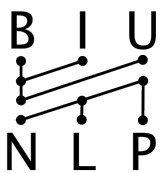
How well can Neural Text Representations Address Lexical Composition?

Vered Shwartz
Natural Language Processing Lab, Bar-Ilan University

AMLD, January 28, 2019

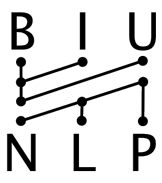


How well can Neural Text Representations Address Lexical Composition?

SPOILER ALERT
NOT TOO WELL

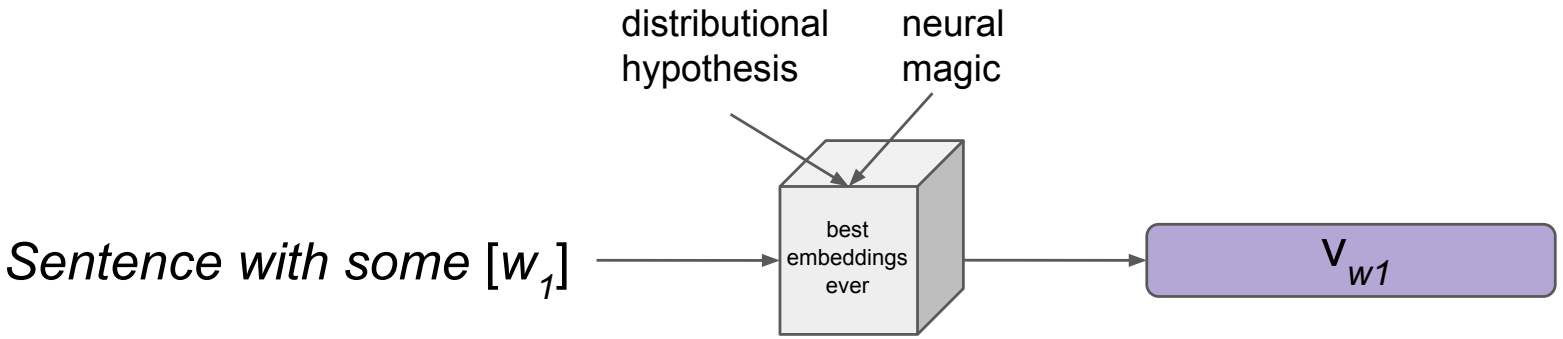
Vered Shwartz
Natural Language Processing Lab, Bar-Ilan University

AMLD, January 28, 2019



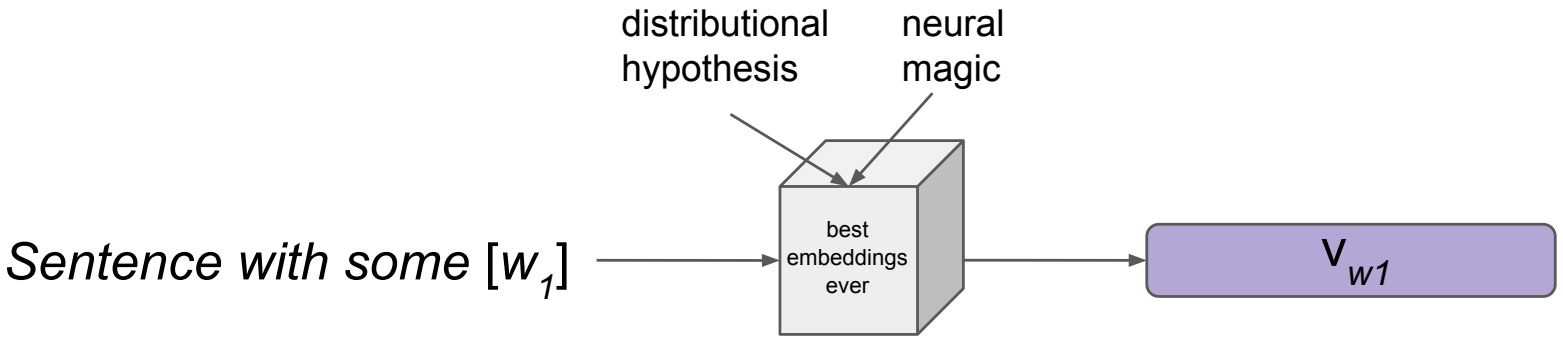
Representing Phrases

- Word representations are pretty much sorted out



Representing Phrases

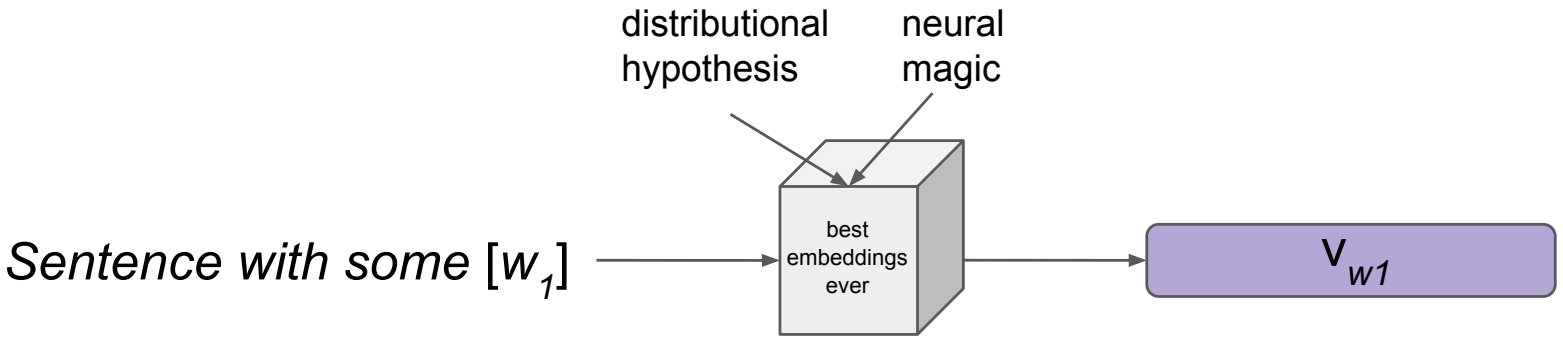
- Word representations are pretty much sorted out



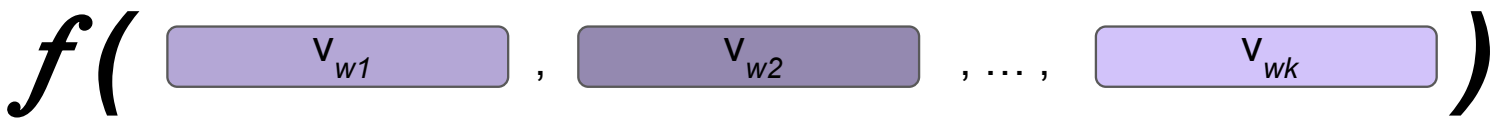
- **How to represent a phrase p = $w_1 \ldots w_k$?**

