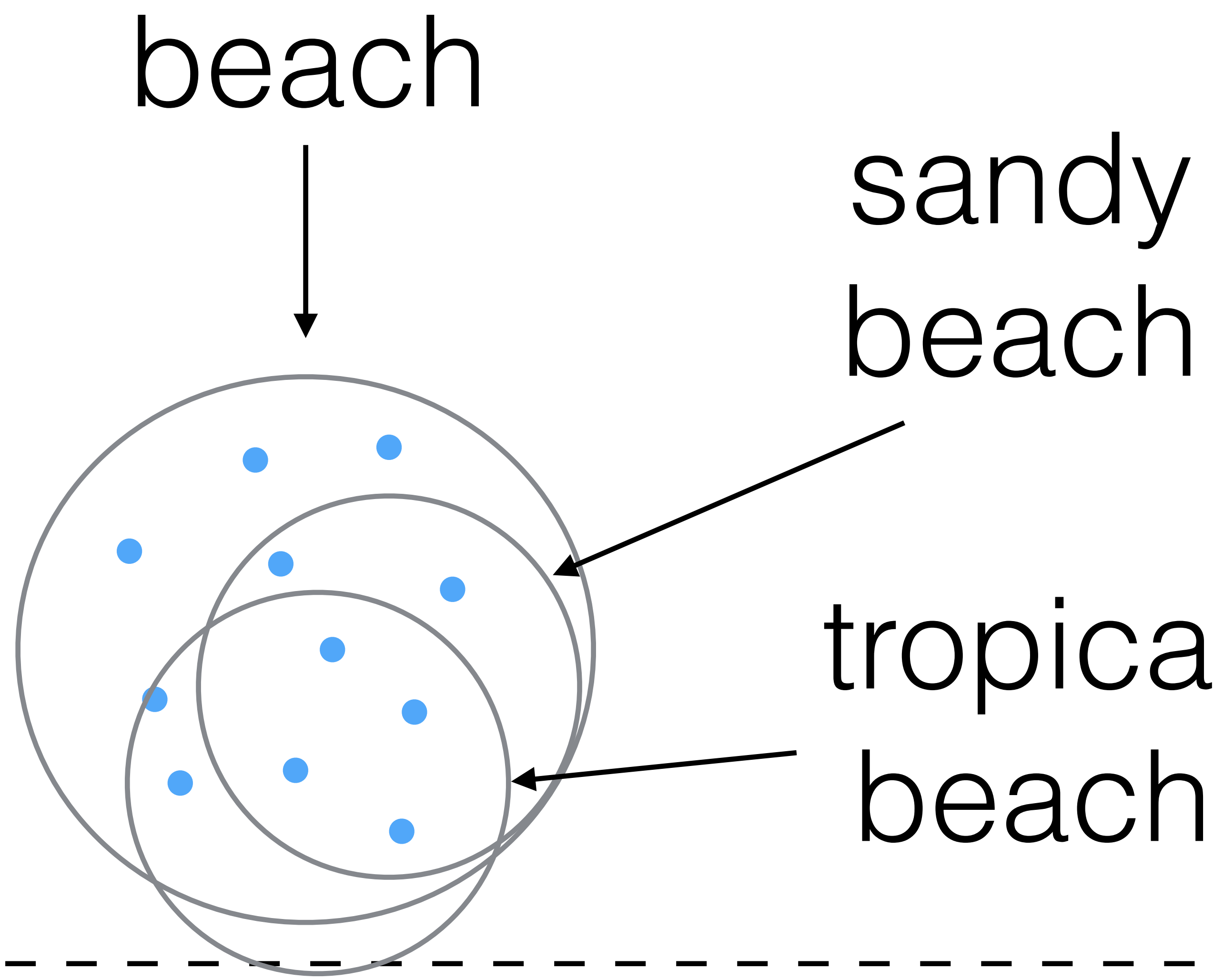


What “belongs” in the representation of a word?

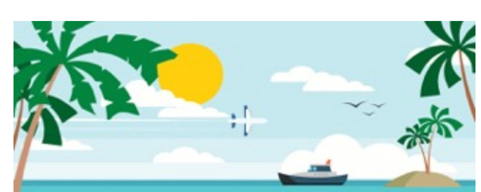




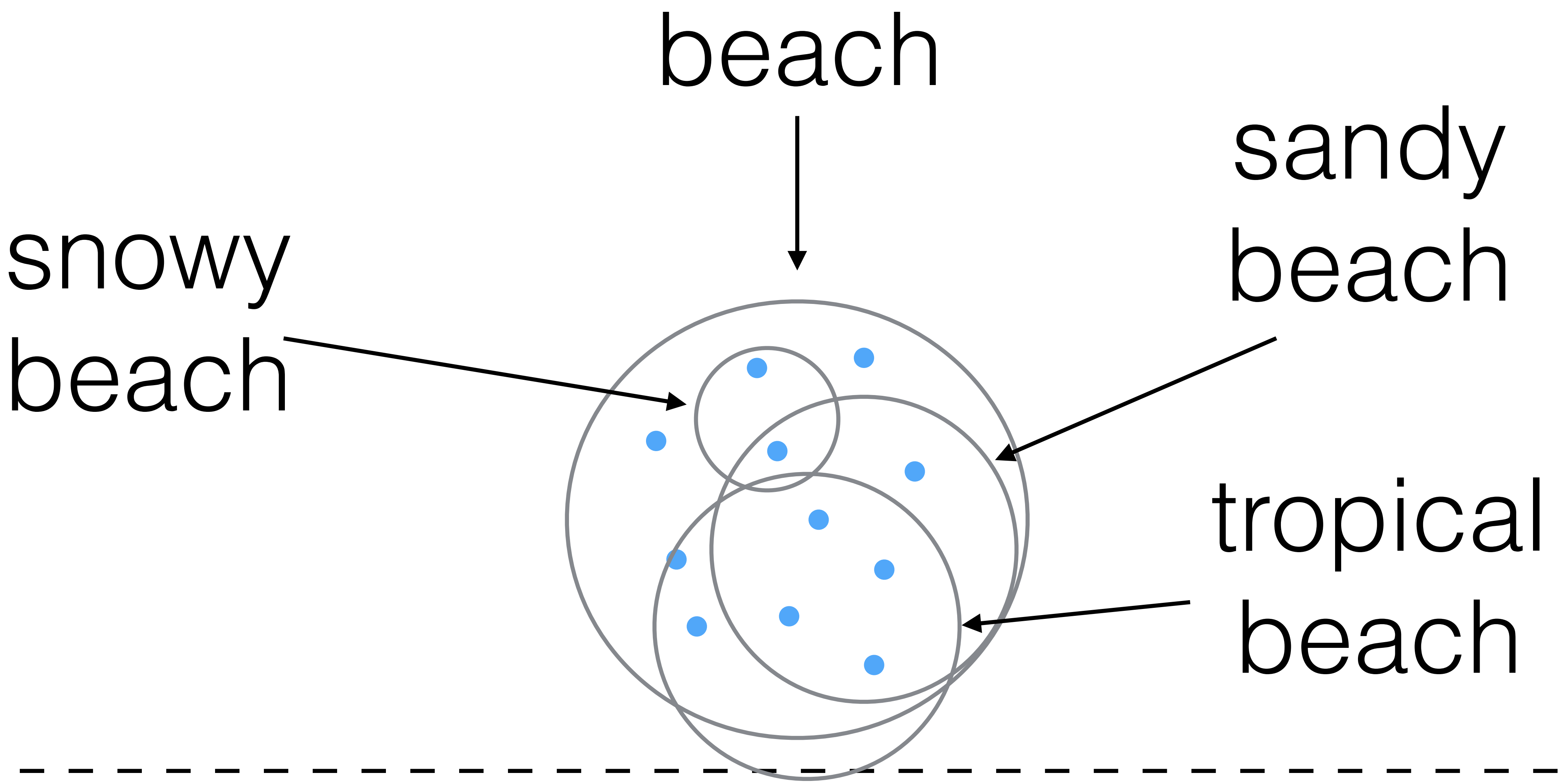
What “belongs” in the representation of a word?

