Shades of meaning composition: Defining compositionality goals in NLU

Allyson Ettinger

The Challenge of Compositionality for AI

June 29 panel

Why compositionality matters for AI?

Language understanding

The magenta tiger recited the ballad but did not forgive the vice principal

COMPOSITION

↓

MEANING

Compositionality stands as the critical alternative to infinite memorization.

Some obvious things

1) Memorization is a part of human language understanding

2) Humans can understand phrases/sentences that are novel, strange, improbable, and nonsensical

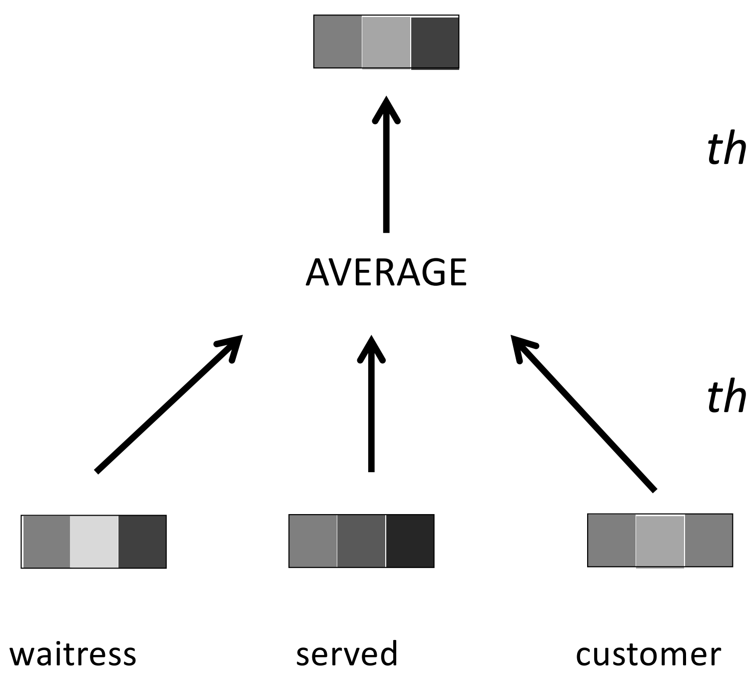
Refining our goals

Default definition of compositionality:

“Meaning of the whole is a function of the meanings of the parts”

… trivially satisfied and not terribly helpful for solving NLU

Trivial compositionality



the waitress served the customer

??

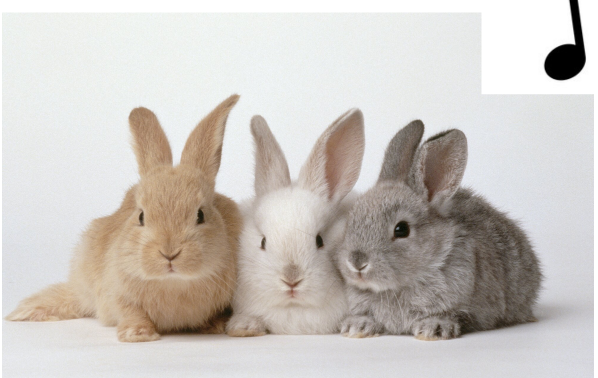
the customer served the waitress

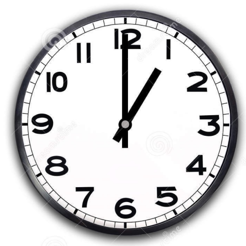
Refining our goals

What we need is *accurate*, *human-like* extraction of *compositional meanings* from language inputs

Meaning extraction

Three singing rabbits walked into the local bar last Wednesday afternoon





Monday	
Tuesday	
Wednesday	
Thursday	
Friday	
Saturday & Sunday	

Shades of composition

- “Syntactic angles” vs “Semantic angles”
- “Supervised angles” vs “Pre-trained NLU angles”

Shades of composition

- **“Syntactic angles” vs “Semantic angles”**
- “Supervised angles” vs “Pre-trained NLU angles”

Shades of composition

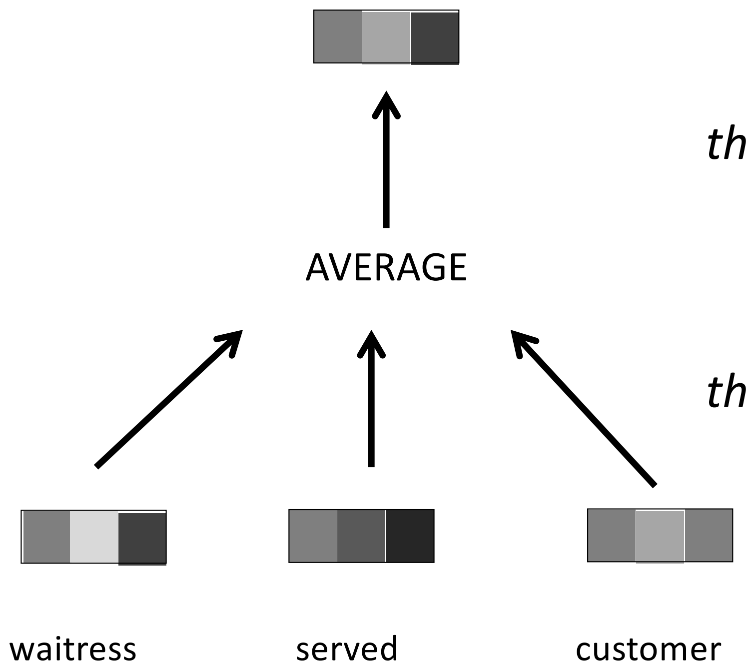
- **“Syntactic angles”** vs “Semantic angles”

- “Supervised angles” vs “Pre-trained NLU angles”

Syntactic angles

- Ability to bind components of a sentence to their correct roles

Syntactic angles



the waitress served the customer

??

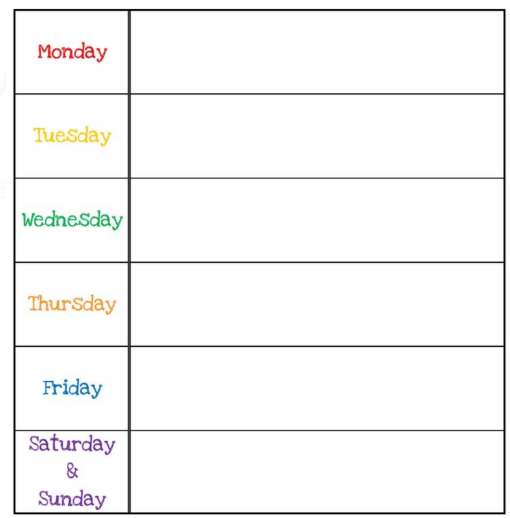
the customer served the waitress

Syntactic angles

Three singing rabbits walked into the local bar last Wednesday afternoon

MODIFIER OF RABBITS

TIME/DAY OF WALKING



AGENTS OF WALKING

LOCATION/DESTINATION OF WALKING

Shades of composition

- **“Syntactic angles”** vs “Semantic angles”
- “Supervised angles” vs “Pre-trained NLU angles”

