1. NLG Decoding Strategy
   1. Greedy Decoding

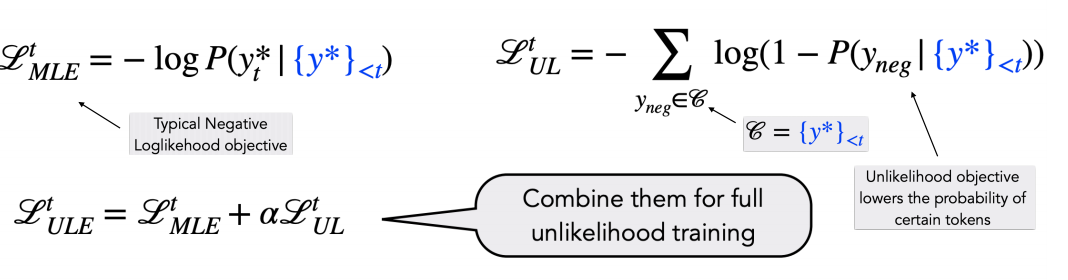
每次选择当前解码出的最大概率的字当作下一次的输入。计算量小，但是生成的句子可能不是最好的那一个，也有可能陷入循环。

* 1. Beam Decoding

Greedy Decoding的改进，每次解码选择当前概率与历史概率加和最大的K个字，用这K个字当作下一次的输入，当获得<STOP>时视为得到一个hypothesis。解码停止的条件可以是事前定义好的时间步T或者是已获得hypothesis的数量n。

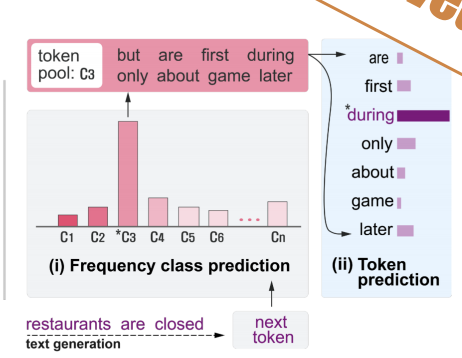
1. Repetitive Problem
   1. Unlikelihood Training

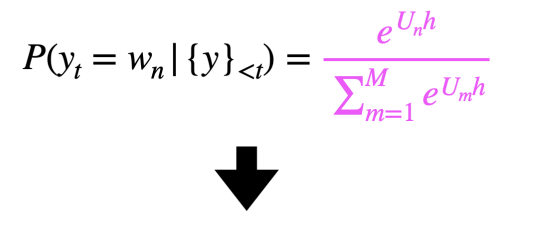
为了解决生成语句重复的问题，在极大似然估计的基础设添加了惩罚项，是已经生成过的词。



* 1. F2 Softmax

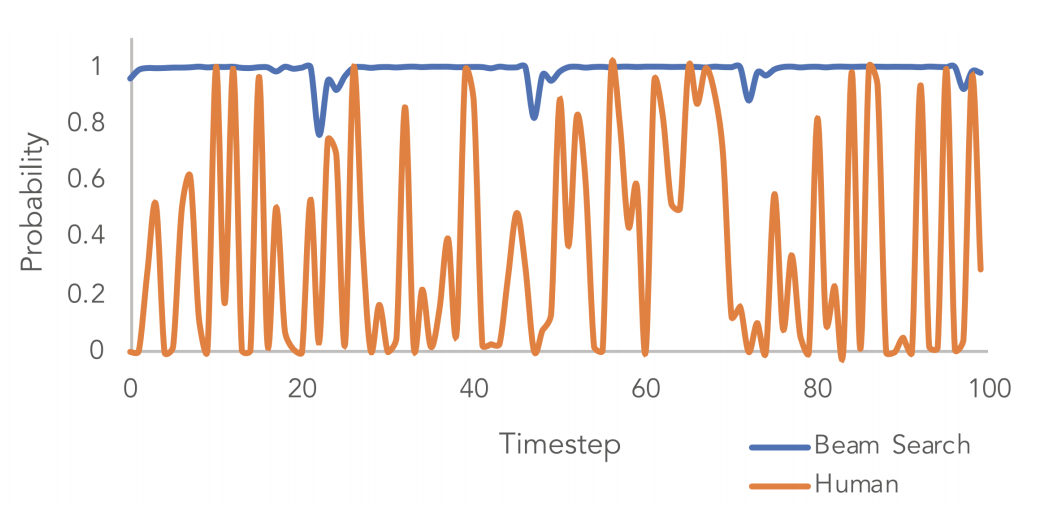
将softmax进行了拆分，先根据词频划分不同的类别，再在每个类别中划分具体的单词。



原来的概率计算公式由变为 两个部分。

1. Sampling Strategy
   1. Top-k sampling

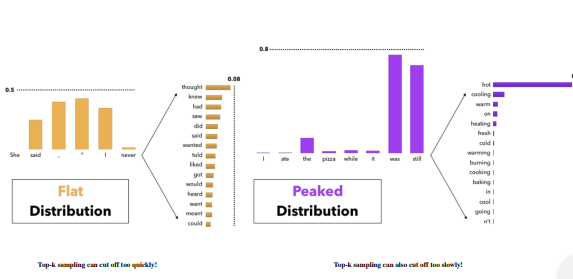
由于Beam Decoding总是选择概率最大的几个词进行生成，与人类说话的情况不同。



因此引入随机取样。

Top-k sampling选择每个时间步中概率最大的K个词，从中进行随机选取。

* 1. Top-p(Nucleus) Sampling

由于Top-k sampling的采样在预测分布比较平缓时丢弃概率较大的词，或者比较集中时保留概率较小的词，

因此Top-p(Nucleus) Sampling设定了一个阈值p，将解码概率从大到小排列，选择相加值大于p的前几个词作为采样对象。

