At Loose Ends:

Challenges and Opportunities in Lexical Composition

Vered Shwartz

Natural Language Processing Lab, Bar-Ilan University

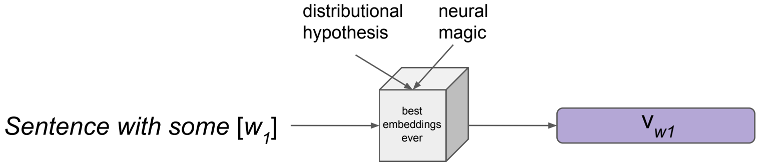
Talk @ EPFL, January 30, 2019

B I U
N L P

Bar-Ilan University

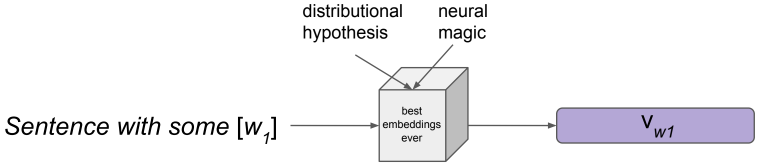
Representing Phrases

- Word representations are pretty much sorted out



Representing Phrases

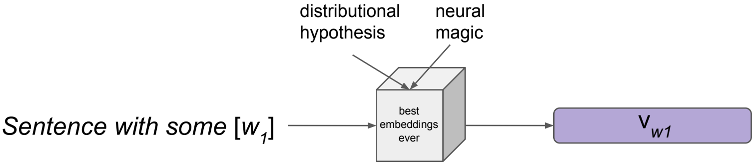
- Word representations are pretty much sorted out



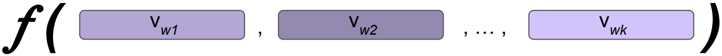
- **How to represent a phrase $\mathbf{p} = \mathbf{w_1}...\mathbf{w_k}$?**

Representing Phrases

- Word representations are pretty much sorted out

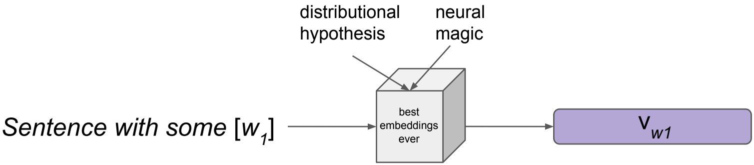


- **How to represent a phrase $\mathbf{p = w_1 \ldots w_k}$?**
- Most straightforward:

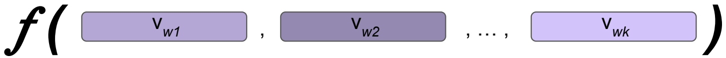


Representing Phrases

- Word representations are pretty much sorted out



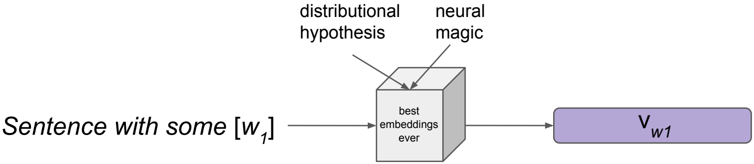
- **How to represent a phrase $\mathbf{p} = \mathbf{w_1} \ldots \mathbf{w_k}$?**
- Most straightforward:



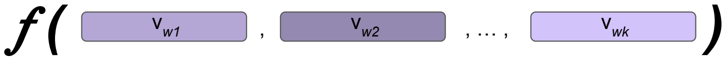
- "The whole is greater than the sum of its parts"

Representing Phrases

- Word representations are pretty much sorted out



- **How to represent a phrase $\mathbf{p = w_1 \ldots w_k}$?**
- Most straightforward:



- "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

- In noun compounds

Implicit Meaning

- In noun compounds

- In adjective-noun compositions

A **simple substance** is any sample of one of the known elements found in the Periodic Table of the Elements. Elements are made up of atoms of the same kind, and cannot be decomposed by any chemical means into any other simpler elements.

In this talk

1. **Testing Existing Text Representations**

   Can they handle the complexity of phrases?

In this talk

1. **Testing Existing Text Representations**

   Can they handle the complexity of phrases?

2. **Paraphrasing Noun-Compounds**

   A model for explicating noun compounds through paraphrases

In this talk

1. **Testing Existing Text Representations**
   Can they handle the complexity of phrases?

2. **Paraphrasing Noun-Compounds**
   A model for explicating noun compounds through paraphrases

3. **Future Directions**
   Thoughts about the future of phrase representations

Still a Pain in the Neck:

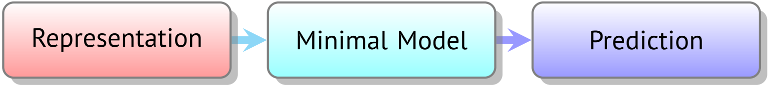
Evaluating Text Representations on Lexical Composition

Vered Shwartz and Ido Dagan

(in submission)

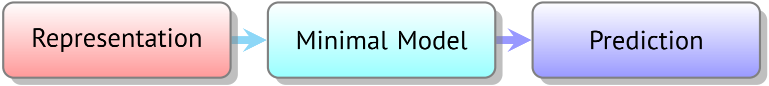
Can existing representations address these phenomena?
Probing Tasks

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



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]

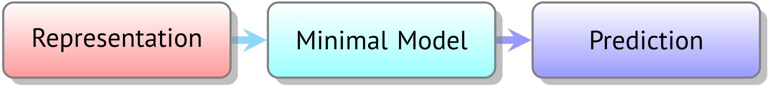


SkipThoughts(*s*)
InferSent(*s*)
...

What is *s*'s length?
Is *w* in *s*?
...

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]



SkipThoughts(*s*)
InferSent(*s*)
...

What is *s*'s length?
Is *w* in *s*?
...

- We follow the same for phrases, with various representations

Representations

Word Embeddings	Sentence Embeddings	Contextualized Word Embeddings
word2vec	SkipThoughts	ELMo
GloVe	InferSent*	OpenAI GPT
fastText	GenSen*	BERT

Representations

Word Embeddings	Sentence Embeddings	Contextualized Word Embeddings
word2vec GloVe fastText	SkipThoughts InferSent* GenSen*	ELMo OpenAI GPT BERT
- vector per word - context-agnostic	- vector per sentence	- vector per word - context-sensitive

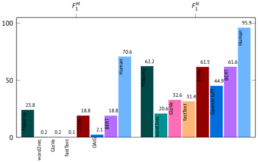
Representations

Word Embeddings	Sentence Embeddings	Contextualized Word Embeddings
word2vec GloVe fastText	SkipThoughts InferSent* GenSen*	ELMo OpenAI GPT BERT
- vector per word - context-agnostic	- vector per sentence	- vector per word - context-sensitive - named after characters from Sesame Street