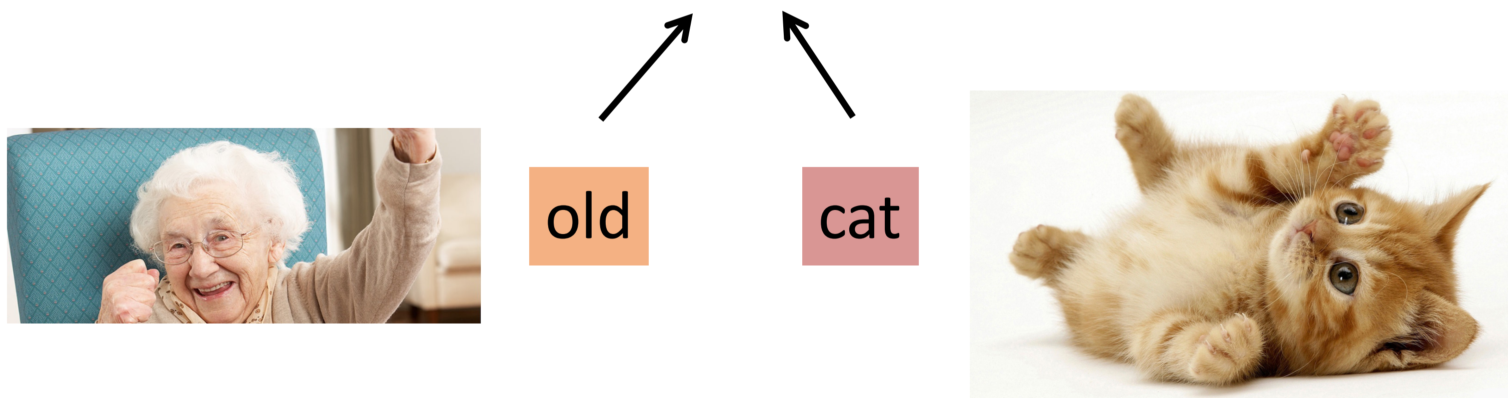
Shades of composition

- “Syntactic angles” vs **“Semantic angles”**
- “Supervised angles” vs “Pre-trained NLU angles”

Semantic angles

- Beyond roles and order of composition
- Are we capturing correct features of composed phrases/sentences?
- As well as implications of those meanings for a given task?

Semantic angles



old cat

old

cat

old cat

old

cat

Semantic angles

Three singing bars walked into the local rabbit last Wednesday afternoon

Semantic angles

Sebastian lives in France. The capital of Sebastian's country is ___

Pandia & Ettinger (2021). *Sorting through the noise: Testing robustness of information processing in pre-trained language models*

Shades of composition

- **“Syntactic angles” vs “Semantic angles”**
- “Supervised angles” vs “Pre-trained NLU angles”

Shades of composition

- “Syntactic angles” vs “Semantic angles”
- **“Supervised angles” vs “Pre-trained NLU angles”**

Shades of composition

- “Syntactic angles” vs “Semantic angles”
- **“Supervised angles”** vs “Pre-trained NLU angles”

Supervised angles

- Define supervised task setting, and test whether trained models show compositional generalization
- Question of focus: can/do current neural models learn supervised tasks such that they generalize compositionally at test time?
- Advantage: full knowledge of what models saw in training, and how test items relate to / force generalization beyond

Supervised angles

- Focused question about particular task/model/dataset
- Not necessarily tied to naturalistic NLU per se

Shades of composition

- “Syntactic angles” vs ”Semantic angles”
- **“Supervised angles”** vs “Pre-trained NLU angles”

Shades of composition

- “Syntactic angles” vs ”Semantic angles”
- “Supervised angles” vs **“Pre-trained NLU angles”**

Pre-trained NLU angles

- Progress in recent years has been driven by pre-trained language models
- Do these successes reflect learning of effective compositional meaning extraction during LM-based pre-training?

Pre-trained NLU angles

- Advantage: allows us to tackle critical compositionality questions about models widely in use by the community

- Challenge: no longer have full control/knowledge with respect to content of training data

Tackling pre-trained NLU angle

- How to address the problem of testing for effective compositionality when we don't control the training data?

1) Define information that should be represented / behaviors that should be produced if effective compositional meaning is being captured
2) Hypothesize and control for potential *heuristics/confounds* that might give illusion of success without proper compositional meaning

Three examples

1. Semantic role in sentence encoders: BOW control
2. Phrasal meaning in transformer LMs: word overlap control
3. Context meaning for prediction in transformer LMs: distractor control

Three examples

1. **Semantic role in sentence encoders: BOW control**

2. Phrasal meaning in transformer LMs: word overlap control

3. Context meaning for prediction in transformer LMs: distractor control

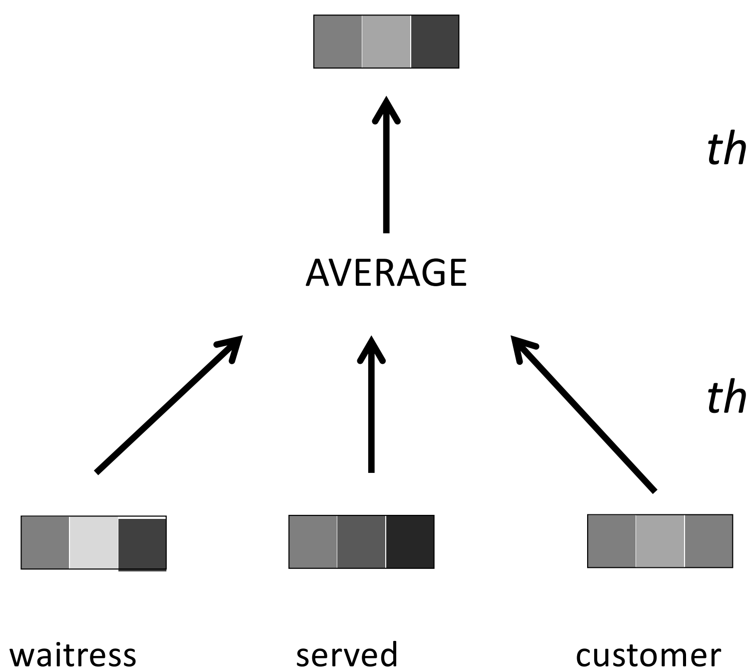
The problem

- Are pre-trained sentence encoders systematically capturing semantic role information?
- Design classification probes for semantic role information encoded in sentence embeddings

Controlling confounds: general statistics

- May see high classification performance because embeddings are sensitive to general statistics of how words tend to combine
- Rather than systematic understanding of *this* sentence