Representing Phrases

- Word representations are pretty much sorted out

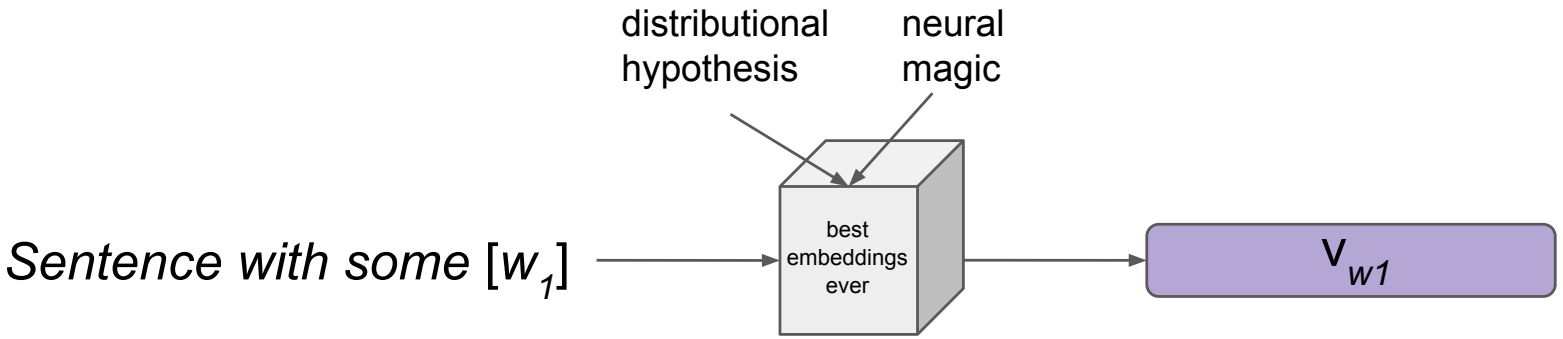


- **How to represent a phrase p = $w_1 \ldots w_k$?**
- Most straightforward:



Representing Phrases

- Word representations are pretty much sorted out



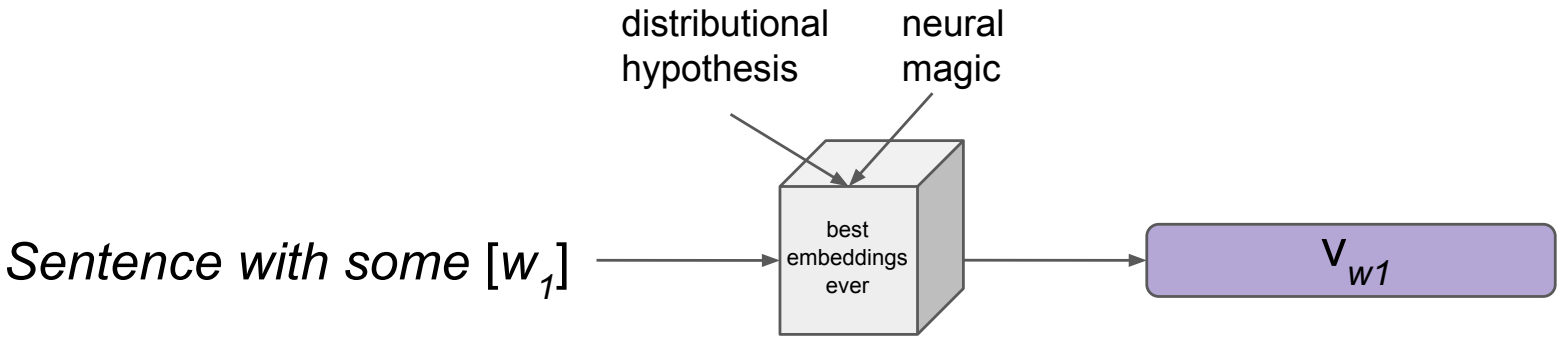
- **How to represent a phrase p = $w_1 \ldots w_k$?**
- Most straightforward:

$$f(\ v_{w1}\ ,\ v_{w2}\ ,\ \ldots\ ,\ v_{wk}\ )$$

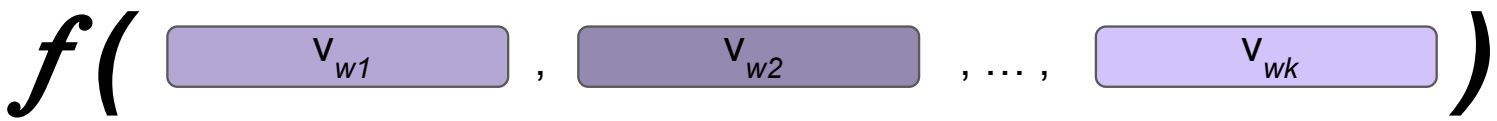
- "The whole is greater than the sum of its parts"

Representing Phrases

- Word representations are pretty much sorted out



- **How to represent a phrase 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

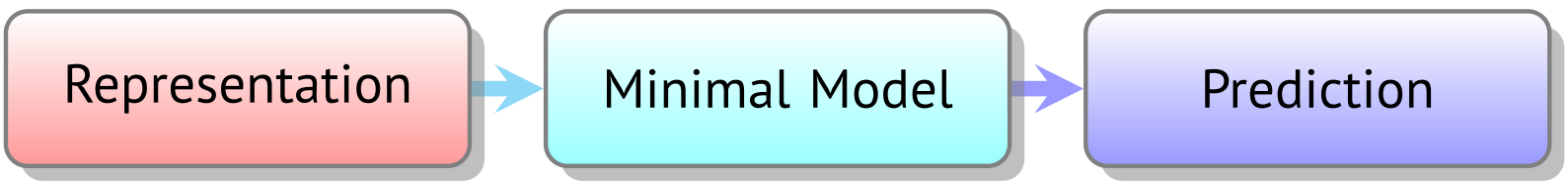
- In adjective-noun compositions

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.

Can existing representations address these phenomena?

Probing Tasks

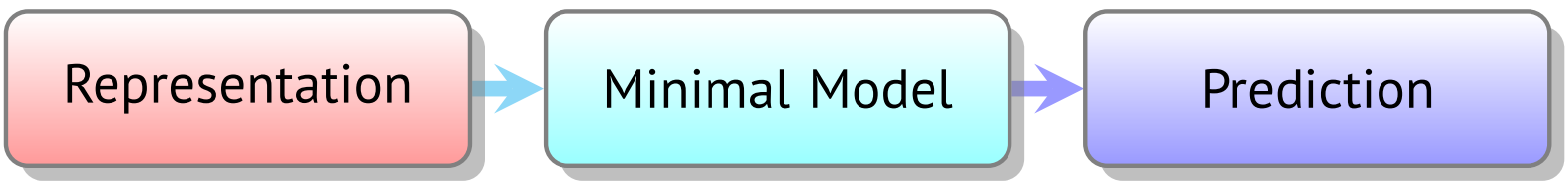
- Simple tasks designed to test a single linguistic property [Adi et al., 2017, Conneau et al., 2018]



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]



