

Types of SLM

- **“Standard” SLMs** all attempt to formally define $p(s) = p(w_1 \dots w_M)$
 - Different ways to refine this definition lead to different types of LMs (= different ways to “generate” text data)
 - Pure statistical vs. Linguistically motivated
 - Many variants come from different ways to capture dependency between words
- **“Non-standard” SLMs** may attempt to define a probability on a transformed form of a text object
 - Only model presence or absence of terms in a text sequence without worrying about different frequencies
 - Model co-occurring word pairs in text
 - ...