Control: Bag-of-words check



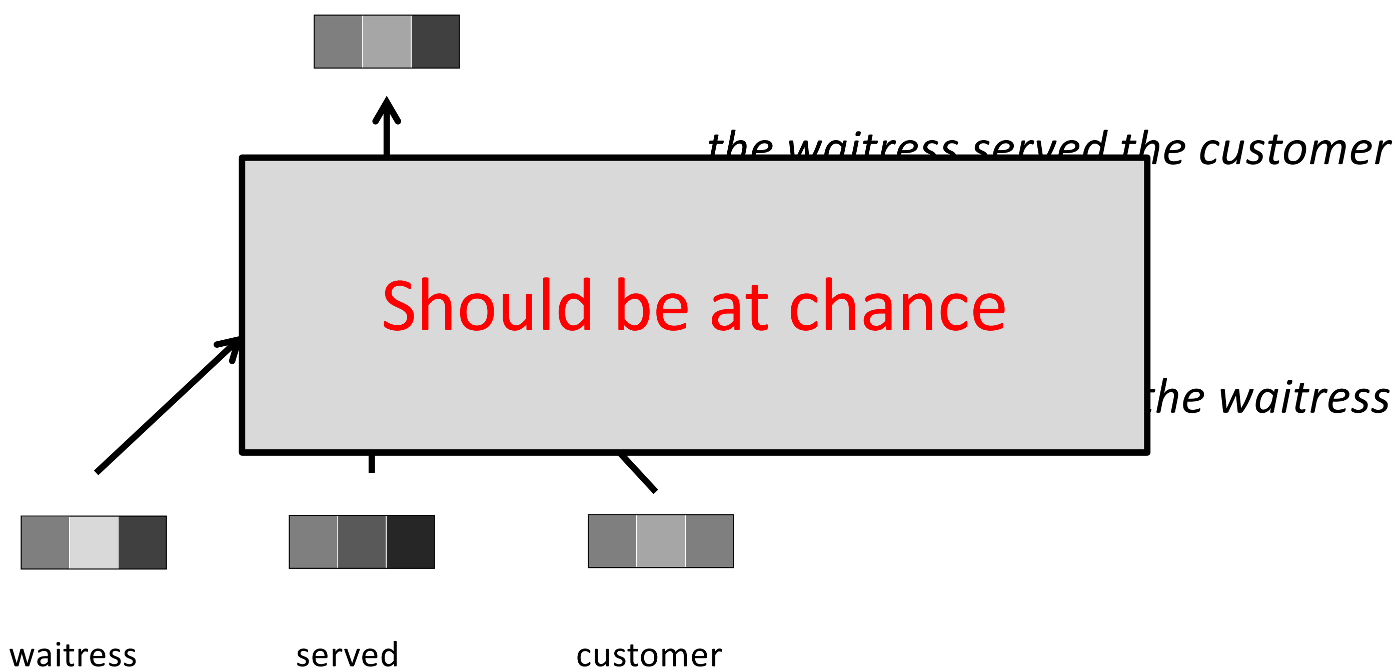
the waitress served the customer

??

the customer served the waitress

Ettinger et al. (2018). *Assessing Composition in Sentence Vector Representations.*

Control: Bag-of-words check



Ettinger et al. (2018). *Assessing Composition in Sentence Vector Representations.*

Results: classification accuracy

	CONTENT	ORDER	ROLE
BOW	100.0	55.0	51.3
SDAE	100.0	92.9	63.7
ST-UNI	100.0	93.2	62.3
ST-BI	96.6	88.7	63.2
InferSent	100.0	86.4	50.1

Ettinger et al. (2018). *Assessing Composition in Sentence Vector Representations*

Three examples

1. **Semantic role in sentence encoders: BOW control**

2. Phrasal meaning in transformer LMs: word overlap control

3. Context meaning for prediction in transformer LMs: distractor control

Three examples

1. Semantic role in sentence encoders: BOW control

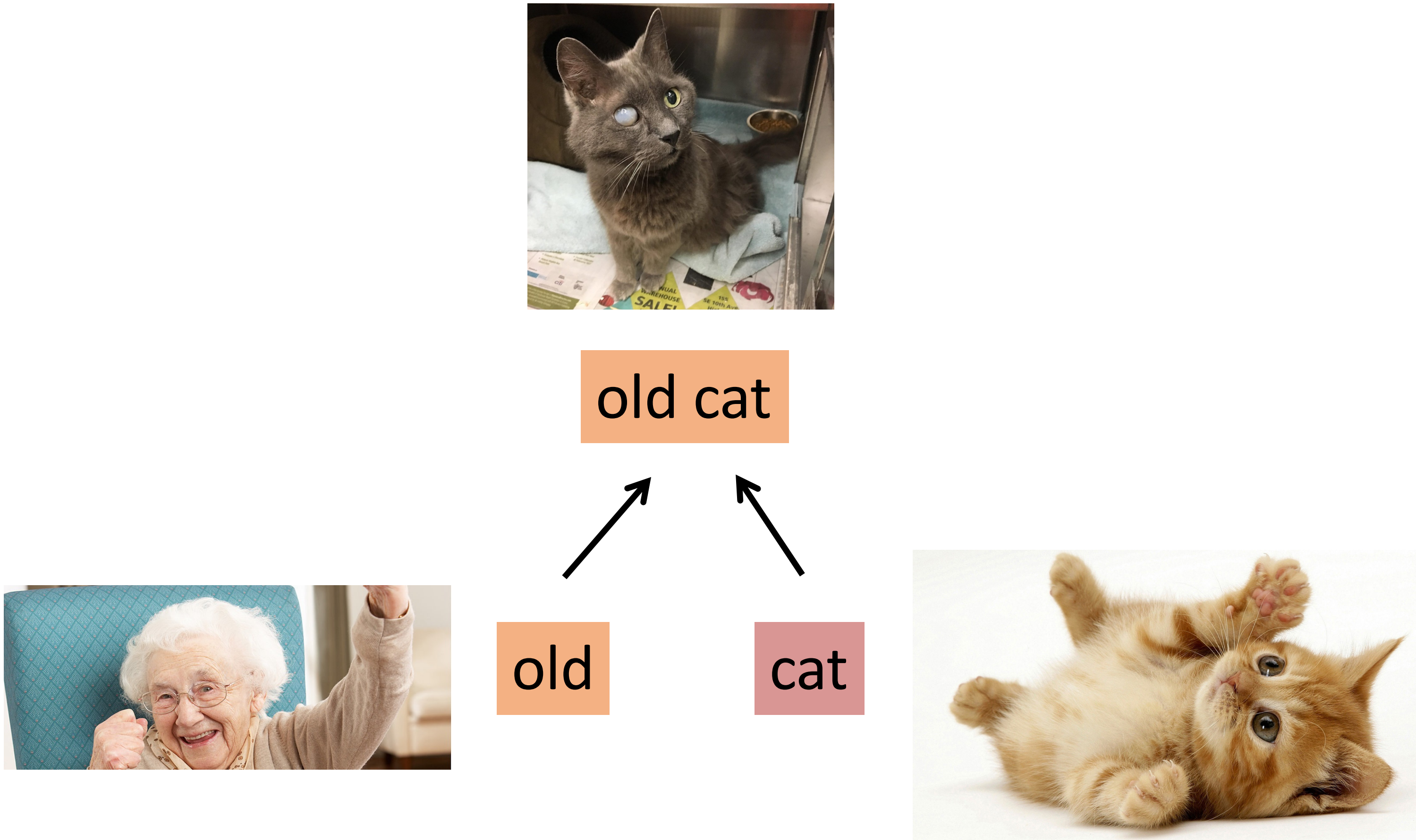
2. **Phrasal meaning in transformer LMs: word overlap control**

