# Not a piece of cake: 
# Lexical Composition and Implicit Information

**Vered Shwartz**

December 2019

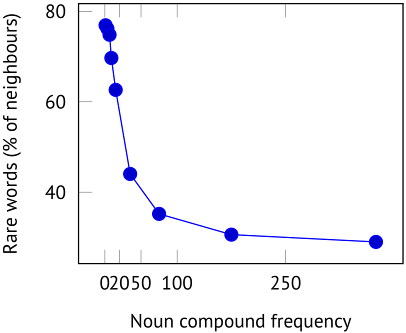
AI2

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

# How to obtain meaningful phrase representations?

# Distributional Representations

## Distributional embeddings of rare terms are of low quality




| *syndicate representative* |
| --- |
| geloios |
| t.franse |
| adopter(s |
| ahchie |
| anquish |

Similar observations for adjective-noun compositions [Boleda et al., 2013].


COMPOSE
ALL THE WORDS!
imgflip.com

Let me do it for you!

# In this talk

1. How well do contextualized embeddings represent phrases?

# In this talk

1. How well do contextualized embeddings represent phrases?
2. What is the best noun compound representation?

# In this talk

1. How well do contextualized embeddings represent phrases?
2. What is the best noun compound representation?
3. How to reveal implicit noun compound relations?

1. **How well do contextualized embeddings represent phrases?**
2. What is the best noun compound representation?
3. How to reveal implicit noun compound relations?

**Still a Pain in the Neck: Evaluating Text Representations on Lexical Composition.** Vered Shwartz and Ido Dagan. TACL 2019

# Issues with compositional representations

$$\mathbf{f}\left(\boxed{\vec{v}_{w_1}}, \boxed{\vec{v}_{w_2}}, \ldots, \boxed{\vec{v}_{w_k}}\right)$$

# Issues with compositional representations

$$\mathbf{f}\left( \boxed{\vec{v}_{w_1}}, \boxed{\vec{v}_{w_2}}, \ldots, \boxed{\vec{v}_{w_k}} \right)$$

"The whole is greater than the sum of its parts"

# Issues with compositional representations

$$\mathbf{f}\left( \vec{v}_{w_1}, \vec{v}_{w_2}, \ldots, \vec{v}_{w_k} \right)$$

"The whole is greater than the sum of its parts"

1. Meaning shift
2. Implicit meaning

# Meaning Shift

A constituent word may be used in a non-literal way

# Meaning Shift

A constituent word may be used in a non-literal way


VPC meanings differ from their verbs' meanings

# Implicit Meaning

Noun compounds

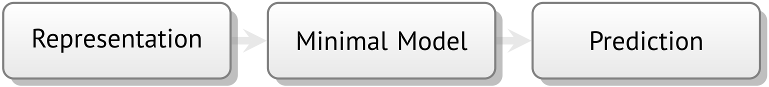
# Implicit Meaning

Adjective-noun compositions

Noun compounds

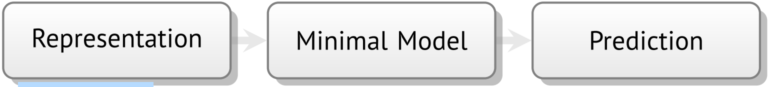

# Can existing representations address these phenomena?
## Probing Tasks

Simple tasks designed to test a single linguistic property [Adi et al., 2017, Conneau et al., 2018]



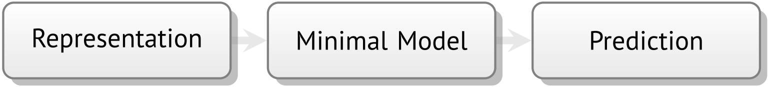
# Probing Tasks
## Representations

Standard / Contextualized




word2vec
GloVe
fastText
ELMo
GPT
BERT

# Probing Tasks
## Classifiers




1. Embed
2. Encode
3. Predict

# Classifiers

1. **Embed**: each representation
2. **Encode**: none / biLSTM / self-attention
3. **Predict**:

   $\vec{x}$ = vector of target span, additional inputs
   $\vec{o} = softmax(W \cdot ReLU(dropout(h(\vec{x}))))$

# Probing Tasks
## Tasks

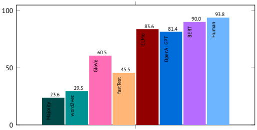
Meaning shift / Implicit meaning

Representation → Minimal Model → Prediction

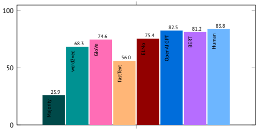
- VPC Classification
- LVC Classification
- NC Literality
- NC Relation
- AN Relation
- Phrase Type

# Tasks and Results

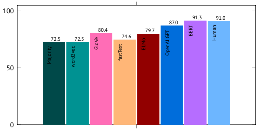
Verb-particle Classification



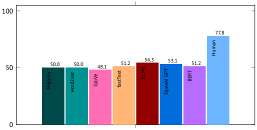

Light-Verb Construction



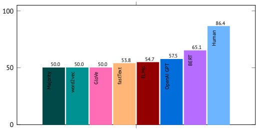

Noun Compound Literality




Noun Compound Relations




Adjective-Noun Attributes




(1) Meaning shift - human-like performance for contextualized
(2) Implicit meaning - far from humans

# Meaning Shift Tasks

# Verb-Particle Classification

## Task Definition

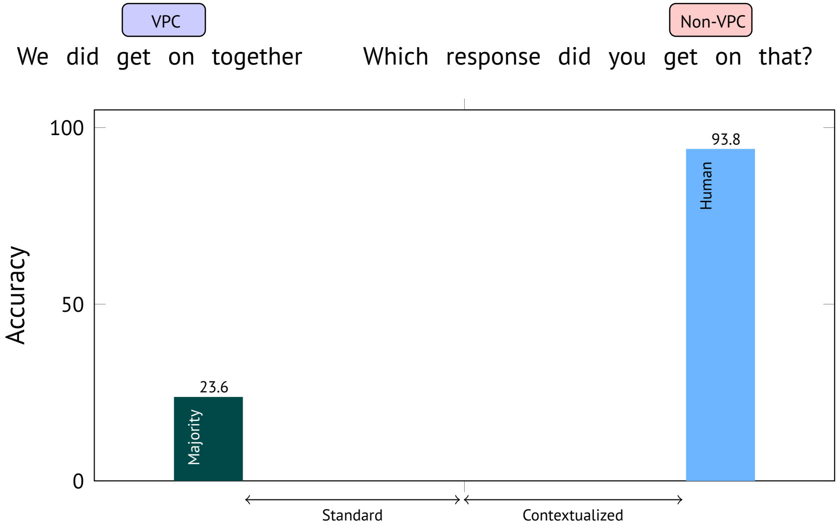
VPC

We did get on together

Non-VPC

Which response did you get on that?

# Results



VPC
Non-VPC
We did get on together
Which response did you get on that?
100
50
0
Accuracy
23.6
Majority
93.8
Human
Standard
Contextualized

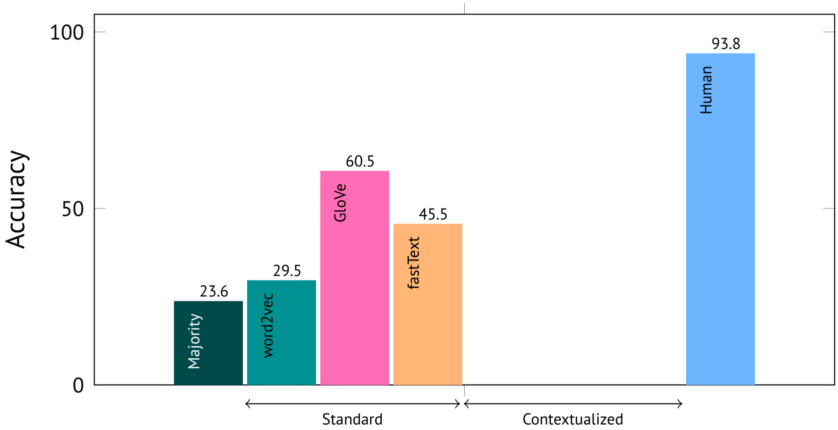
# Results

VPC

We did get on together

Non-VPC

Which response did you get on that?



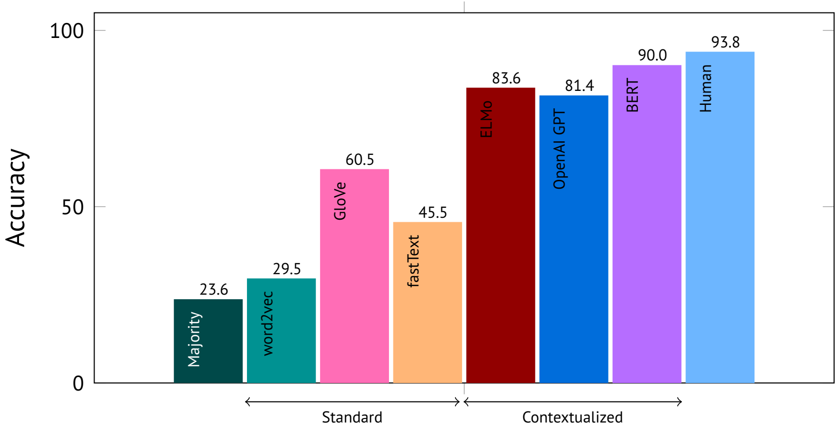
# Results

VPC

We did get on together

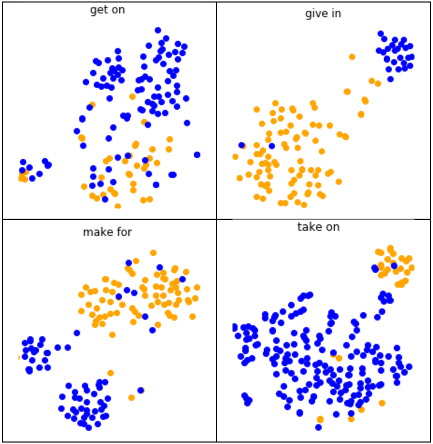
Non-VPC

Which response did you get on that?



