

How do we use it to generate language?

Idea:

- Feed the previous output as the next input
- Take *argmax* at each step (greedily) or use *beam search*