

































Language Modeling: N-Grams

Unigrams (SWB):

- · Most Common: "I", "and", "the", "you", "a"
- Rank-100: "she", "an", "going"
- · Least Common: "Abraham", "Alastair", "Acura"

Bigrams (SWB):

- Most Common: "you know", "yeah SENT!", "!SENT um hum", "I think"
- Rank-100: "do it", "that we", "don't think"
- Least Common: "raw fish", "moisture content", "Reagan Bush"

Trigrams (SWB):

- Most Common: "!SENT um hum SENT!", "a lot of". "I don't know"
- Rank-100: "it was a", "you know that"
- Least Common: "you have parents", "you seen Brooklyn"

LML Speech Recognition 2008







19











