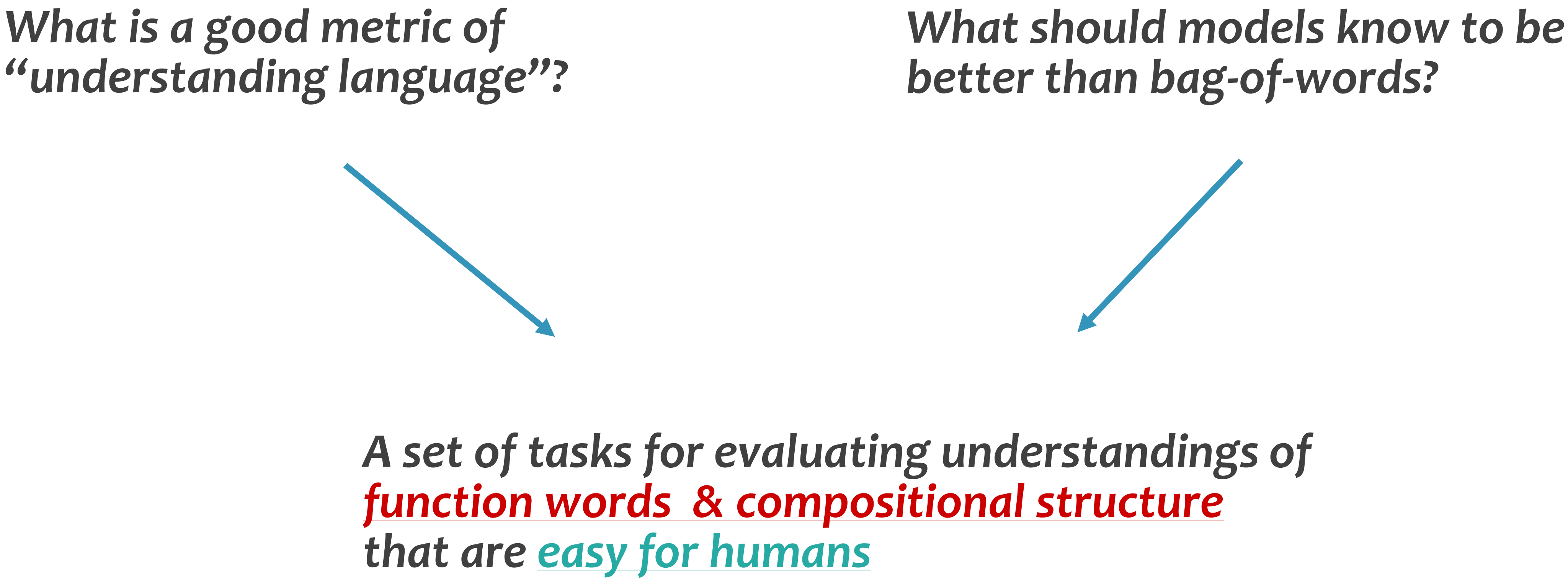
Linguistically informed tasks for evaluating structure encoded by sentence representations (Najoung Kim, Benjamin Van Durme, Ellie Pavlick & Paul Smolensky)



What is a good metric of "understanding language"?

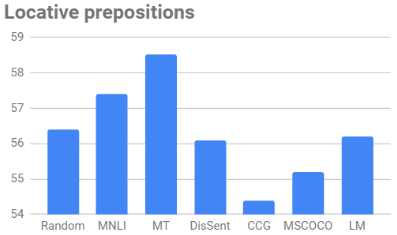
What should models know to be better than bag-of-words?

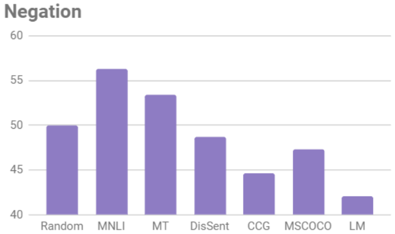
A set of tasks for evaluating understandings of ***function words & compositional structure*** *that are* ***easy for humans***

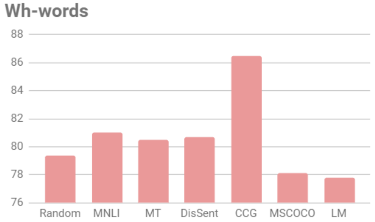
Linguistically informed tasks for evaluating structure encoded by sentence representations (Najoung Kim, Benjamin Van Durme, Ellie Pavlick & Paul Smolensky)