# Verb-Particle Classification

## Analysis



# Noun Compound Literality
## Task Definition

Non-Literal Literal

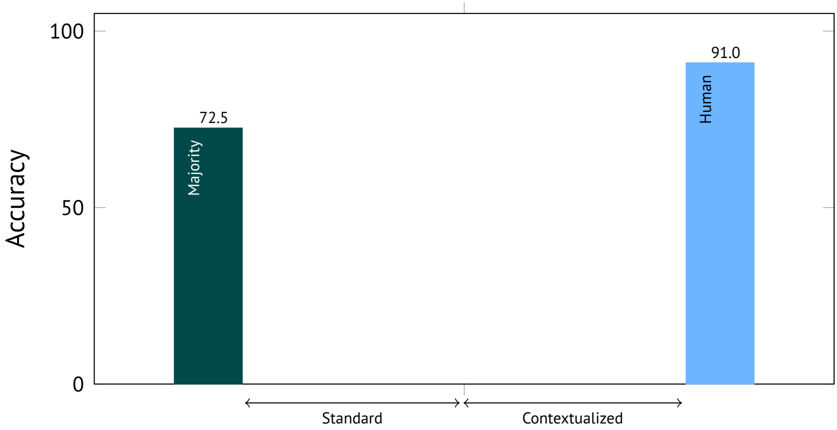
The crash course in litigation made me a better lawyer

# Noun Compound Literality

## Results

Non-Literal Literal

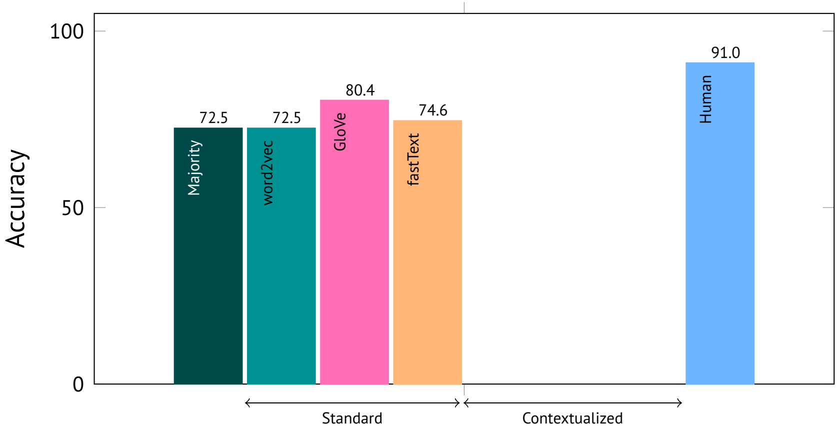
The crash course in litigation made me a better lawyer



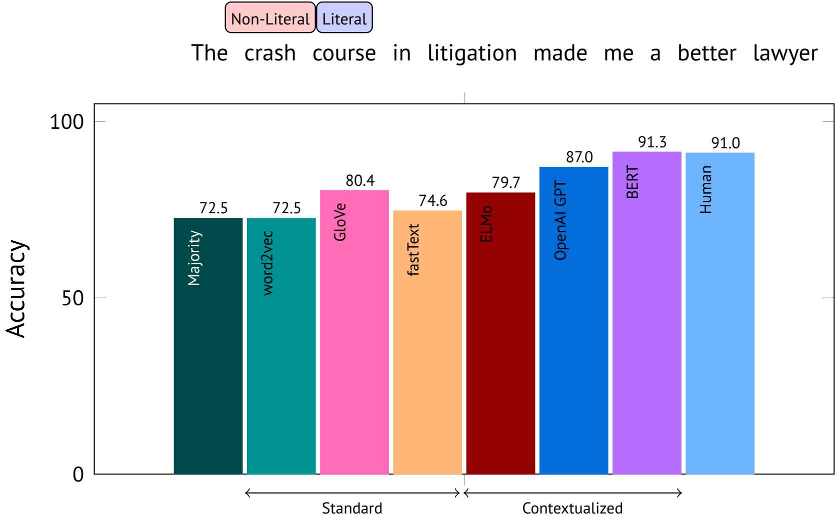
# Noun Compound Literality
## Results

Non-Literal Literal

The crash course in litigation made me a better lawyer



# Noun Compound Literality
## Results



# Noun Compound Literality

## Detecting meaning shift → modeling meaning?

| **ELMo** | **OpenAI GPT** | **BERT** |
|---|---|---|
| The Queen and her husband were on a train **trip** from Sydney to Orange. | | |
| ride | to | travelling |
| carriage | headed | running |
| journey | heading | journey |
| heading | that | going |
| carrying | and | headed |

# Noun Compound Literality

## Detecting meaning shift $\rightarrow$ modeling meaning?

| ELMo | OpenAI GPT | BERT |
|---|---|---|
| The Queen and her husband were on a train **trip** from Sydney to Orange. | | |
| ride | to | travelling |
| carriage | headed | running |
| journey | heading | journey |
| heading | that | going |
| carrying | and | headed |

| | | |
|---|---|---|
| Creating a guilt **trip** in another person may be considered to be psychological manipulation... | | |
| tolerance | that | reaction |
| fest | so | feeling |
| avoidance | trip | attachment |
| onus | he | sensation |
| association | she | note |

# Noun Compound Literality

## Non Decomposable Compounds

| ELMo | OpenAI GPT | BERT |
|---|---|---|
| ...I believe you are a snake **oil** salesman, a narcissist... | | |
| auto | in | oil |
| egg | and | pit |
| hunter | that | bite |
| rogue | charmer | jar |

Substitutes for the entire phrase.

# Implicit Meaning Tasks

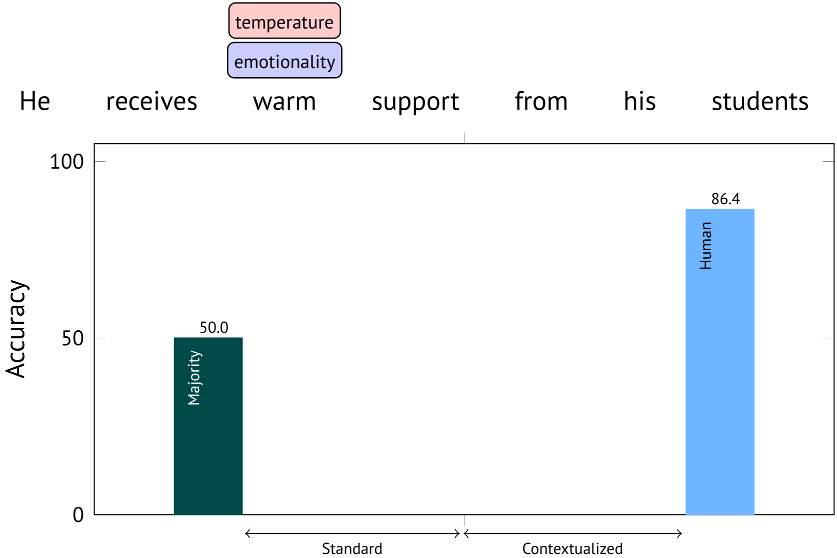
# Adjective-Noun Attributes
## Task Definition

temperature

emotionality

He receives warm support from his students

# Adjective-Noun Attributes
## Results

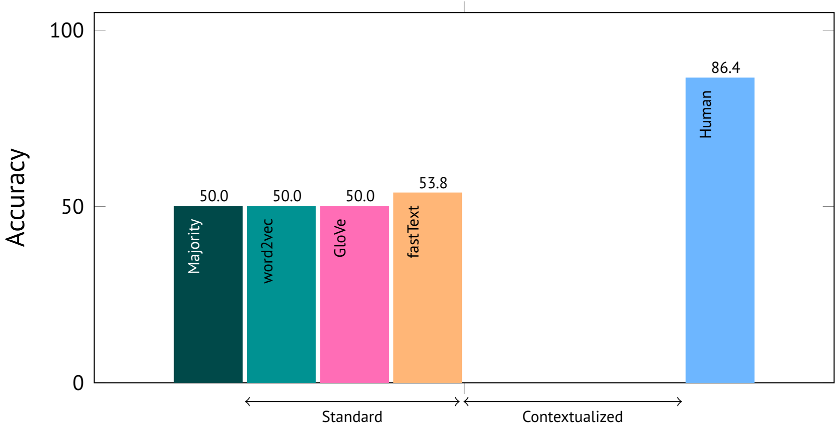


# Adjective-Noun Attributes
## Results

temperature

emotionality

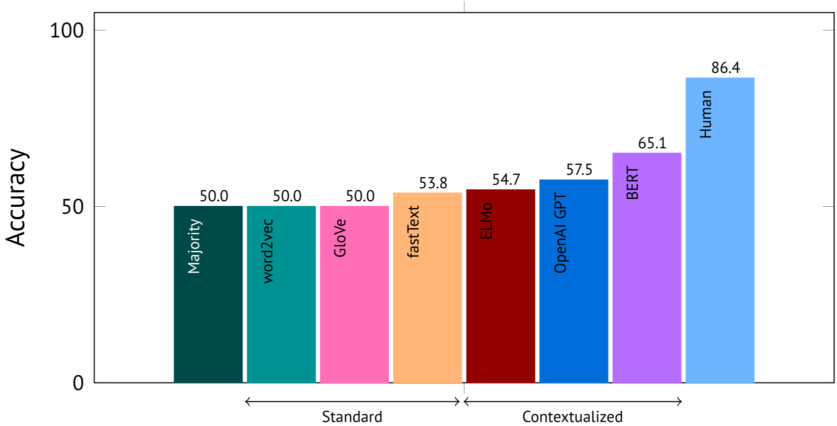
He receives warm support from his students



# Adjective-Noun Attributes
## Results

temperature
emotionality

He receives warm support from his students



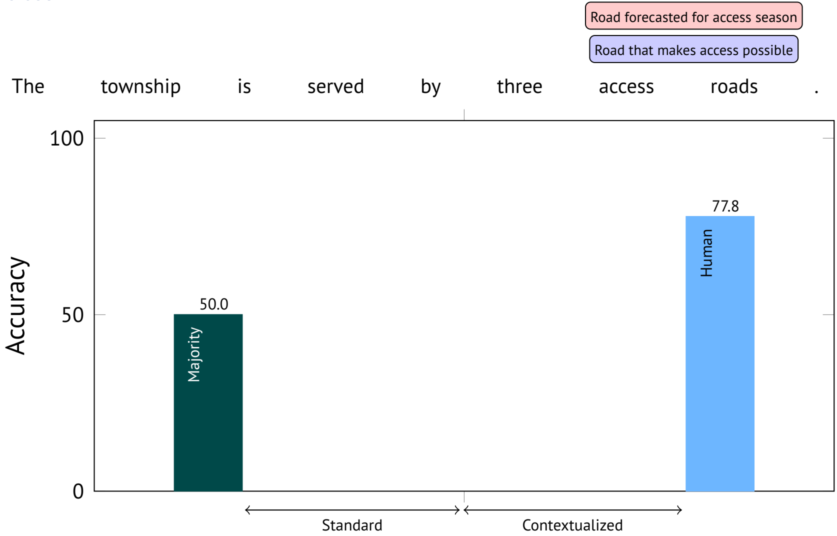
# Noun Compound Relations
## Task Definition

Road forecasted for access season

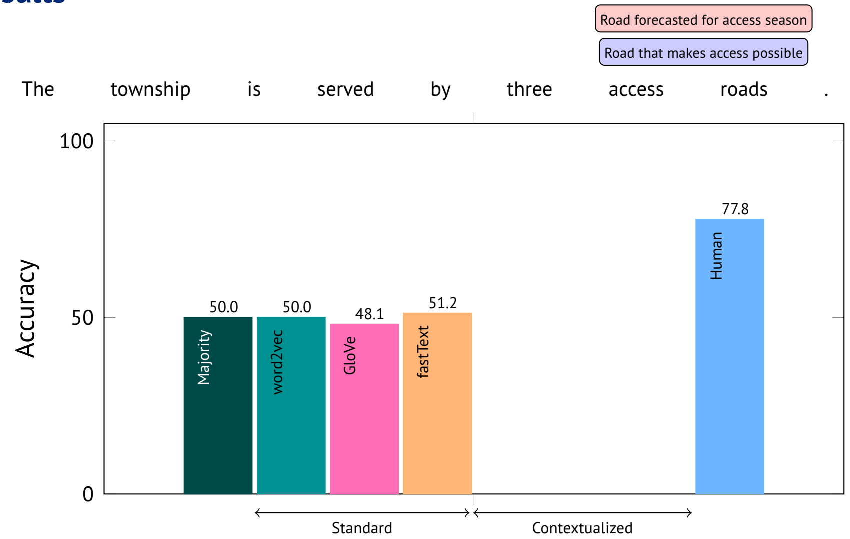
Road that makes access possible

The township is served by three access roads .

