Paraphrase to Explicate:
Revealing Implicit Noun-Compound Relations

Vered Shwartz and Ido Dagan

Natural Language Processing Lab, Bar-Ilan University

July 17, 2018

Noun Compounds

- Two or more nouns function as a unit to create a new concept
  - *hot dog*, *hot dog bun*, *hot dog bun package*...

Noun Compounds

- Two or more nouns function as a unit to create a new concept
  - *hot dog*, *hot dog bun*, *hot dog bun package*...
  - We focus on two-word compounds

Noun Compounds

- Two or more nouns function as a unit to create a new concept
  - *hot dog*, *hot dog bun*, *hot dog bun package*...
  - We focus on two-word compounds

- Express implicit relationship between the constituent nouns:

Noun Compounds

- Two or more nouns function as a unit to create a new concept
  - *hot dog*, *hot dog bun*, *hot dog bun package*...
  - We focus on two-word compounds
- Express implicit relationship between the constituent nouns:
  - *apple cake*: cake *made of* apples

Noun Compounds

- Two or more nouns function as a unit to create a new concept
  - *hot dog*, *hot dog bun*, *hot dog bun package*...
  - We focus on two-word compounds

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

Noun Compounds

- Two or more nouns function as a unit to create a new concept
  - *hot dog*, *hot dog bun*, *hot dog bun package*...
  - We focus on two-word 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

- Two or more nouns function as a unit to create a new concept
  - *hot dog*, *hot dog bun*, *hot dog bun package*...
  - We focus on two-word 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!

We are good at Interpreting Noun-Compounds

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

1

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*"?

1

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?
  - ...

1

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?
  - ...


[1]

[1] from http://www.bazekalim.com

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
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

TOTAL 237 | UNIQUE 119 +

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

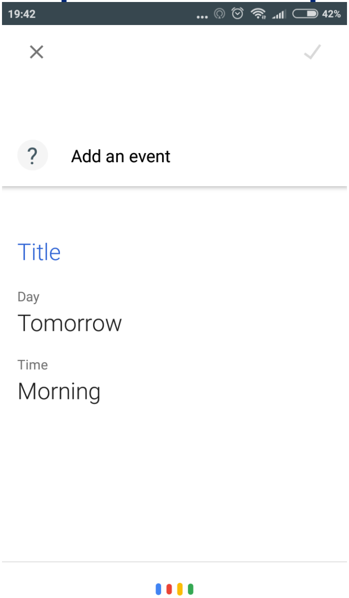
SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...] COMPARE

	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]

We need Computers to Interpret Noun-Compounds

create a morning meeting



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*?

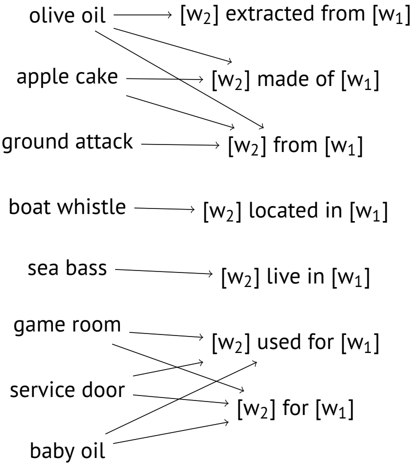
Relation Classification
apple cake → ingredient
birthday cake → time

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

Noun-Compound 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 (1/2)

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

Prior Methods (1/2)

- 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 (1/2)

- 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 (1/2)

- 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)

Prior Methods (2/2)

1. MELODI [Van de Cruys et al., 2013]:
   - Represent NC by applying a function to its constituent distributional vectors: vec(*apple cake*) = f(vec(*apple*), vec(*cake*))

Prior Methods (2/2)

1. MELODI [Van de Cruys et al., 2013]:
   - Represent NC by applying a function to its constituent distributional vectors: vec(*apple cake*) = f(vec(*apple*), vec(*cake*))
   - Predict paraphrase templates given NC vector

Prior Methods (2/2)

1. MELODI [Van de Cruys et al., 2013]:
   - Represent NC by applying a function to its constituent distributional vectors: vec(*apple cake*) = f(vec(*apple*), vec(*cake*))
   - Predict paraphrase templates given NC vector
   - Generalizes for similar unseen NCs, e.g. *pear tart*

Prior Methods (2/2)

