Not a piece of cake: Lexical Composition and Implicit Information

Vered Shwartz

December 2019
How to obtain meaningful phrase representations?
Distributional Representations

Distributional embeddings of rare terms are of low quality

![Graph showing rare words vs. noun compound frequency]

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

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

ALL THE WORDS!
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?

Issues with compositional representations

\[ f(\vec{v}_{w_1}, \vec{v}_{w_2}, \ldots, \vec{v}_{w_k}) \]
Issues with compositional representations

\[ f(\vec{v}_{w_1}, \vec{v}_{w_2}, \ldots, \vec{v}_{w_k}) \]

“The whole is greater than the sum of its parts”
Issues with compositional representations

\[
f(\vec{v}_{w_1}, \vec{v}_{w_2}, \ldots, \vec{v}_{w_k})
\]

“The whole is greater than the sum of its parts”

1. Meaning shift
2. Implicit meaning
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

@_you_had_one_job1
Implicit Meaning

Noun compounds

Adjective-noun compositions
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

Representation → Minimal Model → Prediction

- 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**:
   \[
   \tilde{x} = \text{vector of target span, additional inputs} \\
   \tilde{o} = \text{softmax}(W \cdot \text{ReLU}(\text{dropout}(h(\tilde{x}))))
   \]
Probing Tasks

Tasks

Meaning shift / Implicit meaning

Representation → Minimal Model → Prediction

- VPC Classification
- LVC Classification
- NC Literality
- NC Relation
- AN Relation
- Phrase Type
(1) Meaning shift - human-like performance for contextualized
(2) Implicit meaning - far from humans
Meaning Shift Tasks
Verb-Particle Classification

Task Definition

We did get on together Which response did you get on that?

VPC

Non-VPC
We did get on together. Which response did you get on that?

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPC</td>
<td>0</td>
</tr>
<tr>
<td>Non-VPC</td>
<td>50</td>
</tr>
<tr>
<td>Majority</td>
<td>23.6</td>
</tr>
<tr>
<td>word2vec</td>
<td>29.5</td>
</tr>
<tr>
<td>GloVe</td>
<td>60.5</td>
</tr>
<tr>
<td>fastText</td>
<td>45.5</td>
</tr>
<tr>
<td>ELMo</td>
<td>83.6</td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>81.4</td>
</tr>
<tr>
<td>BERT</td>
<td>90.0</td>
</tr>
<tr>
<td>Human</td>
<td>93.8</td>
</tr>
</tbody>
</table>
We did get on together Which response did you get on that?

<table>
<thead>
<tr>
<th>Method</th>
<th>VPC</th>
<th>Non-VPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>23.6</td>
<td></td>
</tr>
<tr>
<td>word2vec</td>
<td>29.5</td>
<td></td>
</tr>
<tr>
<td>GloVe</td>
<td>60.5</td>
<td></td>
</tr>
<tr>
<td>fastText</td>
<td>45.5</td>
<td></td>
</tr>
<tr>
<td>ELMo</td>
<td>83.6</td>
<td></td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>81.4</td>
<td></td>
</tr>
<tr>
<td>BERT</td>
<td>90.0</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>93.8</td>
<td></td>
</tr>
</tbody>
</table>
Results

We did get on together Which response did you get on that?

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>23.6</td>
</tr>
<tr>
<td>word2vec</td>
<td>29.5</td>
</tr>
<tr>
<td>GloVe</td>
<td>60.5</td>
</tr>
<tr>
<td>fastText</td>
<td>45.5</td>
</tr>
<tr>
<td>ELMo</td>
<td>83.6</td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>81.4</td>
</tr>
<tr>
<td>BERT</td>
<td>90.0</td>
</tr>
<tr>
<td>Human</td>
<td>93.8</td>
</tr>
</tbody>
</table>

Accuracy

- Standard
- Contextualized
Verb-Particle Classification
Analysis

get on

give in

make for

take on
Noun Compound Literality
Task Definition

The crash course in litigation made me a better lawyer
Noun Compound Literality

Results

The crash course in litigation made me a better lawyer

<table>
<thead>
<tr>
<th></th>
<th>Non-Literal</th>
<th>Literal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>72.5</td>
<td>91.0</td>
</tr>
</tbody>
</table>