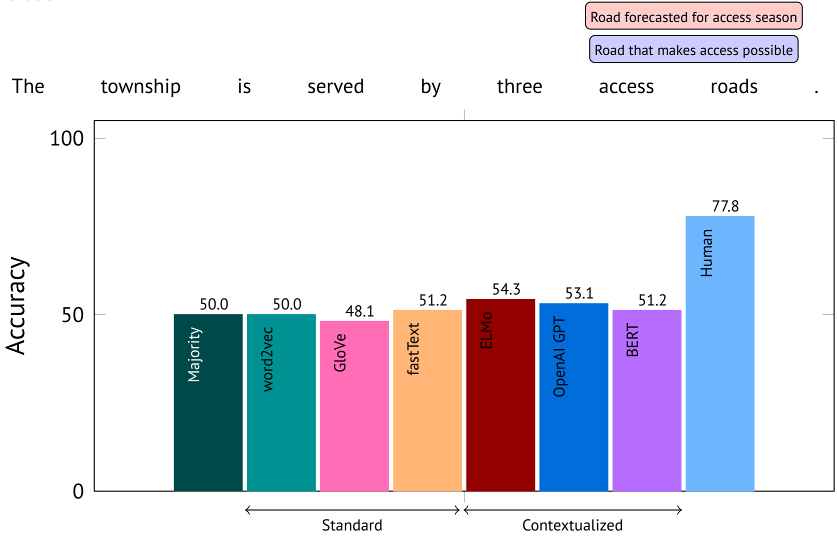
# Noun Compound Relations
## Results



# Noun Compound Relations
## Results



# Noun Compound Relations
## Results



# Still a pain in the neck
## Recap

- Detecting meaning shift is a *piece of cake*

# Still a pain in the neck
## Recap

- Detecting meaning shift is a *piece of cake*
  for contextualized word embeddings

# Still a pain in the neck
## Recap

- Detecting meaning shift is a *piece of cake*
  for contextualized word embeddings

- Modeling the shifted, rare sense is not a *walk in the park*

# Still a pain in the neck

## Recap

- Detecting meaning shift is a *piece of cake*

  for contextualized word embeddings

- Modeling the shifted, rare sense is not a *walk in the park*

- Modeling implicit information is a real *pain in the neck*

# Still a pain in the neck
## Recap

- Context matters: trivially for meaning shift but also for revealing implicit meaning

# Still a pain in the neck
## Recap

- Context matters: trivially for meaning shift but also for revealing implicit meaning
  - **Noun Compounds** [Netzer and Elhadad, 1998]: context can override frequent interpretations ("*the market bench*").

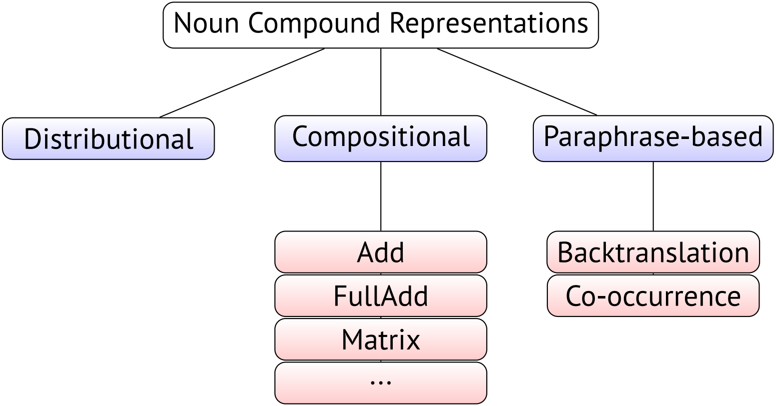
# Still a pain in the neck
## Recap

- Context matters: trivially for meaning shift but also for revealing implicit meaning
  - **Noun Compounds** [Netzer and Elhadad, 1998]: context can override frequent interpretations ("*the market bench*").
  - **Adjective Noun Compositions** [Pavlick and Callison-Burch, 2016]: depending on the context some adjectives are trivially inferred ("*little baby*") or contradicting ("*Bush travelled to Michigan to talk about the Japanese economy*").

1. How well do contextualized embeddings represent phrases?
2. **What is the best noun compound representation?**
3. How to reveal implicit noun compound relations?

**A Systematic Comparison of English Noun Compound Representations.** Vered Shwartz. MWE-WN 2019

# Approaches


Noun Compound Representations
Distributional
Compositional
Paraphrase-based
Add
FullAdd
Matrix
...
Backtranslation
Co-occurrence

# Compositional Representations

$$\mathbf{f}\,\Big(\ \vec{v}_{w_1}\ ,\ \vec{v}_{w_2}\ ,\ \ldots\ ,\ \vec{v}_{w_k}\ \Big)$$

- $f(w_1\ w_2) = \alpha \cdot v_{w_1} + \beta \cdot v_{w_2}$ [Mitchell and Lapata, 2010]
- $f(w_1\ w_2) = Av_{w_1} + Bv_{w_2}$ [Zanzotto et al., 2010, Dinu et al., 2013]
- $f(w_1\ w_2) = tanh(W \cdot [v_{w_1}; v_{w_2}])$ [Socher et al., 2012]
- ...

# Compositional Representations

**f (** $\vec{v}_{w_1}$ **,** $\vec{v}_{w_2}$ **, … ,** $\vec{v}_{w_k}$ **)**

- $f(w_1\ w_2) = \alpha \cdot v_{w_1} + \beta \cdot v_{w_2}$ [Mitchell and Lapata, 2010]
- $f(w_1\ w_2) = Av_{w_1} + Bv_{w_2}$ [Zanzotto et al., 2010, Dinu et al., 2013]
- $f(w_1\ w_2) = tanh(W \cdot [v_{w_1}; v_{w_2}])$ [Socher et al., 2012]
- ...

**Generalization at the constituent level, e.g.:**

*syndicate representative*
f(worker, representative)
f(player, representative)
f(crack, dealer)
f(company, spokesman)
f(industry, commissioner)

# Paraphrase-based Representations

$f(w_1 w_2) \approx f(paraphrase)$

- **Backtranslation**: [Wieting et al., 2015]
  *baby oil* $\rightarrow$ *huile pour bébé* $\rightarrow$ *oil for baby*
- **Co-occurrence** of the constituents, e.g. *cake made of apples*
- ...

# Paraphrase-based Representations

$f(w_1 w_2) \approx f(paraphrase)$

- **Backtranslation**: [Wieting et al., 2015]
  *baby oil* $\rightarrow$ *huile pour bébé* $\rightarrow$ *oil for baby*
- **Co-occurrence** of the constituents, e.g. *cake made of apples*
- ...

**Generalization at the constituent level, e.g.:**

*syndicate representative*

f(worker, representative)
f(union, representative)
f(group, manager)
f(employee, representative)
f(student, representative)

# What is the best representation?
## [Dima, 2016]

- **FullAdd** ($Av_{w_1} + Bv_{w_2}$) vs. **Matrix** ($tanh(W \cdot [v_{w_1}; v_{w_2}])$)

# What is the best representation?
## [Dima, 2016]

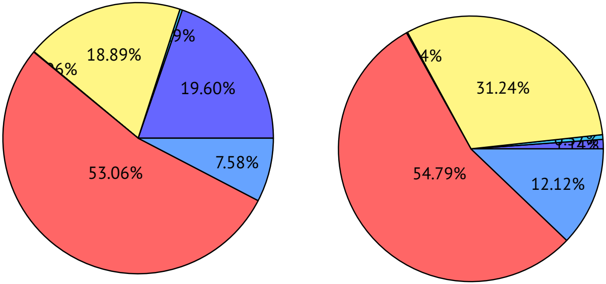
- **FullAdd** ($Av_{w_1} + Bv_{w_2}$) vs. **Matrix** ($tanh(W \cdot [v_{w_1}; v_{w_2}])$)
- Good performance is achieved even with $f(w_1, w_2) = [w_1; w_2]$
- No substantial gain from compositional representations due to lexical memorization

# Our work

| | |
|---|---|
| **Nearest Neighbours** | types of neighbours for rare/frequent compounds |
| **Attribute Prediction** | is *cheese wheel* round? |
| **Relation Classification** | what is the relationship in *baby oil*? |

# Main Takeouts
## No superior representation

- Many neighbours are either incorrect or trivial:




**Matrix** (rare) **Backtranslation** (rare)

- Rare words
- Share constituents with the target compound
- Other noun compounds
- Other words

# Main Takeouts

## No superior representation

- **Attributes**: paraphrase-based
  but with bad generalization capacity: *tomato soup* is round

# Main Takeouts
## No superior representation

- **Attributes**: paraphrase-based
  but with bad generalization capacity: *tomato soup* is round
- **Relations**: compositional + small window
  but with bad absolute performance in strict evaluation setups

# Main Takeouts

## No superior representation

- **Attributes**: paraphrase-based
  but with bad generalization capacity: *tomato soup* is round
- **Relations**: compositional + small window
  but with bad absolute performance in strict evaluation setups
- [Dima et al., 2019]: more composition functions!

1. How well do contextualized embeddings represent phrases?
2. What is the best noun compound representation?
3. **How to reveal implicit noun compound relations?**

**Olive Oil Is Made of Olives, Baby Oil Is Made for Babies: Interpreting Noun Compounds Using Paraphrases in a Neural Model.** Vered Shwartz and Chris Waterson. NAACL 2018

**Paraphrase to Explicate: Revealing Implicit Noun-Compound Relations.** Vered Shwartz and Ido Dagan. ACL 2018

# Noun Compounds

- Express implicit relationship between the constituent nouns:

# Noun Compounds

- Express implicit relationship between the constituent nouns:
  - *apple cake*: cake *made of* apples

# Noun Compounds

- Express implicit relationship between the constituent nouns:
  - *apple cake*: cake *made of* apples
  - *birthday cake*: cake *eaten on a* birthday

# Noun Compounds

- Express implicit relationship between the constituent nouns:
  - *apple cake*: cake *made of* apples
  - *birthday cake*: cake *eaten on a* birthday

- They are like “text compression devices” [Nakov, 2013]

# Noun Compounds

- Express implicit relationship between the constituent nouns:
  - *apple cake*: cake *made of* apples
  - *birthday cake*: cake *eaten on a* birthday

- They are like “text compression devices” [Nakov, 2013]
- We’re pretty good at decompressing them!