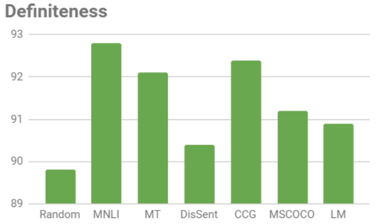
Result 1

A set of evaluation tasks for function word & structure (e.g., *prepositions, negation, wh-words, definite articles*) that differentiate between pretraining tasks


Locative prepositions

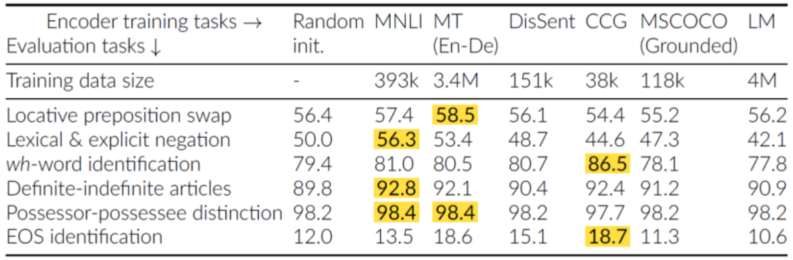

Negation


Wh-words

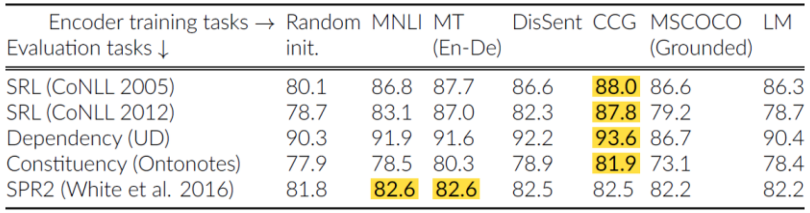

Definiteness

Result 2

Better downstream task performance ↛ Better evaluation task performance



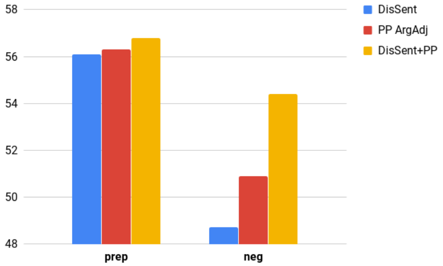
Encoder training tasks → Evaluation tasks ↓	Random init.	MNLI	MT (En-De)	DisSent	CCG	MSCOCO (Grounded)	LM
Training data size	-	393k	3.4M	151k	38k	118k	4M
Locative preposition swap	56.4	57.4	58.5	56.1	54.4	55.2	56.2
Lexical & explicit negation	50.0	56.3	53.4	48.7	44.6	47.3	42.1
wh-word identification	79.4	81.0	80.5	80.7	86.5	78.1	77.8
Definite-indefinite articles	89.8	92.8	92.1	90.4	92.4	91.2	90.9
Possessor-possessee distinction	98.2	98.4	98.4	98.2	97.7	98.2	98.2
EOS identification	12.0	13.5	18.6	15.1	18.7	11.3	10.6

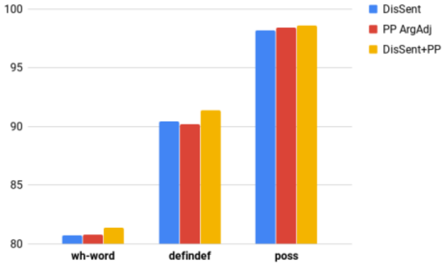


Encoder training tasks → Evaluation tasks ↓	Random init.	MNLI	MT (En-De)	DisSent	CCG	MSCOCO (Grounded)	LM
SRL (CoNLL 2005)	80.1	86.8	87.7	86.6	88.0	86.6	86.3
SRL (CoNLL 2012)	78.7	83.1	87.0	82.3	87.8	79.2	78.7
Dependency (UD)	90.3	91.9	91.6	92.2	93.6	86.7	90.4
Constituency (Ontonotes)	77.9	78.5	80.3	78.9	81.9	73.1	78.4
SPR2 (White et al. 2016)	81.8	82.6	82.6	82.5	82.5	82.2	82.2

Result 3

Room for linguistic theory! MTL on a theoretically motivated task helps multiple evaluation task performances.





9/28/2018 WeCNLP Summit, Menlo Park, CA