- **Majority**: 72.5
- **Human**: 91.0

Models:
- word2vec
- GloVe
- fastText
- ELMo
- OpenAI GPT
- BERT
Noun Compound Literality

Results

The crash course in litigation made me a better lawyer.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Majority</th>
<th>word2vec</th>
<th>GloVe</th>
<th>fastText</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>72.5</td>
<td>72.5</td>
<td>80.4</td>
<td>74.6</td>
<td>91.0</td>
</tr>
</tbody>
</table>

The chart shows the accuracy of various models in identifying the literal meaning of the phrase "The crash course in litigation made me a better lawyer."
Noun Compound Literality

Results

The crash course in litigation made me a better lawyer

Accuracy

Non-Literal
Literal

The chart shows the accuracy of various models in classifying noun compounds as literal or non-literal. The models include:

- Majority
- word2vec
- GloVe
- fastText
- ELMo
- OpenAI GPT
- BERT
- Human

The accuracy values range from 72.5% to 91.3%, with BERT and human achieving the highest accuracy of 91.3% and 91.0%, respectively.
## Noun Compound Literality
Detecting meaning shift → modeling meaning?

<table>
<thead>
<tr>
<th>ELMo</th>
<th>OpenAI GPT</th>
<th>BERT</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Queen and her husband were on a train <strong>trip</strong> from Sydney to Orange.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ride</td>
<td>to</td>
<td>travelling</td>
</tr>
<tr>
<td>carriage</td>
<td>headed</td>
<td>running</td>
</tr>
<tr>
<td>journey</td>
<td>heading</td>
<td>journey</td>
</tr>
<tr>
<td>heading</td>
<td>that</td>
<td>going</td>
</tr>
<tr>
<td>carrying</td>
<td>and</td>
<td>headed</td>
</tr>
</tbody>
</table>
# Noun Compound Literality

Detecting meaning shift → modeling meaning?

<table>
<thead>
<tr>
<th>ELMo</th>
<th>OpenAI GPT</th>
<th>BERT</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Queen and her husband were on a train <strong>trip</strong> from Sydney to Orange.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ride</td>
<td>to</td>
<td>travelling</td>
</tr>
<tr>
<td>carriage</td>
<td>headed</td>
<td>running</td>
</tr>
<tr>
<td>journey</td>
<td>heading</td>
<td>journey</td>
</tr>
<tr>
<td>heading</td>
<td>that</td>
<td>going</td>
</tr>
<tr>
<td>carrying</td>
<td>and</td>
<td>headed</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>ELMo</th>
<th>OpenAI GPT</th>
<th>BERT</th>
</tr>
</thead>
<tbody>
<tr>
<td>...I believe you are a snake oil salesman, a narcissist...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>auto</td>
<td>in</td>
<td>oil</td>
</tr>
<tr>
<td>egg</td>
<td>and</td>
<td>pit</td>
</tr>
<tr>
<td>hunter</td>
<td>that</td>
<td>bite</td>
</tr>
<tr>
<td>rogue</td>
<td>charmer</td>
<td>jar</td>
</tr>
</tbody>
</table>

Substitutes for the entire phrase.
Implicit Meaning Tasks
He receives warm support from his students.
Adjective-Noun Attributes

Results

He receives warm support from his students

Accuracy

Standard Contextualized

50.0

Majority

50.0

53.8

fastText

54.7

ELMo

65.1

OpenAI GPT

86.4

Human

temperature
emotionality
He receives warm support from his students.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>50.0</td>
</tr>
<tr>
<td>word2vec</td>
<td>50.0</td>
</tr>
<tr>
<td>GloVe</td>
<td>50.0</td>
</tr>
<tr>
<td>fastText</td>
<td>53.8</td>
</tr>
<tr>
<td>ELMo</td>
<td>54.7</td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>57.5</td>
</tr>
<tr>
<td>BERT</td>
<td>65.1</td>
</tr>
<tr>
<td>Human</td>
<td>86.4</td>
</tr>
</tbody>
</table>
He receives warm support from his students.

Accuracy:
- Majority: 50.0
- word2vec: 50.0
- GloVe: 50.0
- fastText: 53.8
- ELMo: 54.7
- OpenAI GPT: 57.5
- BERT: 65.1
- Human: 86.4

Contextualized vs. Standard:
- Temperature
- Emotionality
Noun Compound Relations
Task Definition

The township is served by three access roads.