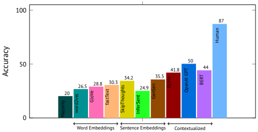
* supervised

Tasks and Results

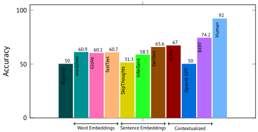
Phrase Type



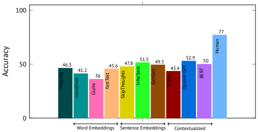
Noun Compound Literality



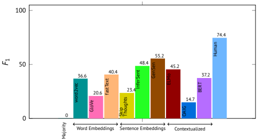
Noun Compound Relations



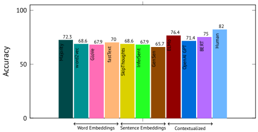
Adjective-Noun Relations



Adjective-Noun Entailment



Verb-particle Classification

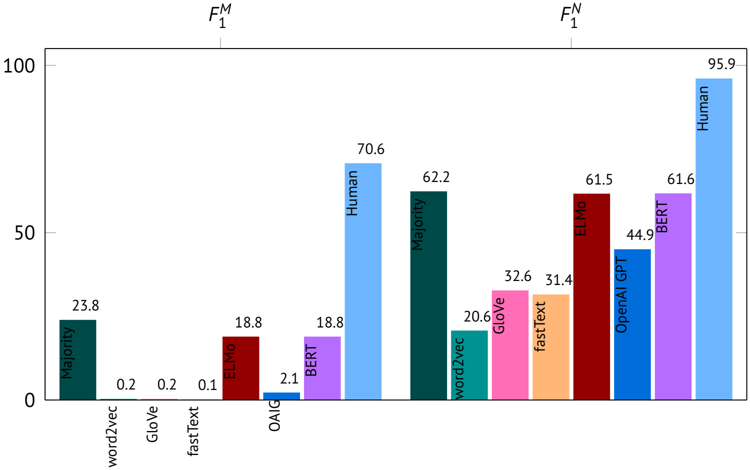


1. Phrase Type

O	B-MW_VPC	I-MW_VPC	B-MW_NC	I-MW_NC	O	O	O	O
Authorities	meted	out	summary	justice	in	cases	as	this

1. Phrase Type

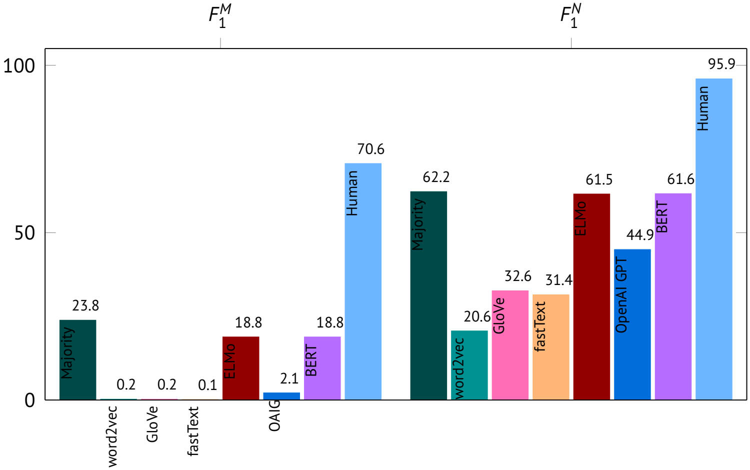
O	B-MW_VPC	I-MW_VPC	B-MW_NC	I-MW_NC	O	O	O	O
Authorities	meted	out	summary	justice	in	cases	as	this



(1) Failure to recognize phrase type

1. Phrase Type

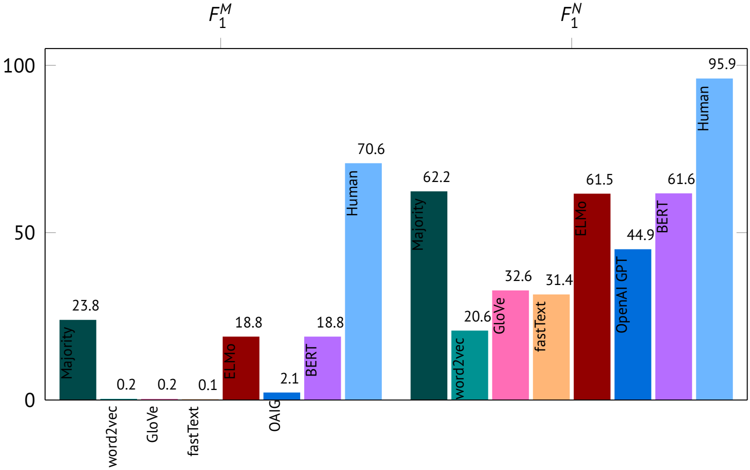
O	B-MW_VPC	I-MW_VPC	B-MW_NC	I-MW_NC	O	O	O	O
Authorities	meted	out	summary	justice	in	cases	as	this



(1) Failure to recognize phrase type; (2) Named entities are easier

1. Phrase Type

O	B-MW_VPC	I-MW_VPC	B-MW_NC	I-MW_NC	O	O	O	O
Authorities	meted	out	summary	justice	in	cases	as	this



(1) Failure to recognize phrase type; (2) Named entities are easier; (3) Context helps