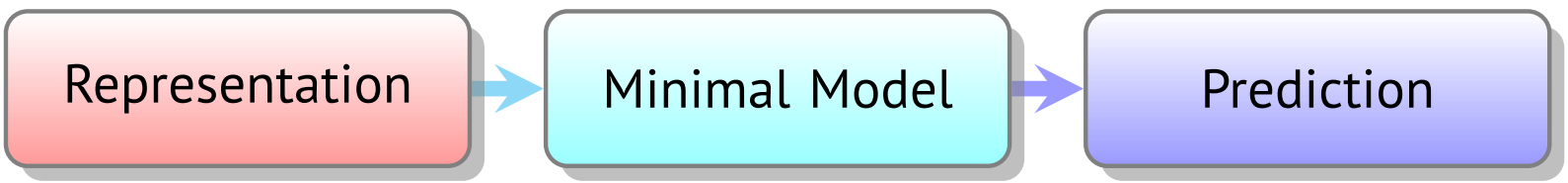
SkipThoughts(*s*)
InferSent(*s*)
…

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

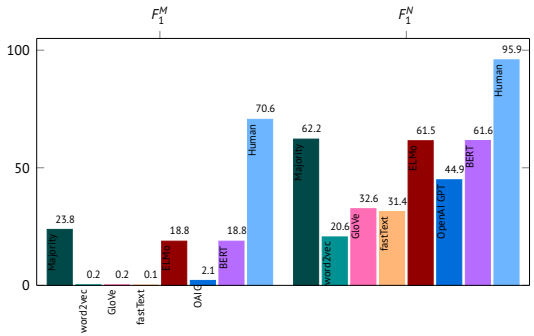
Representations

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

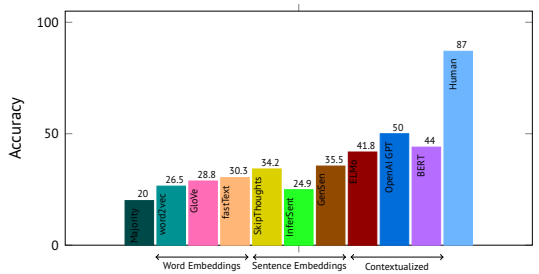
* supervised

Tasks and Results

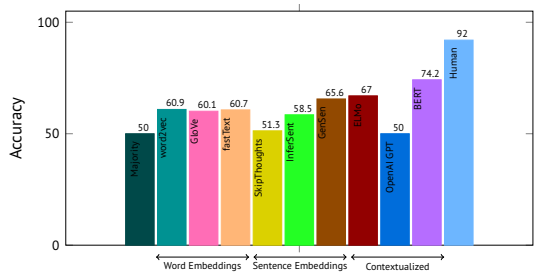
Phrase Type



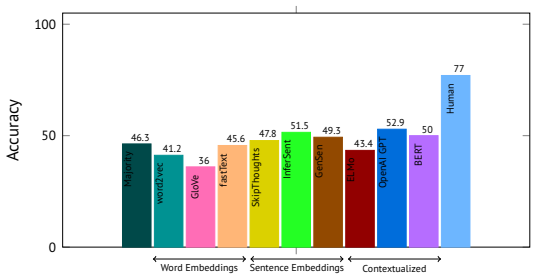
Noun Compound Literality



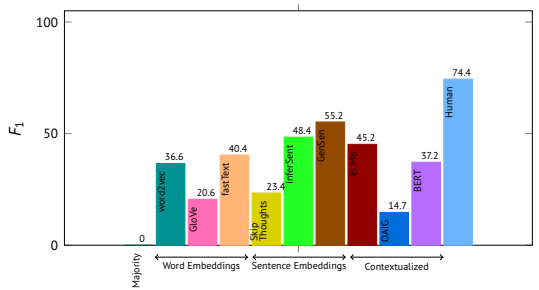
Noun Compound Relations



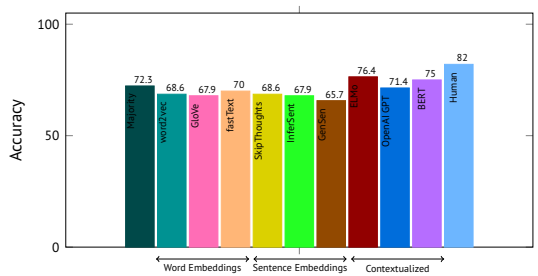
Adjective-Noun Relations



Adjective-Noun Entailment



Verb-particle Classification



Noun Compound Literality

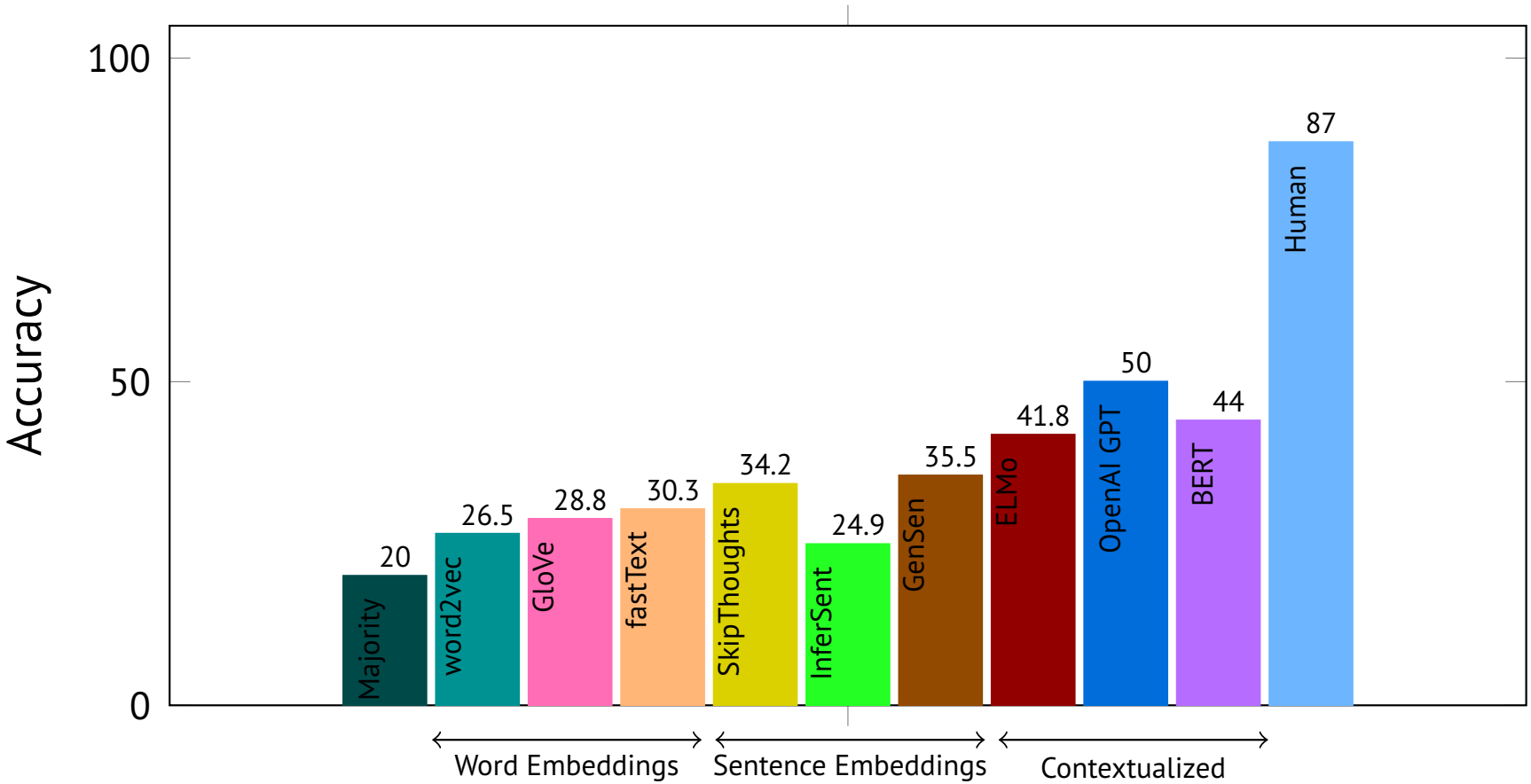
Non-Literal | Literal

The crash course in litigation made me a better lawyer