1. MELODI [Van de Cruys et al., 2013]:
   - Represent NC by applying a function to its constituent distributional vectors: vec(*apple cake*) = f(vec(*apple*), vec(*cake*))
   - Predict paraphrase templates given NC vector
   - Generalizes for similar unseen NCs, e.g. *pear tart*

2. IIITH [Surtani et al., 2013]:
   - Learn "is-a" relations between paraphrases: e.g. "[w_2] extracted from [w_1]" $\subset$ "[w_2] made of [w_1]"

Prior Methods (2/2)

1. MELODI [Van de Cruys et al., 2013]:
   - Represent NC by applying a function to its constituent distributional vectors: vec(*apple cake*) = f(vec(*apple*), vec(*cake*))
   - Predict paraphrase templates given NC vector
   - Generalizes for similar unseen NCs, e.g. *pear tart*

2. IIITH [Surtani et al., 2013]:
   - Learn "is-a" relations between paraphrases: e.g. "[w_2] extracted from [w_1]" $\subset$ "[w_2] made of [w_1]"

Our solution: multi-task learning to address both problems

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]"}

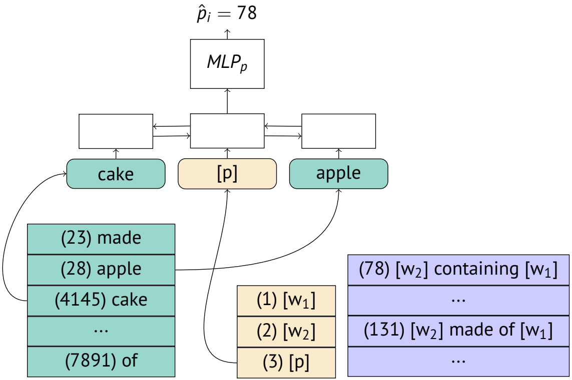
1. Predict a paraphrase p for a given NC $w_1 w_2$:
   What is the relation between *apple* and *cake*?

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

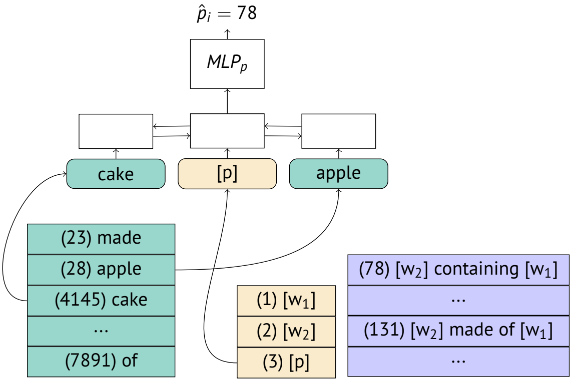
What is the relation between *apple* and *cake*?



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

Main Task (1): Predicting Paraphrases

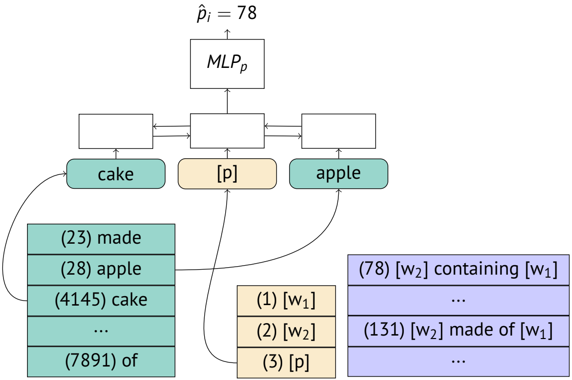
What is the relation between *apple* and *cake*?



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

Main Task (1): Predicting Paraphrases

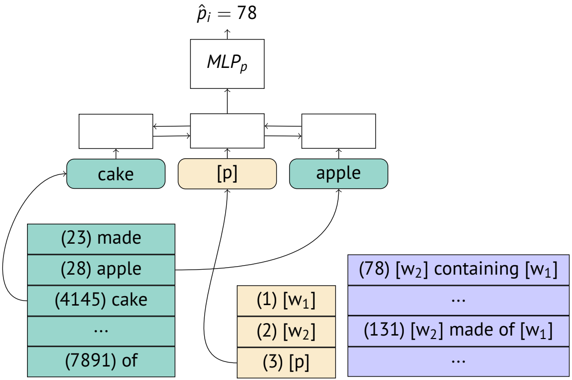
What is the relation between *apple* and *cake*?