2. Noun Compound Literality

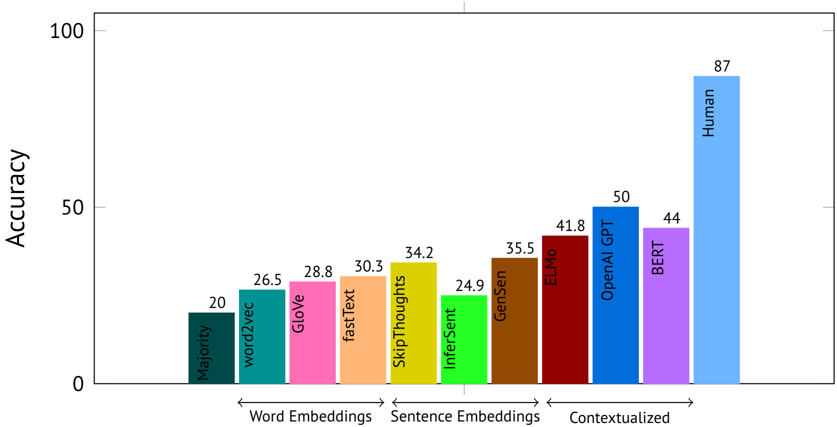
Non-Literal Literal

The crash course in litigation made me a better lawyer

2. Noun Compound Literality

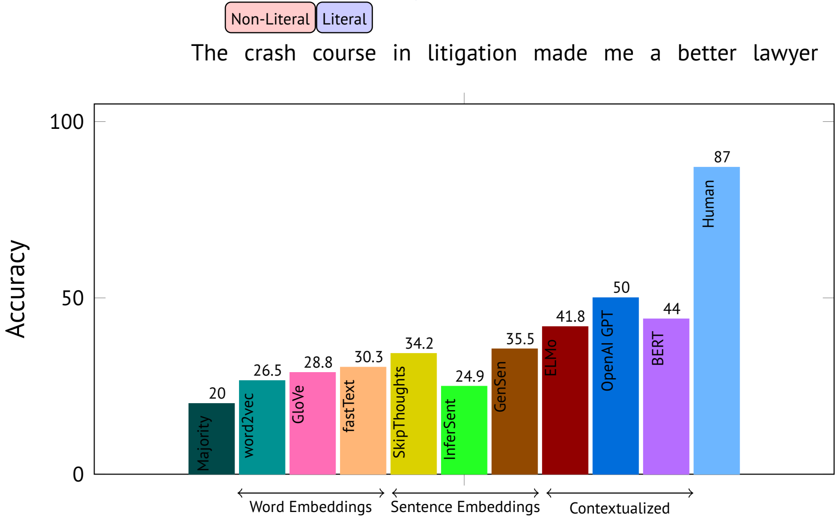
Non-Literal | Literal

The crash course in litigation made me a better lawyer



(1) word embeddings < sentence embeddings < contextualized

2. Noun Compound Literality



(1) word embeddings < sentence embeddings < contextualized; (2) Far from humans

2. Noun Compound Literality
Analysis

ELMo	OpenAI GPT	BERT
A search team located the [crash]$_L$ site and found small amounts of human remains.		
landfill	body	archaeological
wreckage	place	burial
Web	man	wreck
crash	missing	excavation
burial	location	grave
After a [crash]$_N$ course in tactics and maneuvers, the squadron was off to the war...		
crash	few	short
changing	while	successful
collision	moment	rigorous
training	long	brief
reversed	couple	training

2. Noun Compound Literality
Analysis

ELMo	OpenAI GPT	BERT
A search team located the [crash]$_L$ site and found small amounts of human remains.		
landfill	body	archaeological
wreckage	place	burial
Web	man	wreck
crash	missing	excavation
burial	location	grave
After a [crash]$_N$ course in tactics and maneuvers, the squadron was off to the war...		
crash	few	short
changing	while	successful
collision	moment	rigorous
training	long	brief
reversed	couple	training

(1) Literal: fewer errors

2. Noun Compound Literality
Analysis

ELMo	OpenAI GPT	BERT
A search team located the [crash]$_L$ site and found small amounts of human remains.		
landfill	body	archaeological
wreckage	place	burial
Web	man	wreck
crash	missing	excavation
burial	location	grave
After a [crash]$_N$ course in tactics and maneuvers, the squadron was off to the war...		
crash	few	short
changing	while	successful
collision	moment	rigorous
training	long	brief
reversed	couple	training

(1) Literal: fewer errors
(2) BERT > ELMo, both reasonable
(3) OpenAI GPT errs due to uni-directionality

2. Noun Compound Literality
Analysis

ELMo	OpenAI GPT	BERT
Growing up with a [silver]$_N$ spoon in his mouth, he was always cheerful...		
silver	mother	wooden
rubber	father	greasy
iron	lot	big
tin	big	silver
wooden	man	little

Things get tougher when both constituent nouns are non-literal!

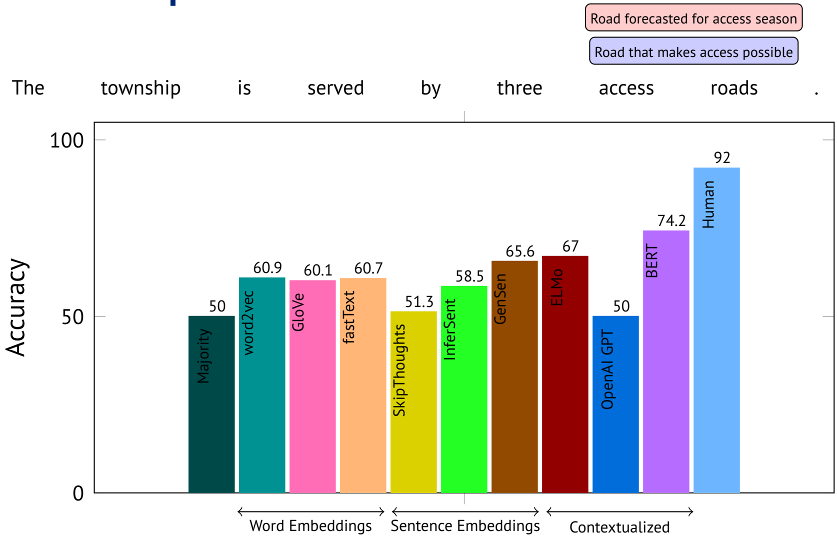
3. Noun Compound Relations

Road forecasted for access season

Road that makes access possible

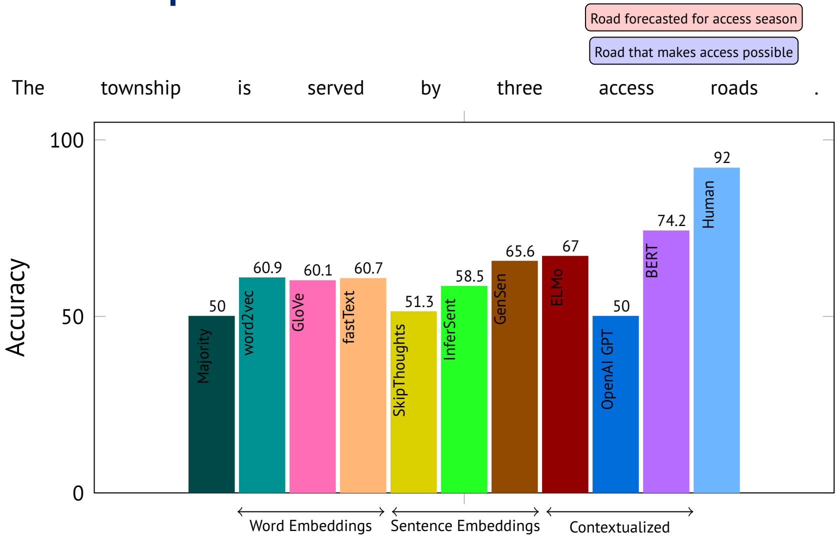
The township is served by three access roads .