Noun Compound Literality

Non-Literal | Literal

The crash course in litigation made me a better lawyer

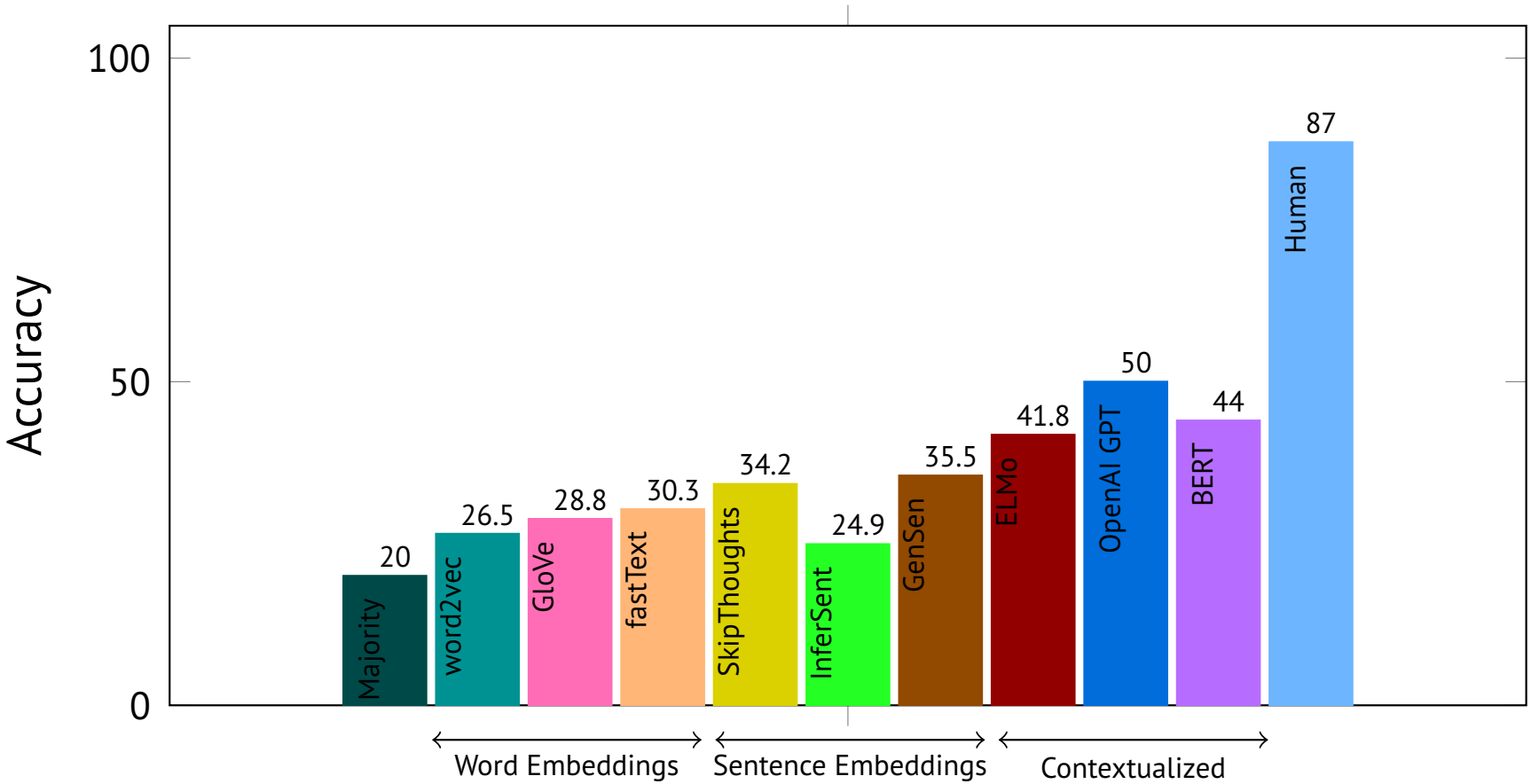


(1) word embeddings < sentence embeddings < contextualized

Noun Compound Literality

Non-Literal | Literal

The crash course in litigation made me a better lawyer



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

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

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

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

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!

Noun Compound Relations

Road forecasted for access season

Road that makes access possible

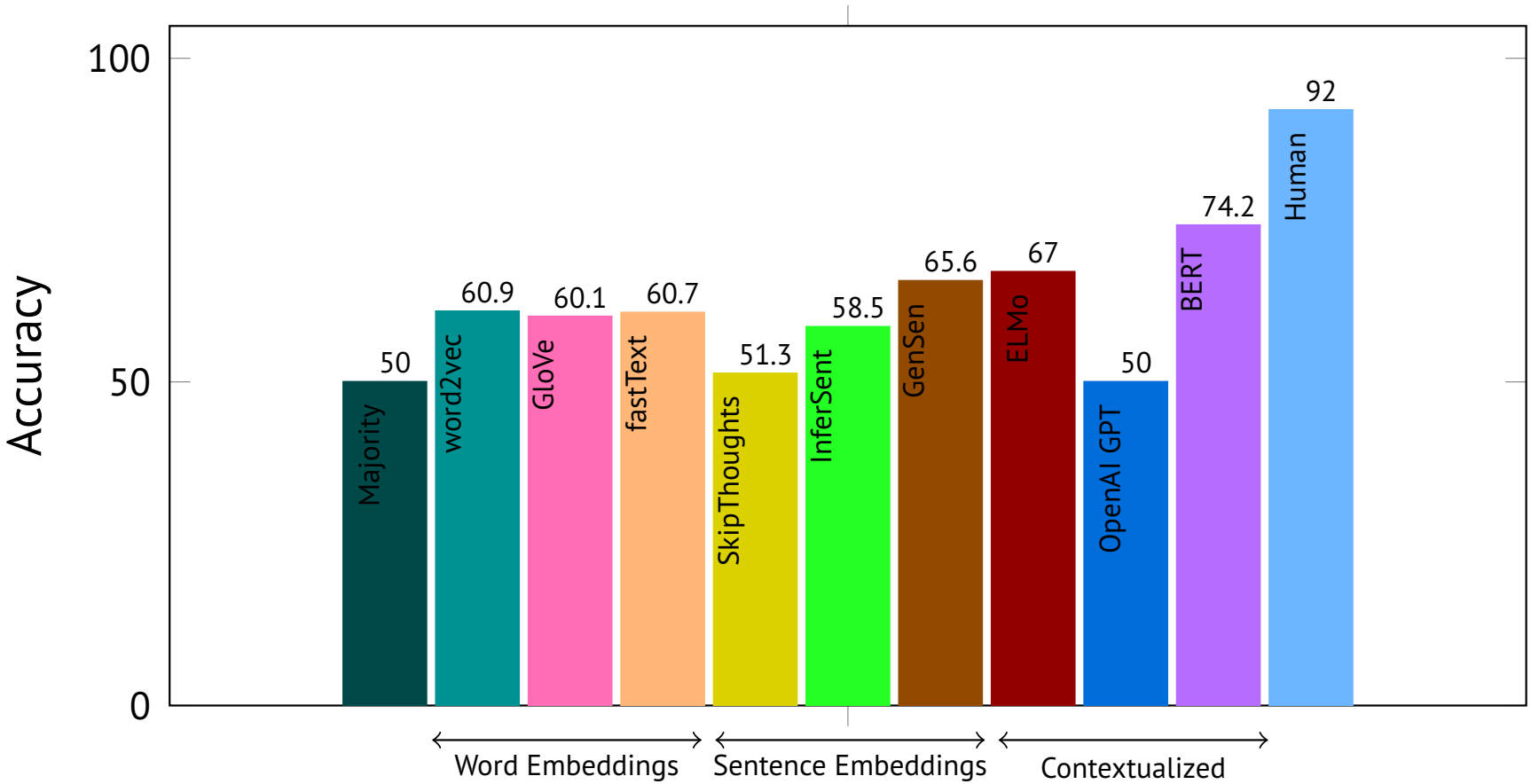
The township is served by three access roads .