- 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

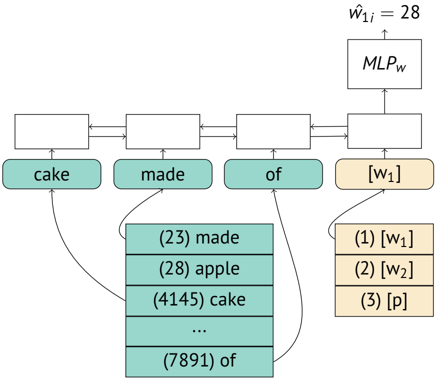
What is the relation between *apple* and *cake*?



- 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

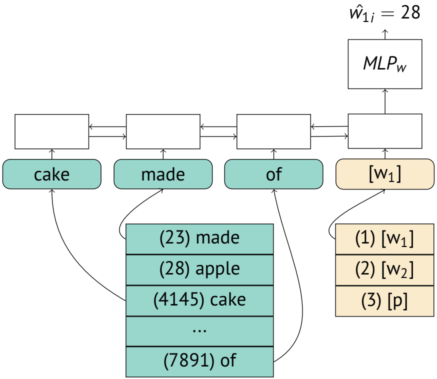
What can *cake* be made of?



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

Helper Task (2): Predicting Missing Constituents

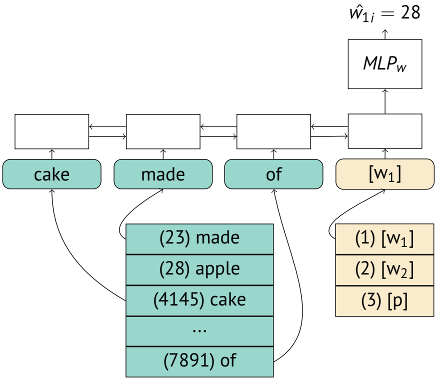
What can *cake* be made of?



- 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

Training Data

- Collected from Google N-grams

Training Data

- Collected from Google N-grams
- Input:
  - Set of NCs
  - Templates of POS tags (e.g. "[w_2] verb prep [w_1]")

Training Data

- Collected from Google N-grams
- Input:
  - Set of NCs
  - Templates of POS tags (e.g. "[w_2] verb prep [w_1]")
- Weighting by frequency and length

Training Data

- Collected from Google N-grams
- Input:
  - Set of NCs
  - Templates of POS tags (e.g. “[w_2] verb prep [w_1]”)
- Weighting by frequency and length
- 140k instances

