Using Paraphrases to Study Properties of Contextual Embeddings

Laura Burdick, Jonathan K. Kummerfeld, Rada Mihalcea

lburdick@umich.edu
http://laura-burdick.github.io

Big Idea

We use **paraphrases** to analyze contextualized embeddings

Why are paraphrases special?

- They naturally encode phrase semantics…

- …and word semantics

note	of	the	information	provided	by
0.96	0.96	0.95	0.94	0.79	0.82
note	of	the	information	contained	in

- The Paraphrase Database 2.0 (PPDB)
  - Ganitkevitch et al., 2013; Pavlick et al., 2015
- Word alignment information, automatically generated quality rating, some human quality ratings

Outline

Phrase-level Embeddings

Can BERT distinguish between two phrases that are paraphrases and two phrases that are unrelated?

- Use phrase-level embeddings
  - Average together word embeddings to get a phrase embedding
  - Take cosine similarity between two phrase embeddings
  - Compare cosine similarities to human annotations (Spearman's correlation)

Phrase-level Embeddings

Can BERT distinguish between two phrases that are paraphrases and two phrases that are unrelated?

- Experiment details
  - Uncased base model of BERT
  - 25,736 phrase pairs with human annotations
  - Compare BERT with w2v trained on Wikipedia

Phrase-level Embeddings
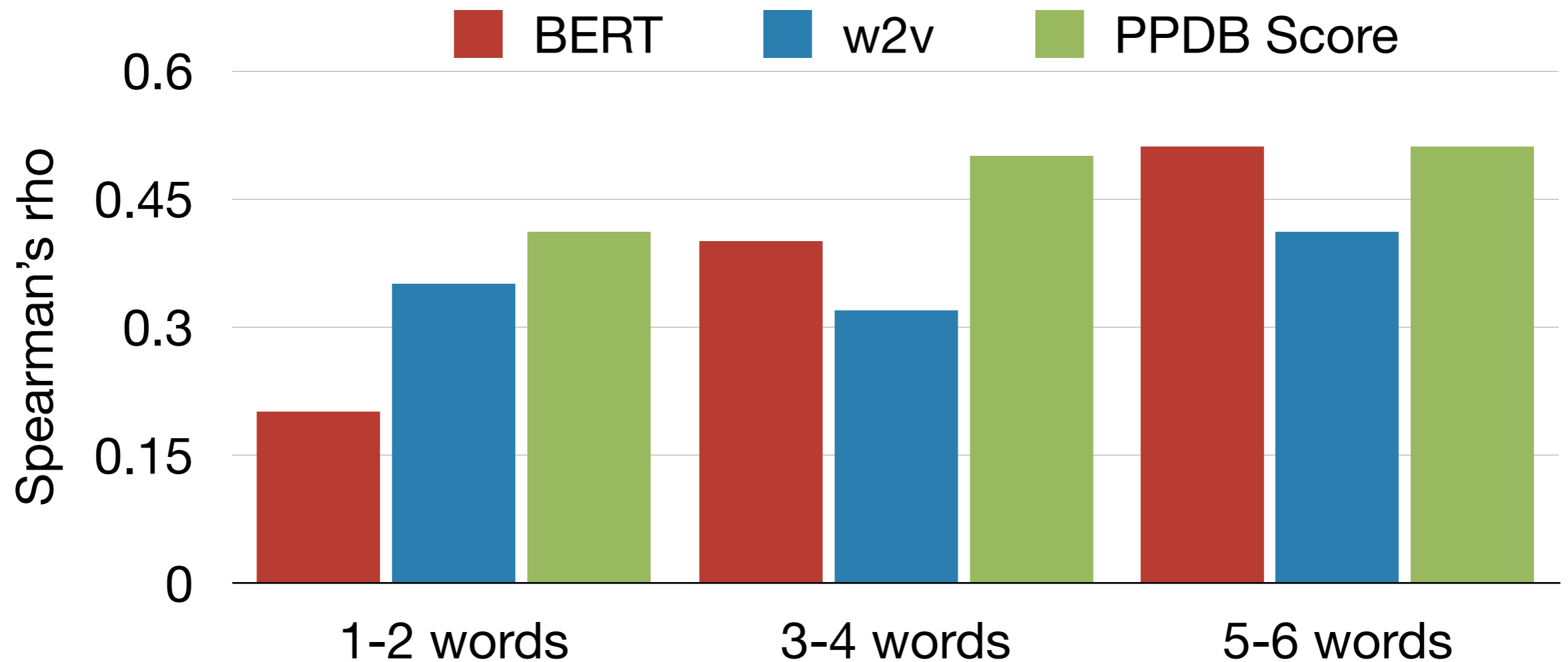
- BERT does better with longer paraphrases
- With longest paraphrases, BERT is comparable to PPDB score