# Noun-Compound Interpretation Tasks

**Bracketing**
*[[pumpkin spice] latte]*

# Noun-Compound Interpretation Tasks

**Bracketing**
*[[pumpkin spice] latte]*

**Compositionality Prediction**
is *spelling bee* related to *bee*?

# Noun-Compound Interpretation Tasks

**Bracketing**
*[[pumpkin spice] latte]*

**Compositionality Prediction**
is *spelling bee* related to *bee*?

**Relation Classification**
*apple cake* $\rightarrow$ ingredient
*birthday cake* $\rightarrow$ time

# Noun-Compound Interpretation Tasks

**Bracketing**
*[[pumpkin spice] latte]*

**Compositionality Prediction**
is *spelling bee* related to *bee*?

**Relation Classification**
*apple cake* → ingredient
*birthday cake* → time

**Paraphrasing**
*cake* made of *apples*
*cake* eaten on a *birthday*

# Noun Compound Relation Classification

- The task is similar to **semantic relation classification**

# Noun Compound Relation Classification

- The task is similar to **semantic relation classification**
- Difference: we are interested in the relation between *olive* and *oil* in the context of the noun-compound, not in general

# Noun Compound Relation Classification

- The task is similar to **semantic relation classification**
- Difference: we are interested in the relation between *olive* and *oil* in the context of the noun-compound, not in general
- We apply lessons learned from semantic relation classification to noun-compound interpretation:

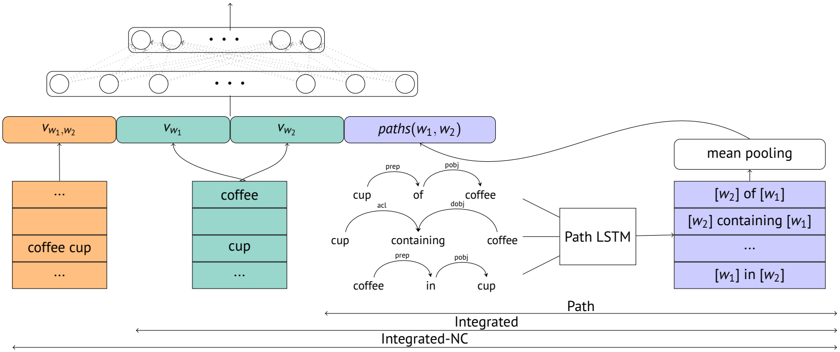
# Noun Compound Relation Classification

- The task is similar to **semantic relation classification**
- Difference: we are interested in the relation between *olive* and *oil* in the context of the noun-compound, not in general
- We apply lessons learned from semantic relation classification to noun-compound interpretation:
  - Represent NCs using their joint non-NC corpus occurrences features [Shwartz et al., 2016]

# Noun Compound Relation Classification

- The task is similar to **semantic relation classification**
- Difference: we are interested in the relation between *olive* and *oil* in the context of the noun-compound, not in general
- We apply lessons learned from semantic relation classification to noun-compound interpretation:
  - Represent NCs using their joint non-NC corpus occurrences features [Shwartz et al., 2016]
  - Split the dataset lexically

# Overall Architecture



# Evaluation - Datasets

- Dataset: [Tratz, 2011]

| | | |
|---|---|---|
| **Purpose/Activity Group** | | |
| PERFORM&ENGAGE_IN | 11.5% | cooking pot |
| CREATE-PROVIDE-GENERATE-SELL | 4.8% | nicotine patch |
| OBTAIN&ACCESS&SEEK | 0.9% | shrimp boat |
| MITIGATE&OPPOSE | 0.8% | flak jacket |
| ORGANIZE&SUPERVISE&AUTHORITY | 1.6% | ethics authority |
| PURPOSE | 1.9% | chicken spit |
| **Ownership, Experience, Employment, Use** | | |
| OWNER-USER | 2.1% | family estate |
| EXPERIENCER-OF-EXPERIENCE | 0.5% | family greed |
| EMPLOYER | 2.3% | team doctor |
| USER_RECIPIENT | 1.0% | voter pamphlet |
| **Temporal Group** | | |
| TIME-OF1 | 2.2% | night work |
| TIME-OF2 | 0.5% | birth date |
| **Location and Whole+Part/Member of** | | |
| LOCATION | 5.2% | hillside home |
| WHOLE+PART_OR_MEMBER_OF | 1.7% | robot arm |

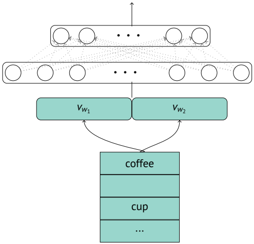
# Evaluation - Datasets

- Dataset: [Tratz, 2011]

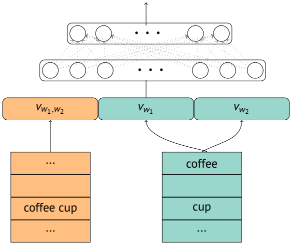
| | | |
|---|---|---|
| **Purpose/Activity Group** | | |
| PERFORM&ENGAGE_IN | 11.5% | cooking pot |
| CREATE-PROVIDE-GENERATE-SELL | 4.8% | nicotine patch |
| OBTAIN&ACCESS&SEEK | 0.9% | shrimp boat |
| MITIGATE&OPPOSE | 0.8% | flak jacket |
| ORGANIZE&SUPERVISE&AUTHORITY | 1.6% | ethics authority |
| PURPOSE | 1.9% | chicken spit |
| **Ownership, Experience, Employment, Use** | | |
| OWNER-USER | 2.1% | family estate |
| EXPERIENCER-OF-EXPERIENCE | 0.5% | family greed |
| EMPLOYER | 2.3% | team doctor |
| USER_RECIPIENT | 1.0% | voter pamphlet |
| **Temporal Group** | | |
| TIME-OF1 | 2.2% | night work |
| TIME-OF2 | 0.5% | birth date |
| **Location and Whole+Part/Member of** | | |
| LOCATION | 5.2% | hillside home |
| WHOLE+PART_OR_MEMBER_OF | 1.7% | robot arm |

- Datatset splits:
  - Random 75:20:5 (like previous work)
  - Lexical-full [Levy et al., 2015]
  - Lexical-head
  - Lexical-mod

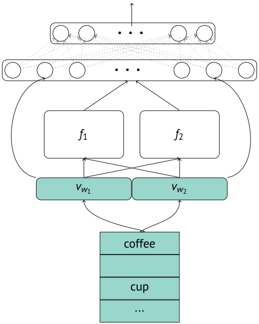
# Evaluation - Baselines




Dist




Dist-NC




Compositional [Dima, 2016]

# Evaluation - Results

| Dataset | Split | Best Baseline | Path | Int | Int-NC |
|---|---|---|---|---|---|
| Tratz-fine | Rand | **0.725** | 0.538 | 0.714 | 0.692 |
| | $Lex_{head}$ | 0.458 | 0.448 | **0.510** | 0.478 |
| | $Lex_{mod}$ | 0.607 | 0.472 | **0.613** | 0.600 |
| | $Lex_{full}$ | 0.363 | 0.423 | 0.421 | **0.429** |
| Tratz-coarse | Rand | **0.775** | 0.586 | 0.736 | 0.712 |
| | $Lex_{head}$ | 0.538 | 0.518 | **0.569** | 0.548 |
| | $Lex_{mod}$ | 0.645 | 0.548 | **0.646** | 0.632 |
| | $Lex_{full}$ | 0.409 | 0.472 | **0.475** | 0.478 |

- Random split: distributional/compositional baselines outperform all other methods, by memorizing words.

# Evaluation - Results

| Dataset | Split | Best Baseline | Path | Int | Int-NC |
|---|---|---|---|---|---|
| `Tratz-fine` | Rand | **0.725** | 0.538 | 0.714 | 0.692 |
| | $\text{Lex}_{\textit{head}}$ | 0.458 | 0.448 | **0.510** | 0.478 |
| | $\text{Lex}_{\textit{mod}}$ | 0.607 | 0.472 | **0.613** | 0.600 |
| | $\text{Lex}_{\textit{full}}$ | 0.363 | 0.423 | 0.421 | **0.429** |
| `Tratz-coarse` | Rand | **0.775** | 0.586 | 0.736 | 0.712 |
| | $\text{Lex}_{\textit{head}}$ | 0.538 | 0.518 | **0.569** | 0.548 |
| | $\text{Lex}_{\textit{mod}}$ | 0.645 | 0.548 | **0.646** | 0.632 |
| | $\text{Lex}_{\textit{full}}$ | 0.409 | 0.472 | **0.475** | 0.478 |

- Lexical split: our methods perform better.

# Evaluation - Results

| Dataset | Split | Best Baseline | Path | Int | Int-NC |
|---|---|---|---|---|---|
| Tratz-fine | Rand | **0.725** | 0.538 | 0.714 | 0.692 |
| | $Lex_{head}$ | 0.458 | 0.448 | **0.510** | 0.478 |
| | $Lex_{mod}$ | 0.607 | 0.472 | **0.613** | 0.600 |
| | $Lex_{full}$ | 0.363 | 0.423 | 0.421 | **0.429** |
| Tratz-coarse | Rand | **0.775** | 0.586 | 0.736 | 0.712 |
| | $Lex_{head}$ | 0.538 | 0.518 | **0.569** | 0.548 |
| | $Lex_{mod}$ | 0.645 | 0.548 | **0.646** | 0.632 |
| | $Lex_{full}$ | 0.409 | 0.472 | **0.475** | 0.478 |

- The performance gap is larger in lexical-full.

# Analysis

## Which relations can the path-based model learn?

| relation | path | examples |
|---|---|---|
| measure | [$w_2$] varies by [$w_1$] | *state limit* |
| | 2,560 [$w_1$] portion of [$w_2$] | *acre estate* |
| personal title | [$w_2$] Anderson [$w_1$]/title | *Mrs. Brown* |
| | [$w_2$] Sheridan [$w_1$]/title | *Gen. Johnson* |
| create-provide-generate-sell | [$w_2$] produce [$w_1$] | *food producer* |
| | [$w_2$] manufacture [$w_1$] | *engine plant* |
| time-of1 | [$w_2$] begin [$w_1$] | *morning program* |
| | [$w_2$] held Saturday [$w_1$] | *afternoon meeting* |
| substance-material - ingredient | [$w_2$] made of wood and [$w_1$] | *marble table* |
| | [$w_2$] material includes type of [$w_1$] | *steel pipe* |

# Analysis

## Which relations CAN'T the path-based model learn?

- lexicalized has no indicative paths! (e.g. *soap opera*)

# Analysis

## Which relations CAN'T the path-based model learn?

- lexicalized has no indicative paths! (e.g. *soap opera*)
- partial_attribute_transfer (e.g. *bullet train*) has few indicative paths (e.g. "*train as fast as a bullet*")

# Noun Compound Relation Classification

## Recap

- Joint corpus occurrences improve the performance in strict evaluation setups $\vee$

# Noun Compound Relation Classification
## Recap

- Joint corpus occurrences improve the performance in strict evaluation setups ✓
- Assumes compositionality ×

# Noun Compound Relation Classification

## Recap

- Joint corpus occurrences improve the performance in strict evaluation setups $\checkmark$
- Assumes compositionality $\times$

- Lexical splits help prevent lexical memorization $\checkmark$

# Noun Compound Relation Classification
## Recap

- Joint corpus occurrences improve the performance in strict evaluation setups V
- Assumes compositionality ×

- Lexical splits help prevent lexical memorization V
- The dataset is noisy, it's difficult to label each NC to a single relationship ×

# Noun-Compound Interpretation Tasks

**Bracketing**
*[[pumpkin spice] latte]*

**Compositionality Prediction**
is *spelling bee* related to *bee*?

**Relation Classification**
*apple cake* → ingredient
*birthday cake* → time

**Paraphrasing**
*cake* made of *apples*
*cake* eaten on a *birthday*

# We are good at Interpreting Noun-Compounds

- We easily interpret noun-compounds
    - Even when we see them for the first time

# We are good at Interpreting Noun-Compounds

- We easily interpret noun-compounds
  - Even when we see them for the first time
- What is a "*parsley cake*"?

