



Greedy: $P(\text{Since an sandwich}) = 4.5 \times 10^{-6}$
 Poss. input: $P(\text{Sense in language}) = 3.2 \times 10^{-5}$
 Best: $P(\text{Since in language}) = 1.8 \times 10^{-4}$

Problem with "predict next word" as a strategy for LM

In addition, no access to "higher-level" classes or structure