Outline

Word-level Embeddings

|	**Same**	**Different**
Aligned	**adopted** by the general assembly at / **adopted** by the assembly at	, with a special **focus** on / , with special **emphasis** on
Unaligned	**okay** , so everything 's fine / you guys **okay** over there	between the canadian **government** and / between the government of **canada** and

Word-level Embeddings

Does BERT recognize that aligned words are more similar than unaligned words?

- Experiment details
  - Only use highest quality paraphrases in PPDB
  - Randomly sample 2,500 words from each category
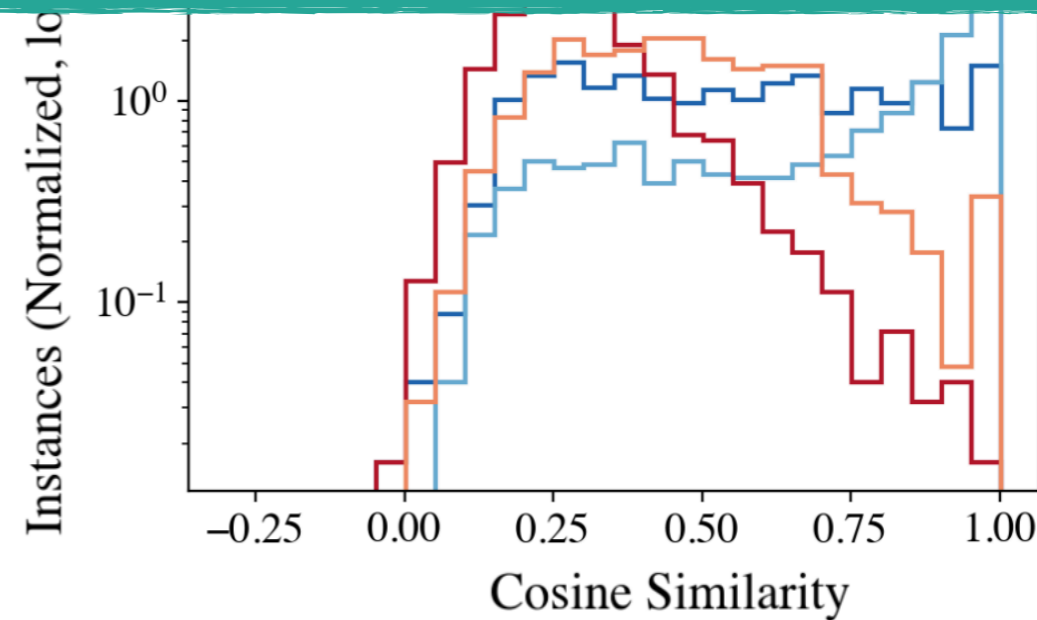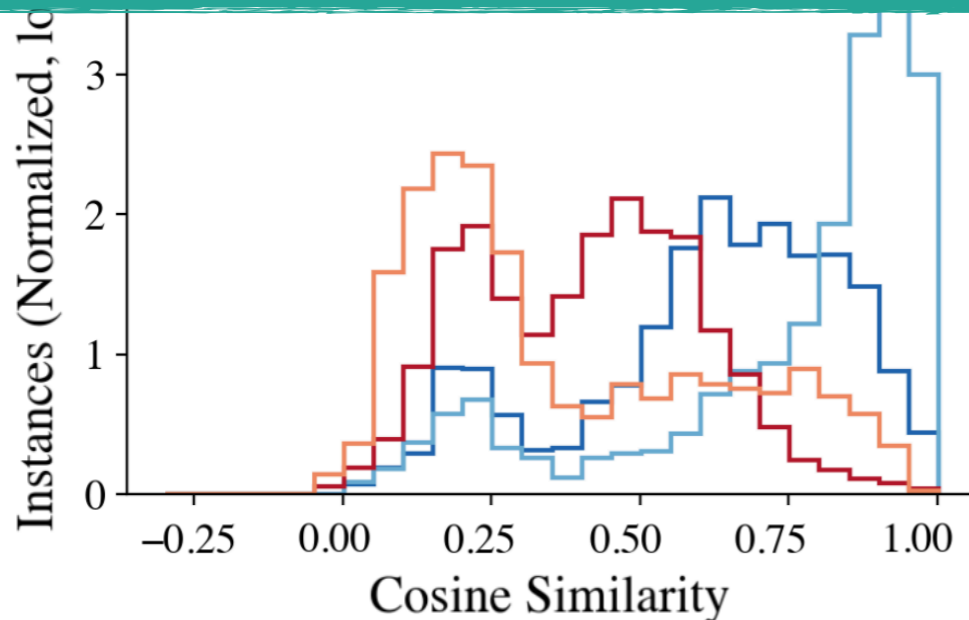 - For aligned words, only consider 1-1 alignment