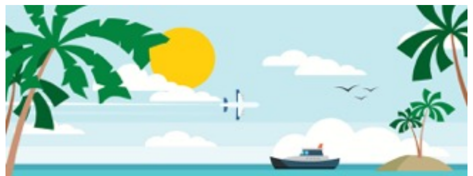




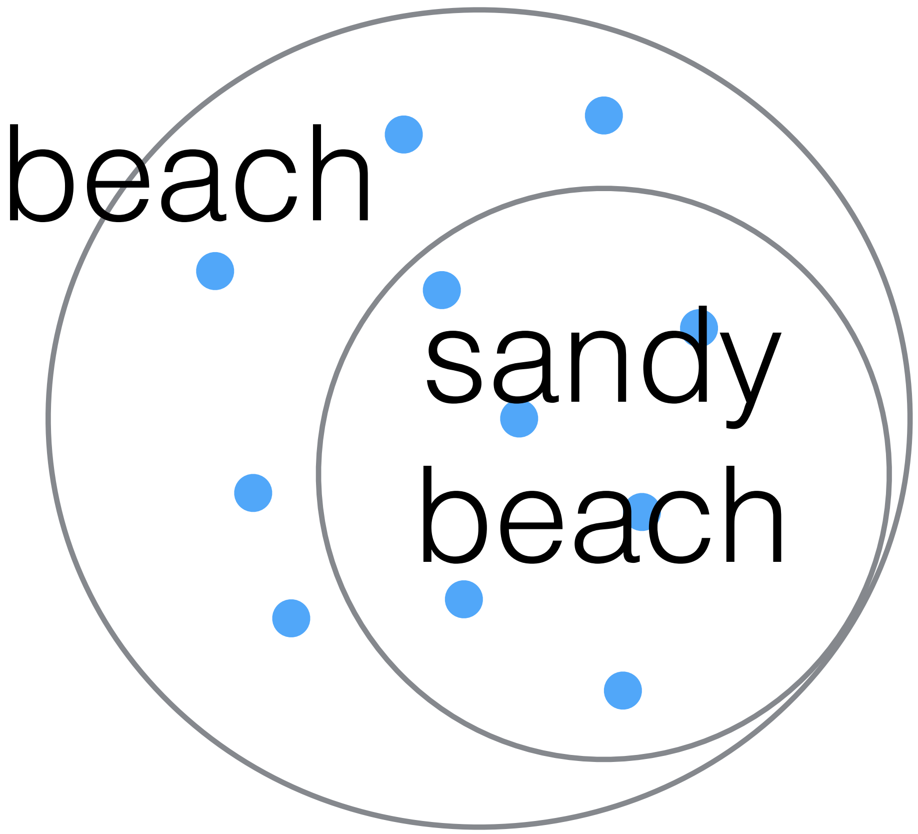



What “belongs” in the representation of a word?



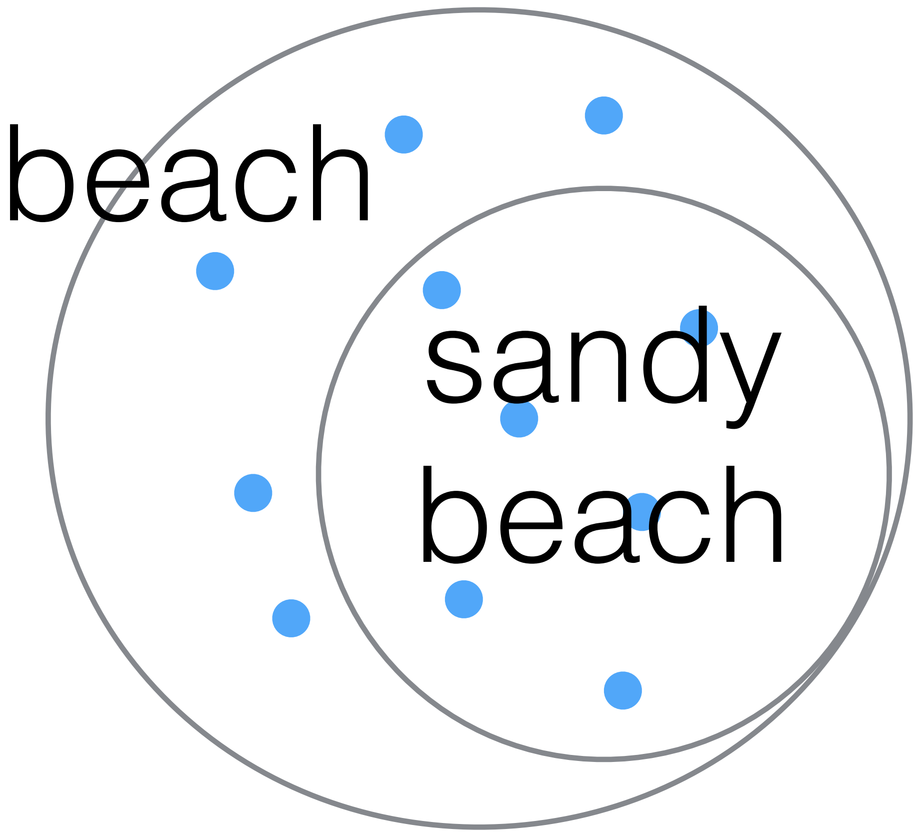


Set-Theoretic Semantics



A little boy doing a hand stand on the beach.

Set-Theoretic Semantics

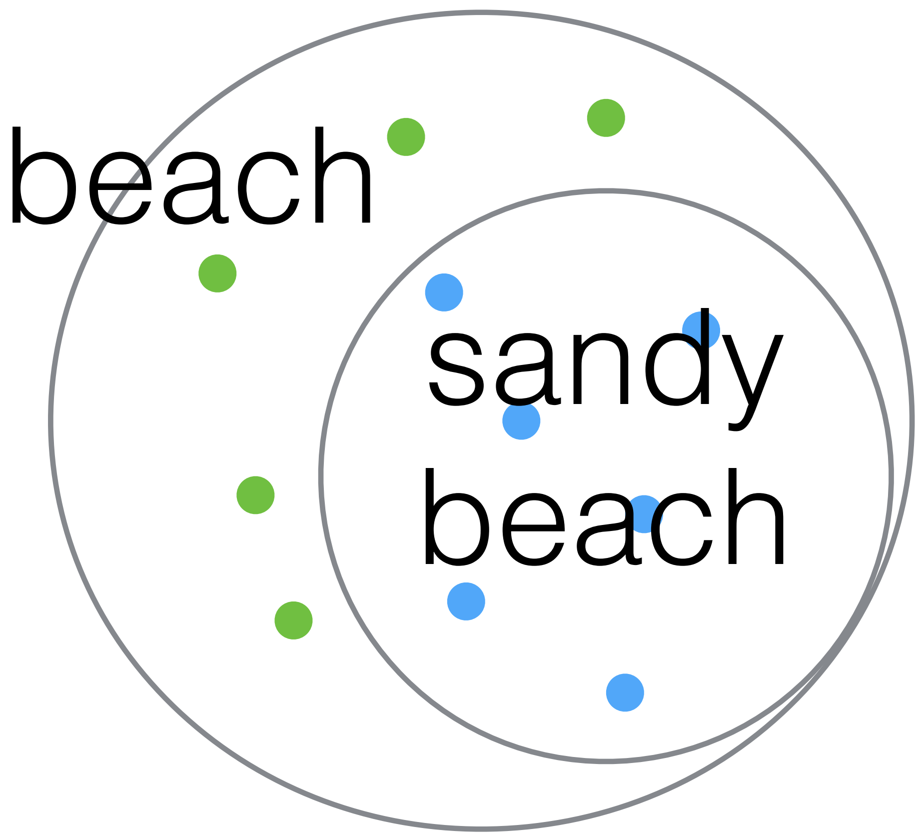


A little boy doing a hand stand on the beach.

A little boy doing a hand stand on the sandy beach.

Set-Theoretic Semantics

$\exists x(\text{beach}(x) \land \neg \text{sandy\_beach}(x))$

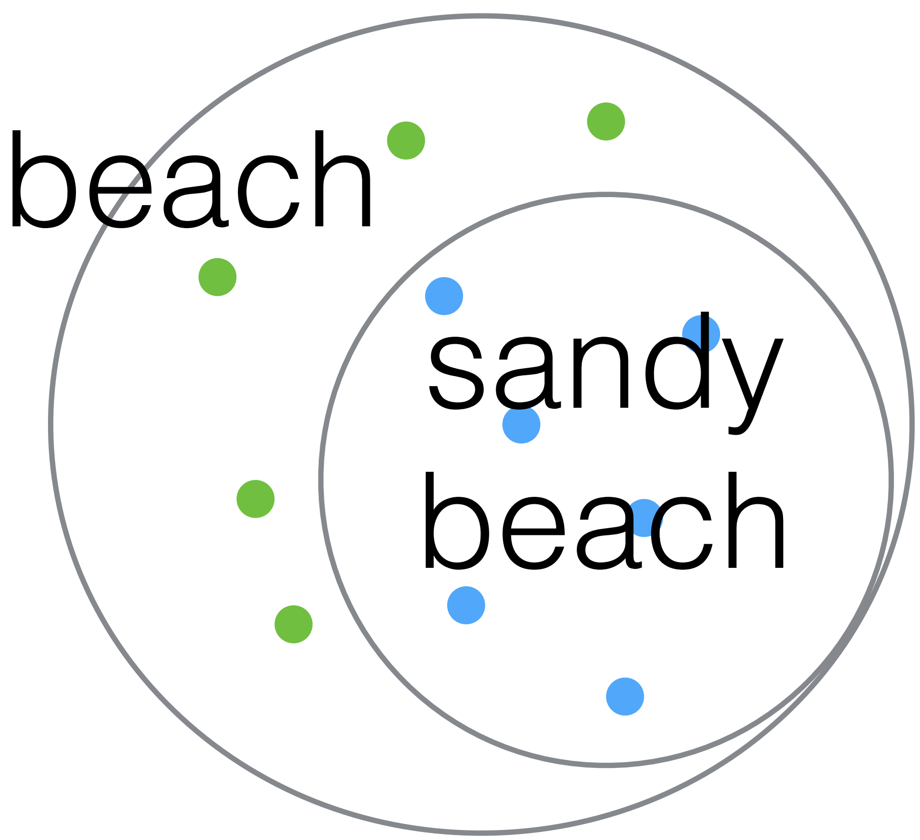


A little boy doing a hand stand on the beach.