Road forecasted for access season
Road that makes access possible
Noun Compound Relations

Results

The township is served by three access roads.

Accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>50.0</td>
</tr>
<tr>
<td>word2vec</td>
<td>50.0</td>
</tr>
<tr>
<td>GloVe</td>
<td>48.1</td>
</tr>
<tr>
<td>fastText</td>
<td>51.2</td>
</tr>
<tr>
<td>ELMo</td>
<td>54.3</td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>53.1</td>
</tr>
<tr>
<td>BERT</td>
<td>51.2</td>
</tr>
<tr>
<td>Human</td>
<td>77.8</td>
</tr>
</tbody>
</table>
The township is served by three access roads.

Accuracy:
- Majority: 50.0
- word2vec: 50.0
- GloVe: 48.1
- fastText: 51.2
- Human: 77.8

Label:
- Road forecasted for access season
- Road that makes access possible
The township is served by three access roads.

<table>
<thead>
<tr>
<th>Model</th>
<th>Standard</th>
<th>Contextualized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>50.0</td>
<td></td>
</tr>
<tr>
<td>word2vec</td>
<td>50.0</td>
<td></td>
</tr>
<tr>
<td>GloVe</td>
<td>48.1</td>
<td></td>
</tr>
<tr>
<td>fastText</td>
<td>51.2</td>
<td></td>
</tr>
<tr>
<td>ELMo</td>
<td>54.3</td>
<td></td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>53.1</td>
<td></td>
</tr>
<tr>
<td>BERT</td>
<td>51.2</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td></td>
<td>77.8</td>
</tr>
</tbody>
</table>

Accuracy

Road forecasted for access season
Road that makes access possible
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*
Context matters: trivially for meaning shift but also for revealing implicit meaning.
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
  - Add
  - FullAdd
  - Matrix
  - ...

- Compositional

- Paraphrase-based
  - Backtranslation
  - Co-occurrence
Compositional Representations

\[ f(\vec{v}_{w_1}, \vec{v}_{w_2}, \ldots, \vec{v}_{w_k}) \]

- \( f(w_1 w_2) = \alpha \cdot \vec{v}_{w_1} + \beta \cdot \vec{v}_{w_2} \) [Mitchell and Lapata, 2010]
- \( f(w_1 w_2) = A\vec{v}_{w_1} + B\vec{v}_{w_2} \) [Zanzotto et al., 2010, Dinu et al., 2013]
- \( f(w_1 w_2) = tanh(W \cdot [\vec{v}_{w_1}; \vec{v}_{w_2}]) \) [Socher et al., 2012]
- ...
Compositional Representations

\[
f(\begin{array}{c}
  \vec{v}_{w_1} \\
  \vec{v}_{w_2} \\
  \vdots \\
  \vec{v}_{w_k}
\end{array})
\]

- \[f(w_1 \ w_2) = \alpha \cdot \vec{v}_{w_1} + \beta \cdot \vec{v}_{w_2}\] [Mitchell and Lapata, 2010]
- \[f(w_1 \ w_2) = A\vec{v}_{w_1} + B\vec{v}_{w_2}\] [Zanzotto et al., 2010, Dinu et al., 2013]
- \[f(w_1 \ w_2) = \tanh(W \cdot [\vec{v}_{w_1}; \vec{v}_{w_2}])\] [Socher et al., 2012]
- ...

Generalization at the constituent level, e.g.:

syndicate representative

\[
\begin{align*}
  f(\text{worker}, \text{representative}) \\
  f(\text{player}, \text{representative}) \\
  f(\text{crack}, \text{dealer}) \\
  f(\text{company}, \text{spokesman}) \\
  f(\text{industry}, \text{commissioner})
\end{align*}
\]
Paraphrase-based Representations

\( f(w_1w_2) \approx f(\text{paraphrase}) \)

- **Backtranslation**: [Wieting et al., 2015]
  
  baby oil → **huile pour bébé** → oil for baby

- **Co-occurrence** of the constituents, e.g. *cake made of apples*

...
Paraphrase-based Representations

\[ f(w_1 w_2) \approx f(\text{paraphrase}) \]