Word-level Embeddings
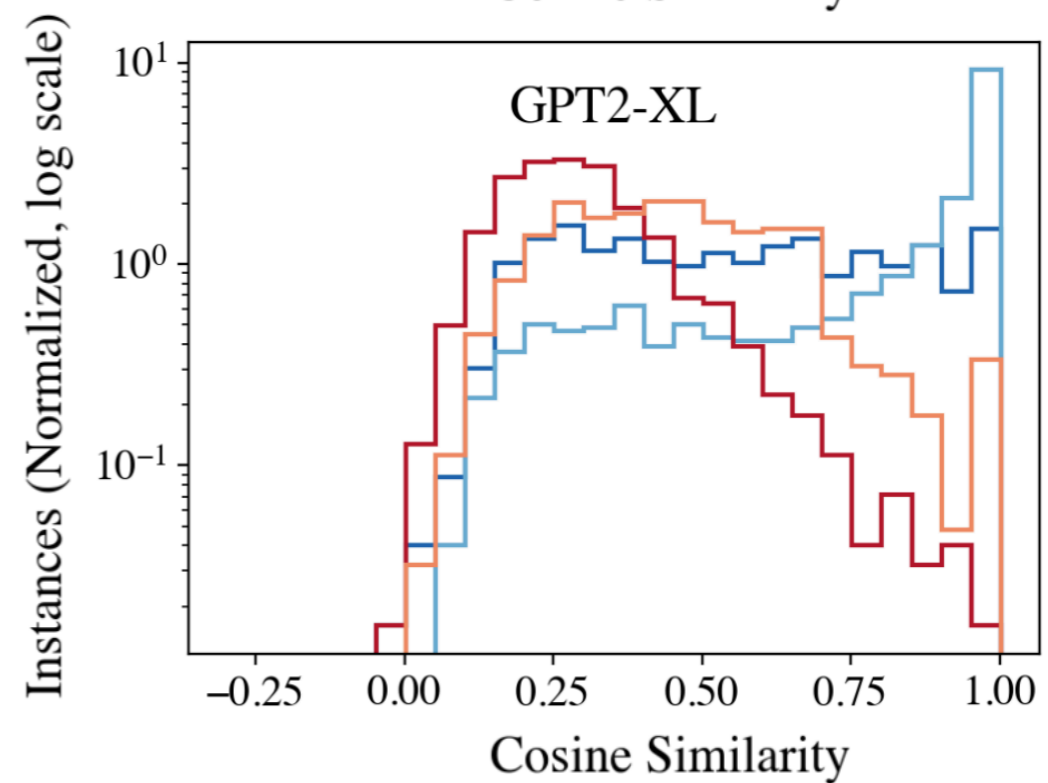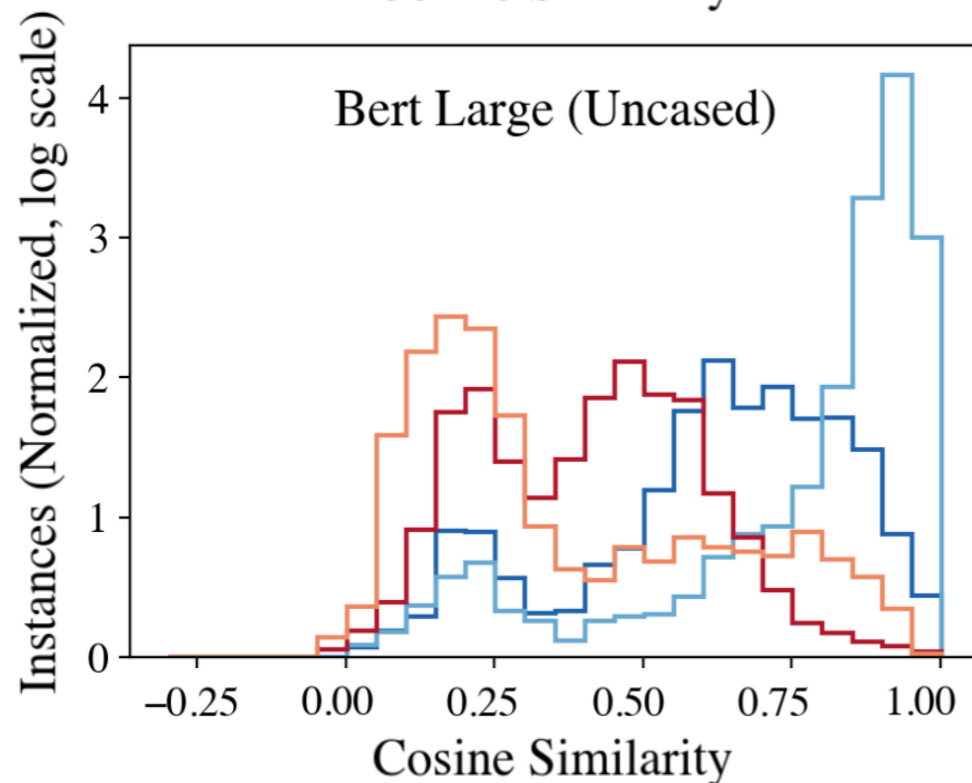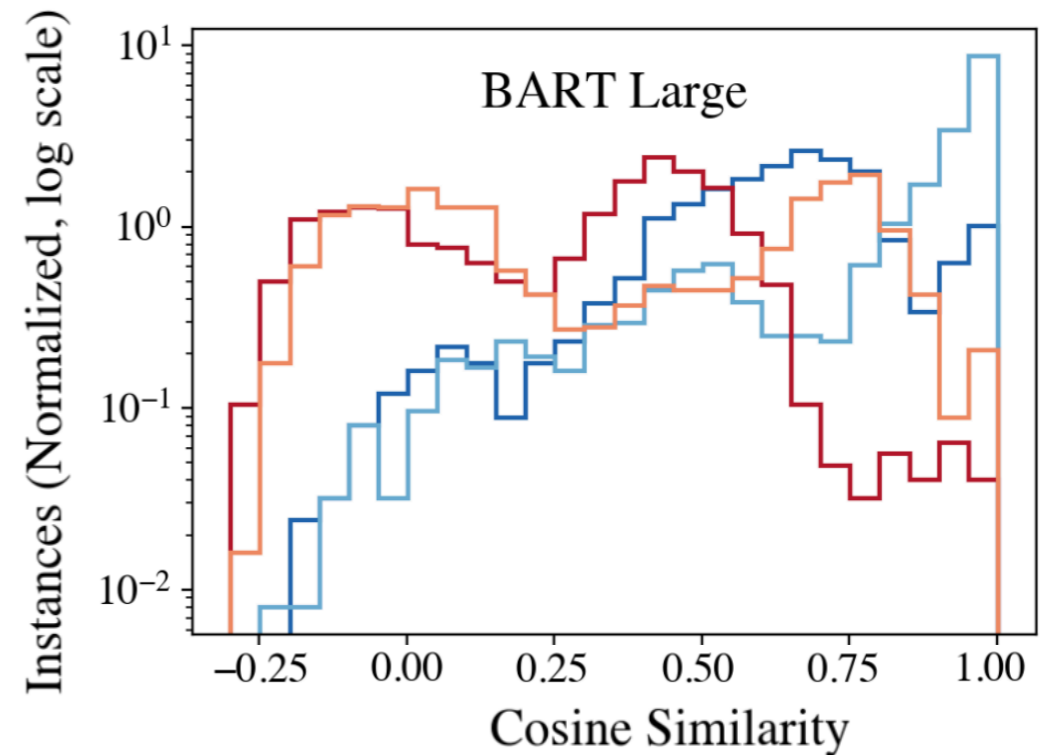