A little boy doing a hand stand on the sandy beach.

Set-Theoretic Semantics

∃x(beach(x)∧¬sandy_beach(x))



A little boy doing a hand stand on the beach.

$\not\downarrow$

A little boy doing a hand stand on the sandy beach.

Set-theoretic semantics does not allow this inference.

Set-Theoretic Semantics?

A little boy doing a hand stand on the beach.

A little boy doing a hand stand on the sandy beach.

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Set-Theoretic Semantics?

A little boy doing a hand stand on the beach.

+

A little boy doing a hand stand on the sandy beach.

↓

Human Annotators

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Set-Theoretic Semantics?

A little boy doing a hand stand on the beach.

+

A little boy doing a hand stand on the sandy beach.

↓

Human Annotators

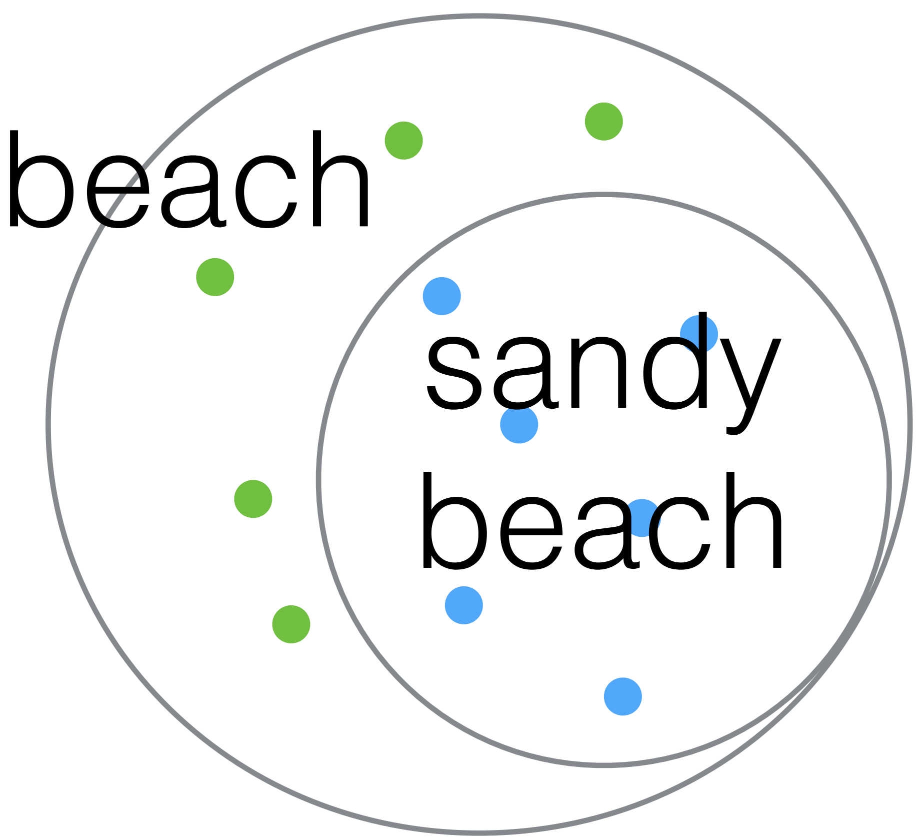
↓

Yes

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Set-Theoretic Semantics?

$$\exists x\,(\mathtt{beach}(x) \land \neg \mathtt{sandy\_beach}(x))$$



A little boy doing a hand stand on the beach.

$$\not\Downarrow$$

A little boy doing a hand stand on the sandy beach.

Set-theoretic semantics does not allow this inference.

…ge”:
Compositional Entailment in Adjective Nouns
Pavlick and Callison-Burch (2016)

Set-Theoretic Semantics?

$\cancel{\exists x \, (beach(x) \wedge \neg sandy\_beach(x))}$

Human subjects unanimously agree that this inference in valid.

A little boy doing a hand stand on the beach.

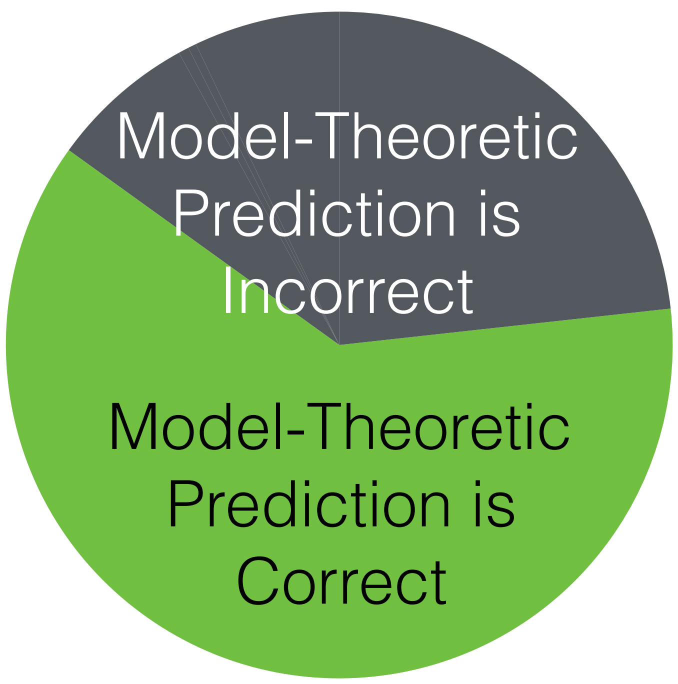
↓

A little boy doing a hand stand on the sandy beach.

~~Set-theoretic semantics does not allow this inference.~~

…ge”:
Compositional Entailment in Adjective Nouns
Pavlick and Callison-Burch (2016)

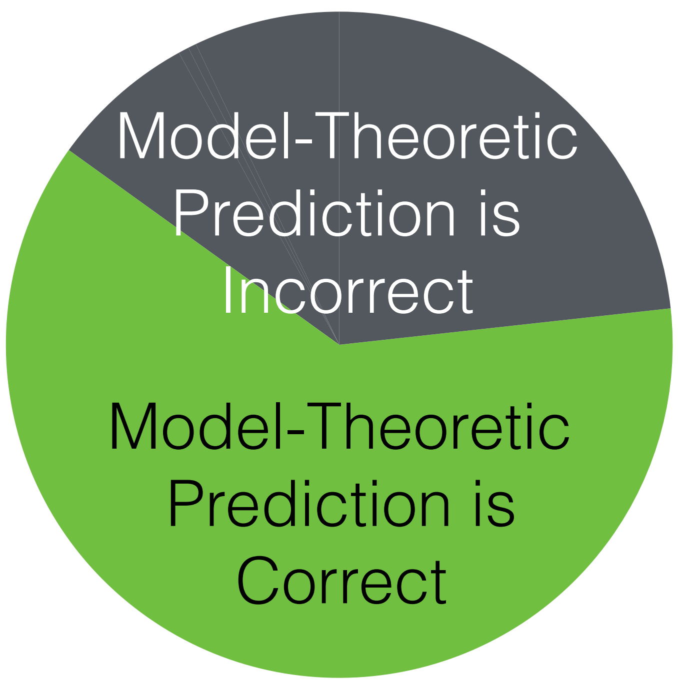
Set-Theoretic Semantics?



Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

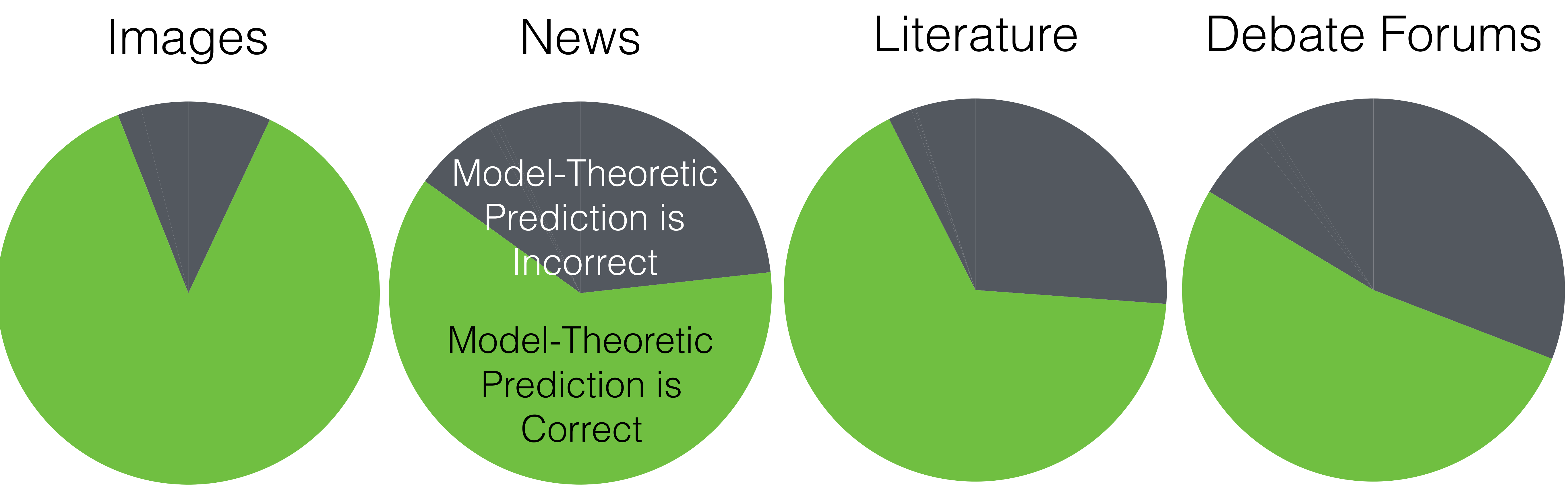
Set-Theoretic Semantics?

News



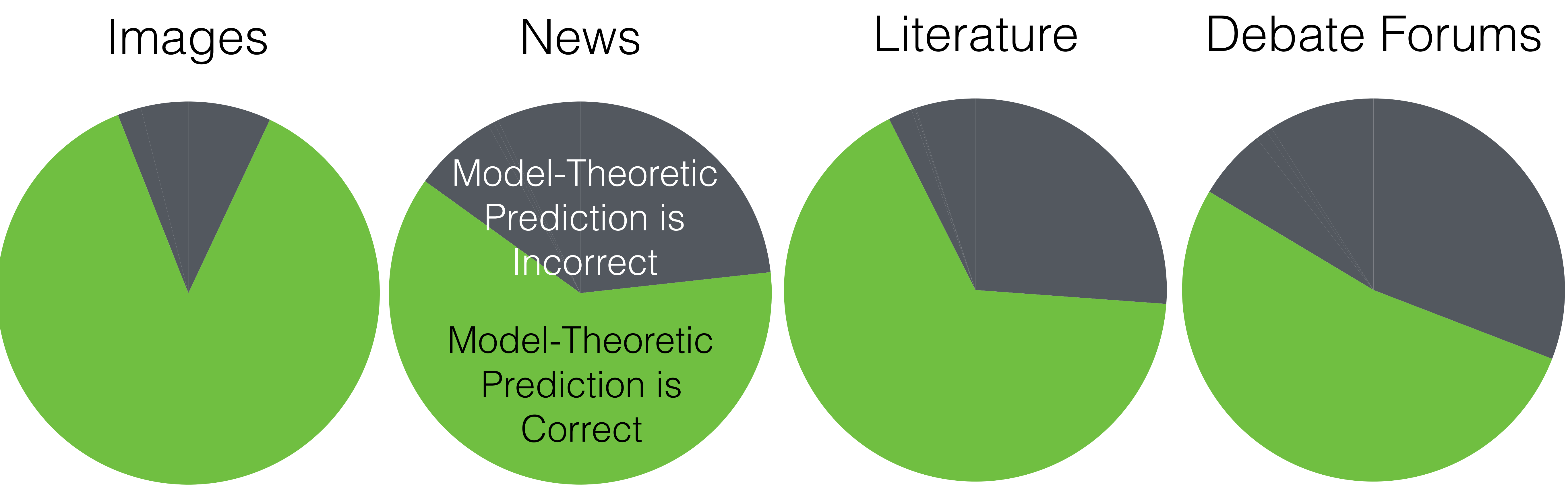
Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Set-Theoretic Semantics?



Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Set-Theoretic Semantics?



In the worst case, Model-Theoretic representation makes incorrect predictions 47% of the time!

lems” are “huge”:
Adjective-Nouns
Pavlick and Callison-Burch (2016)

Human Inferences

P entails H	P contradicts H
Somehow, I feel there will be a lack of evidence forthcoming **evidence -> credible evidence**	Bush travels Monday to Michigan to make remarks on the economy. **economy -/-> Japanese economy**
Penfield Evans grasped his hand and shook it warmly. **hand -> outstretched hand**	Government is the only thing holding back large corporations. **government -/-> small government**
His body is found a week later. **body -> dead body**	A child rides on a man's shoulders. **man -/-> homeless man**

Pavlick and Callison-Burch (2016)

What “belongs” in the representation of a word?

evidence -> credible evidence

hand -> outstretched hand

body -> dead body

economy -/-> Japanese economy

government -/-> small government

man -/-> homeless man

What “belongs” in the representation of a word?

evidence -> is credible?

body -> is dead?

government -> isn’t small?

man -> isn’t homeless?

hand -> is outstretched?

economy -> isn’t Japanese?

What “belongs” in the representation of a word?

Semantics

evidence -> is credible?

- -

body -> is dead?

government -> isn’t small?

man -> isn’t homeless?

hand -> is outstretched?

economy -> isn’t Japanese?

Pragmatics

What “belongs” in the representation of a word?

Semantics

evidence -> is credible?

body -> is dead?

- -

government -> isn’t small?

man -> isn’t homeless?

hand -> is outstretched?

economy -> isn’t Japanese?

Pragmatics

What “belongs” in the representation of a word?

Semantics

evidence -> is credible?

body -> is dead?

government -> isn’t small?

Pragmatics

man -> isn’t homeless?

hand -> is outstretched?

economy -> isn’t Japanese?

Semantics

evidence -> is credible?

body -> is dead?

government -> isn't small?

Pragmatics

man -> isn't homeless?

hand -> is outstretched?

economy -> isn't Japanese?

Should we care about linguistics?

Semantics

evidence -> is credible?

body -> is dead?

government -> isn't small?

Pragmatics

man -> isn't homeless?

hand -> is outstretched?

economy -> isn't Japanese?

Recognizing Textual Entailment Task

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Recognizing Textual Entailment Task

A group of hikers walk a path that leads from a sandy beach towards a hill

+

The hikers are walking outside

↓

RTE System

↓

True

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task

A hiker walking on a path at the foot of snow capped mountains

+

A hiker walking on a sandy path at the foot of snow capped mountains

↓

RTE System

↓

False

Most “babies” are “little” and most “problems” are “huge”: Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task

- 5,378 add-one pairs

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task

- 5,378 add-one pairs
- 4,991 for training (4,481 training, 510 dev)

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task

- 5,378 add-one pairs
- 4,991 for training (4,481 training, 510 dev)
- 387 test (removed pairs with low human agreement)