3. Noun Compound Relations



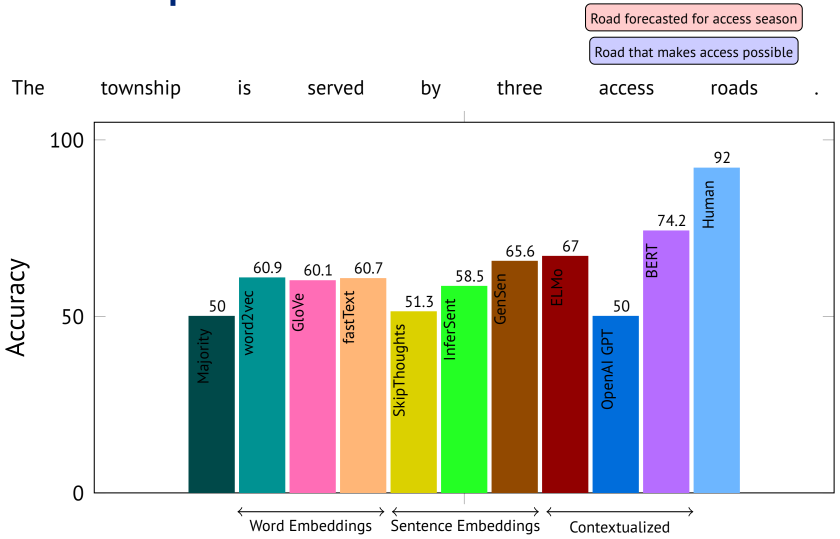
(1) word embeddings < sentence embeddings < contextualized

3. Noun Compound Relations



(1) word embeddings < sentence embeddings < contextualized; (2) Far from humans

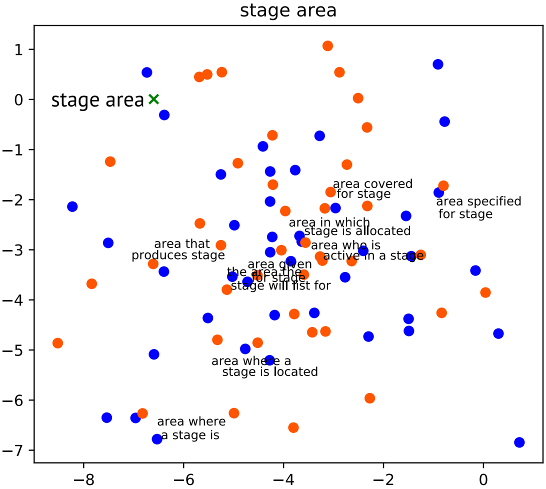
3. Noun Compound Relations



(1) word embeddings < sentence embeddings < contextualized; (2) Far from humans; (3) Open AI GPT fails

3. Noun Compound Relations

Analysis



No clear signal from BERT. Capturing implicit information is challenging!

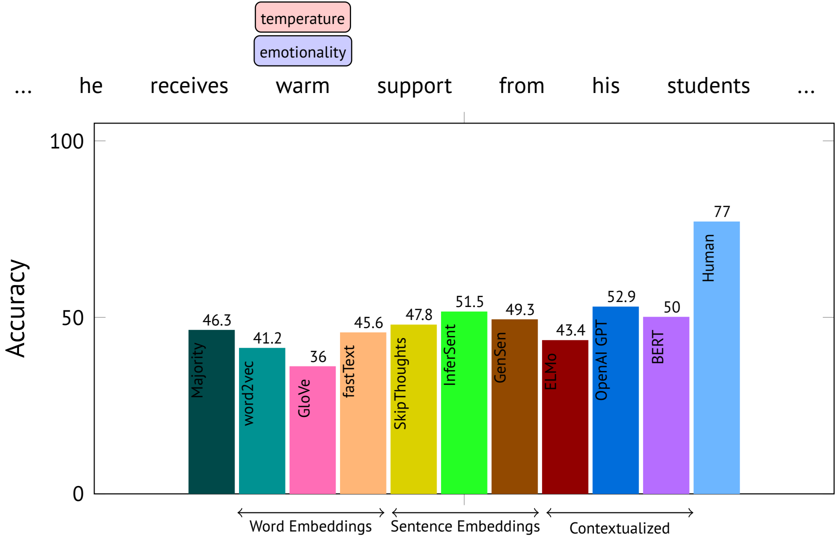
4. Adjective-Noun Relations

temperature

emotionality

... he receives warm support from his students ...

4. Adjective-Noun Relations



Best model performs only slightly better than majority

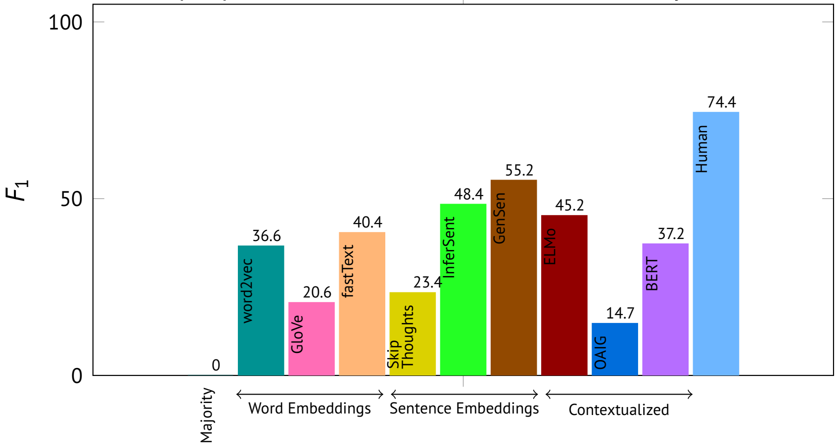
5. Adjective-Noun Entailment

Most people die in the class to which they were born $\rightarrow$

*Most people die in the **social** class to which they were born*

5. Adjective-Noun Entailment

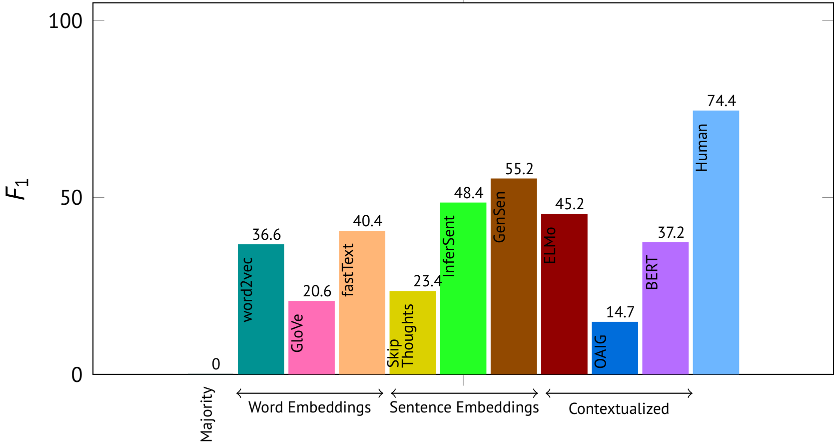
Most people die in the class to which they were born $\rightarrow$
Most people die in the ***social*** *class to which they were born*



(1) Bad performance for all models

5. Adjective-Noun Entailment

Most people die in the class to which they were born $\rightarrow$
*Most people die in the **social** class to which they were born*



(1) Bad performance for all models
(2) Best: sentence embeddings trained on RTE

6. Verb-Particle Classification

VPC

We did get on together

Non-VPC

Which response did you get on that?