# We are good at Interpreting Noun-Compounds

- We easily interpret noun-compounds
  - Even when we see them for the first time

- What is a "*parsley cake*"?
  - cake *eaten on a* parsley?
  - cake *with* parsley?
  - cake *for* parsley?
  - …

# We are good at Interpreting Noun-Compounds

- We easily interpret noun-compounds
  - Even when we see them for the first time
- What is a "*parsley cake*"?
  - cake *eaten on a* parsley?
  - $\boxed{\text{cake } \textit{with} \text{ parsley?}}$
  - cake *for* parsley?
  - ...

# Generalizing Existing Knowledge

- What can cake be made of?

Corpus of Contemporary American English

SEARCH | FREQUENCY | CONTEXT | HELP

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...]    COMPARE

| | | CONTEXT ALL FORMS (SAMPLE): 100 200 500 | FREQ | TOTAL 237 \| UNIQUE 119 + |
|---|---|---|---|---|
| 1 | | CAKE WITH CHOCOLATE | 31 | |
| 2 | | CAKE WITH LEMON | 13 | |
| 3 | | CAKE WITH STRAWBERRIES | 10 | |
| 4 | | CAKE WITH CANDLES | 7 | |
| 5 | | CAKE WITH CARAMEL | 7 | |
| 6 | | CAKE WITH FROSTING | 6 | |
| 7 | | CAKE WITH VANILLA | 6 | |
| 8 | | CAKE WITH BERRIES | 5 | |
| 9 | | CAKE WITH EGGS | 4 | |
| 10 | | CAKE WITH TOWEL | 4 | |
| 11 | | CAKE WITH RASPBERRY | 3 | |
| 12 | | CAKE WITH ICE | 3 | |
| 13 | | CAKE WITH MARSHMALLOW | 3 | |
| 14 | | CAKE WITH HONEY | 3 | |
| 15 | | CAKE WITH CINNAMON | 3 | |
| 16 | | CAKE WITH COFFEE | 3 | |
| 17 | | CAKE WITH BUTTER | 3 | |
| 18 | | CAKE WITH YOGURT | 3 | |
| 19 | | CAKE WITH ALMOND | 2 | |
| 20 | | CAKE WITH BLUEBERRIES | 2 | |
| 21 | | CAKE WITH COCONUT | 2 | |
| 22 | | CAKE WITH CITRUS | 2 | |
| 23 | | CAKE WITH BUTTERCREAM | 2 | |
| 24 | | CAKE WITH CREME | 2 | |
| 25 | | CAKE WITH CREAM | 2 | |
| 26 | | CAKE WITH DULCE | 2 | |
| 27 | | CAKE WITH CUSTARD | 2 | |
| 28 | | CAKE WITH FRUIT | 2 | |
| 29 | | CAKE WITH CONFECTIONERS | 2 | |
| 30 | | CAKE WITH ORANGE | 2 | |

# Generalizing Existing Knowledge

- What can cake be made of?

Corpus of Contemporary American English

SEARCH | FREQUENCY | CONTEXT | HELP

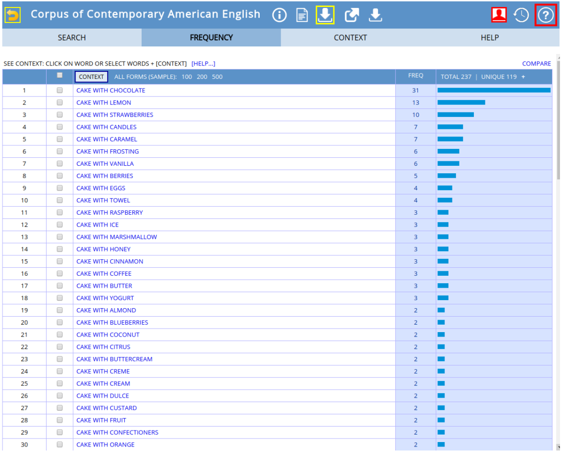
SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...] COMPARE

| | | CONTEXT ALL FORMS (SAMPLE): 100 200 500 | FREQ | TOTAL 237 \| UNIQUE 119 + |
|---|---|---|---|---|
| 1 | | CAKE WITH CHOCOLATE | 31 | |
| 2 | | CAKE WITH LEMON | 13 | |
| 3 | | CAKE WITH STRAWBERRIES | 10 | |
| 4 | | CAKE WITH CANDLES | 7 | |
| 5 | | CAKE WITH CARAMEL | 7 | |
| 6 | | CAKE WITH FROSTING | 6 | |
| 7 | | CAKE WITH VANILLA | 6 | |
| 8 | | CAKE WITH BERRIES | 5 | |
| 9 | | CAKE WITH EGGS | 4 | |
| 10 | | CAKE WITH TOWEL | 4 | |
| 11 | | CAKE WITH RASPBERRY | 3 | |
| 12 | | CAKE WITH ICE | 3 | |
| 13 | | CAKE WITH MARSHMALLOW | 3 | |
| 14 | | CAKE WITH HONEY | 3 | |
| 15 | | CAKE WITH CINNAMON | 3 | |
| 16 | | CAKE WITH COFFEE | 3 | |
| 17 | | CAKE WITH BUTTER | 3 | |
| 18 | | CAKE WITH YOGURT | 3 | |
| 19 | | CAKE WITH ALMOND | 2 | |
| 20 | | CAKE WITH BLUEBERRIES | 2 | |
| 21 | | CAKE WITH COCONUT | 2 | |
| 22 | | CAKE WITH CITRUS | 2 | |
| 23 | | CAKE WITH BUTTERCREAM | 2 | |
| 24 | | CAKE WITH CREME | 2 | |
| 25 | | CAKE WITH CREAM | 2 | |
| 26 | | CAKE WITH DULCE | 2 | |
| 27 | | CAKE WITH CUSTARD | 2 | |
| 28 | | CAKE WITH FRUIT | 2 | |
| 29 | | CAKE WITH CONFECTIONERS | 2 | |
| 30 | | CAKE WITH ORANGE | 2 | |

- Parsley (sort of) fits into this distribution

# Generalizing Existing Knowledge

- What can cake be made of?



Corpus of Contemporary American English

SEARCH | FREQUENCY | CONTEXT | HELP

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...] COMPARE

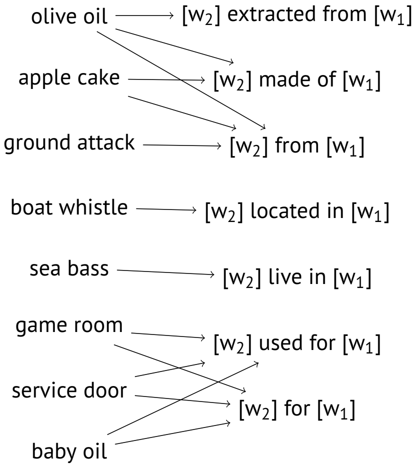
| | | CONTEXT ALL FORMS (SAMPLE): 100 200 500 | FREQ | TOTAL 237 \| UNIQUE 119 + |
|---|---|---|---|---|
| 1 | | CAKE WITH CHOCOLATE | 31 | |
| 2 | | CAKE WITH LEMON | 13 | |
| 3 | | CAKE WITH STRAWBERRIES | 10 | |
| 4 | | CAKE WITH CANDLES | 7 | |
| 5 | | CAKE WITH CARAMEL | 7 | |
| 6 | | CAKE WITH FROSTING | 6 | |
| 7 | | CAKE WITH VANILLA | 6 | |
| 8 | | CAKE WITH BERRIES | 5 | |
| 9 | | CAKE WITH EGGS | 4 | |
| 10 | | CAKE WITH TOWEL | 4 | |
| 11 | | CAKE WITH RASPBERRY | 3 | |
| 12 | | CAKE WITH ICE | 3 | |
| 13 | | CAKE WITH MARSHMALLOW | 3 | |
| 14 | | CAKE WITH HONEY | 3 | |
| 15 | | CAKE WITH CINNAMON | 3 | |
| 16 | | CAKE WITH COFFEE | 3 | |
| 17 | | CAKE WITH BUTTER | 3 | |
| 18 | | CAKE WITH YOGURT | 3 | |
| 19 | | CAKE WITH ALMOND | 2 | |
| 20 | | CAKE WITH BLUEBERRIES | 2 | |
| 21 | | CAKE WITH COCONUT | 2 | |
| 22 | | CAKE WITH CITRUS | 2 | |
| 23 | | CAKE WITH BUTTERCREAM | 2 | |
| 24 | | CAKE WITH CREME | 2 | |
| 25 | | CAKE WITH CREAM | 2 | |
| 26 | | CAKE WITH DULCE | 2 | |
| 27 | | CAKE WITH CUSTARD | 2 | |
| 28 | | CAKE WITH FRUIT | 2 | |
| 29 | | CAKE WITH CONFECTIONERS | 2 | |
| 30 | | CAKE WITH ORANGE | 2 | |

- Parsley (sort of) fits into this distribution
- Similar to “selectional preferences” [Pantel et al., 2007]

# Noun-Compound Paraphrasing

# Motivation

Given a noun-compound $w_1 w_2$, express the relation between the head $w_2$ and the modifier $w_1$ with multiple prepositional and verbal paraphrases [Nakov and Hearst, 2006]



# Evaluation Setting

- Available dataset: SemEval 2013 task 4 [Hendrickx et al., 2013]

## Evaluation Setting

- Available dataset: SemEval 2013 task 4 [Hendrickx et al., 2013]
- A *ranking* rather than a *retrieval* task
  - Systems get a list of noun compounds

# Evaluation Setting

- Available dataset: SemEval 2013 task 4 [Hendrickx et al., 2013]
- A *ranking* rather than a *retrieval* task
  - Systems get a list of noun compounds
  - Extract paraphrases from free text

# Evaluation Setting

- Available dataset: SemEval 2013 task 4 [Hendrickx et al., 2013]
- A *ranking* rather than a *retrieval* task
  - Systems get a list of noun compounds
  - Extract paraphrases from free text
  - Rank them

# Evaluation Setting

- Available dataset: SemEval 2013 task 4 [Hendrickx et al., 2013]
- A *ranking* rather than a *retrieval* task
  - Systems get a list of noun compounds
  - Extract paraphrases from free text
  - Rank them
- Evaluated for correlation with human judgments
  - Gold paraphrase score: how many annotators suggested it?