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task

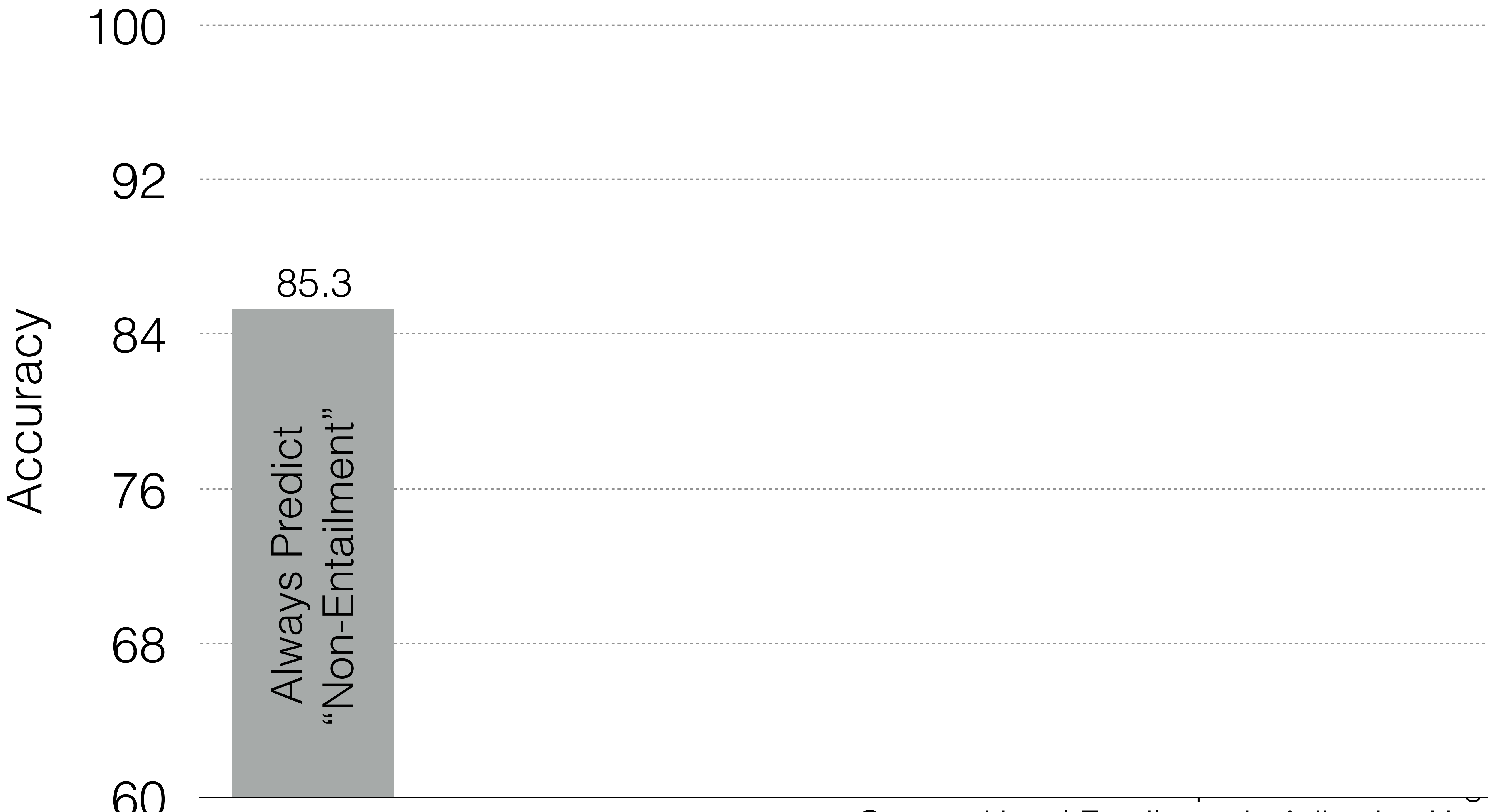
- 5,378 add-one pairs
- 4,991 for training (4,481 training, 510 dev)
- 387 test (removed pairs with low human agreement)
- 500K general RTE pairs from SNLI

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task

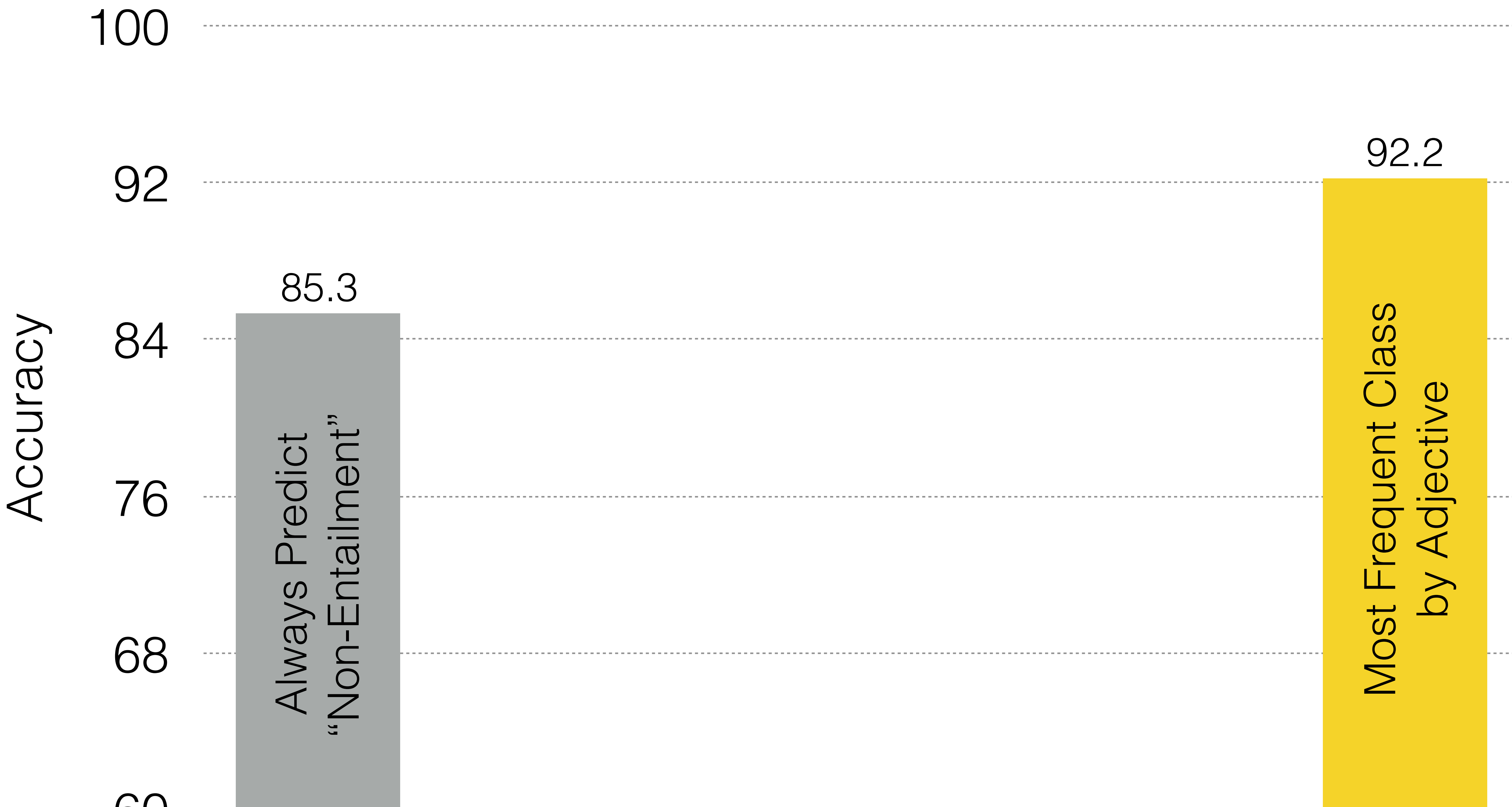
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task