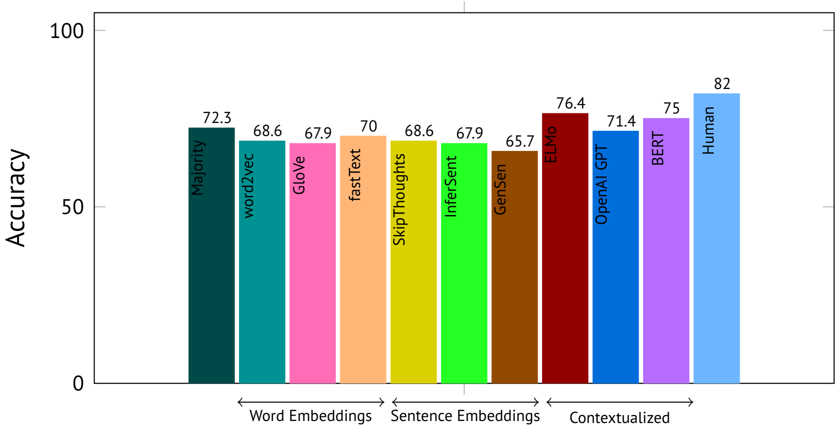
6. Verb-Particle Classification

VPC

We did get on together

Non-VPC

Which response did you get on that?



Similar performance for all models.

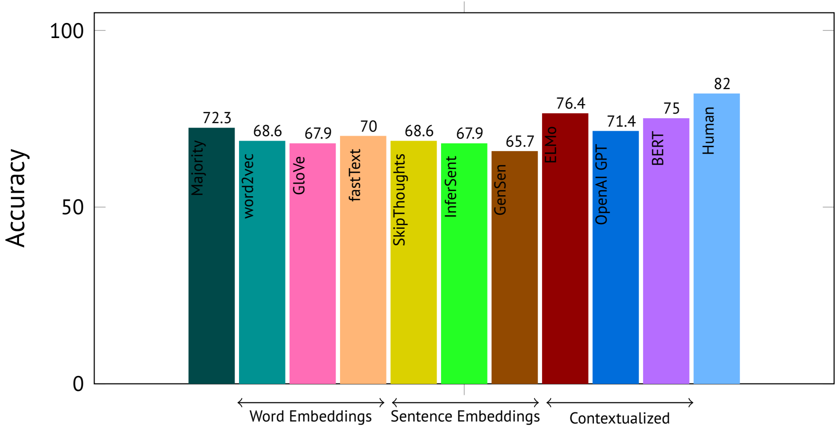
6. Verb-Particle Classification

VPC

We did get on together

Non-VPC

Which response did you get on that?

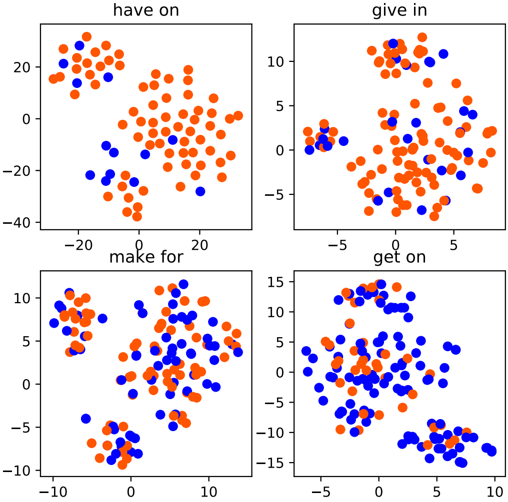


Similar performance for all models.

Is the good performance merely due to label imbalance?

6. Verb-Particle Classification

Analysis



Weak signal from ELMo. Mostly performs well due to label imbalance.

Paraphrase to Explicate:
Revealing Implicit Noun-Compound Relations

Vered Shwartz and Ido Dagan

(ACL 2018)

Interpreting Noun-Compounds

- Noun compounds are “text compression devices” [Nakov, 2013]

Interpreting Noun-Compounds

- Noun compounds are “text compression devices” [Nakov, 2013]
- We’re pretty good at decompressing them, even when we see them for the first time

Interpreting Noun-Compounds

- Noun compounds are “text compression devices” [Nakov, 2013]
- We’re pretty good at decompressing them, even when we see them for the first time

What is a “*parsley cake*”?

Interpreting Noun-Compounds

- Noun compounds are “text compression devices” [Nakov, 2013]
- We’re pretty good at decompressing them, 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?
- ...

Interpreting Noun-Compounds

- Noun compounds are "text compression devices" [Nakov, 2013]
- We're pretty good at decompressing them, 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?
- ...

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

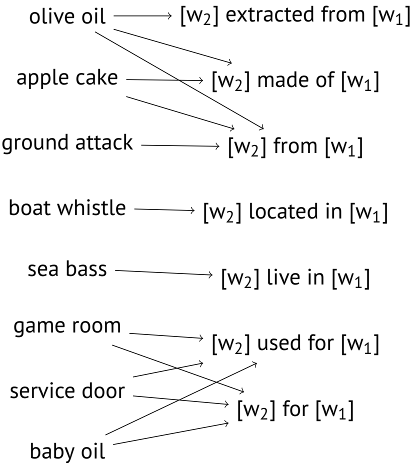
	CONTEXT ALL FORMS (SAMPLE): 100 200 500	FREQ
1	CAKE WITH CHOCOLATE	31
2	CAKE WITH LEMON	13
3	CAKE WITH STRAWBERRIES	10
4	CAKE WITH CANDLES	7
5	CAKE WITH CARAMEL	7
6	CAKE WITH FROSTING	6
7	CAKE WITH VANILLA	6
8	CAKE WITH BERRIES	5
9	CAKE WITH EGGS	4
10	CAKE WITH TOWEL	4
11	CAKE WITH RASPBERRY	3
12	CAKE 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 +

- Parsley (sort of) fits into this distribution

Noun-Compound Paraphrasing

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]



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 using **compositional distributional representations**

Prior Methods (2/2)

1. MELODI [Van de Cruys et al., 2013]:
   - Represent NC using **compositional distributional representations**
   - Predict paraphrase templates given NC vector

Prior Methods (2/2)

1. MELODI [Van de Cruys et al., 2013]:
   - Represent NC using **compositional distributional representations**
   - 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 using **compositional distributional representations**
   - 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 using **compositional distributional representations**
   - 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

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]"}

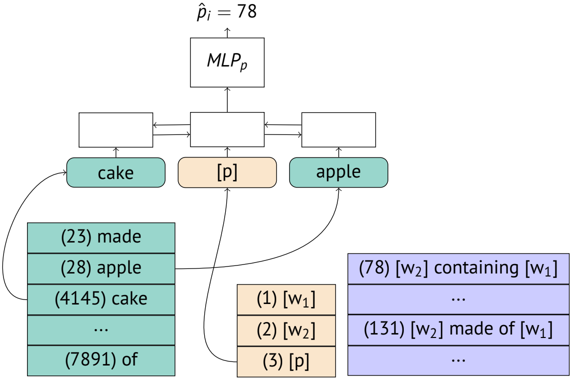
1. Predict a paraphrase p for a given NC $w_1 w_2$:
   What is the relation between *apple* and *cake*?