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

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

- *syndicate representative*
  
  \[
  \begin{align*}
  f(\text{worker, representative}) \\
  f(\text{union, representative}) \\
  f(\text{group, manager}) \\
  f(\text{employee, representative}) \\
  f(\text{student, representative})
  \end{align*}
  \]
What is the best representation?
[Dima, 2016]

- **FullAdd** ($A v_{w_1} + B v_{w_2}$) 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.
What is the best representation?
[Dima, 2016]

- **FullAdd** \((A v_{w_1} + B v_{w_2})\) 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:

- 53.06% Rare words
- 18.89% Share constituents with the target compound
- 19.60% Other noun compounds
- 7.58% Other words

Matrix (rare)

- 54.79% Rare words
- 12.12% Share constituents with the target compound
- 31.24% Other noun compounds
- 4% Other words

Backtranslation (rare)
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?


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!
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 → ingredient

birthday cake → 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
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

$\hat{v}_{w_1, w_2}
\hat{v}_{w_1}
\hat{v}_{w_2}$

$\text{paths}(w_1, w_2)$

mean pooling

$[w_2] \text{ of } [w_1]$

$[w_2] \text{ containing } [w_1]$

$[w_1] \text{ in } [w_2]$

Path LSTM

Path

Integrated-NC

Integrated
### Evaluation - Datasets

#### Dataset: [Tratz, 2011]

<table>
<thead>
<tr>
<th>Purpose/Activity Group</th>
<th>Percentage</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERFORM&amp;ENGAGE..IN</td>
<td>11.5%</td>
<td>cooking pot</td>
</tr>
<tr>
<td>CREATE-PROVIDE-GENERATE-SELL</td>
<td>4.8%</td>
<td>nicotine patch</td>
</tr>
<tr>
<td>OBTAIN&amp;ACCESS&amp;SEEK</td>
<td>0.9%</td>
<td>shrimp boat</td>
</tr>
<tr>
<td>MITIGATE&amp;OPPOSE</td>
<td>0.8%</td>
<td>flak jacket</td>
</tr>
<tr>
<td>ORGANIZE&amp;SUPERVISE&amp;AUTHORITY</td>
<td>1.6%</td>
<td>ethics authority</td>
</tr>
<tr>
<td>PURPOSE</td>
<td>1.9%</td>
<td>chicken spit</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ownership, Experience, Employment, Use</th>
<th>Percentage</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>OWNER-USER</td>
<td>2.1%</td>
<td>family estate</td>
</tr>
<tr>
<td>EXPERIENCER-OF-EXPERIENCE</td>
<td>0.5%</td>
<td>family greed</td>
</tr>
<tr>
<td>EMPLOYER</td>
<td>2.3%</td>
<td>team doctor</td>
</tr>
<tr>
<td>USER..RECIPIENT</td>
<td>1.0%</td>
<td>voter pamphlet</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Temporal Group</th>
<th>Percentage</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME-OF1</td>
<td>2.2%</td>
<td>night work</td>
</tr>
<tr>
<td>TIME-OF2</td>
<td>0.5%</td>
<td>birth date</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location and Whole+Part/Member of</th>
<th>Percentage</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOCATION</td>
<td>5.2%</td>
<td>hillside home</td>
</tr>
<tr>
<td>WHOLE+PART..OR..MEMBER..OF</td>
<td>1.7%</td>
<td>robot arm</td>
</tr>
</tbody>
</table>
Evaluation - Datasets

- Dataset: [Tratz, 2011]