# Prior Methods

- Based on constituent co-occurrences: "*cake* made of *apple*"

# Prior Methods

- Based on constituent co-occurrences: "*cake* made of *apple*"

- **Problems:**
  1. Many unseen compounds, no paraphrases in the corpus
     - rare: *parsley cake* or highly lexicalized: *ice cream*

# Prior Methods

- Based on constituent co-occurrences: “*cake* made of *apple*”

- **Problems:**
  1. Many unseen compounds, no paraphrases in the corpus
     - rare: *parsley cake* or highly lexicalized: *ice cream*

  2. Many compounds with just a few paraphrases
     - Can we infer “*cake* containing *apple*” given “*cake* made of *apple*”?

# Prior Methods

- Based on constituent co-occurrences: "*cake* made of *apple*"
- **Problems:**
  1. Many unseen compounds, no paraphrases in the corpus
     - rare: *parsley cake* or highly lexicalized: *ice cream*
  2. Many compounds with just a few paraphrases
     - Can we infer "*cake* containing *apple*" given "*cake* made of *apple*"?
- Prior work provides partial solutions to either (1) or (2)

# Model

# Multi-task Reformulation

- Training example {$w_1$ = apple, $w_2$ = cake, $p$ = "[$w_2$] made of [$w_1$]"}

# Multi-task Reformulation

- Training example {$w_1$ = apple, $w_2$ = cake, $p$ = "[$w_2$] made of [$w_1$]"}

  1. Predict a paraphrase $p$ for a given NC $w_1 w_2$:
     What is the relation between *apple* and *cake*?

# Multi-task Reformulation

- Training example {$w_1$ = apple, $w_2$ = cake, $p$ = “[$w_2$] made of [$w_1$]”}

  1. Predict a paraphrase $p$ for a given NC $w_1 w_2$:
     What is the relation between *apple* and *cake*?

  2. Predict $w_1$ given a paraphrase $p$ and $w_2$:
     What can *cake* be made of?

# Multi-task Reformulation

- Training example {$w_1$ = apple, $w_2$ = cake, $p$ = "[w$_2$] made of [w$_1$]"}

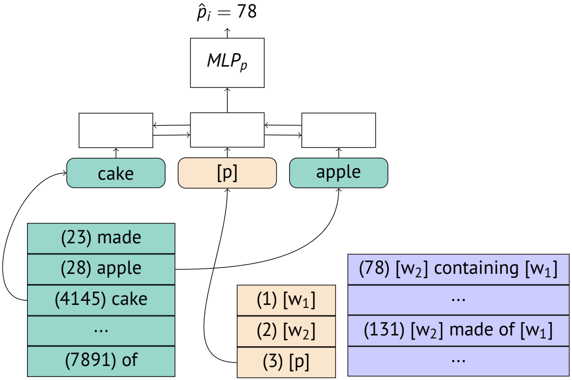
1. Predict a paraphrase $p$ for a given NC $w_1 w_2$:
   What is the relation between *apple* and *cake*?

2. Predict $w_1$ given a paraphrase $p$ and $w_2$:
   What can *cake* be made of?

3. Predict $w_2$ given a paraphrase $p$ and $w_1$:
   What can be made of *apple*?

# Main Task (1): Predicting Paraphrases

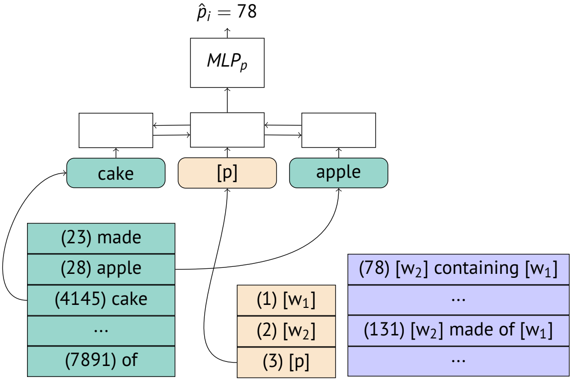
## What is the relation between *apple* and *cake*?




- Encode placeholder [p] in “cake [p] apple” using biLSTM

# Main Task (1): Predicting Paraphrases

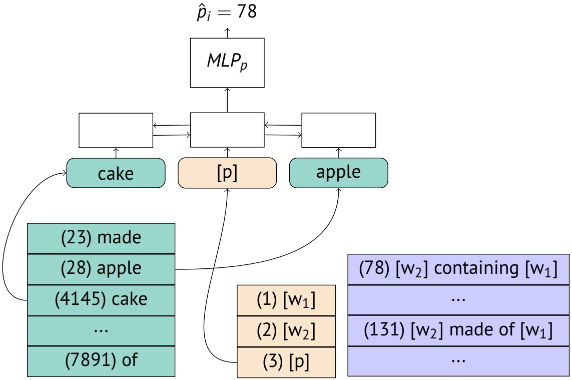
## What is the relation between *apple* and *cake*?




- Encode placeholder [p] in “cake [p] apple” using biLSTM
- Predict an index in the paraphrase vocabulary

# Main Task (1): Predicting Paraphrases

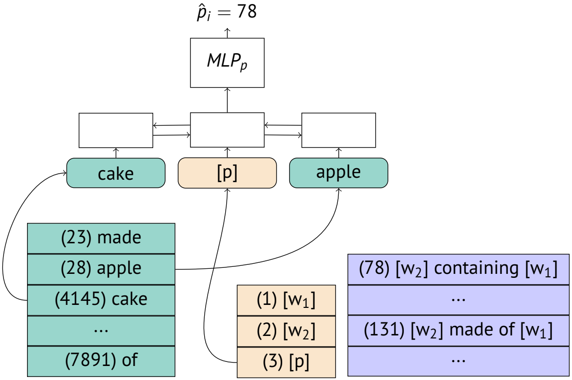
## What is the relation between *apple* and *cake*?




- Encode placeholder [p] in “cake [p] apple” using biLSTM
- Predict an index in the paraphrase vocabulary
- Fixed word embeddings, learned placeholder embeddings

# Main Task (1): Predicting Paraphrases

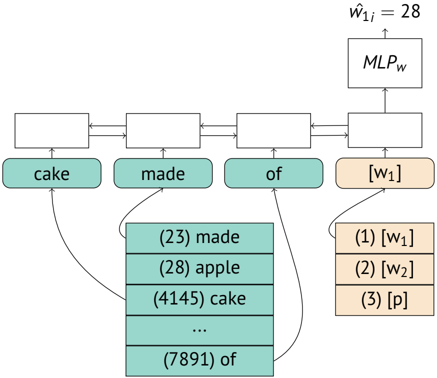
## What is the relation between *apple* and *cake*?




- Encode placeholder [p] in “cake [p] apple” using biLSTM
- Predict an index in the paraphrase vocabulary
- Fixed word embeddings, learned placeholder embeddings
- (1) Generalizes NCs: *pear tart* expected to yield similar results

# Helper Task (2): Predicting Missing Constituents

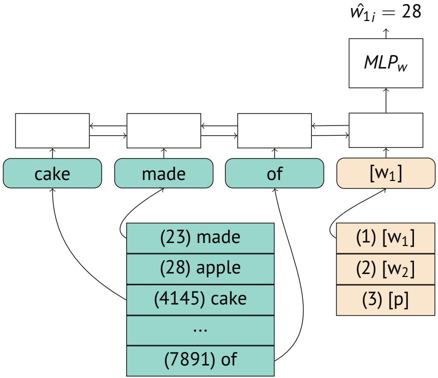
## What can *cake* be made of?




- Encode placeholder in “cake made of $[w_1]$” using biLSTM

# Helper Task (2): Predicting Missing Constituents

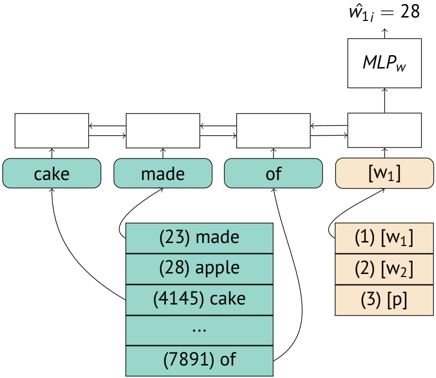
## What can *cake* be made of?




- Encode placeholder in "cake made of [$w_1$]" using biLSTM
- Predict an index in the word vocabulary

# Helper Task (2): Predicting Missing Constituents

## What can *cake* be made of?




- Encode placeholder in "cake made of [$w_1$]" using biLSTM
- Predict an index in the word vocabulary
- (2) Generalizes paraphrases:

  "[$w_2$] containing [$w_1$]" expected to yield similar results

# Evaluation

# Ranking Model

- Predict top *k* paraphrases for each noun compound

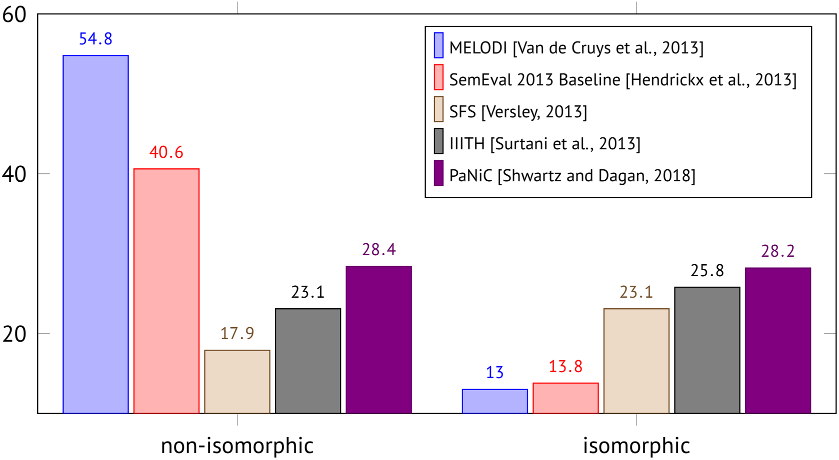
# Ranking Model

- Predict top *k* paraphrases for each noun compound
- Learn to re-rank the paraphrases
  - to better correlate with human judgments