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

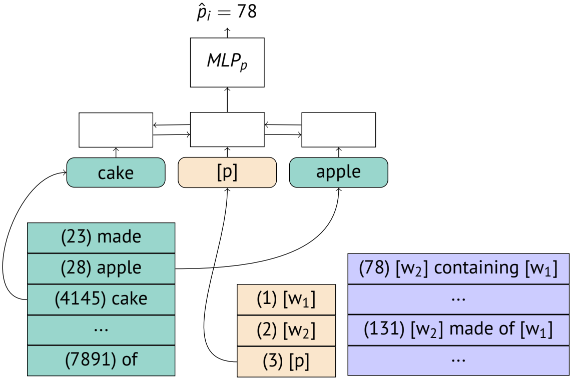
What is the relation between *apple* and *cake*?



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

Main Task (1): Predicting Paraphrases

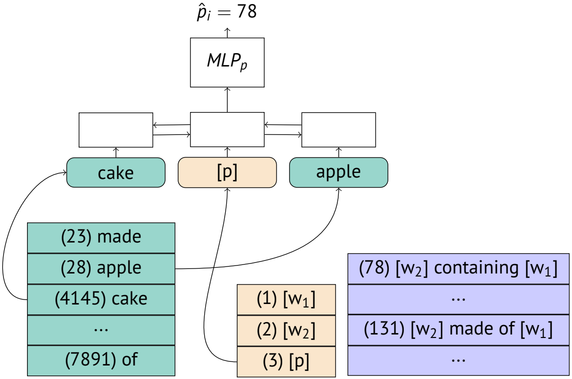
What is the relation between *apple* and *cake*?



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

Main Task (1): Predicting Paraphrases

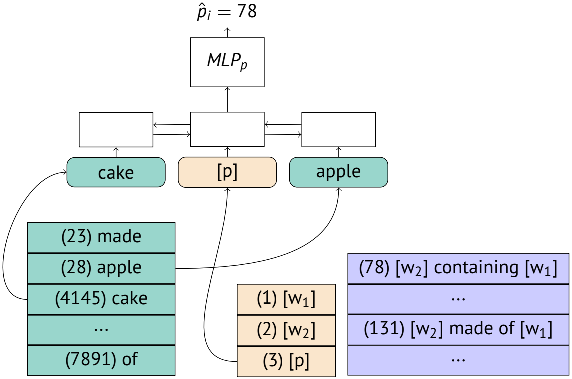
What is the relation between *apple* and *cake*?



- 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

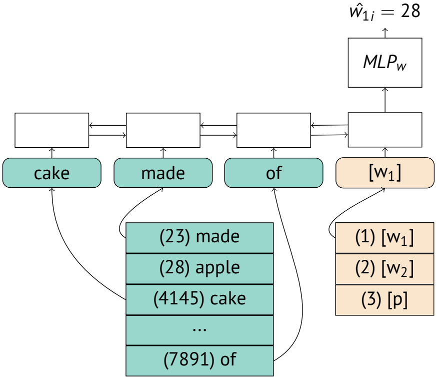
What is the relation between *apple* and *cake*?



- 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

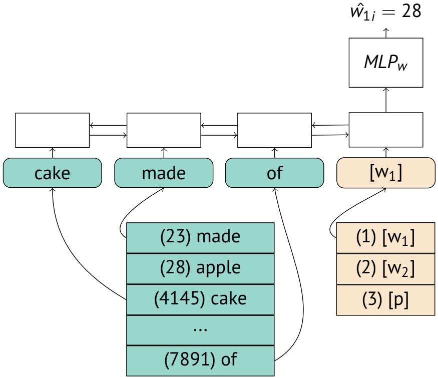
What can *cake* be made of?



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

Helper Task (2): Predicting Missing Constituents

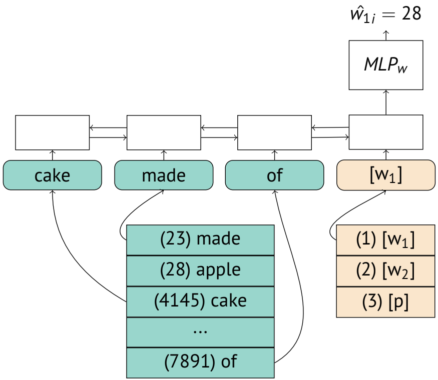
What can *cake* be made of?



- 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

Evaluation Setting

- Available dataset: SemEval 2013 task 4 [Hendrickx et al., 2013]
- Semi-supervised: infer templates of POS tags (e.g. "[w_2] verb prep [w_1]") from training data, use Google N-grams to generate training data

Evaluation Setting

- Available dataset: SemEval 2013 task 4 [Hendrickx et al., 2013]
- Semi-supervised: infer templates of POS tags (e.g. "[w_2] verb prep [w_1]") from training data, use Google N-grams to generate training data

- A *ranking* rather than a *retrieval* task
  - Systems expected to return a ranked list of paraphrases for each noun compound

Evaluation Setting

- Available dataset: SemEval 2013 task 4 [Hendrickx et al., 2013]
- Semi-supervised: infer templates of POS tags (e.g. "[w_2] verb prep [w_1]") from training data, use Google N-grams to generate training data

- A *ranking* rather than a *retrieval* task
  - Systems expected to return a ranked list of paraphrases for each noun compound
  - We implemented a ranking model that re-ranks the top *k* paraphrases retrieved by the model

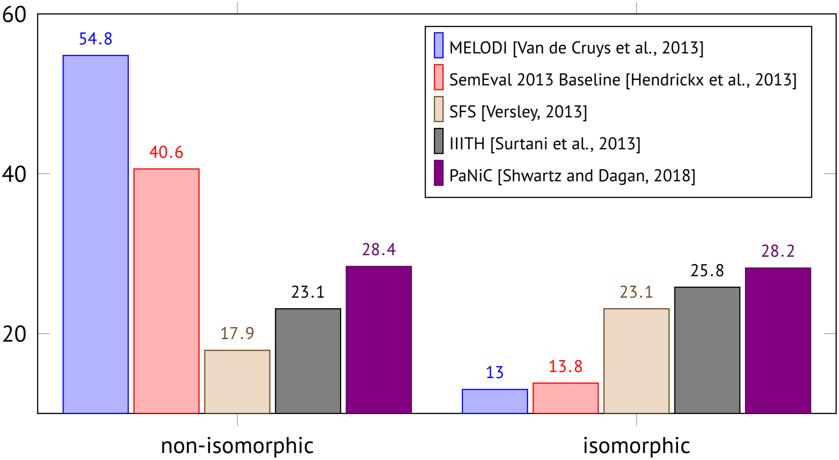
Evaluation Setting

- Available dataset: SemEval 2013 task 4 [Hendrickx et al., 2013]
- Semi-supervised: infer templates of POS tags (e.g. "[w_2] verb prep [w_1]") from training data, use Google N-grams to generate training data