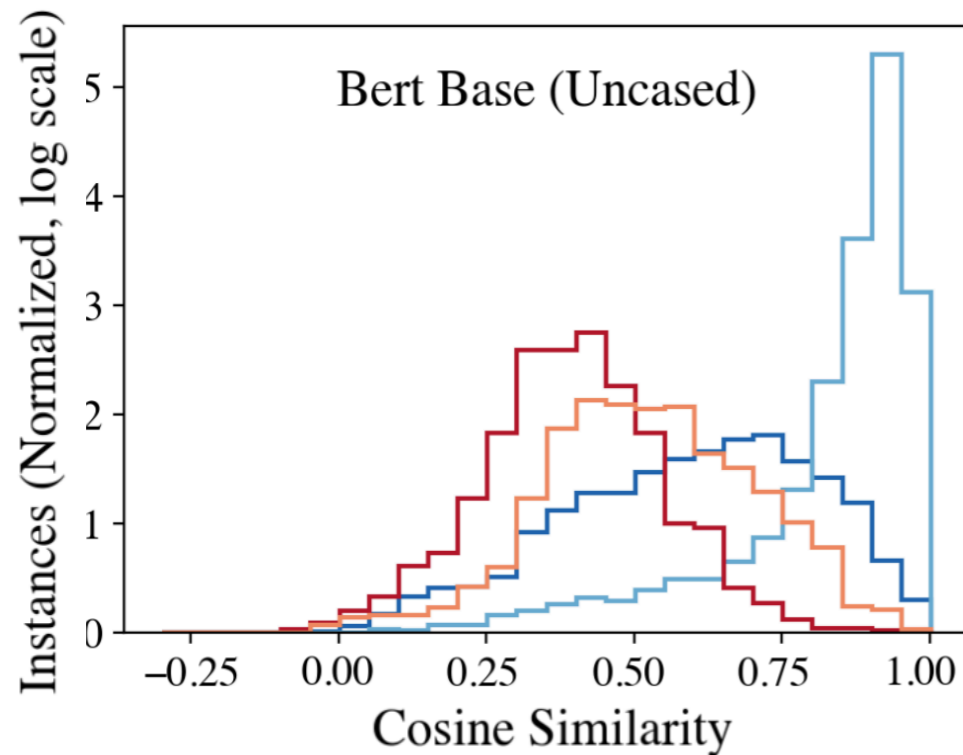
Word-level Embeddings
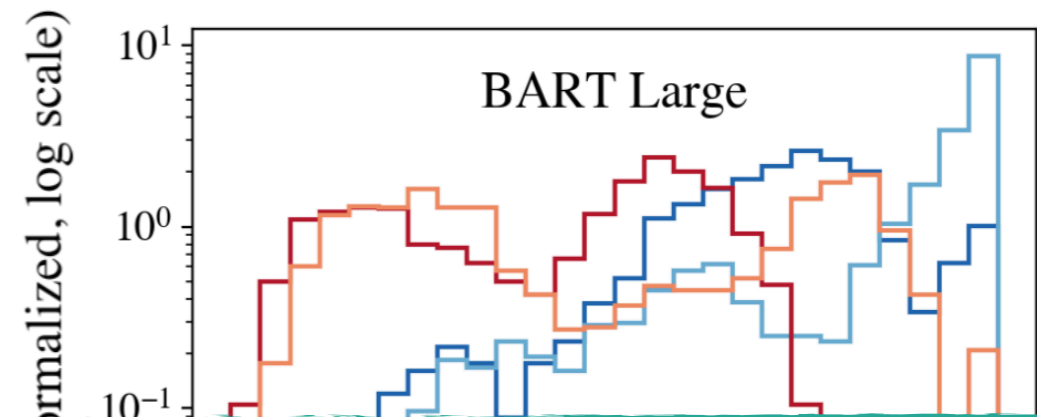