3. Context meaning for prediction in transformer LMs: distractor control

Phrase-level composition

old

cat


old cat
old
cat

The problem

- Are transformer LM representations capturing nuances of phrase meaning?
- Extract representations and compare against human judgments based on 1) similarity correlations, and 2) paraphrase classification

Controlling confounds: word overlap

- High correlations or paraphrase classification accuracy could be influenced by simple sensitivity to amount of word overlap
- Introduce control such that amount of word overlap is removed as a cue for similarity/paraphrase status

Yu & Ettinger (2020). *Assessing Phrasal Representation and Composition in Transformers.*

Similarity correlations

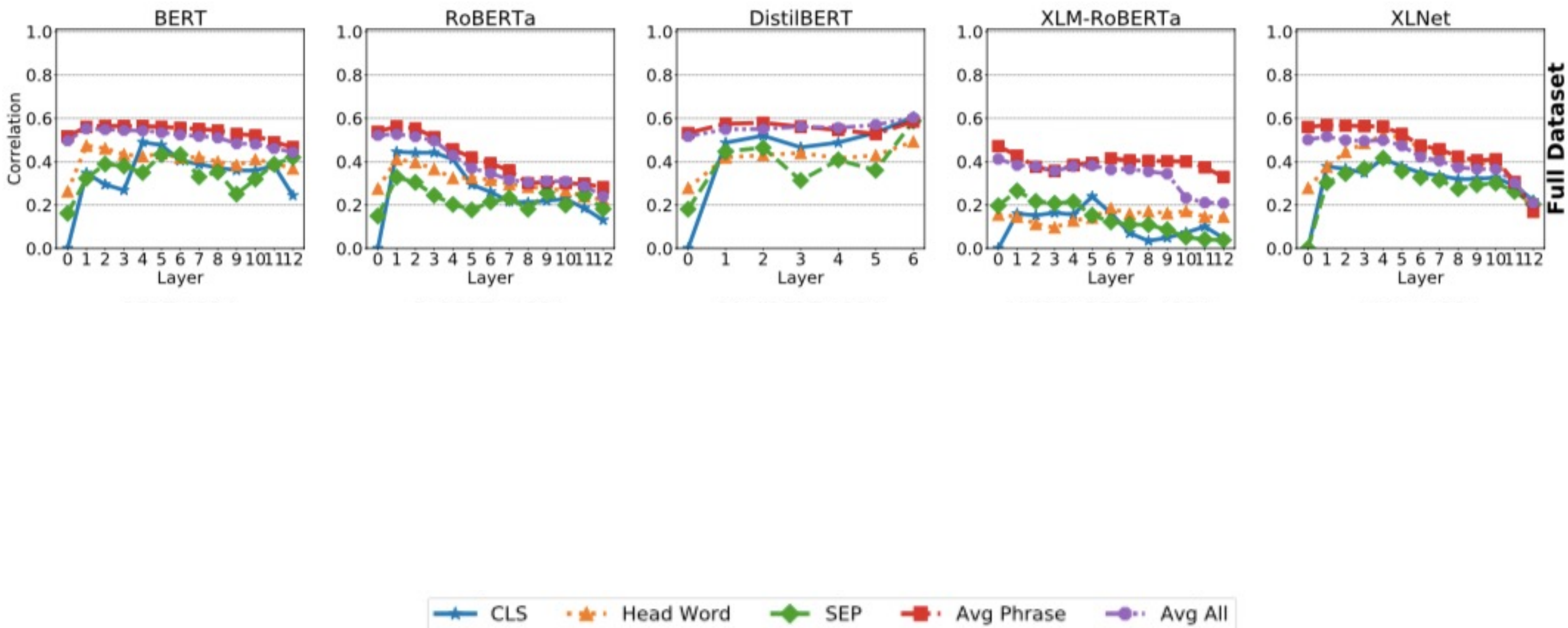
Normal Examples	
Source Phrase	Target Phrase & Score
average person	ordinary citizen (0.724)
	person average (0.518)
	country (0.255)
AB-BA Examples	
Source Phrase	Target Phrase & Score
law school	school law (0.382)
adult female	female adult (0.812)
arms control	control arms (0.473)

BiRD dataset (Asaadi et al., 2019)

Similarity correlations

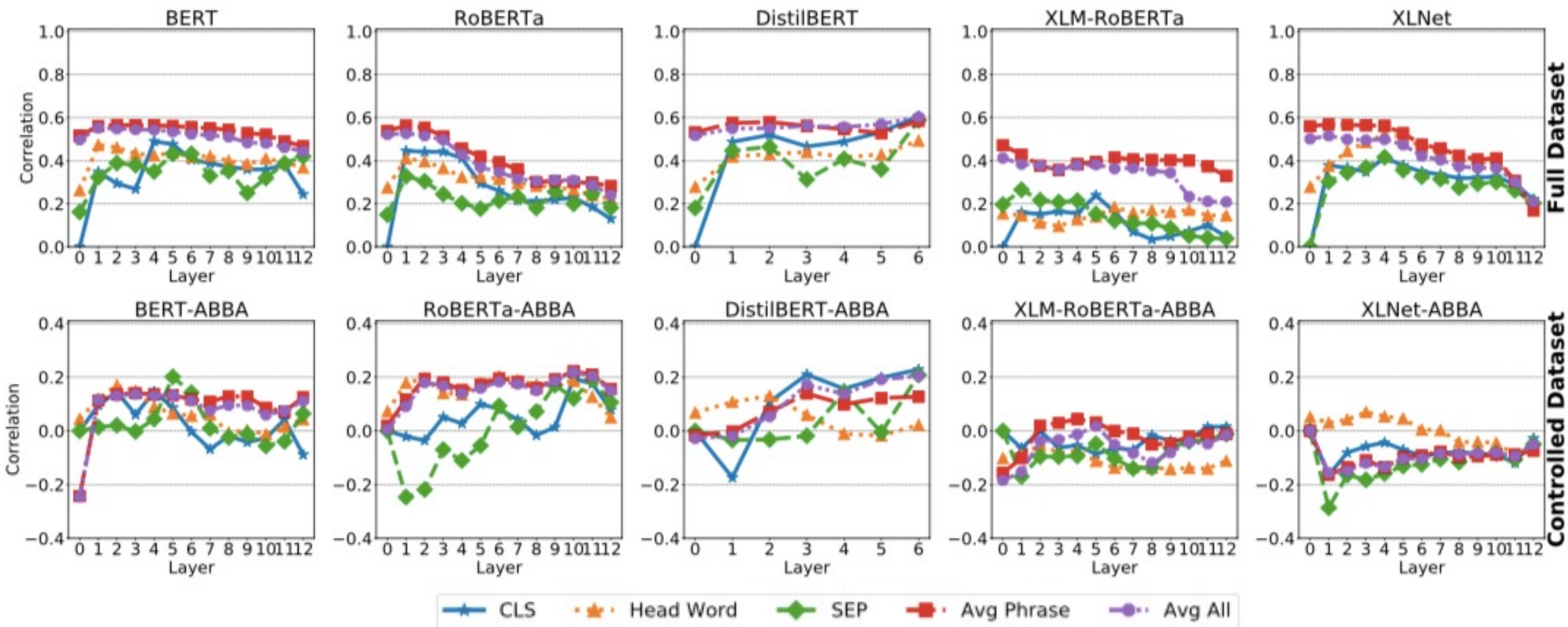
Yu & Ettinger (2020). *Assessing Phrasal Representation and Composition in Transformers.*