- A *ranking* rather than a *retrieval* task
  - Systems expected to return a ranked list of paraphrases for each noun compound
  - We implemented a ranking model that re-ranks the top *k* paraphrases retrieved by the model

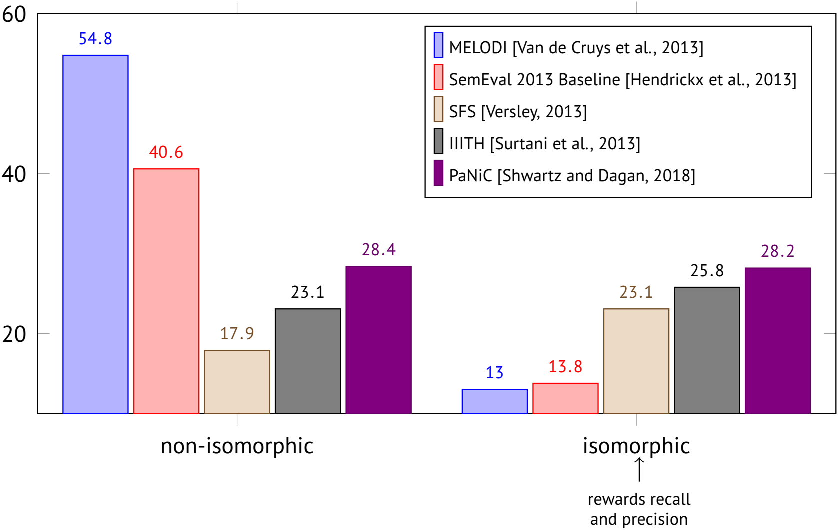
- Evaluation: based on n-gram overlap, provided evaluation script
  - Gold paraphrase score: how many annotators suggested it?

Results

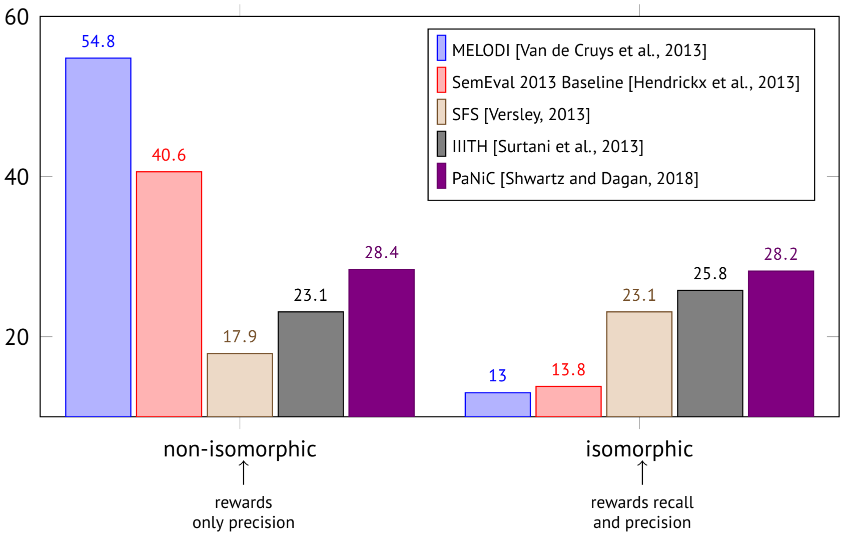


60
40
20
54.8
40.6
17.9
23.1
28.4
13
13.8
23.1
25.8
28.2
MELODI [Van de Cruys et al., 2013]
SemEval 2013 Baseline [Hendrickx et al., 2013]
SFS [Versley, 2013]
IIITH [Surtani et al., 2013]
PaNiC [Shwartz and Dagan, 2018]
non-isomorphic
isomorphic

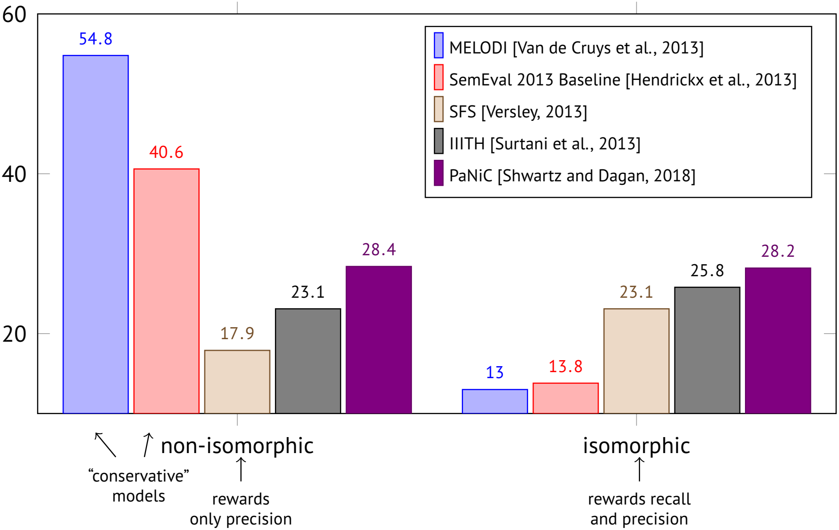
Results


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

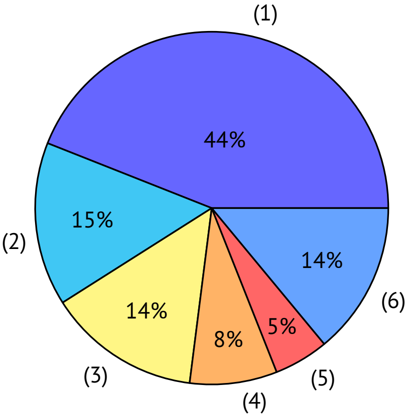
Results


60
40
20
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
rewards
only precision
isomorphic
rewards recall
and precision

Results

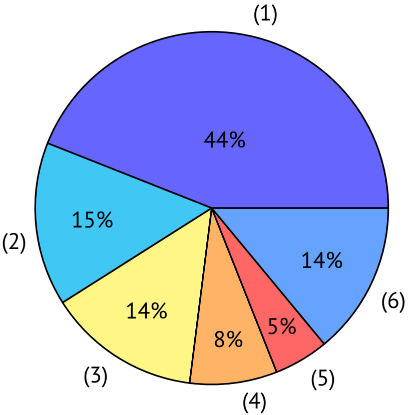

60
40
20
54.8
40.6
17.9
23.1
28.4
13
13.8
23.1
25.8
28.2
MELODI [Van de Cruys et al., 2013]
SemEval 2013 Baseline [Hendrickx et al., 2013]
SFS [Versley, 2013]
IIITH [Surtani et al., 2013]
PaNiC [Shwartz and Dagan, 2018]
non-isomorphic
isomorphic
“conservative” models
rewards only precision
rewards recall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