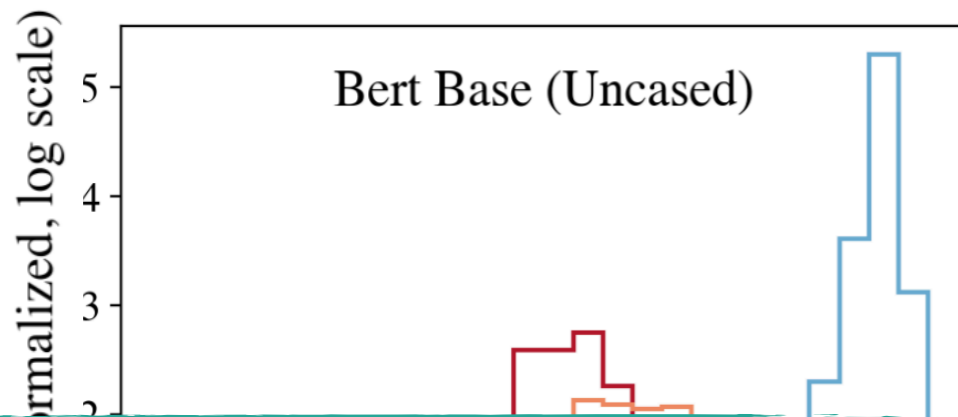
Contextual embeddings consistently handle aligned words in paraphrases, but there are substantial