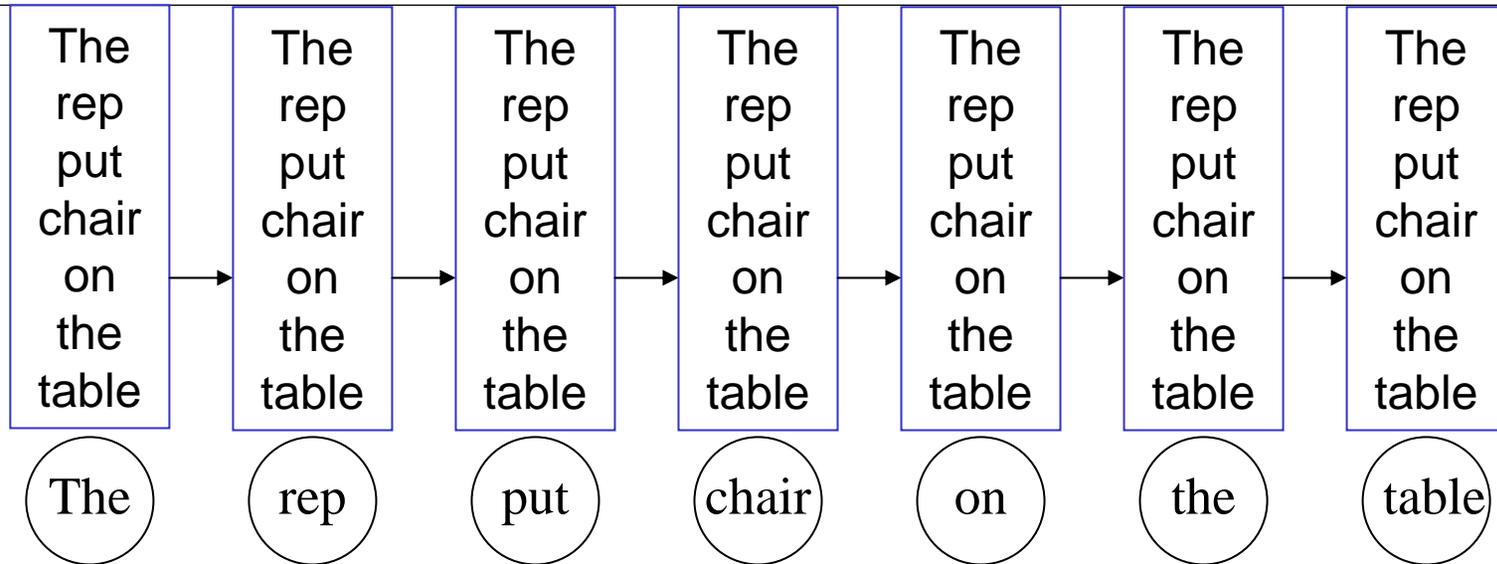


Lab3: Word Sorting from a Bag of Words



$$\arg \max_w \log P(W) \approx \arg \max_{w_1^N} P(w_1 | Sen_B) * P(Sen_E | w_N) \sum_{k=2}^N \log P(w_k | w_{k-1}) \quad \text{bigram approximation}$$

Given the first 5K WSJ sentences, estimate the word bigrams, and use them to sort the last 20 sentences with the Viterbi algorithm assuming the original order is unknown. You can use all 10K sentences to figure out the vocabulary size, but you need to consider unknown words if some words in the last 100 sentences contains words not in the first 5K sentences. Compare the three cases on Slide 41 of Lectures 9-10 by computing $P(W)$ in all three cases, and their sorting results. For MLE you need to handle the zero cell problem with your own assumptions. How many of the results are exactly right? Give an explanation if the results are wrong?