Noun Compound Relations

Road forecasted for access season

Road that makes access possible

The township is served by three access roads .



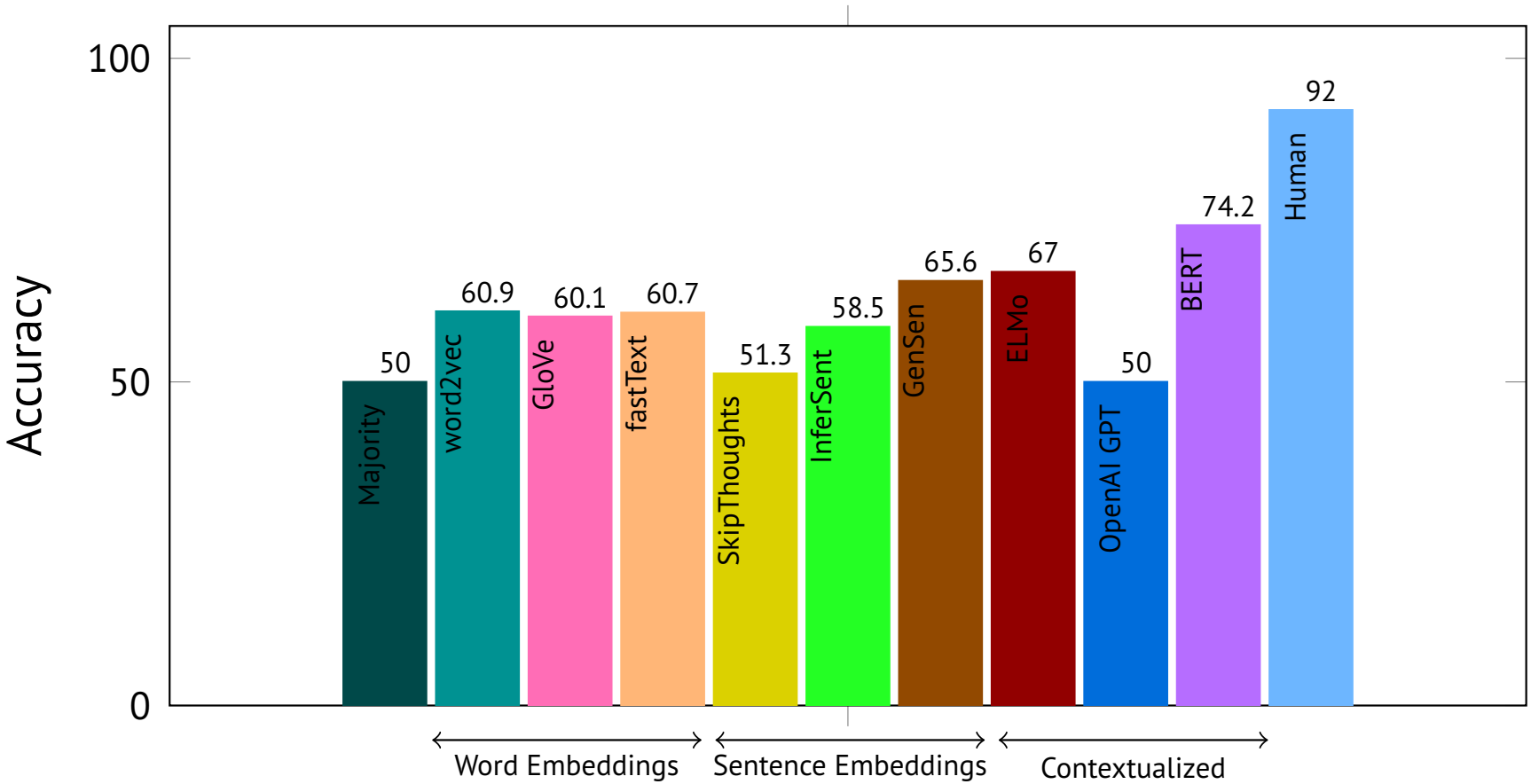
(1) word embeddings < sentence embeddings < contextualized

Noun Compound Relations

Road forecasted for access season

Road that makes access possible

The township is served by three access roads .