variations across models in how peaked the distributions of same-aligned words are.
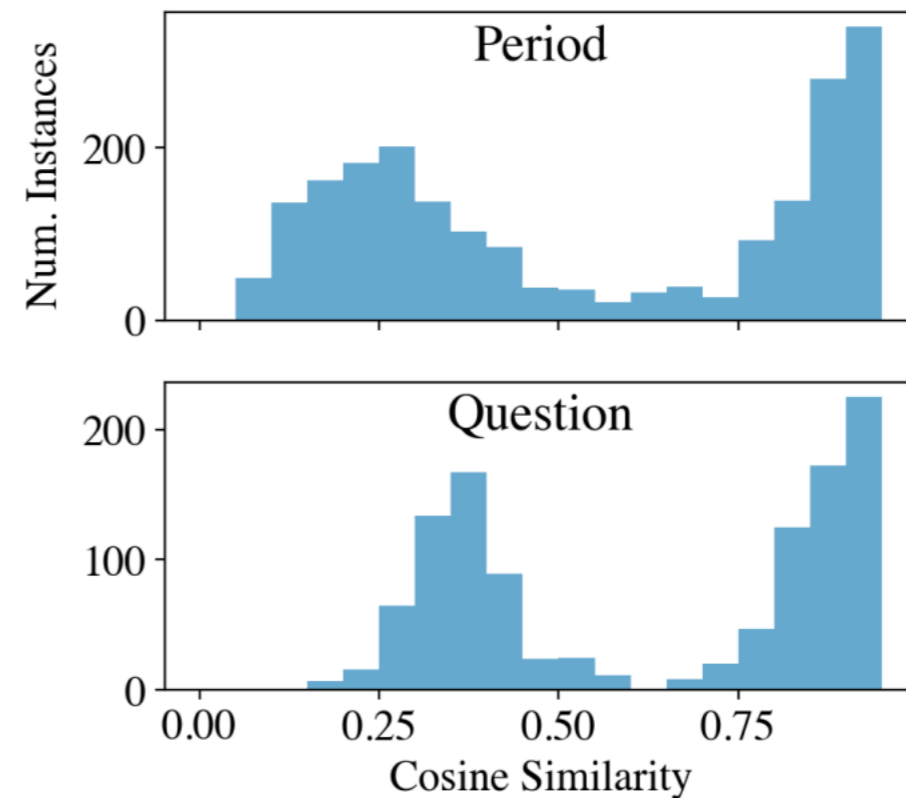
Outline