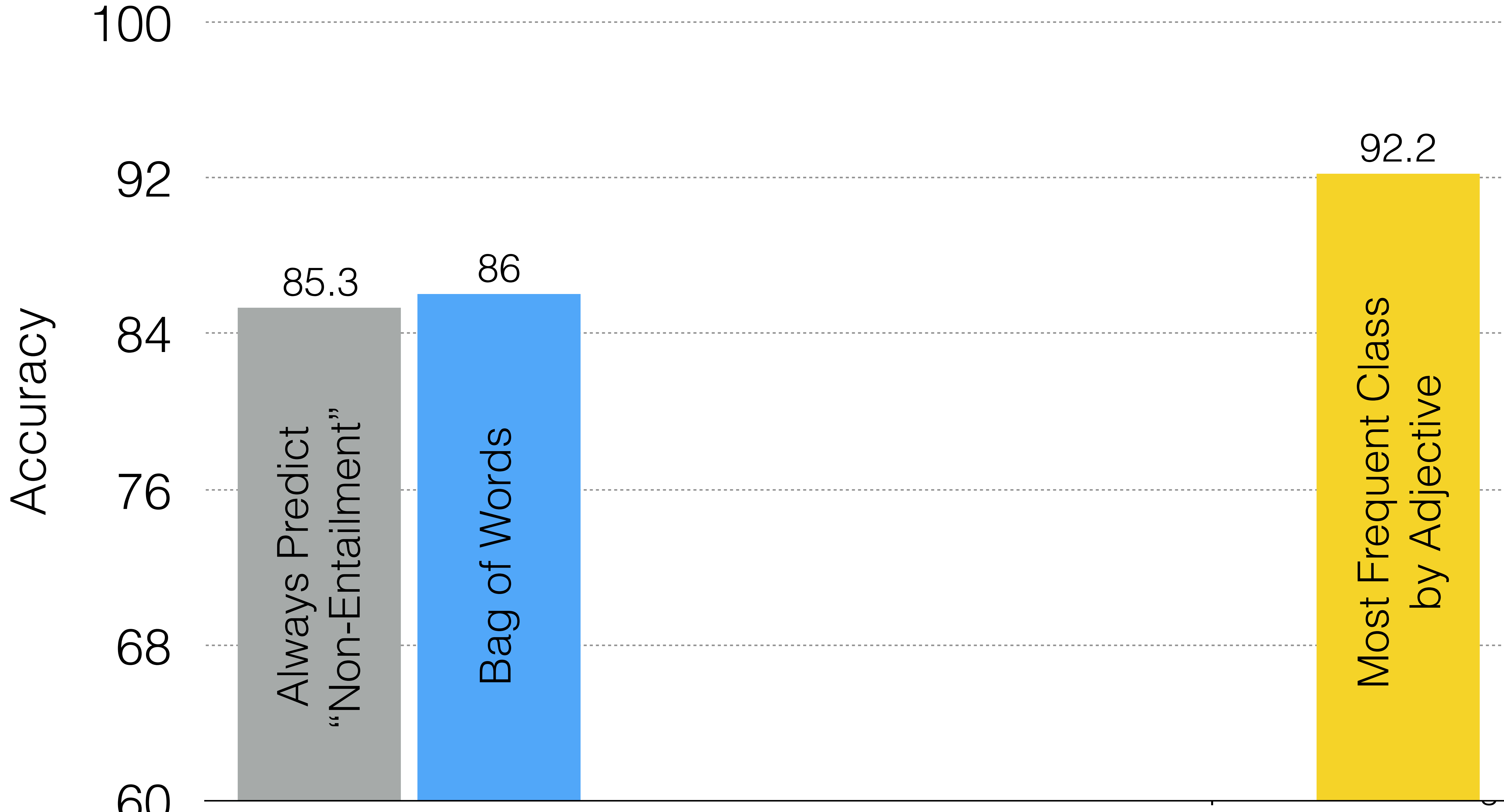
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task



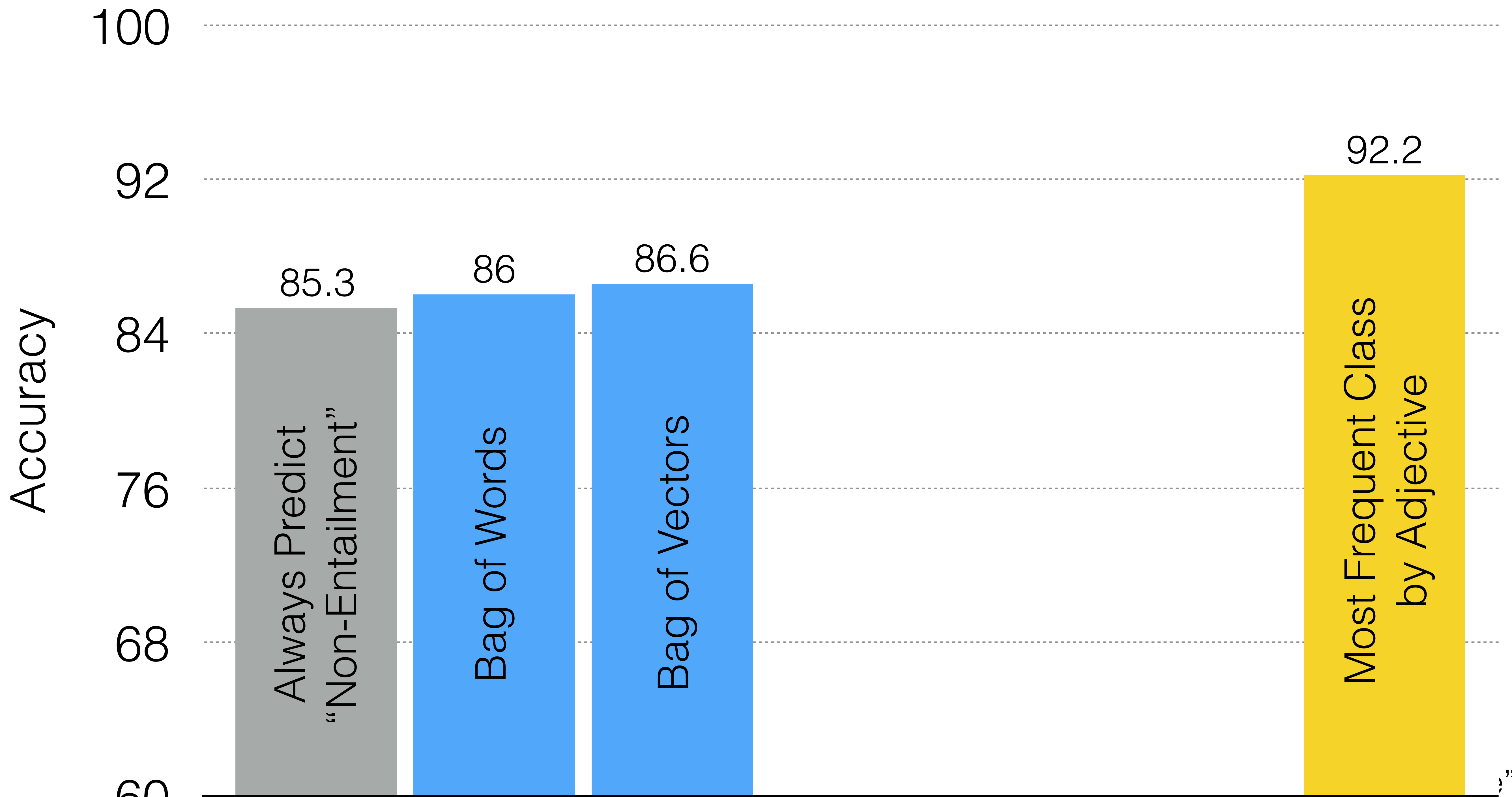
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task



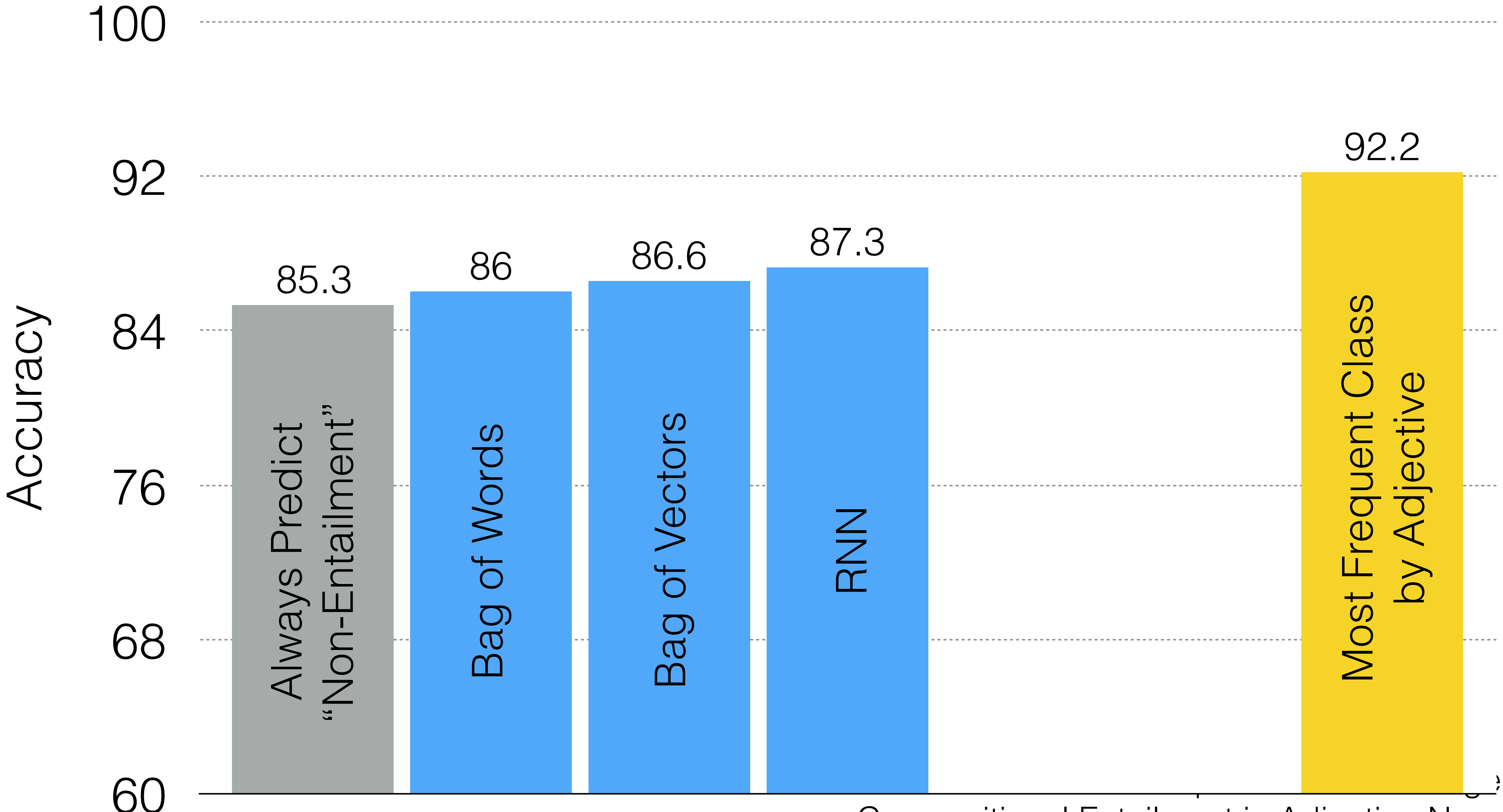
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task