Similarity correlations



Yu & Ettinger (2020). *Assessing Phrasal Representation and Composition in Transformers.*

Similarity correlations

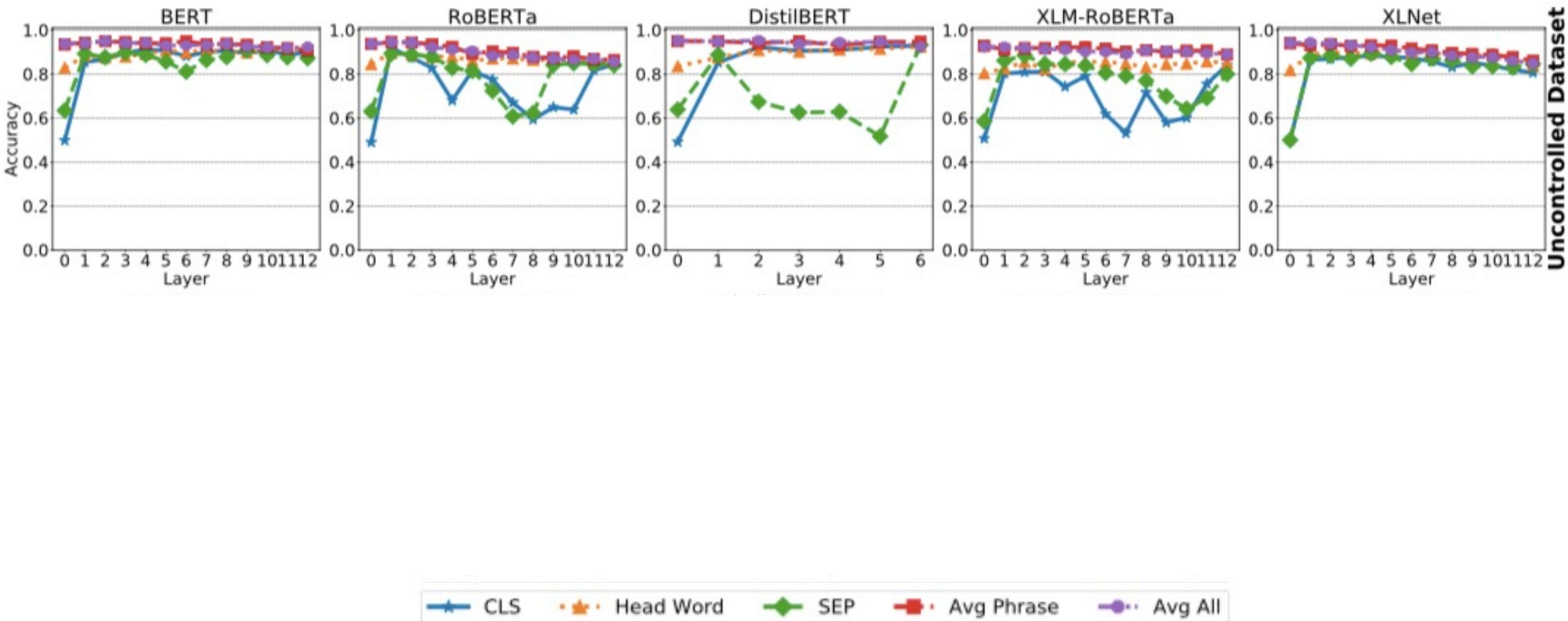


Yu & Ettinger (2020). *Assessing Phrasal Representation and Composition in Transformers.*

Paraphrase classification

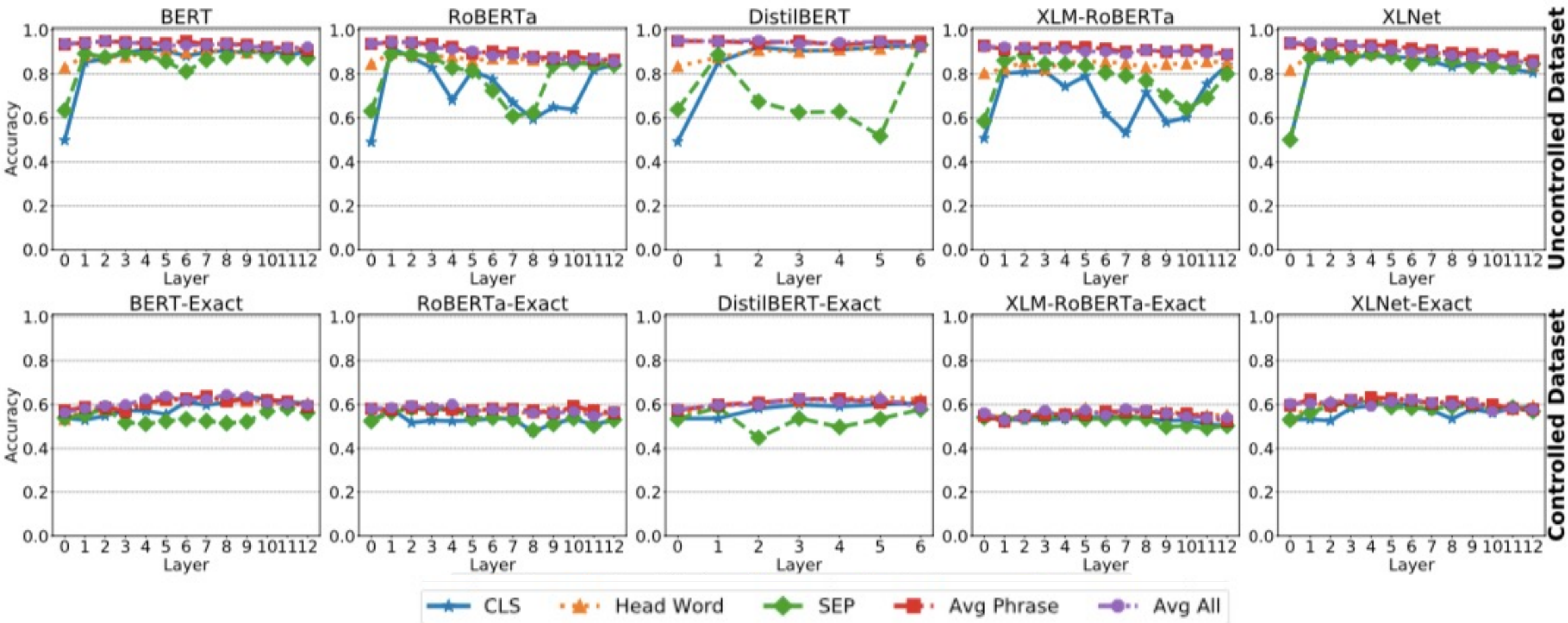
Yu & Ettinger (2020). *Assessing Phrasal Representation and Composition in Transformers.*

