

What is a Statistical Language Model (LM)?

- A probability distribution over word sequences
 - $p(\text{"*Today is Wednesday*"}) \approx 0.001$
 - $p(\text{"*Today Wednesday is*"}) \approx 0.0000000000000001$
 - $p(\text{"*The eigenvalue is positive*"}) \approx 0.00001$
- Context-dependent!
- Can also be regarded as a probabilistic mechanism for "generating"
 - 'so called a "generative" model
 - Today is Wednesday**
Today Wednesday is
 - The eigenvalue is positive**