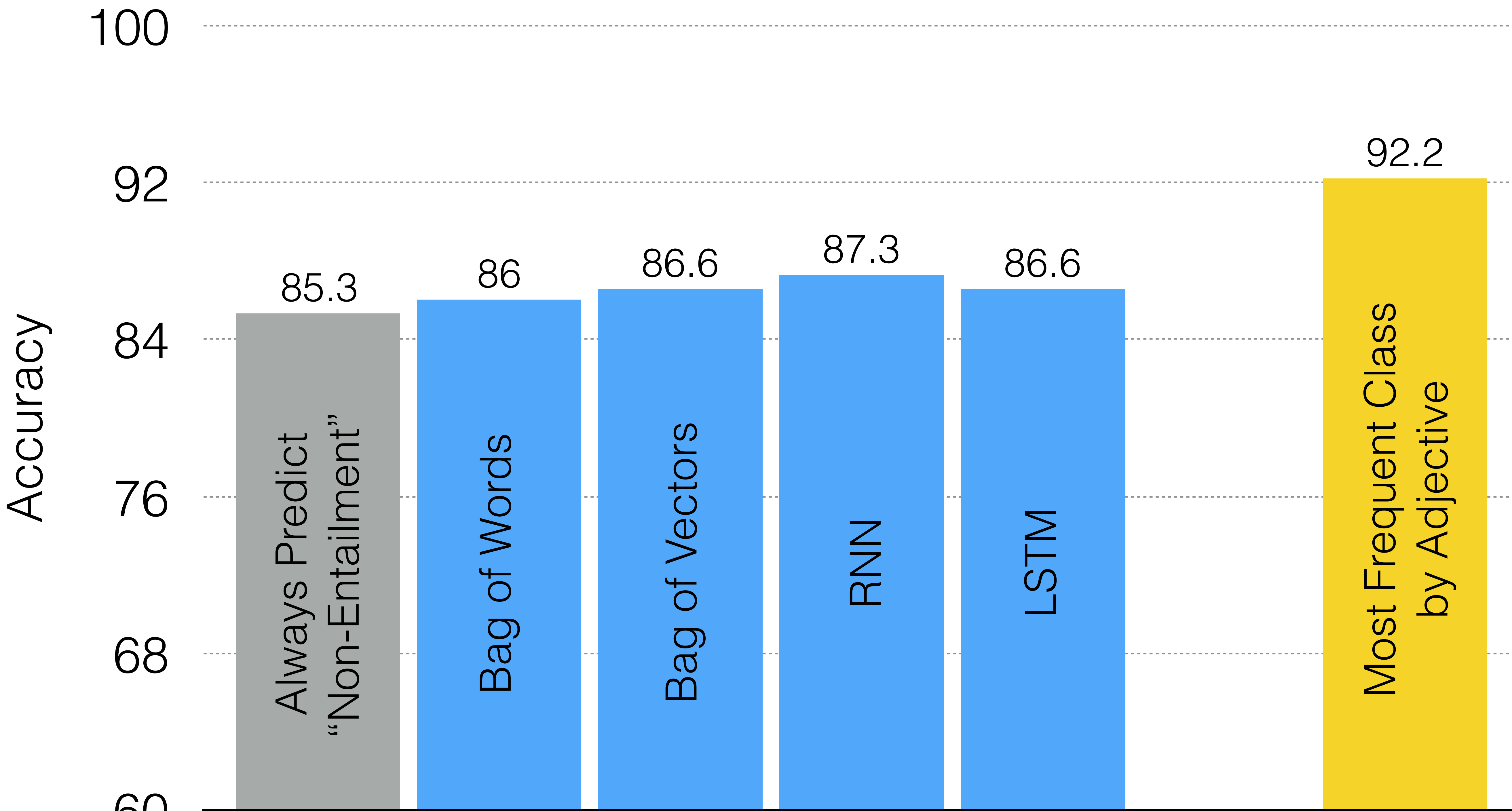
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task



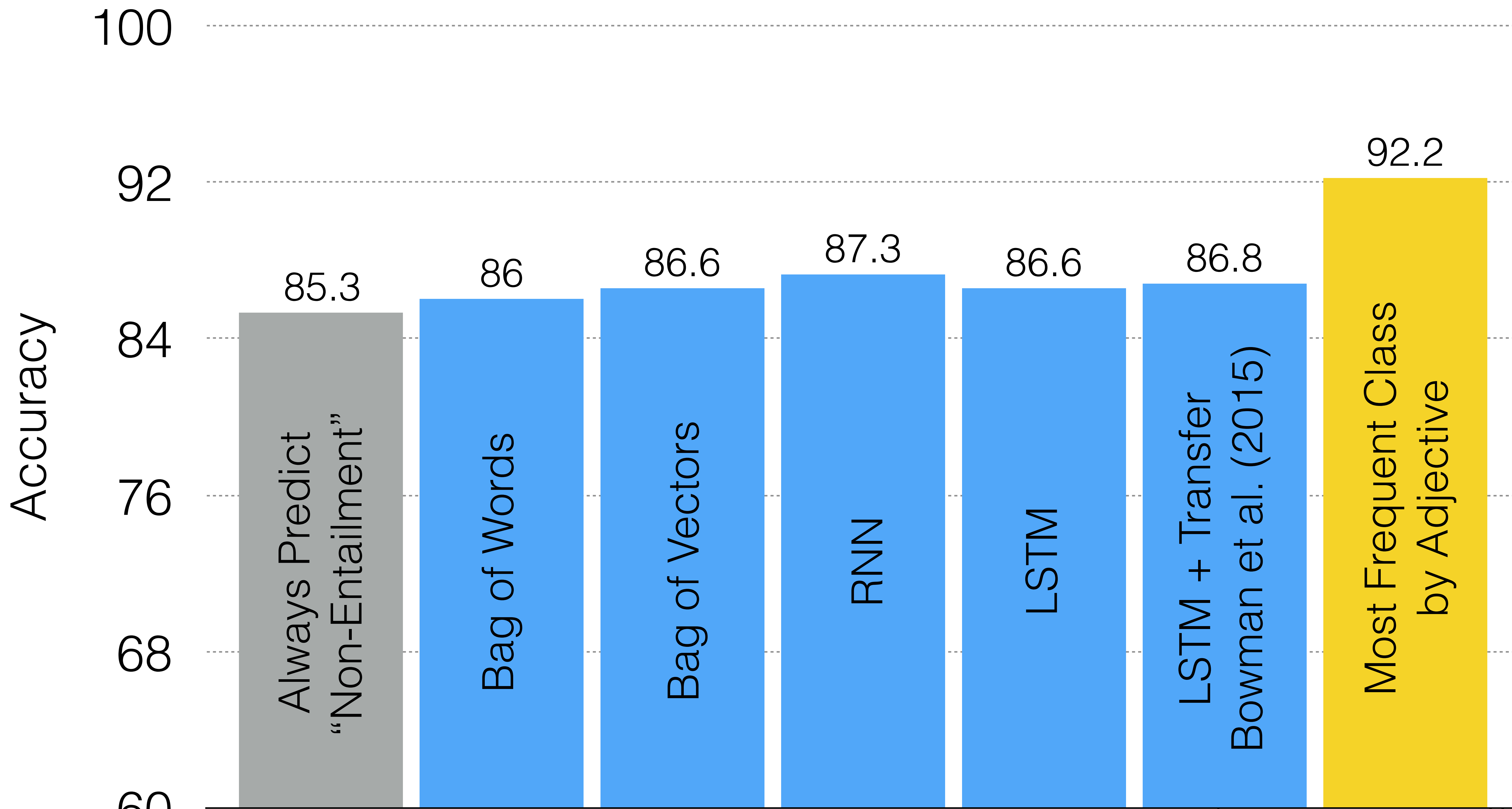
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task



