



[Lec 19]

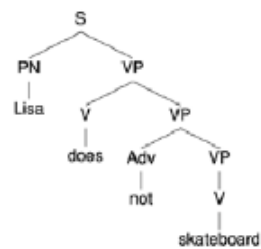
Extremly large language models and GPT-3

- GPT2: Language models: unsupervised multitask learners
 - not needed fine-tuning
 - provides good outputs
 - GPT3: Extremely large - many parameters
 - sparse attention
 - meta-learning method of training
- pro) good at language modeling (ex: story completion)
- included “in-context” learning: can be applied to new tasks

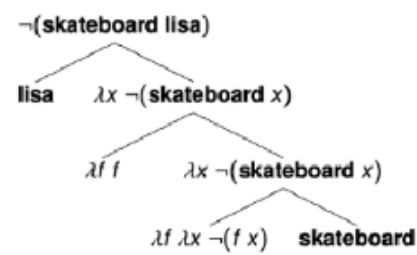
Compositional representations and systematic generalization

- Systemacity - Applicable patterns in sentences humans understand
- Compositionality - meaning of words & its composition determine a sentence
- Tree Reconstruction Error : Measure of compositionality
 - models gets more compositional after deep lengths of training
 - pre-training helps for generalization but doesn't solve the problem

Lisa does not skateboard =
 $\langle \text{Lisa}, \langle \text{does}, \langle \text{not}, \text{skateboard} \rangle \rangle \rangle$



$m(\text{Lisa does not skateboard}) =$
 $\langle m(\text{Lisa}), \langle m(\text{does}), \langle m(\text{not}), m(\text{skateboard}) \rangle \rangle \rangle$



Improving how we evaluate models in NLP

- Dynamic benchmarks