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

- Dataset 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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Split</th>
<th>Best Baseline</th>
<th>Path</th>
<th>Int</th>
<th>Int-NC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tratz-fine</strong></td>
<td>Rand</td>
<td>0.725</td>
<td>0.538</td>
<td>0.714</td>
<td>0.692</td>
</tr>
<tr>
<td></td>
<td>Lex&lt;sub&gt;head&lt;/sub&gt;</td>
<td>0.458</td>
<td>0.448</td>
<td>0.510</td>
<td>0.478</td>
</tr>
<tr>
<td></td>
<td>Lex&lt;sub&gt;mod&lt;/sub&gt;</td>
<td>0.607</td>
<td>0.472</td>
<td>0.613</td>
<td>0.600</td>
</tr>
<tr>
<td></td>
<td>Lex&lt;sub&gt;full&lt;/sub&gt;</td>
<td>0.363</td>
<td>0.423</td>
<td>0.421</td>
<td>0.429</td>
</tr>
<tr>
<td><strong>Tratz-coarse</strong></td>
<td>Rand</td>
<td><strong>0.775</strong></td>
<td>0.586</td>
<td>0.736</td>
<td>0.712</td>
</tr>
<tr>
<td></td>
<td>Lex&lt;sub&gt;head&lt;/sub&gt;</td>
<td>0.538</td>
<td>0.518</td>
<td><strong>0.569</strong></td>
<td>0.548</td>
</tr>
<tr>
<td></td>
<td>Lex&lt;sub&gt;mod&lt;/sub&gt;</td>
<td>0.645</td>
<td>0.548</td>
<td><strong>0.646</strong></td>
<td>0.632</td>
</tr>
<tr>
<td></td>
<td>Lex&lt;sub&gt;full&lt;/sub&gt;</td>
<td>0.409</td>
<td>0.472</td>
<td><strong>0.475</strong></td>
<td>0.478</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Split</th>
<th>Best Baseline</th>
<th>Path</th>
<th>Int</th>
<th>Int-NC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tratz-fine</strong></td>
<td>Rand</td>
<td>0.725</td>
<td>0.538</td>
<td>0.714</td>
<td>0.692</td>
</tr>
<tr>
<td></td>
<td>Lex$_{head}$</td>
<td>0.458</td>
<td>0.448</td>
<td><strong>0.510</strong></td>
<td>0.478</td>
</tr>
<tr>
<td></td>
<td>Lex$_{mod}$</td>
<td>0.607</td>
<td>0.472</td>
<td><strong>0.613</strong></td>
<td>0.600</td>
</tr>
<tr>
<td></td>
<td>Lex$_{full}$</td>
<td>0.363</td>
<td>0.423</td>
<td>0.421</td>
<td><strong>0.429</strong></td>
</tr>
<tr>
<td><strong>Tratz-coarse</strong></td>
<td>Rand</td>
<td><strong>0.775</strong></td>
<td>0.586</td>
<td>0.736</td>
<td>0.712</td>
</tr>
<tr>
<td></td>
<td>Lex$_{head}$</td>
<td>0.538</td>
<td>0.518</td>
<td><strong>0.569</strong></td>
<td>0.548</td>
</tr>
<tr>
<td></td>
<td>Lex$_{mod}$</td>
<td>0.645</td>
<td>0.548</td>
<td><strong>0.646</strong></td>
<td>0.632</td>
</tr>
<tr>
<td></td>
<td>Lex$_{full}$</td>
<td>0.409</td>
<td>0.472</td>
<td><strong>0.475</strong></td>
<td>0.478</td>
</tr>
</tbody>
</table>

- **Lexical split:** our methods perform better.
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Split</th>
<th>Best Baseline</th>
<th>Path</th>
<th>Int</th>
<th>Int-NC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tratz-fine</td>
<td>Rand</td>
<td><strong>0.725</strong></td>
<td>0.538</td>
<td>0.714</td>
<td>0.692</td>
</tr>
<tr>
<td></td>
<td>Lex\textsubscript{head}</td>
<td>0.458</td>
<td>0.448</td>
<td><strong>0.510</strong></td>
<td>0.478</td>
</tr>
<tr>
<td></td>
<td>Lex\textsubscript{mod}</td>
<td>0.607</td>
<td>0.472</td>
<td><strong>0.613</strong></td>
<td>0.600</td>
</tr>
<tr>
<td></td>
<td>Lex\textsubscript{full}</td>
<td><strong>0.363</strong></td>
<td>0.423</td>
<td>0.421</td>
<td><strong>0.429</strong></td>
</tr>
<tr>
<td>Tratz-coarse</td>
<td>Rand</td>
<td><strong>0.775</strong></td>
<td>0.586</td>
<td>0.736</td>
<td>0.712</td>
</tr>
<tr>
<td></td>
<td>Lex\textsubscript{head}</td>
<td>0.538</td>
<td>0.518</td>
<td><strong>0.569</strong></td>
<td>0.548</td>
</tr>
<tr>
<td></td>
<td>Lex\textsubscript{mod}</td>
<td>0.645</td>
<td>0.548</td>
<td><strong>0.646</strong></td>
<td>0.632</td>
</tr>
<tr>
<td></td>
<td>Lex\textsubscript{full}</td>
<td><strong>0.409</strong></td>
<td>0.472</td>
<td><strong>0.475</strong></td>
<td>0.478</td>
</tr>
</tbody>
</table>

- The performance gap is larger in lexical-full.
Which relations can the path-based model learn?