Evaluation

Ranking Model

- Predict top *k* paraphrases for each noun compound

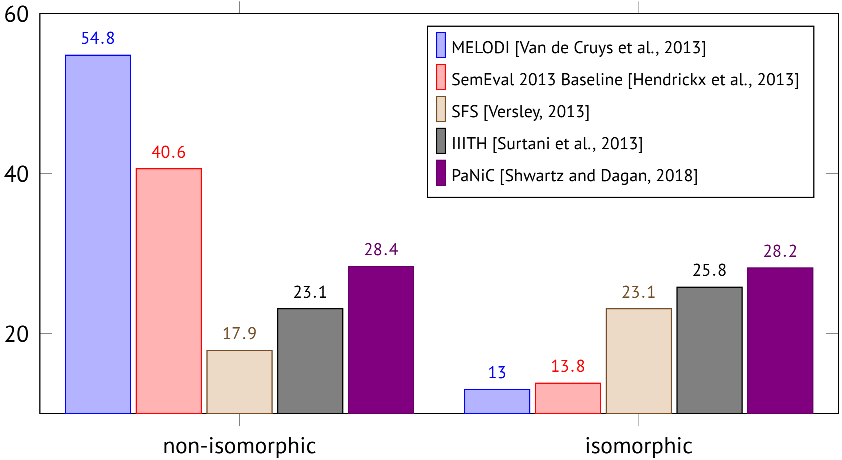
Ranking Model

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

Ranking Model

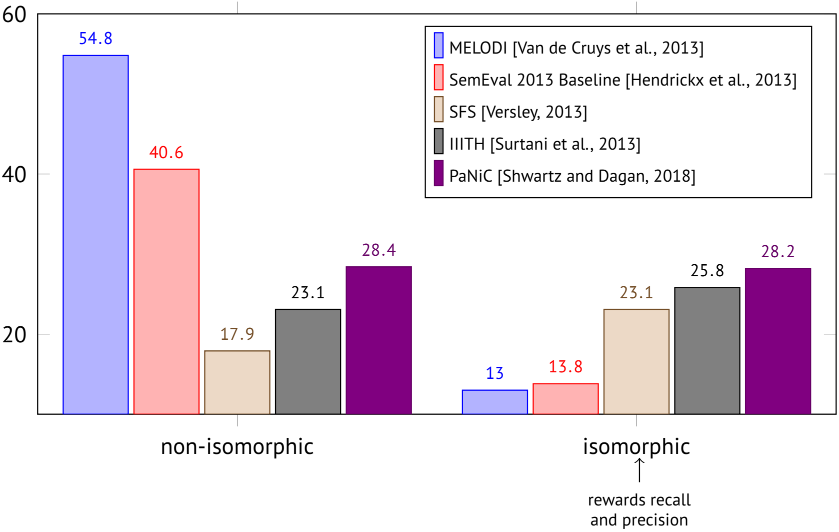
- Predict top *k* paraphrases for each noun compound
- Learn to re-rank the paraphrases
  - to better correlate with human judgments
- SVM pair-wise ranking with the following features:
  - POS tags in the paraphrase
  - Prepositions in the paraphrase
  - Length
  - Special symbols
  - Similarity to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

Results



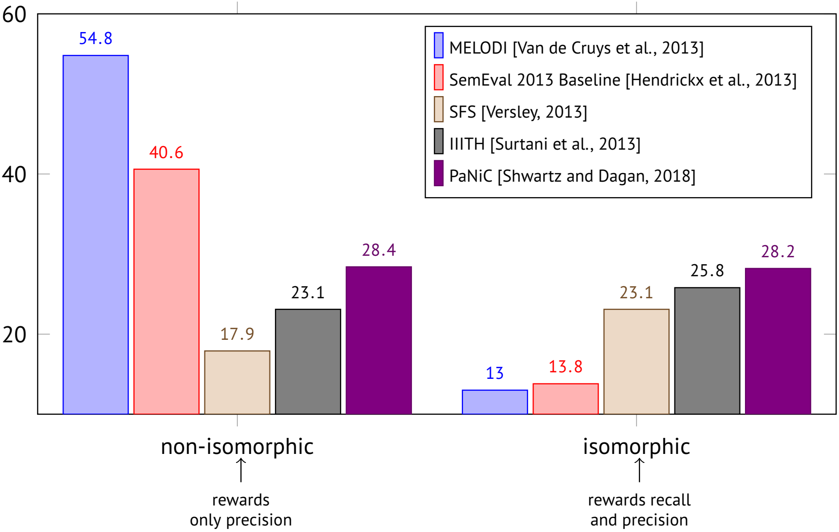
60
40
20
54.8
40.6
17.9
23.1
28.4
13
13.8
23.1
25.8
28.2
MELODI [Van de Cruys et al., 2013]
SemEval 2013 Baseline [Hendrickx et al., 2013]
SFS [Versley, 2013]
IIITH [Surtani et al., 2013]
PaNiC [Shwartz and Dagan, 2018]
non-isomorphic
isomorphic

Results



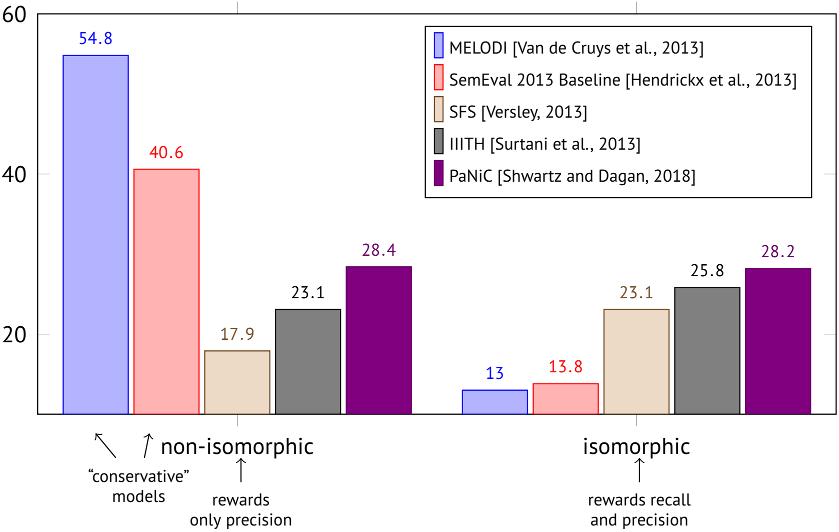
60
40
20
54.8
40.6
17.9
23.1
28.4
13
13.8
23.1
25.8
28.2
MELODI [Van de Cruys et al., 2013]
SemEval 2013 Baseline [Hendrickx et al., 2013]
SFS [Versley, 2013]
IIITH [Surtani et al., 2013]
PaNiC [Shwartz and Dagan, 2018]
non-isomorphic
isomorphic
rewards recall
and precision