Error Analysis
False Positive



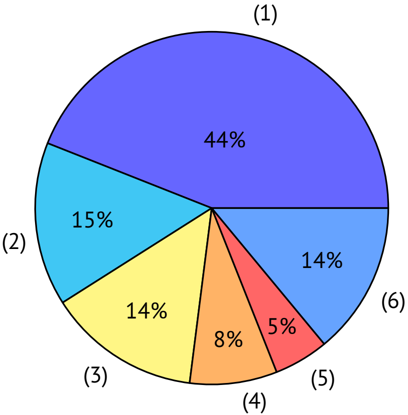
1. Valid, missing from gold-standard ("discussion by group")

Error Analysis
False Positive



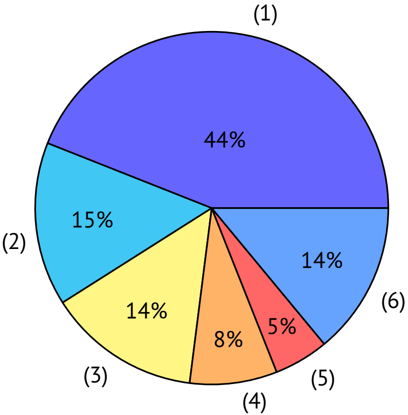
1. Valid, missing from gold-standard
   (“discussion by group”)
2. Too specific
   (“life *of women in* community”)

Error Analysis
False Positive



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

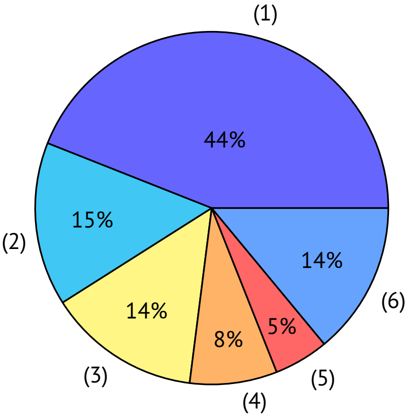
Error Analysis
False Positive



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”

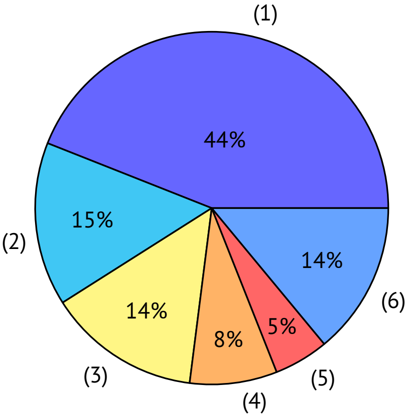
Error Analysis

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

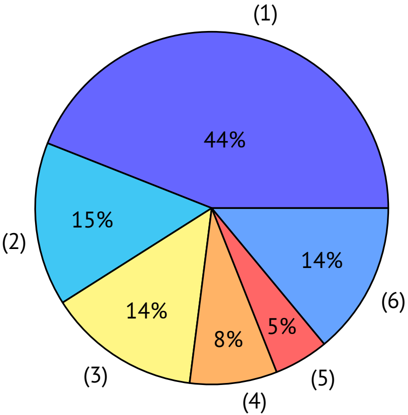
Error Analysis
False Positive



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")

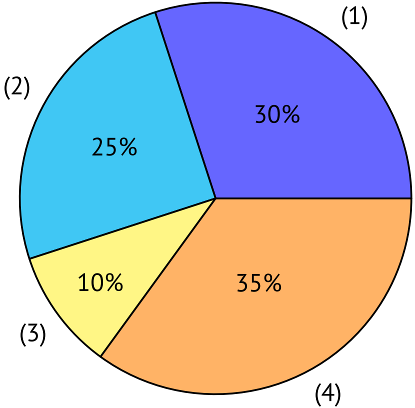
Error Analysis

False Positive



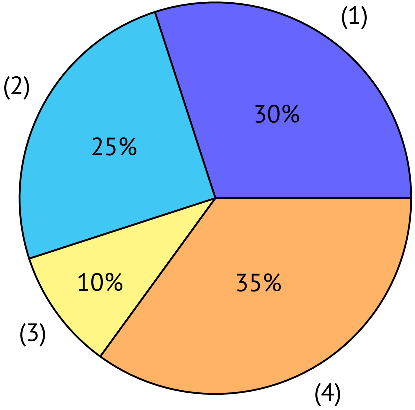
1. Valid, missing from gold-standard ("discussion by group")
2. Too specific ("life *of women in* community")
3. Incorrect prepositions
   E.g., n-grams don't respect syntactic structure: "rinse away the oil from baby 's head" ⇒ "oil from baby"
4. Syntactic errors
5. Borderline grammatical ("force of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

Error Analysis
False Negative



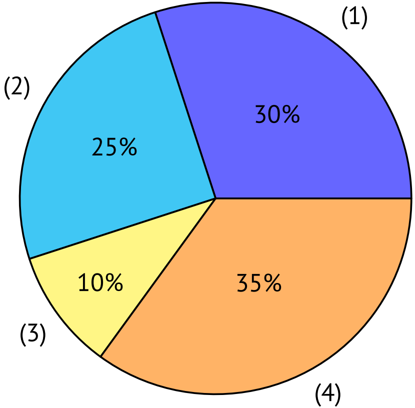
1. Long paraphrase ($n > 5$)

Error Analysis
False Negative



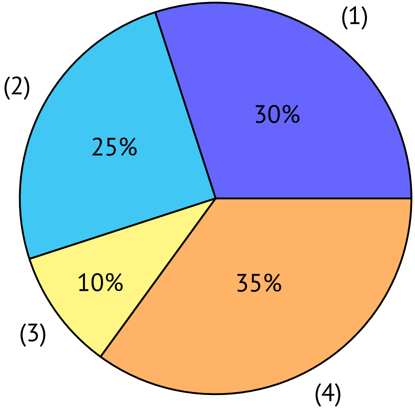
1. Long paraphrase ($n > 5$)
2. Determiners
   ("mutation of **a** gene")

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

Future Directions

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]

Recap

1. **Testing Existing Pre-trained Representations**

   Contextualized word embeddings provide better phrase representations, but there is still a long way to go

Recap

1. **Testing Existing Pre-trained Representations**

   Contextualized word embeddings provide better phrase representations, but there is still a long way to go

2. **Paraphrasing Noun-Compounds**

   Representations of compositional phrases can rely upon and generalize existing knowledge about similar concepts

Recap

1. **Testing Existing Pre-trained Representations**

   Contextualized word embeddings provide better phrase representations, but there is still a long way to go

2. **Paraphrasing Noun-Compounds**

   Representations of compositional phrases can rely upon and generalize existing knowledge about similar concepts

3. **Future Directions**

   To represent phrases like humans do, we need better context and world knowledge modeling

Recap

1. **Testing Existing Pre-trained Representations**

   Contextualized word embeddings provide better phrase representations, but there is still a long way to go

2. **Paraphrasing Noun-Compounds**

   Representations of compositional phrases can rely upon and generalize existing knowledge about similar concepts

3. **Future Directions**

   To represent phrases like humans do, we need better context and world knowledge modeling

Thank you!

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.

[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.

[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.

[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.