

## Transparency Heps Revea When Language Models Learn Meaning



Zhaofeng Wu, William Merrill, Hao Peng, Iz Beltagy, Noah A. Smith zfw@csail.mit.edu

> LMs can learn the meaning of some languages with form-only pretraining, but doesn't do well with non-transparent languages.

> > transparent

## Propositional Logic

Q: Can LMs learn the meaning of PL with form-only pretraining?



![](_page_0_Figure_10.jpeg)

What will future language models understand?. 2021