# Ranking Model

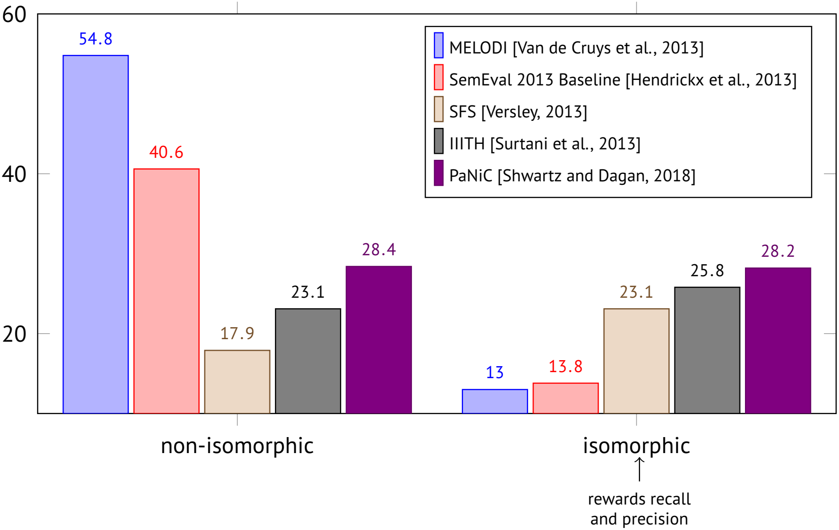
- Predict top *k* paraphrases for each noun compound
- Learn to re-rank the paraphrases
  - to better correlate with human judgments
- SVM pair-wise ranking with the following features:
  - POS tags in the paraphrase
  - Prepositions in the paraphrase
  - Length
  - Special symbols
  - Similarity to predicted paraphrase

# Results

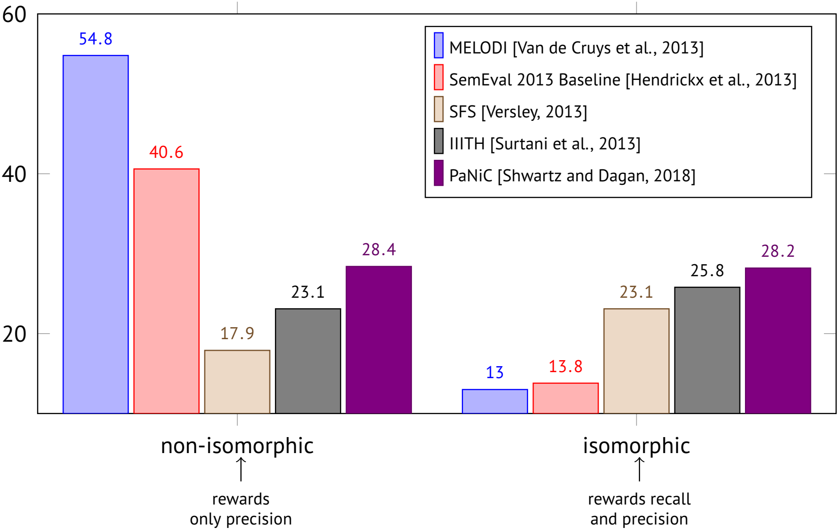


60
40
20
54.8
40.6
17.9
23.1
28.4
13
13.8
23.1
25.8
28.2
MELODI [Van de Cruys et al., 2013]
SemEval 2013 Baseline [Hendrickx et al., 2013]
SFS [Versley, 2013]
IIITH [Surtani et al., 2013]
PaNiC [Shwartz and Dagan, 2018]
non-isomorphic
isomorphic

# Results

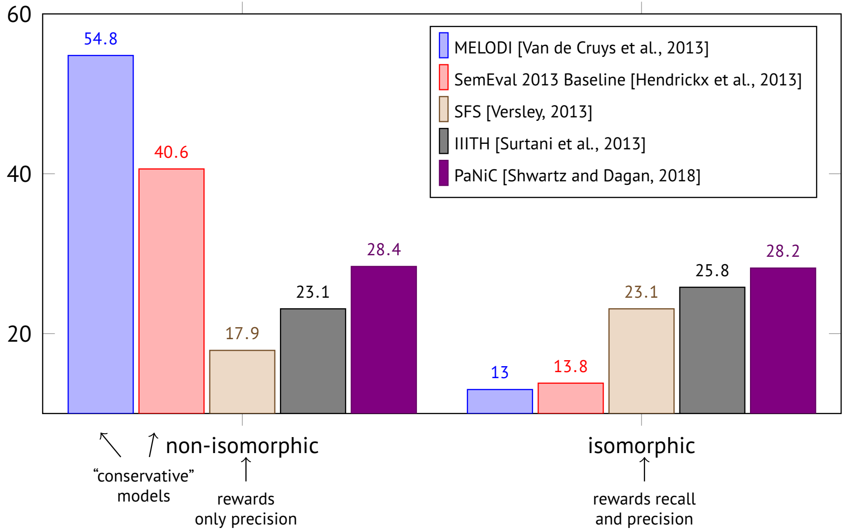

60
40
20
54.8
40.6
17.9
23.1
28.4
13
13.8
23.1
25.8
28.2
MELODI [Van de Cruys et al., 2013]
SemEval 2013 Baseline [Hendrickx et al., 2013]
SFS [Versley, 2013]
IIITH [Surtani et al., 2013]
PaNiC [Shwartz and Dagan, 2018]
non-isomorphic
isomorphic
rewards recall
and precision

# Results



60
40
20
54.8
40.6
17.9
23.1
28.4
13
13.8
23.1
25.8
28.2
MELODI [Van de Cruys et al., 2013]
SemEval 2013 Baseline [Hendrickx et al., 2013]
SFS [Versley, 2013]
IIITH [Surtani et al., 2013]
PaNiC [Shwartz and Dagan, 2018]
non-isomorphic
isomorphic
rewards
only precision
rewards recall
and precision

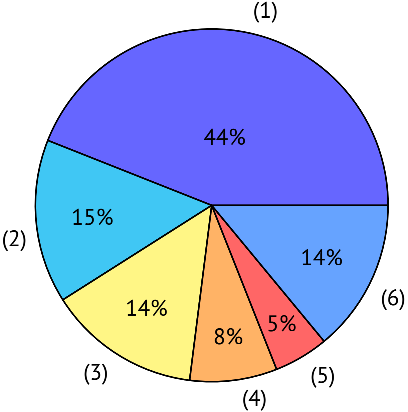
# Results



60
40
20
MELODI [Van de Cruys et al., 2013]
SemEval 2013 Baseline [Hendrickx et al., 2013]
SFS [Versley, 2013]
IIITH [Surtani et al., 2013]
PaNiC [Shwartz and Dagan, 2018]
54.8
40.6
17.9
23.1
28.4
13
13.8
23.1
25.8
28.2
non-isomorphic
isomorphic
"conservative" models
rewards only precision
rewards recall and precision

# Error Analysis

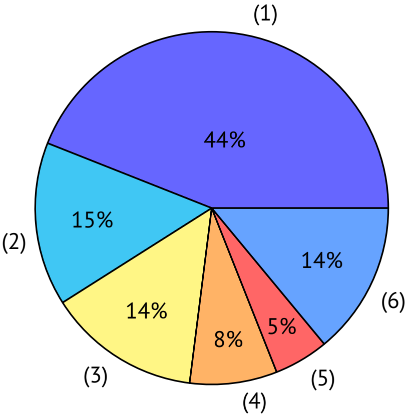
## False Positive




1. Valid, missing from gold-standard ("discussion by group")

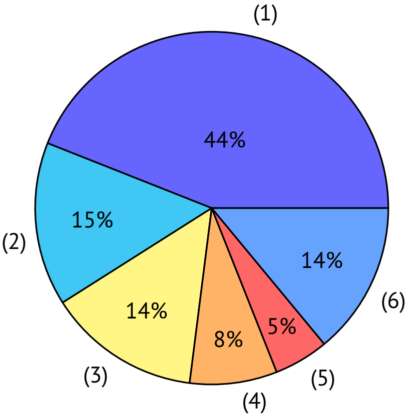
# Error Analysis

## False Positive



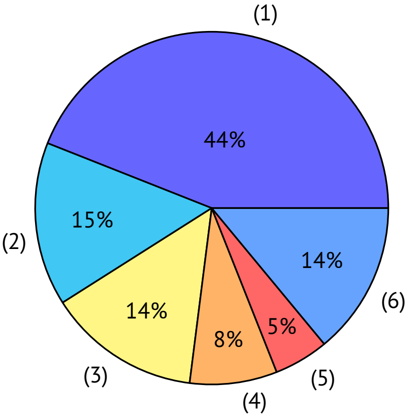

1. Valid, missing from gold-standard
   (“discussion by group”)
2. Too specific
   (“life *of women in* community”)

# Error Analysis
## False Positive




1. Valid, missing from gold-standard
   (“discussion by group”)
2. Too specific
   (“life *of women in* community”)
3. Incorrect prepositions

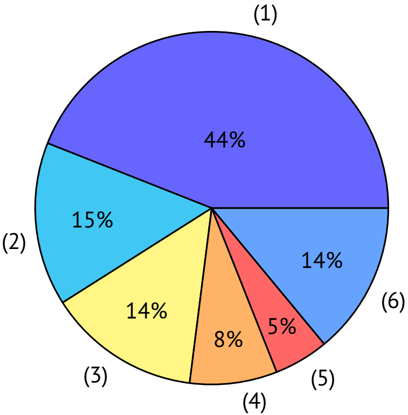
# Error Analysis
## False Positive




1. Valid, missing from gold-standard
   (“discussion by group”)
2. Too specific
   (“life *of women in* community”)
3. Incorrect prepositions
   E.g., n-grams don’t respect syntactic structure: “rinse away the oil from baby ’s head” $\Rightarrow$ “oil from baby”

# Error Analysis

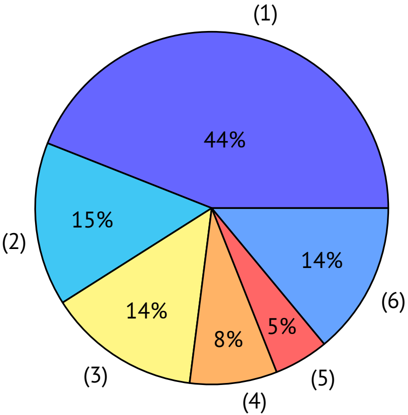
## False Positive




1. Valid, missing from gold-standard
   ("discussion by group")
2. Too specific
   ("life *of women in* community")
3. Incorrect prepositions
   E.g., n-grams don't respect syntactic structure: "rinse away the oil from baby 's head" ⇒ "oil from baby"
4. Syntactic errors

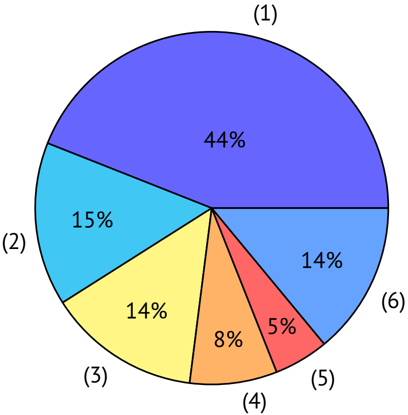
# Error Analysis

## False Positive



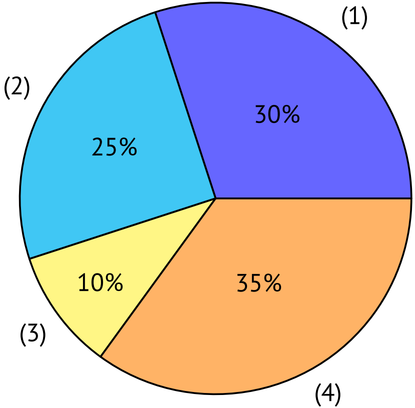

1. Valid, missing from gold-standard
   (“discussion by group”)
2. Too specific
   (“life *of women in* community”)
3. Incorrect prepositions
   E.g., n-grams don’t respect syntactic structure: “rinse away the oil from baby ’s head” ⇒ “oil from baby”
4. Syntactic errors
5. Borderline grammatical
   (“force of coalition forces”)