Paraphrase classification



Yu & Ettinger (2020). *Assessing Phrasal Representation and Composition in Transformers.*

Paraphrase classification



Yu & Ettinger (2020). *Assessing Phrasal Representation and Composition in Transformers.*

Three examples

1. Semantic role in sentence encoders: BOW control
2. **Phrasal meaning in transformer LMs: word overlap control**
3. Context meaning for prediction in pre-trained LMs: distractor control

Three examples

1. Semantic role in sentence encoders: BOW control

2. Phrasal meaning in transformer LMs: word overlap control

3. **Context meaning for prediction in pre-trained LMs: distractor control**

The problem: meaning from context

Sebastian lives in France. The capital of Sebastian's country is ___

Pandia & Ettinger (2021). *Sorting through the noise: Testing robustness of information processing in pre-trained language models*

Controlling confounds: shallow heuristics

- Correct predictions may be reliant on simpler heuristics like "produce a capital associated with recently-mentioned country"

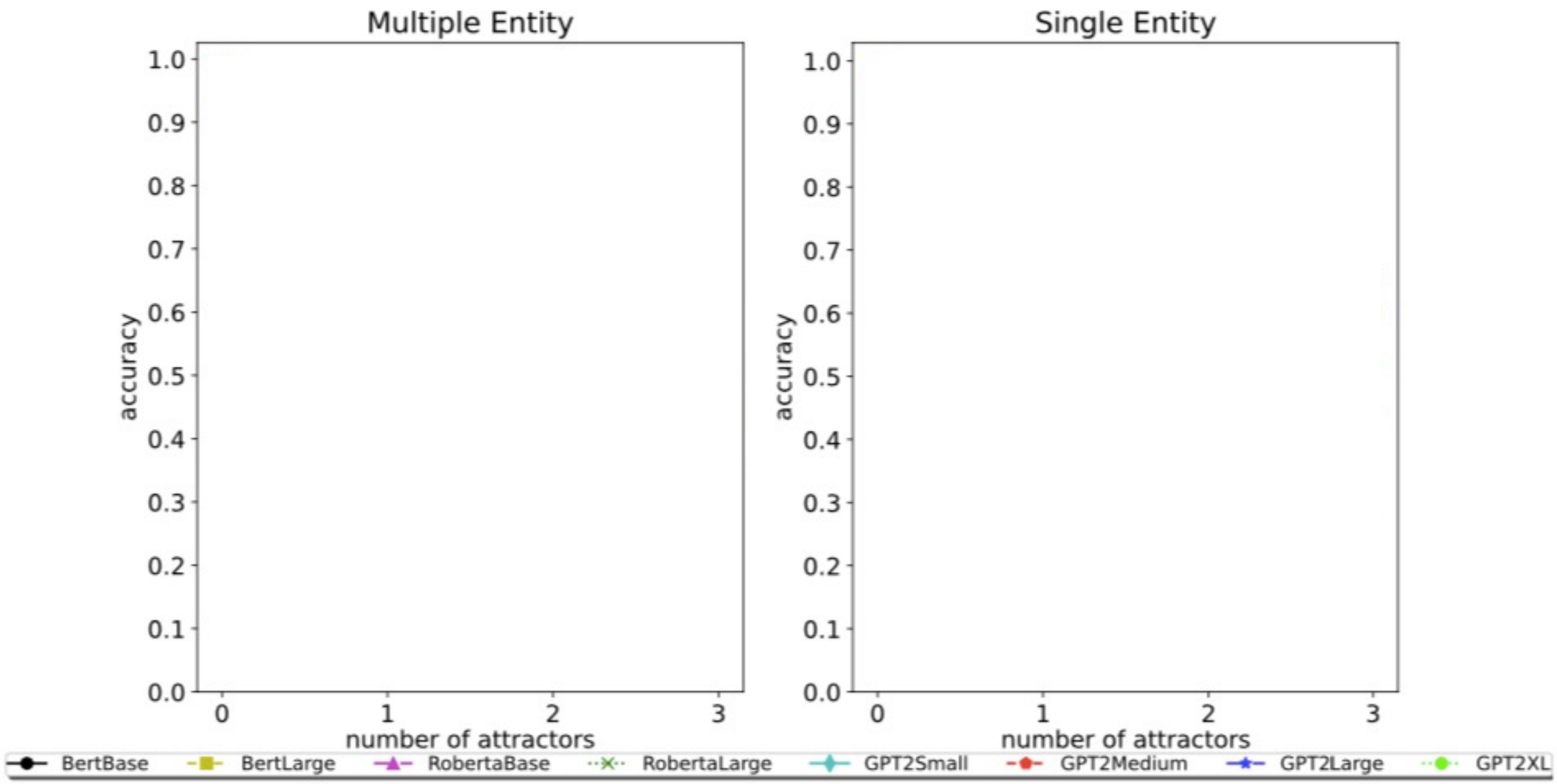
Pandia & Ettinger (2021). *Sorting through the noise: Testing robustness of information processing in pre-trained language models*

Controlling confounds: shallow heuristics

***Sebastian** lives in France, **Rowan** lives in Indonesia, and **Daniel** lives in Chile. The capital of **Sebastian's** country is ___*