<table>
<thead>
<tr>
<th>relation</th>
<th>path</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>measure</td>
<td>$[w_2]$ varies by $[w_1]$</td>
<td>state limit</td>
</tr>
<tr>
<td></td>
<td>$2,560$ $[w_1]$ portion of $[w_2]$</td>
<td>acre estate</td>
</tr>
<tr>
<td>personal title</td>
<td>$[w_2]$ Anderson $[w_1]$/title</td>
<td>Mrs. Brown</td>
</tr>
<tr>
<td></td>
<td>$[w_2]$ Sheridan $[w_1]$/title</td>
<td>Gen. Johnson</td>
</tr>
<tr>
<td>create-provide-generate-sell</td>
<td>$[w_2]$ produce $[w_1]$</td>
<td>food producer</td>
</tr>
<tr>
<td></td>
<td>$[w_2]$ manufacture $[w_1]$</td>
<td>engine plant</td>
</tr>
<tr>
<td>time-of1</td>
<td>$[w_2]$ begin $[w_1]$</td>
<td>morning program</td>
</tr>
<tr>
<td></td>
<td>$[w_2]$ held Saturday $[w_1]$</td>
<td>afternoon meeting</td>
</tr>
<tr>
<td>substance-material-ingredient</td>
<td>$[w_2]$ made of wood and $[w_1]$</td>
<td>marble table</td>
</tr>
<tr>
<td></td>
<td>$[w_2]$ material includes type of $[w_1]$</td>
<td>steel pipe</td>
</tr>
</tbody>
</table>
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”)
Joint corpus occurrences improve the performance in strict evaluation setups ✔
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 ✓
- Assumes compositionality ×

- Lexical splits help prevent lexical memorization ✓
Noun Compound Relation Classification

Recap

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

- Lexical splits help prevent lexical memorization √
- 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?
  - cake with parsley?
  - cake for parsley?
  - ...

![Parsley Cake Image]
Generalizing Existing Knowledge

What can cake be made of?

<table>
<thead>
<tr>
<th>Context</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAKE WITH CHOCOLATE</td>
<td>31</td>
</tr>
<tr>
<td>CAKE WITH LEMON</td>
<td>13</td>
</tr>
<tr>
<td>CAKE WITH STRAWBERRIES</td>
<td>10</td>
</tr>
<tr>
<td>CAKE WITH CANDIES</td>
<td>7</td>
</tr>
<tr>
<td>CAKE WITH CARAMEL</td>
<td>7</td>
</tr>
<tr>
<td>CAKE WITH FROSTING</td>
<td>6</td>
</tr>
<tr>
<td>CAKE WITH VANILLA</td>
<td>6</td>
</tr>
<tr>
<td>CAKE WITH BERRIES</td>
<td>5</td>
</tr>
<tr>
<td>CAKE WITH EGGS</td>
<td>4</td>
</tr>
<tr>
<td>CAKE WITH TOWEL</td>
<td>4</td>
</tr>
<tr>
<td>CAKE WITH RASPBERRY</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH ICE</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH MARSHMALLOWS</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH HONEY</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH CINNAMON</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH COFFEE</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH BUTTER</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH YOGURT</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH ALMOND</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH BLUEBERRIES</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH COCONUT</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH CITRUS</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH BUTTERCREAM</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH CREME</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH CREAM</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH JUICE</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH CUSTARD</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH FRUIT</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH CONFECTIONS</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH ORANGE</td>
<td>2</td>
</tr>
</tbody>
</table>

Similar to "selectional preferences" [Pantel et al., 2007]
Generalizing Existing Knowledge

What can cake be made of?