Punctuation

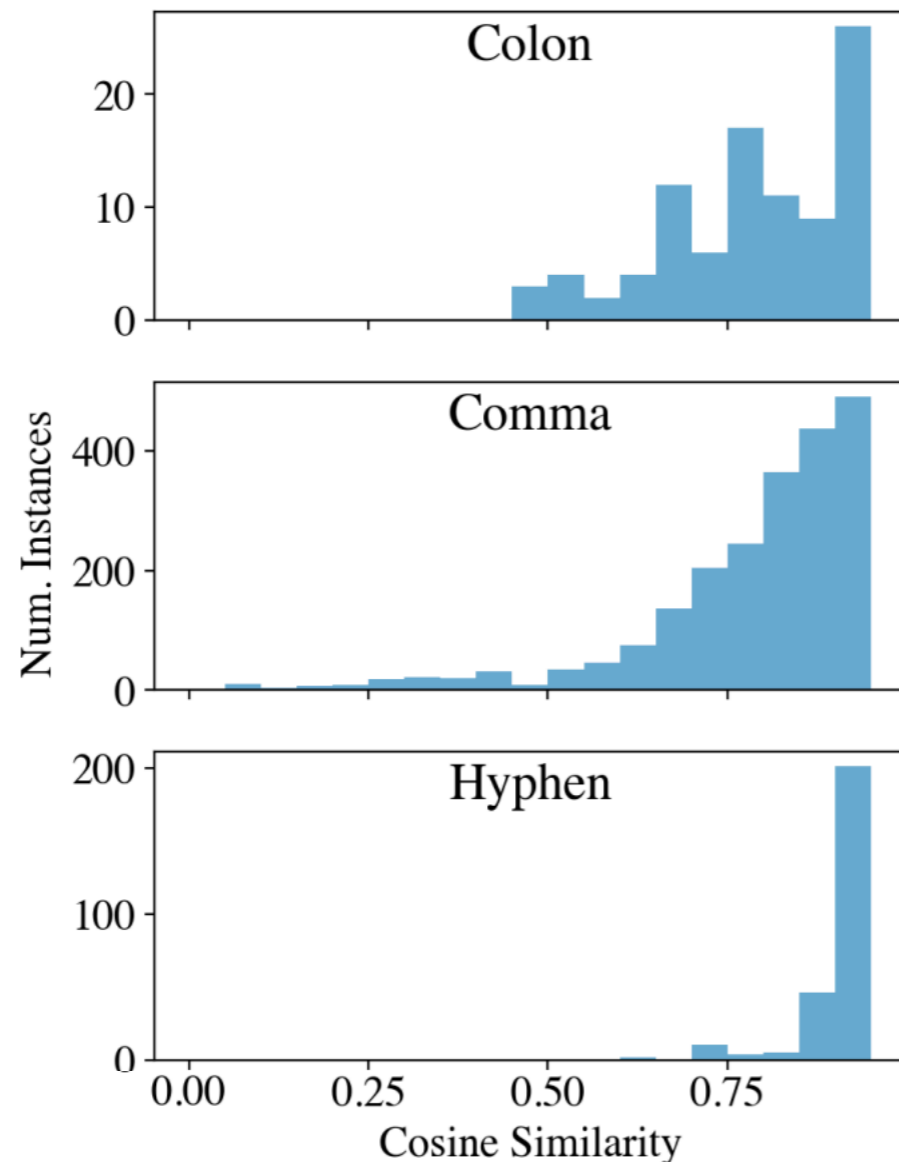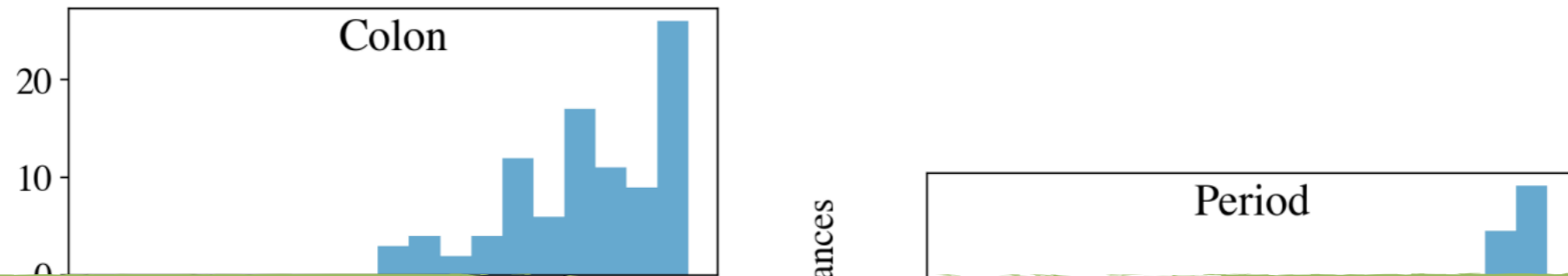Is the distribution of embeddings for punctuation different than the distribution for other words?
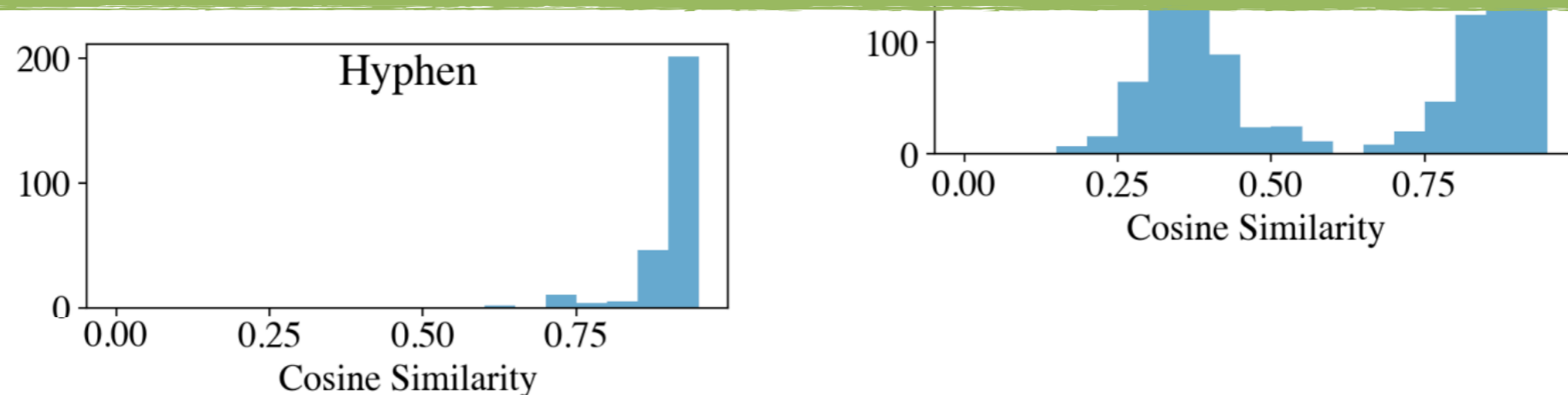
Punctuation

Is the distribution of embeddings for punctuation different than the distribution for other words?

BERT's representation of punctuation is surprisingly context sensitive, with substantial variation even when we control for meaning.

Takeaways

- BERT consistently represents paraphrases.

- We can use paraphrases to explore other representation methods!

- More in the paper…
  - Polysemous vs. non-polysemous words
  - One-word paraphrases (synonyms)
  - How the position of a word in a phrase affects its representation
  - How contextualized representations are at different BERT layers

Thank you!

lburdick@umich.edu
http://laura-burdick.github.io