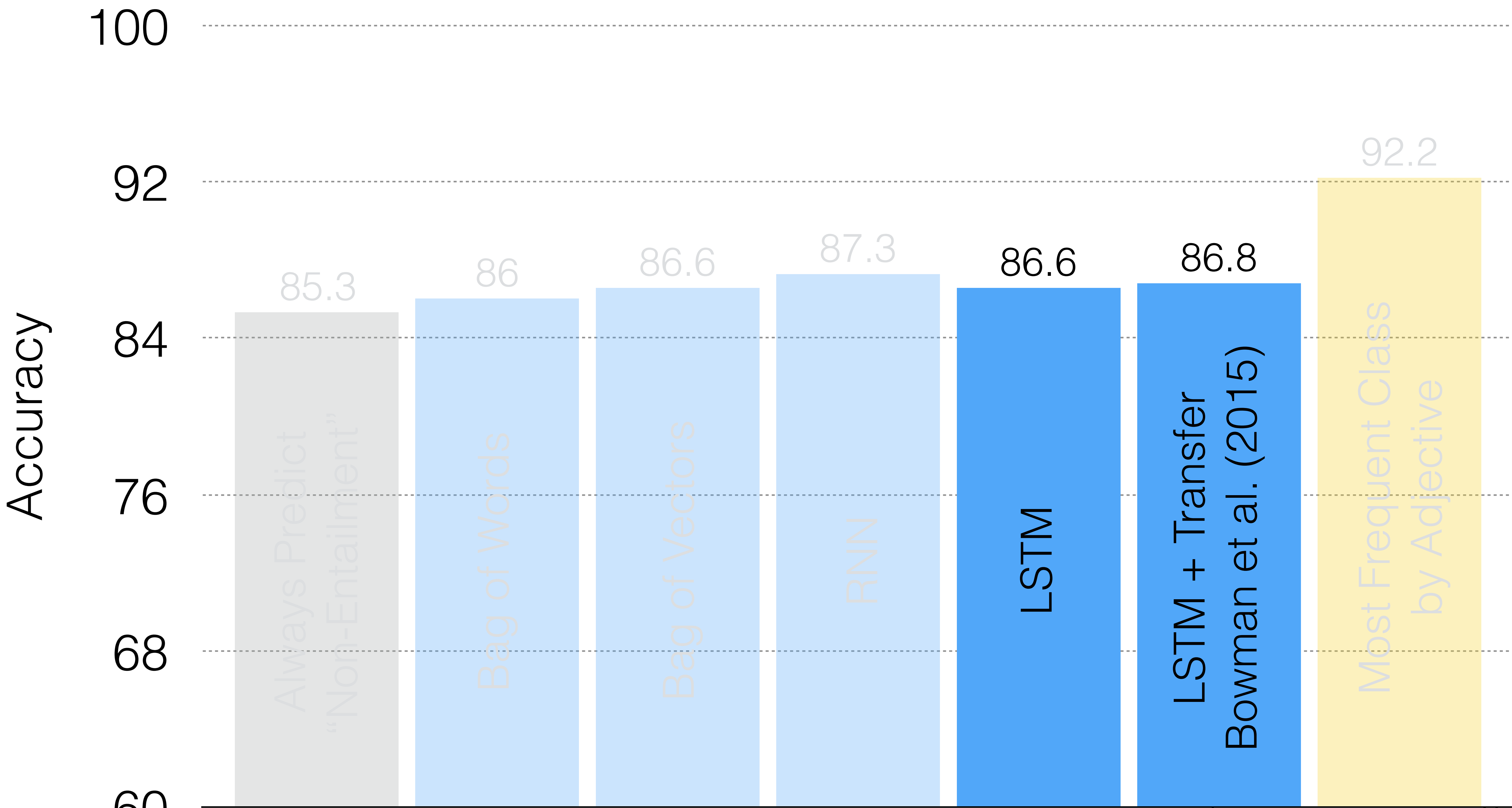
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task



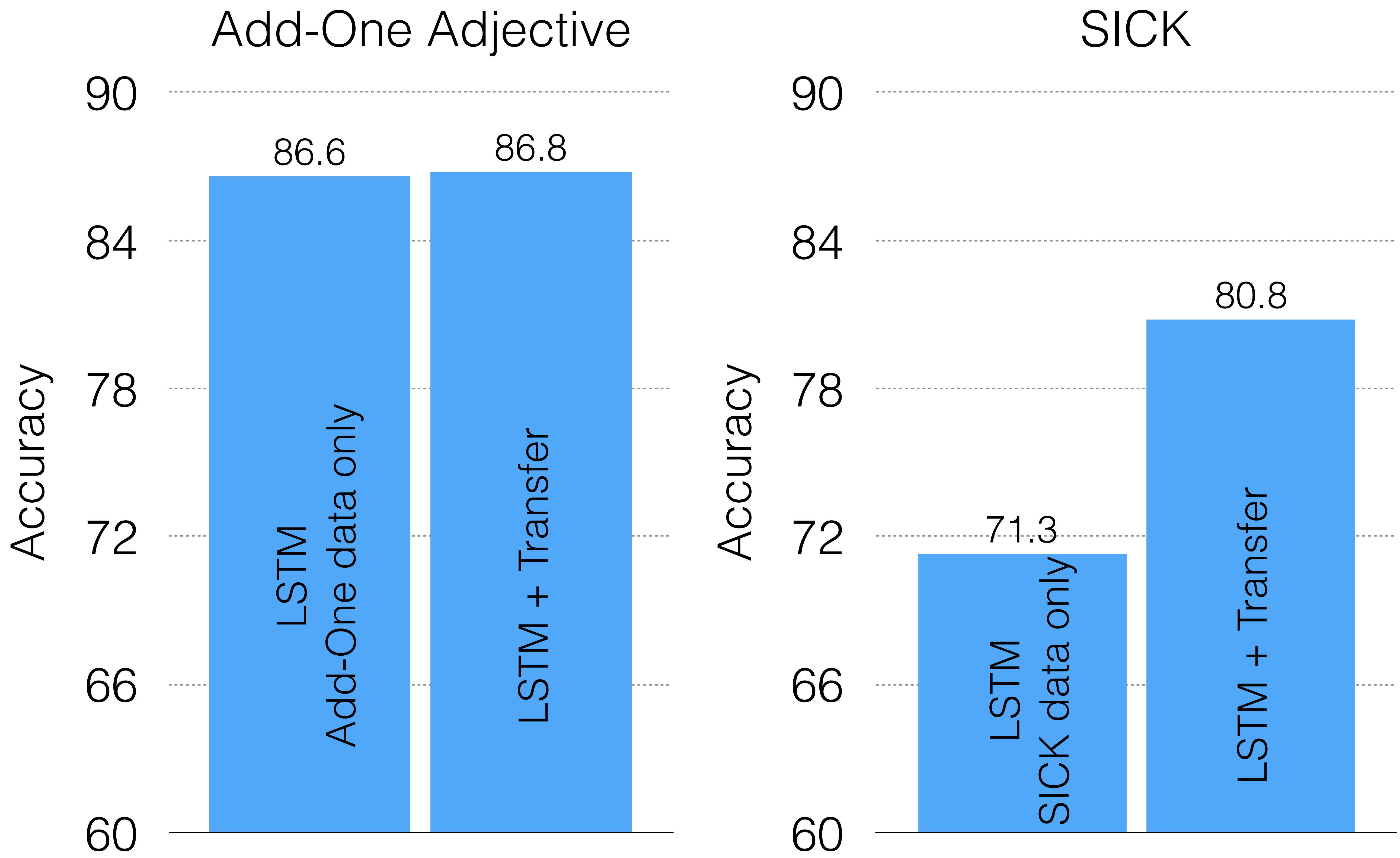
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task



Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task



Takeaways

Takeaways

- Should we care about linguistics? Yes!

Takeaways

- Should we care about linguistics? Yes!
- Because we want to learn task-independent representations of language, which requires asking and answering:

Takeaways

- Should we care about linguistics? Yes!
- Because we want to learn task-independent representations of language, which requires asking and answering:

1. What components of linguistic meaning are “intrinsic”, and what is derived in context/at “runtime”?

Takeaways

- Should we care about linguistics? Yes!
- Because we want to learn task-independent representations of language, which requires asking and answering:

1. What components of linguistic meaning are “intrinsic”, and what is derived in context/at “runtime”?
2. If these representation can’t be trained in end-to-end tasks: how to we know what is the “right” representation? Which tasks should be viewed as “fundamental” and trained/test explicitly, and which ones should come along “for free”?

Thank you!