<table>
<thead>
<tr>
<th>CONCEPT</th>
<th>FREQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAKE WITH CHOCOLATE</td>
<td>31</td>
</tr>
<tr>
<td>CAKE WITH LEMON</td>
<td>13</td>
</tr>
<tr>
<td>CAKE WITH STRAWBERRIES</td>
<td>10</td>
</tr>
<tr>
<td>CAKE WITH CANDIES</td>
<td>7</td>
</tr>
<tr>
<td>CAKE WITH CARAMEL</td>
<td>7</td>
</tr>
<tr>
<td>CAKE WITH FROSTING</td>
<td>6</td>
</tr>
<tr>
<td>CAKE WITH VANILLA</td>
<td>6</td>
</tr>
<tr>
<td>CAKE WITH BERRIES</td>
<td>5</td>
</tr>
<tr>
<td>CAKE WITH EGGS</td>
<td>4</td>
</tr>
<tr>
<td>CAKE WITH TOWEL</td>
<td>4</td>
</tr>
<tr>
<td>CAKE WITH RASPBERRY</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH ICE</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH MARSHMALLOW</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH HONEY</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH CINNAMON</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH COFFEE</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH BUTTER</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH YOGURT</td>
<td>3</td>
</tr>
<tr>
<td>CAKE WITH ALMOND</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH BLUEBERRIES</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH COCONUT</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH CITRUS</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH BUTTERCREAM</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH CREME</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH CREAM</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH JUICE</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH CUSTARD</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH FRUIT</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH CONFECTIONS</td>
<td>2</td>
</tr>
<tr>
<td>CAKE WITH ORANGE</td>
<td>2</td>
</tr>
</tbody>
</table>

Parsley (sort of) fits into this distribution
Generalizing Existing Knowledge

- What can cake be made of?

- Parsley (sort of) fits into this distribution

- Similar to “selectional preferences” [Pantel et al., 2007]
Noun-Compound Paraphrasing
Motivation

Given a noun-compound $w_1w_2$, express the relation between the head $w_2$ and the modifier $w_1$ with multiple prepositional and verbal paraphrases [Nakov and Hearst, 2006]

- olive oil $\rightarrow$ [w$_2$] extracted from [w$_1$]
- apple cake $\rightarrow$ [w$_2$] made of [w$_1$]
- ground attack $\rightarrow$ [w$_2$] from [w$_1$]
- boat whistle $\rightarrow$ [w$_2$] located in [w$_1$]
- sea bass $\rightarrow$ [w$_2$] live in [w$_1$]
- game room $\rightarrow$ [w$_2$] used for [w$_1$]
- service door $\rightarrow$ [w$_2$] for [w$_1$]
- baby oil $\rightarrow$ [w$_2$]
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

- Evaluated for correlation with human judgments
  - Gold paraphrase score: how many annotators suggested it?
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 \( \{ w_1 = \text{apple}, w_2 = \text{cake}, p = \text{“[w_2] made of [w_1]”} \} \)
Multi-task Reformulation

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

1. Predict a paraphrase \(p\) for a given NC \(w_1w_2\):
   What is the relation between \textit{apple} and \textit{cake}?
Multi-task Reformulation

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

1. Predict a paraphrase \(p\) for a given NC \(w_1 w_2\):
   What is the relation between \text{apple} and \text{cake}?

2. Predict \(w_1\) given a paraphrase \(p\) and \(w_2\):
   What can \text{cake} be made of?
Multi-task Reformulation

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

1. Predict a paraphrase \( p \) for a given NC \( w_1w_2 \):
   What is the relation between \text{apple} and \text{cake}?

2. Predict \( w_1 \) given a paraphrase \( p \) and \( w_2 \):
   What can \text{cake} be made of?

3. Predict \( w_2 \) given a paraphrase \( p \) and \( w_1 \):
   What can be made of \text{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

\[ \hat{p}_i = 78 \]

MLP\_p

(23) made
(28) apple
(4145) cake
... 
(7891) of 

(1) \[w_1\]
(2) \[w_2\]
(3) \[p\]

(78) \[w_2\] containing \[w_1\]
... 
(131) \[w_2\] made of \[w_1\]
... 