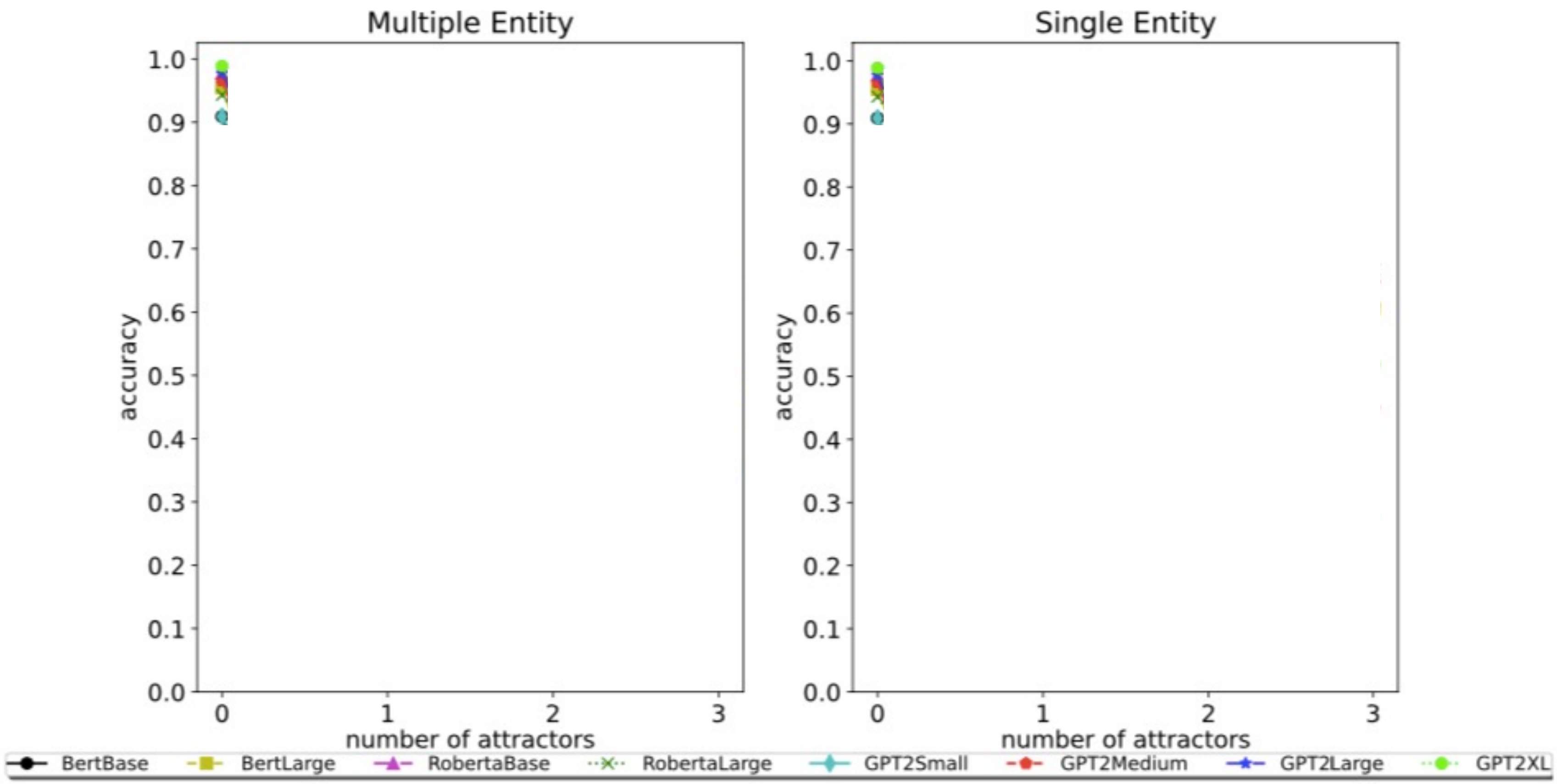
Pandia & Ettinger (2021). *Sorting through the noise: Testing robustness of information processing in pre-trained language models*

Accuracy (correct target prob > other words in semantic set)



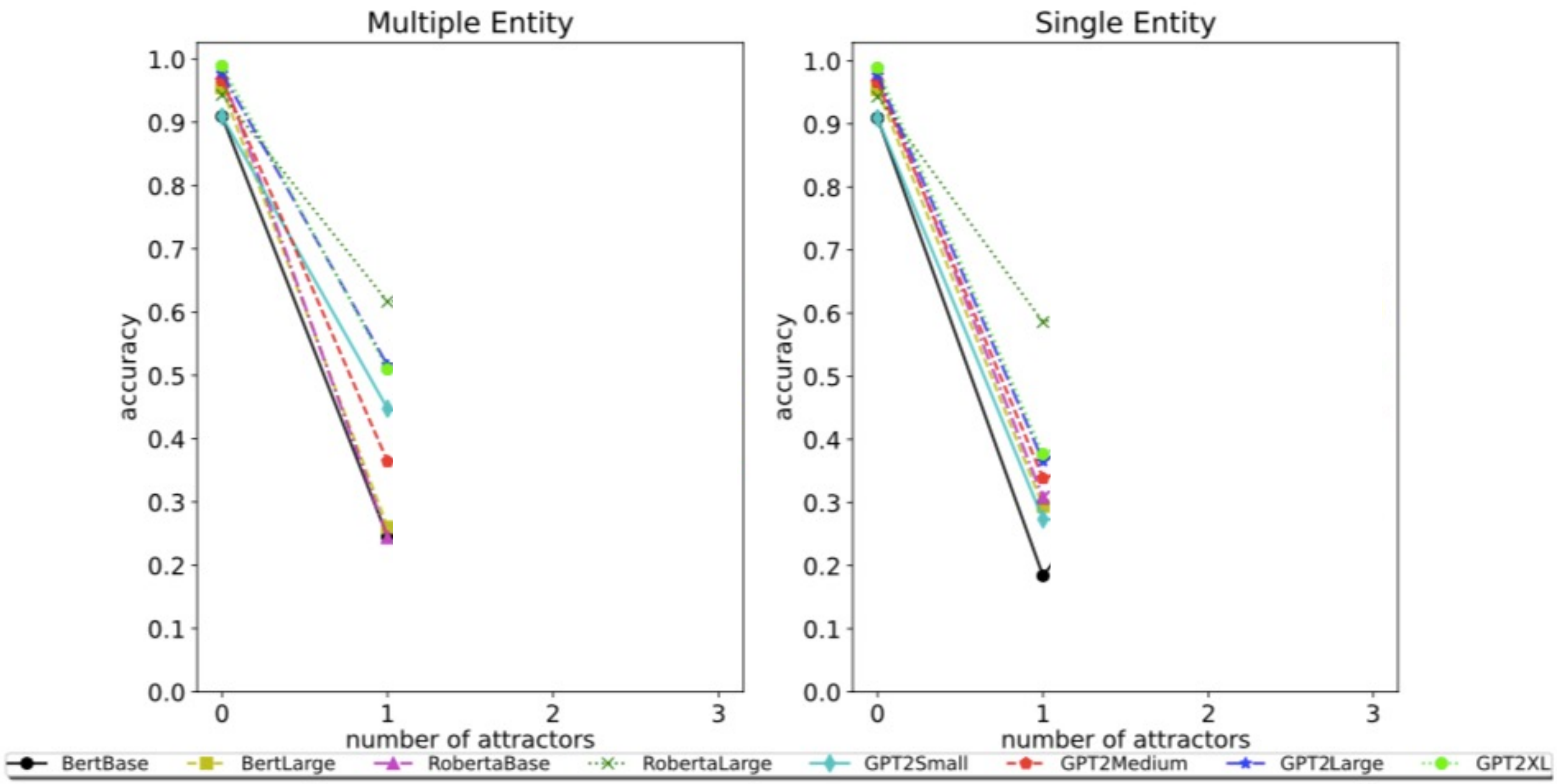
Pandia & Ettinger (2021). *Sorting through the noise: Testing robustness of information processing in pre-trained language models*