Results



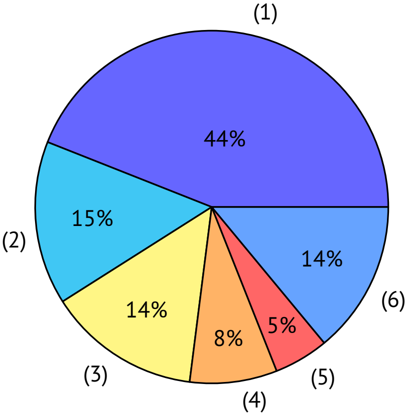
60
40
20
54.8
40.6
17.9
23.1
28.4
13
13.8
23.1
25.8
28.2
MELODI [Van de Cruys et al., 2013]
SemEval 2013 Baseline [Hendrickx et al., 2013]
SFS [Versley, 2013]
IIITH [Surtani et al., 2013]
PaNiC [Shwartz and Dagan, 2018]
non-isomorphic
isomorphic
rewards
only precision
rewards recall
and precision

Results



60
40
20
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
"conservative" models
rewards only precision
rewards recall and precision

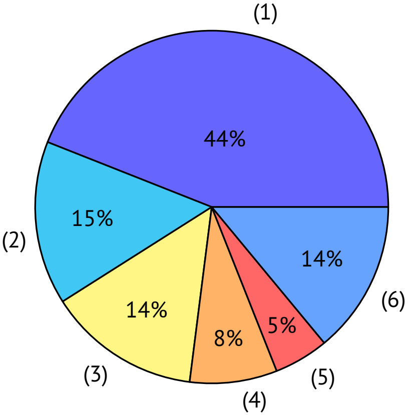
Error Analysis
False Positive



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

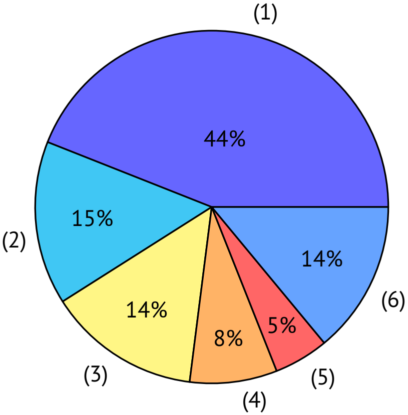
Error Analysis

False Positive



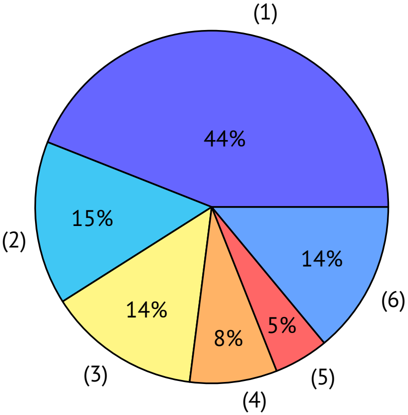
1. Valid, missing from gold-standard ("discussion by group")
2. Too specific ("life *of women in* community")

Error Analysis
False Positive



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

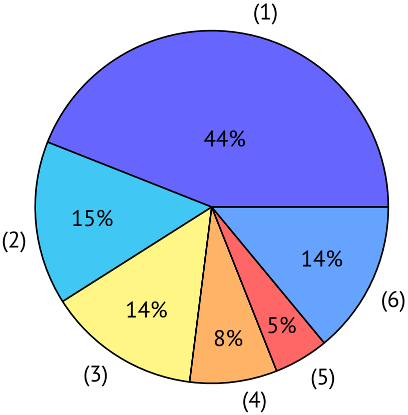
Error Analysis
False Positive



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” ⇒ “oil from baby”