Fixed word embeddings, learned placeholder embeddings

Generalizes NCs:
pear tart
expected to yield similar results
Main Task (1): Predicting Paraphrases
What is the relation between *apple* and *cake*?

\[ \hat{p}_i = 78 \]

- 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 [w₁]” using biLSTM
Helper Task (2): Predicting Missing Constituents

What can *cake* be made of?

- Encode placeholder in “cake made of [w<sub>1</sub>]” 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 [w₁]” using biLSTM
- Predict an index in the word vocabulary
- (2) Generalizes paraphrases:
  “[w₂] containing [w₁]” 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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>non-isomorphic</td>
<td>54.8</td>
<td>40.6</td>
<td>23.1</td>
<td>17.9</td>
<td>28.4</td>
</tr>
<tr>
<td>isomorphic</td>
<td>28.4</td>
<td>23.1</td>
<td>25.8</td>
<td>13.8</td>
<td>28.2</td>
</tr>
</tbody>
</table>
Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Non-isomorphic</th>
<th>Isomorphic</th>
</tr>
</thead>
<tbody>
<tr>
<td>MELODI [Van de Cruys et al., 2013]</td>
<td>54.8</td>
<td>28.4</td>
</tr>
<tr>
<td>SemEval 2013 Baseline [Hendrickx et al., 2013]</td>
<td>40.6</td>
<td>23.1</td>
</tr>
<tr>
<td>SFS [Versley, 2013]</td>
<td>17.9</td>
<td>13</td>
</tr>
<tr>
<td>IIITH [Surtani et al., 2013]</td>
<td>23.1</td>
<td>25.8</td>
</tr>
<tr>
<td>PaNiC [Shwartz and Dagan, 2018]</td>
<td>13.8</td>
<td>28.2</td>
</tr>
</tbody>
</table>

Rewards recall and precision
Results

### Non-isomorphic

<table>
<thead>
<tr>
<th>Model</th>
<th>Rewards Only Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>MELODI [Van de Cruys et al., 2013]</td>
<td>54.8</td>
</tr>
<tr>
<td>SemEval 2013 Baseline [Hendrickx et al., 2013]</td>
<td>40.6</td>
</tr>
<tr>
<td>SFS [Versley, 2013]</td>
<td>17.9</td>
</tr>
<tr>
<td>IIITH [Surtani et al., 2013]</td>
<td>23.1</td>
</tr>
<tr>
<td>PaNiC [Shwartz and Dagan, 2018]</td>
<td>28.4</td>
</tr>
</tbody>
</table>

### Isomorphic

<table>
<thead>
<tr>
<th>Model</th>
<th>Rewards Recall and Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>MELODI [Van de Cruys et al., 2013]</td>
<td>13.8</td>
</tr>
<tr>
<td>SemEval 2013 Baseline [Hendrickx et al., 2013]</td>
<td>13.8</td>
</tr>
<tr>
<td>SFS [Versley, 2013]</td>
<td>23.1</td>
</tr>
<tr>
<td>IIITH [Surtani et al., 2013]</td>
<td>25.8</td>
</tr>
<tr>
<td>PaNiC [Shwartz and Dagan, 2018]</td>
<td>28.2</td>
</tr>
</tbody>
</table>
Results

Non-isomorphic models:
- MELODI [Van de Cruys et al., 2013]: 54.8
- SemEval 2013 Baseline [Hendrickx et al., 2013]: 40.6
- SFS [Versley, 2013]: 23.1
- IIITH [Surtani et al., 2013]: 28.4
- PaNiC [Shwartz and Dagan, 2018]: 25.8

Isomorphic models:
- MELODI [Van de Cruys et al., 2013]: 13.8
- SemEval 2013 Baseline [Hendrickx et al., 2013]: 13.8
- SFS [Versley, 2013]: 23.1
- IIITH [Surtani et al., 2013]: 28.2
- PaNiC [Shwartz and Dagan, 2018]: 28.2

“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
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”)
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 \((n > 5)\)
2. Determiners \("mutation of a gene"\)
3. Inflected constituents \("holding of shares"\)
4. Other errors
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 shares”)
1. Long paraphrase \((n > 5)\)
2. Determiners
   (“mutation of a gene”)
3. Inflected constituents
   (“holding of shares”)
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

We need “extended” contexts
[Asl, 2013]: more successful idiom interpretation with extended contexts (stories)
Inferring from context

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


References II


References III


References IV