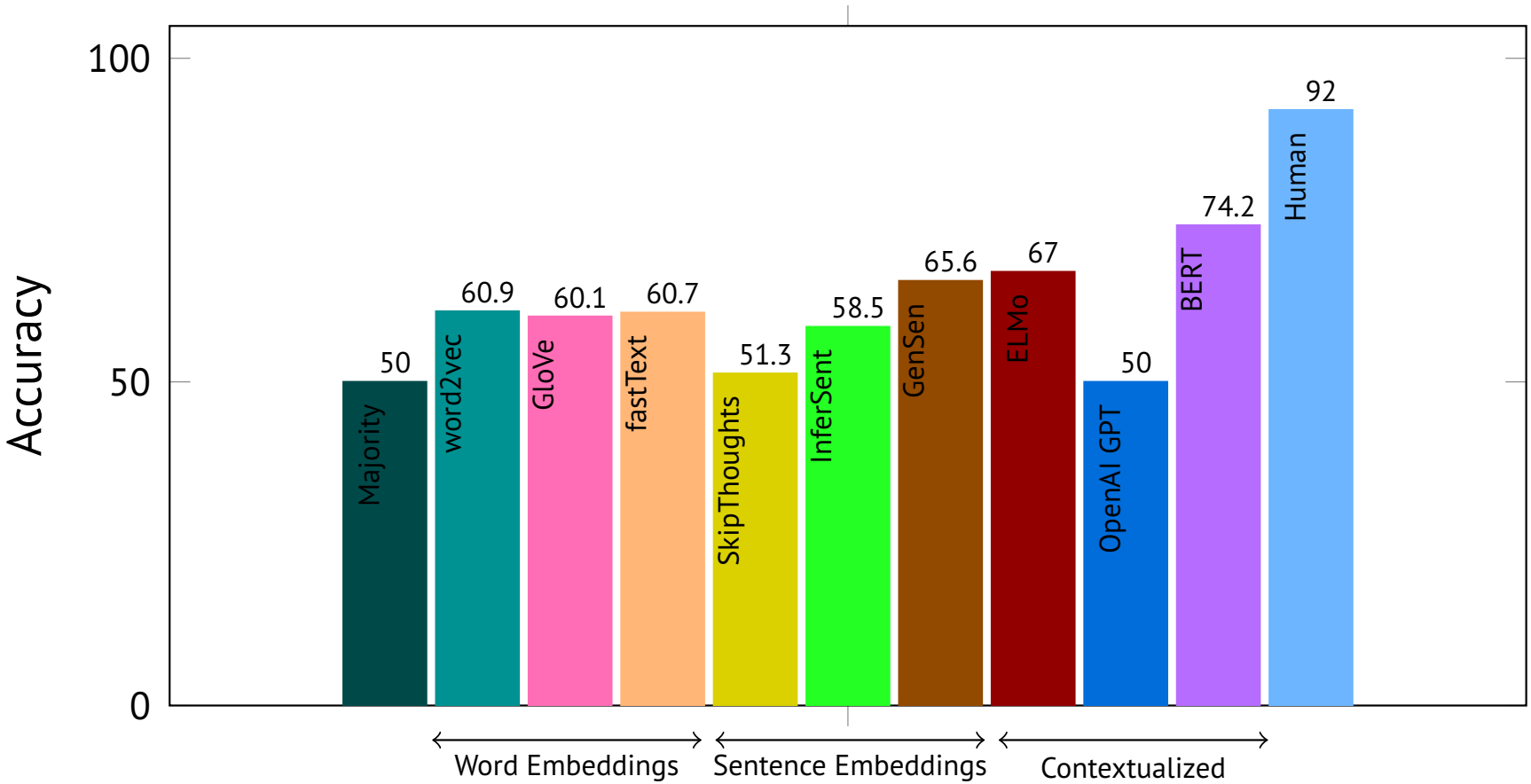
(1) word embeddings < sentence embeddings < contextualized; (2) Far from humans

Noun Compound Relations

Road forecasted for access season

Road that makes access possible

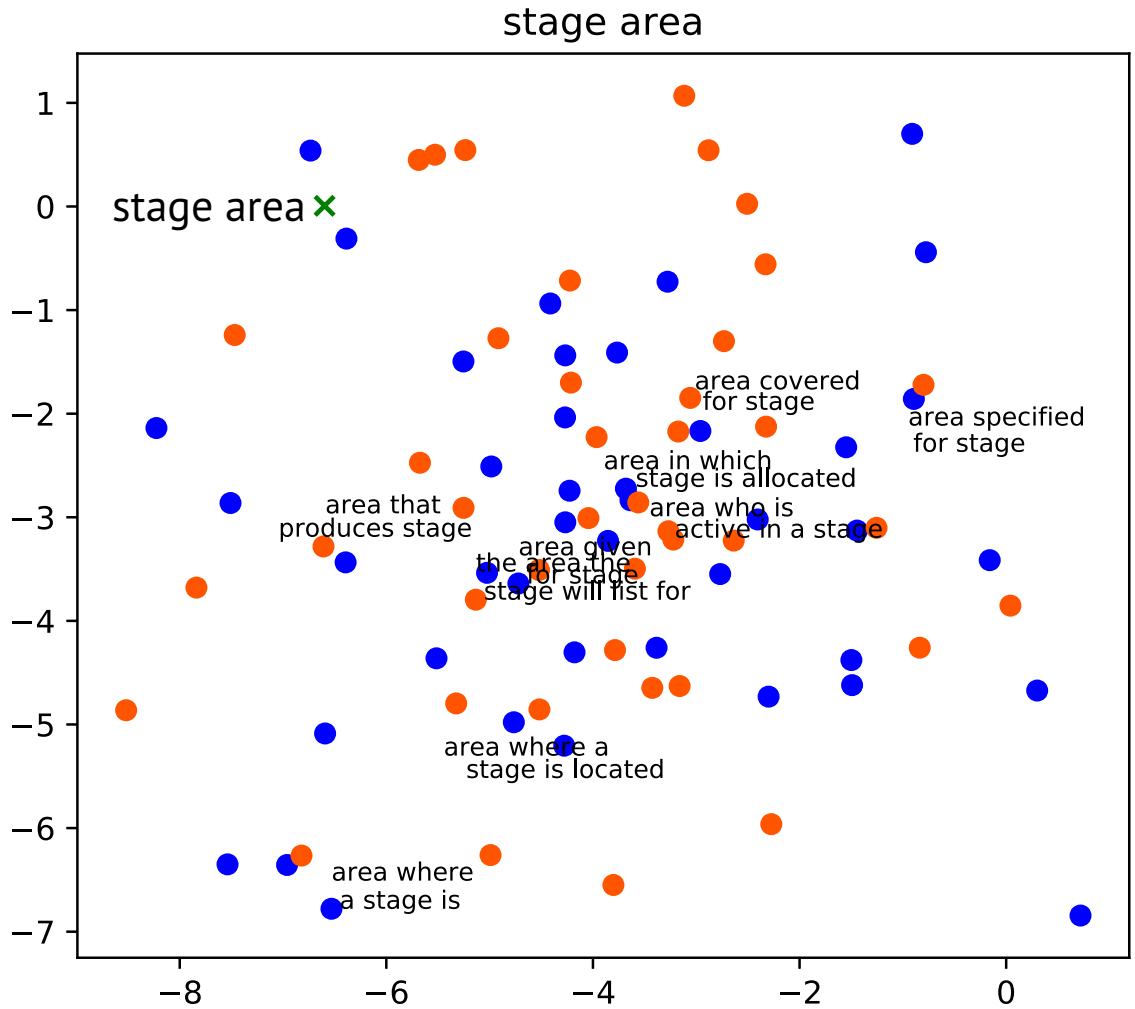
The township is served by three access roads .



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

Noun Compound Relations

Analysis



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

Verb-Particle Classification

VPC

We did get on together

Non-VPC

Which response did you get on that?