Error Analysis

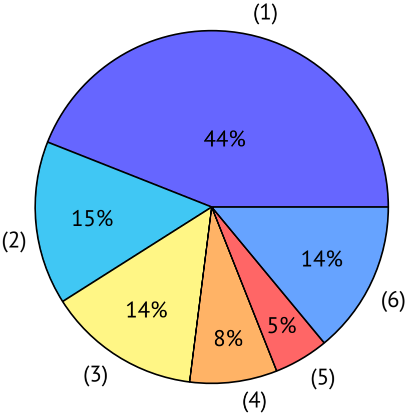
False Positive



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

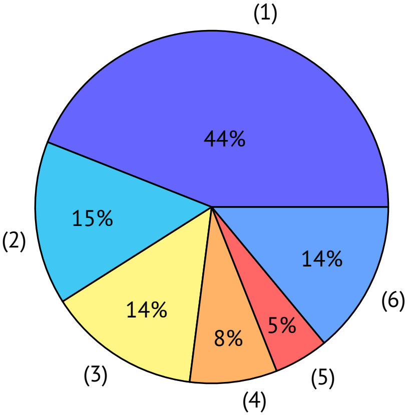
Error Analysis

False Positive



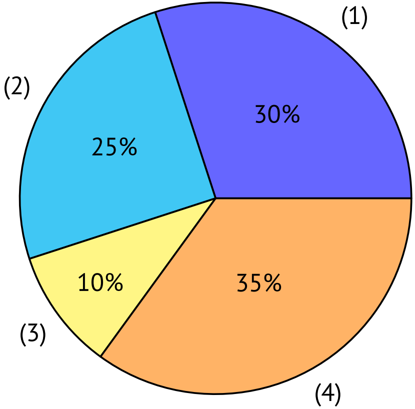
1. Valid, missing from gold-standard
   ("discussion by group")
2. Too specific
   ("life *of women in* community")
3. Incorrect prepositions
   E.g., n-grams don't respect syntactic structure: "rinse away the oil from baby 's head" ⇒ "oil from baby"
4. Syntactic errors
5. Borderline grammatical
   ("force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

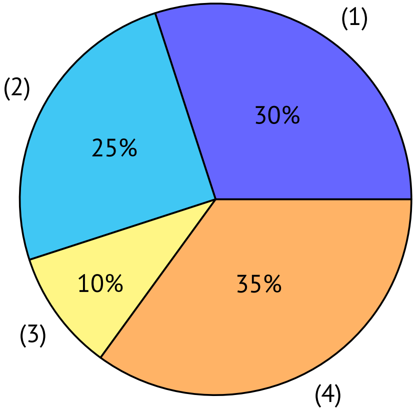
Error Analysis
False Negative



1. Long paraphrase ($n > 5$)

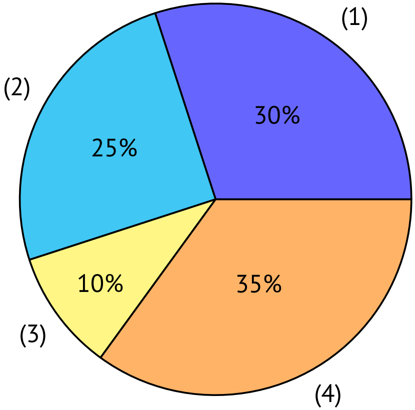
Error Analysis

False Negative



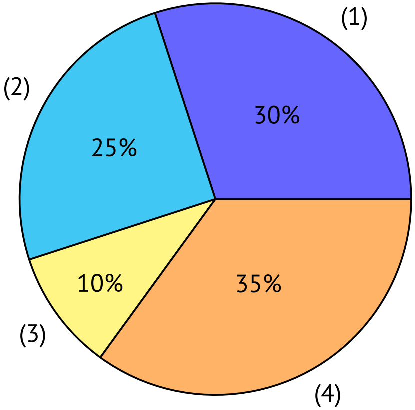
1. Long paraphrase ($n > 5$)
2. Determiners
   ("mutation of **a** gene")

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

Recap

- A model for generating paraphrases for given noun-compounds

Recap

- A model for generating paraphrases for given noun-compounds

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

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

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

Thank you!

References

[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.

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

[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.

[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.

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

[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.

[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.