Accuracy (correct target prob > other words in semantic set)



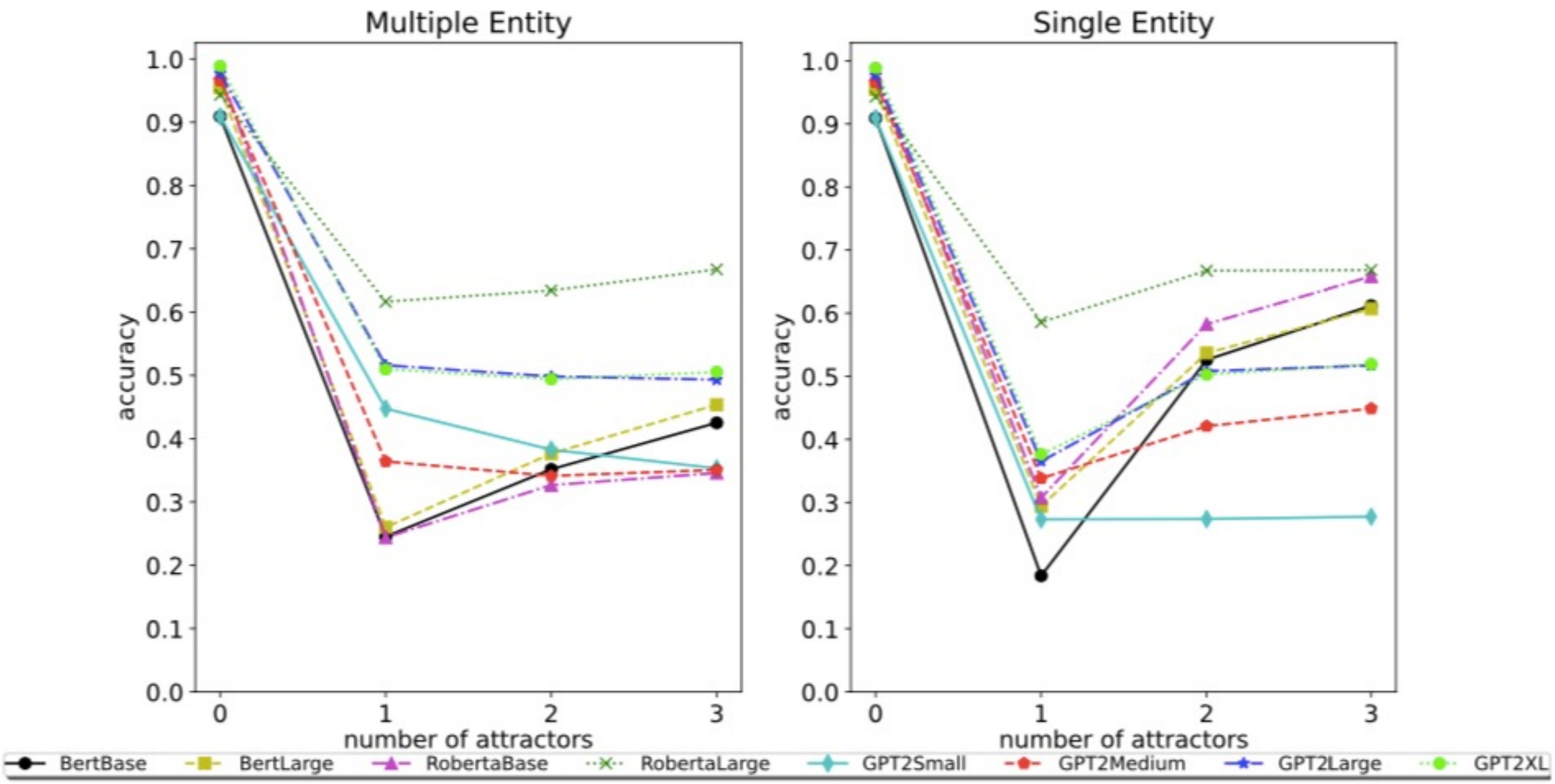
Pandia & Ettinger (2021). *Sorting through the noise: Testing robustness of information processing in pre-trained language models*

Accuracy (correct target prob > other words in semantic set)



Pandia & Ettinger (2021). *Sorting through the noise: Testing robustness of information processing in pre-trained language models*

Accuracy (correct target prob > other words in semantic set)



Pandia & Ettinger (2021). *Sorting through the noise: Testing robustness of information processing in pre-trained language models*

Takeaways

- Confounds can have critical impact on our tests for composition
- Shallow heuristics can give strong illusion of compositional meaning understanding
- Careful control for confounds/heuristics can quickly reveal fundamental limitations in models’ encoding/use of robust, compositional meaning from language inputs

Summarizing: composition needs in NLU

- Composition is the critical alternative to infinite memorization

- For effective NLU, we need accurate, human-like derivation of compositional meanings from language inputs

Syntactic angles

Three singing rabbits walked into the local bar last Wednesday afternoon

MODIFIER OF RABBITS

TIME/DAY OF WALKING

Monday	
Tuesday	
Wednesday	
Thursday	
Friday	
Saturday & Sunday	

AGENTS OF WALKING

LOCATION/DESTINATION OF WALKING

Semantic angles

old

cat

old cat

old

cat

Supervised angles

- Focused tests of compositional generalization in particular supervised settings

Pre-trained NLU angles

- Testing compositional meaning capabilities in pre-trained LMs trained in naturalistic settings

Tackling pre-trained NLU angles

- Definition of what compositional meaning capability would look like in model representations/behaviors
- Careful control of confounds/heuristics that don't constitute systematic compositional meaning
- Can disentangle shallower behaviors from target compositional meaning understanding

Looking forward

Accurate, systematic meaning composition is a critical open problem for NLU!

Thank you!

Lang Yu

Lalchand Pandia

Ahmed Elgohary

Philip Resnik

NSF Award No. 1941160
GRF Grant DGE-1322106
NRT Grant DGE-1449815
Toyota Technological Institute at Chicago

Colin Phillips