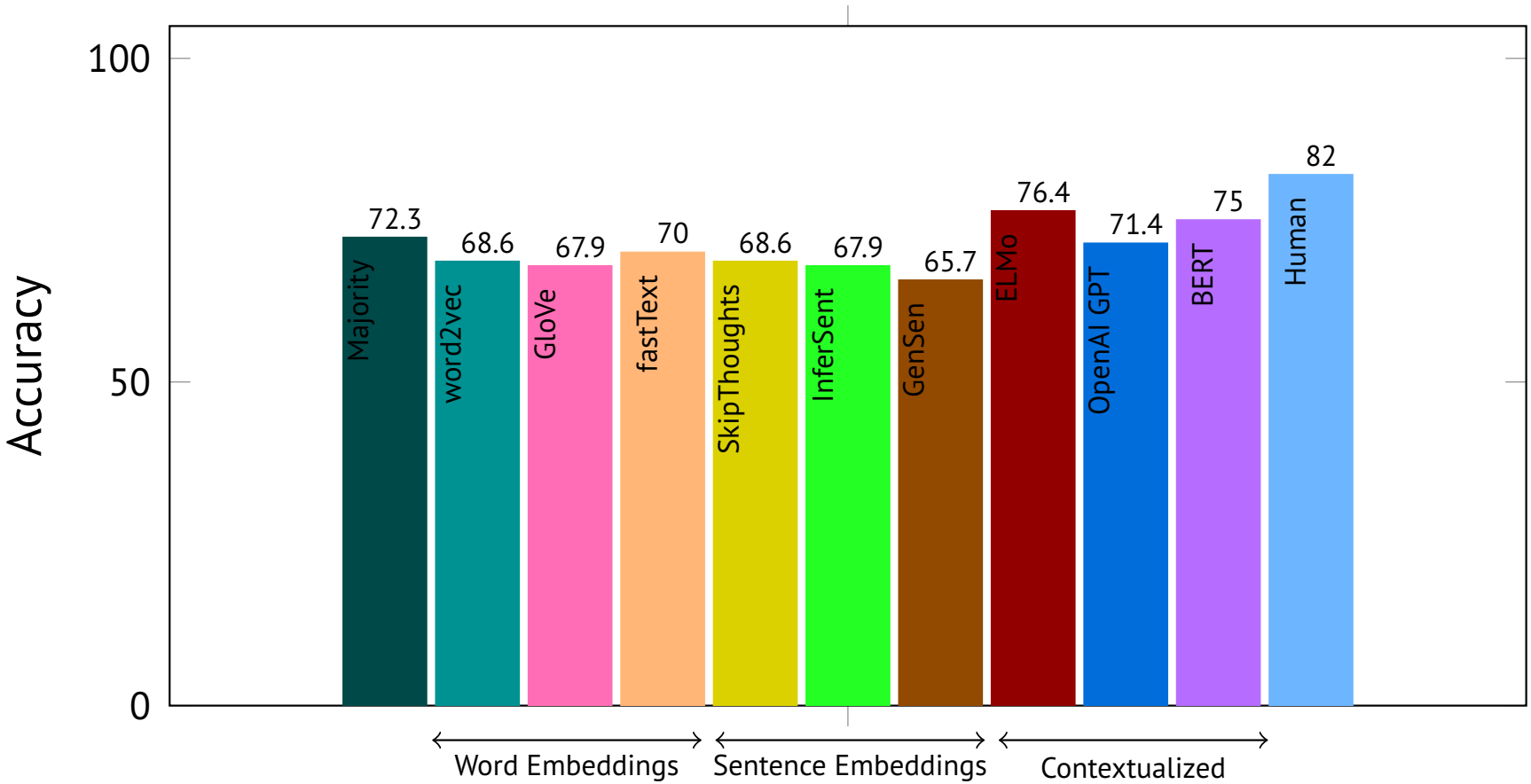
Verb-Particle Classification

VPC

We did get on together

Non-VPC

Which response did you get on that?



Similar performance for all models.

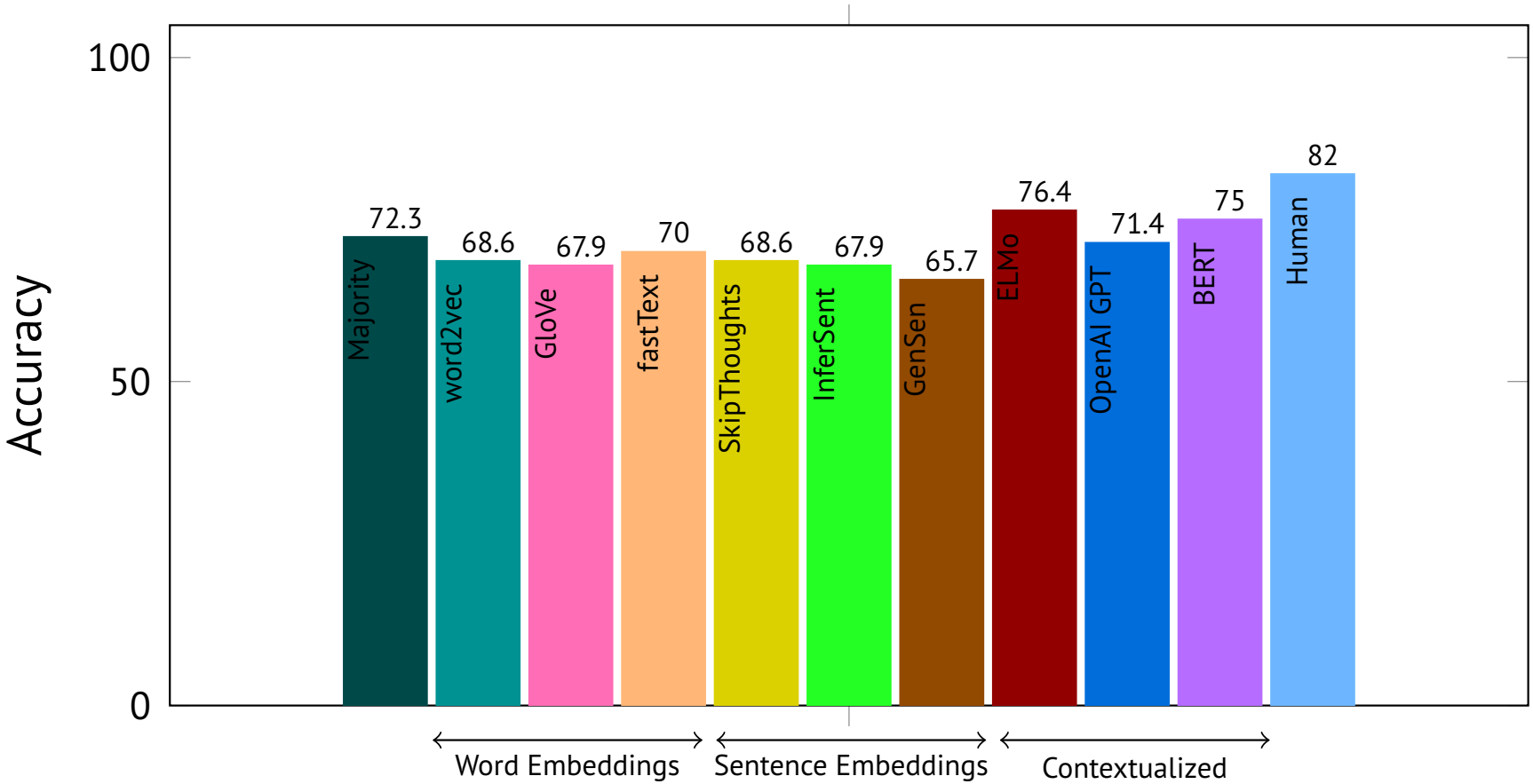
Verb-Particle Classification

VPC

We did get on together

Non-VPC

Which response did you get on that?