# Error Analysis
## False Positive




1. Valid, missing from gold-standard
   (“discussion by group”)
2. Too specific
   (“life *of women in* community”)
3. Incorrect prepositions
   E.g., n-grams don’t respect syntactic structure: “rinse away the oil from baby ’s head” ⇒ “oil from baby”
4. Syntactic errors
5. Borderline grammatical
   (“force of coalition forces”)
6. Other errors

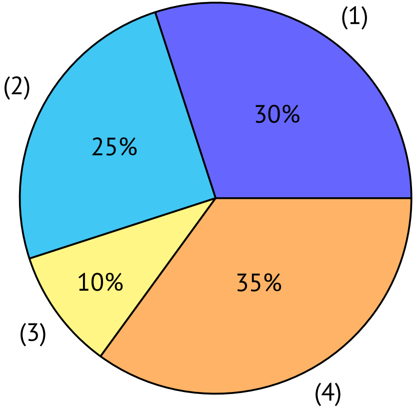
# Error Analysis
## False Negative




1. Long paraphrase ($n > 5$)

# Error Analysis

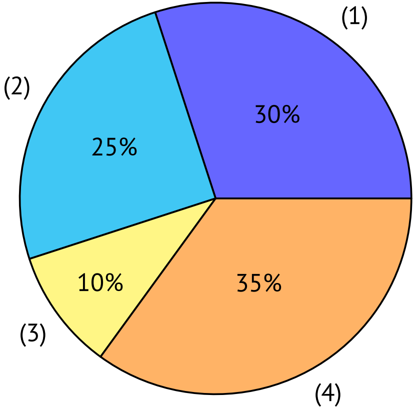
## False Negative




1. Long paraphrase ($n > 5$)
2. Determiners
   ("mutation of **a** gene")

# Error Analysis

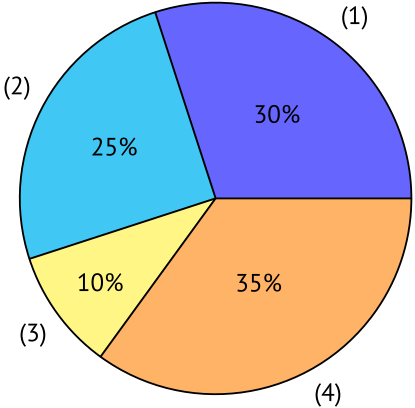
## False Negative




1. Long paraphrase ($n > 5$)
2. Determiners ("mutation of **a** gene")
3. Inflected constituents ("holding of share**s**")

# Error Analysis

## False Negative




1. Long paraphrase ($n > 5$)
2. Determiners
   (“mutation of **a** gene”)
3. Inflected constituents
   (“holding of share**s**”)
4. Other errors

# Noun Compound Paraphrasing
## Recap

- A model for generating paraphrases for given noun-compounds

# Noun Compound Paraphrasing
## Recap

- A model for generating paraphrases for given noun-compounds
- Better generalization abilities:
  - Generalize for unseen noun-compounds
  - Embed semantically-similar paraphrases in proximity

# Noun Compound Paraphrasing

## Recap

- A model for generating paraphrases for given noun-compounds
- Better generalization abilities:
  - Generalize for unseen noun-compounds
  - Embed semantically-similar paraphrases in proximity
- Improved performance in challenging evaluation settings

# Future Directions
## in phrase representations

# Can we learn phrase meanings like humans do?

- [Cooper, 1999]: how do L2 learners process idioms?
  - **Infer from context:** 28% (57% success rate)
  - **Rely on literal meaning:** 19% (22% success rate)
  - ...

# Inferring from context

**Furious Meghan Markle says she won’t fall for dad’s ‘crocodile tears’ after he claimed ‘she’d be better off if he were dead’**

FURIOUS Meghan Markle has said she won't fall for her dad's "crocodile tears" after he claimed "she'd be better off if he were dead".

The Duchess of Sussex reportedly told pals Thomas Markle is using "emotional blackmail" to try and manipulate her but she's had "enough already".

## We need “extended” contexts

[Asl, 2013]: more successful idiom interpretation with extended contexts (stories)

# Inferring from context

**Furious Meghan Markle says she won’t fall for dad’s ‘crocodile tears’ after he claimed ‘she’d be better off if he were dead’**

FURIOUS Meghan Markle has said she won't fall for her dad's "crocodile tears" after he claimed "she'd be better off if he were dead".

The Duchess of Sussex reportedly told pals Thomas Markle is using "emotional blackmail" to try and manipulate her but she's had "enough already".

## We need “extended” contexts

[Asl, 2013]: more successful idiom interpretation with extended contexts (stories)

## We need richer context modeling

- Characters in the story
- Relationships between them
- Dialogues
- …

# Relying on literal meaning

*"Robert knew he was robbing the cradle by dating a sixteen-year-old girl"*

**We need world knowledge**

*"Cradle is something you put the baby in"*

# Relying on literal meaning

*"Robert knew he was robbing the cradle by dating a sixteen-year-old girl"*

**We need world knowledge**
*"Cradle is something you put the baby in"*

**We need to be able to reason**
*"You're stealing a child from a mother"*

*"So **robbing the cradle** is like dating a really young person"*

[Cooper, 1999]

# Thank you!
# Questions?

@VeredShwartz vereds@allenai.org

# References I


[Adi et al., 2017] Adi, Y., Kermany, E., Belinkov, Y., Lavi, O., and Goldberg, Y. (2017). Fine-grained analysis of sentence embeddings using auxiliary prediction tasks. In *Proceedings of ICLR Conference Track*.

[Asl, 2013] Asl, F. M. (2013). The impact of context on learning idioms in efl classes. *TESOL Journal*, 37(1):2.

[Boleda et al., 2013] Boleda, G., Baroni, M., Pham, T. N., and McNally, L. (2013). Intensionality was only alleged: On adjective-noun composition in distributional semantics. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers*, pages 35–46, Potsdam, Germany. Association for Computational Linguistics.

[Conneau et al., 2018] Conneau, A., Kruszewski, G., Lample, G., Barrault, L., and Baroni, M. (2018). What you can cram into a single vector: Probing sentence embeddings for linguistic properties. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2126–2136. Association for Computational Linguistics.

[Cooper, 1999] Cooper, T. C. (1999). Processing of idioms by l2 learners of english. *TESOL quarterly*, 33(2):233–262.

[Dima, 2016] Dima, C. (2016). On the compositionality and semantic interpretation of english noun compounds. In *Proceedings of the 1st Workshop on Representation Learning for NLP*.

# References II


[Dima et al., 2019] Dima, C., de Kok, D., Witte, N., and Hinrichs, E. (2019). No word is an island—a transformation weighting model for semantic composition. *Transactions of the Association for Computational Linguistics*, 7:437–451.

[Dinu et al., 2013] Dinu, G., Pham, N. T., and Baroni, M. (2013). General estimation and evaluation of compositional distributional semantic models. In *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, pages 50–58, Sofia, Bulgaria. Association for Computational Linguistics.

[Hendrickx et al., 2013] Hendrickx, I., Kozareva, Z., Nakov, P., Ó Séaghdha, D., Szpakowicz, S., and Veale, T. (2013). Semeval-2013 task 4: Free paraphrases of noun compounds. In *SemEval*, pages 138–143.

[Levy et al., 2015] Levy, O., Remus, S., Biemann, C., and Dagan, I. (2015). Do supervised distributional methods really learn lexical inference relations? In *NAACL*, pages 970–976.

[Mitchell and Lapata, 2010] Mitchell, J. and Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive science*, 34(8):1388–1429.

[Nakov, 2013] Nakov, P. (2013). On the interpretation of noun compounds: Syntax, semantics, and entailment. *Natural Language Engineering*, 19(03):291–330.

# References III


[Nakov and Hearst, 2006] Nakov, P. and Hearst, M. (2006). Using verbs to characterize noun-noun relations. In *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*, pages 233–244. Springer.

[Netzer and Elhadad, 1998] Netzer, Y. D. and Elhadad, M. (1998). Generation of noun compounds in hebrew: Can syntactic knowledge be fully encapsulated? In *Natural Language Generation*.

[Pantel et al., 2007] Pantel, P., Bhagat, R., Coppola, B., Chklovski, T., and Hovy, E. (2007). ISP: Learning inferential selectional preferences. In *ACL*, pages 564–571.

[Pavlick and Callison-Burch, 2016] Pavlick, E. and Callison-Burch, C. (2016). Most "babies" are "little" and most "problems" are "huge": Compositional entailment in adjective-nouns. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2164–2173, Berlin, Germany. Association for Computational Linguistics.

[Shwartz and Dagan, 2018] Shwartz, V. and Dagan, I. (2018). Paraphrase to explicate: Revealing implicit noun-compound relations. In *ACL*, Melbourne, Australia.

[Shwartz et al., 2016] Shwartz, V., Goldberg, Y., and Dagan, I. (2016). Improving hypernymy detection with an integrated path-based and distributional method. In *ACL*, pages 2389–2398.

# References IV


[Socher et al., 2012] Socher, R., Huval, B., Manning, D. C., and Ng, Y. A. (2012). Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1201–1211. Association for Computational Linguistics.

[Surtani et al., 2013] Surtani, N., Batra, A., Ghosh, U., and Paul, S. (2013). Iiit-h: A corpus-driven co-occurrence based probabilistic model for noun compound paraphrasing. In *SemEval*, pages 153–157.

[Tratz, 2011] Tratz, S. (2011). *Semantically-enriched parsing for natural language understanding*. University of Southern California.

[Van de Cruys et al., 2013] Van de Cruys, T., Afantenos, S., and Muller, P. (2013). Melodi: A supervised distributional approach for free paraphrasing of noun compounds. In *SemEval*, pages 144–147.

[Versley, 2013] Versley, Y. (2013). Sfs-tue: Compound paraphrasing with a language model and discriminative reranking. In *SemEval*, pages 148–152.

[Wieting et al., 2015] Wieting, J., Bansal, M., Gimpel, K., and Livescu, K. (2015). Towards universal paraphrastic sentence embeddings. *CoRR*, abs/1511.08198.

# References V

[Zanzotto et al., 2010] Zanzotto, F. M., Korkontzelos, I., Fallucchi, F., and Manandhar, S. (2010). Estimating linear models for compositional distributional semantics. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 1263–1271. Association for Computational Linguistics.