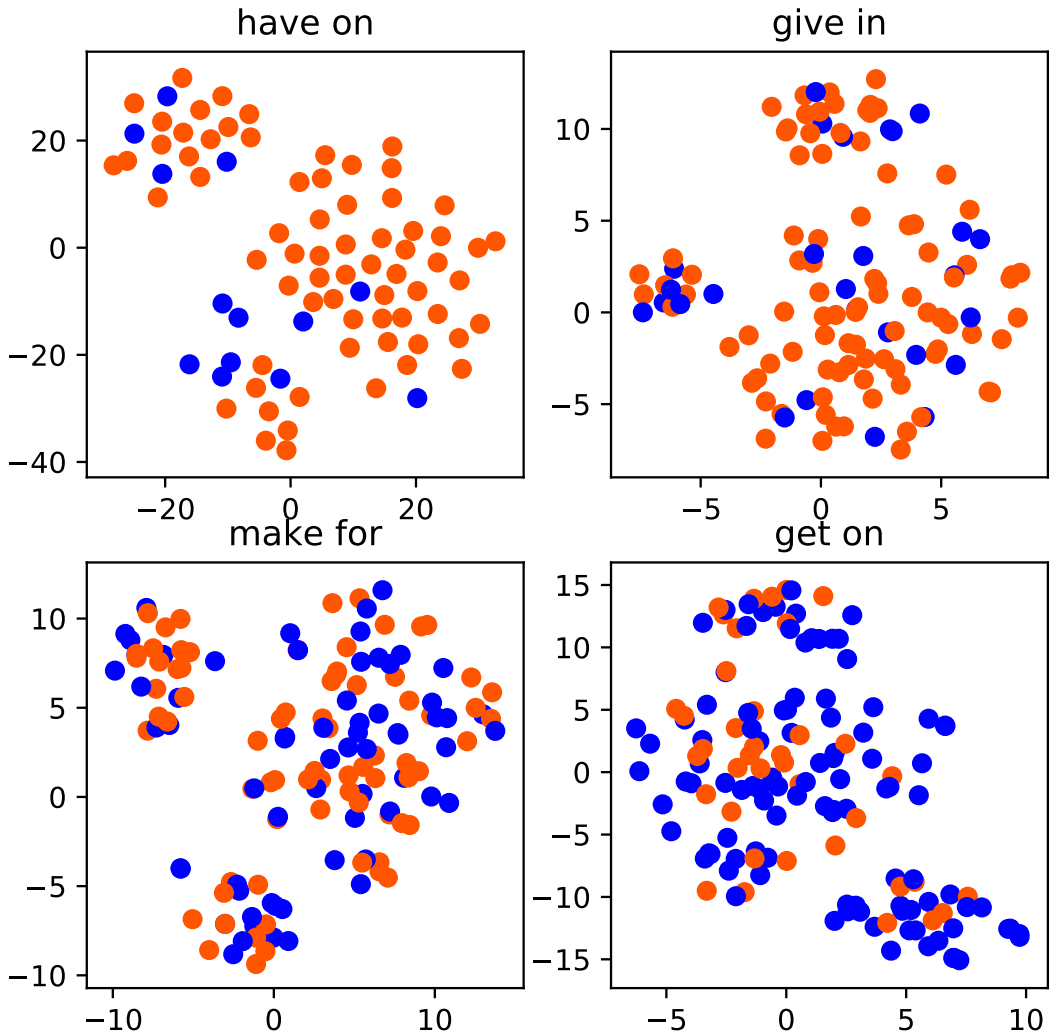


Similar performance for all models.

Is the good performance merely due to label imbalance?

Verb-Particle Classification

Analysis



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

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

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

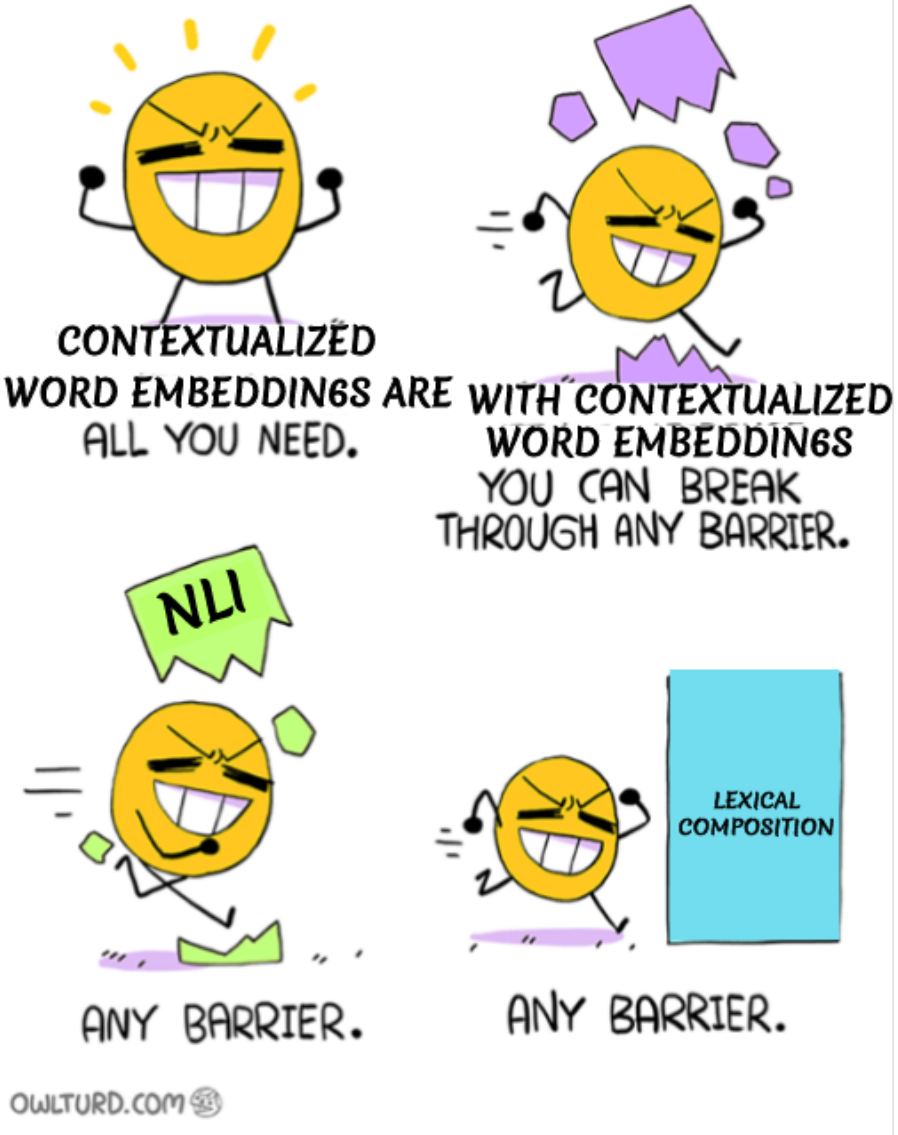
"Cradle is something you put the baby in"

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!

vered1986@gmail.com

@VeredShwartz

References

[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.