The fifirst step is to replace arguments in the event. Both the argument

to be replaced and the new one should have played ever the same role. While the roles are inherited after replacement, so we can still use origin labels for the generated samples.

## 第一步：替换事件中的argument。 无论是原argument还是新argyment都应该发挥相同的作用。 尽管角色是在替换后继承的，但是我们仍然可以将原始标签用于生成的样本。（替换是基于elmo嵌入算语义重叠最高的。）

## （嵌入本身是同一个argument的所有出现提及对elmo的平均值，再基于此算余弦相似度。）

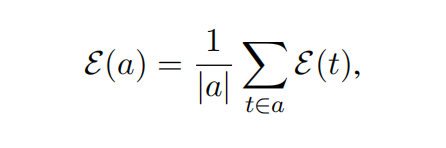
用概率80%代替论点，而用20%的概率保持不变，使表示偏向于实际事件注意，触发器仍然处于uncha状态。避免依赖关系的不理想偏差。

## 第二步：我们用微调的BERT重写辅助词（adjunct word）。

We use cosine similarity between embeddings to measure the similarity of two arguments. And

due to ELMO’s ability to handle the OOV problem, we employ it to embed arguments:

这个是一个词的



我们使用嵌入之间的余弦相似性来度量两个参数的相似性。由于Elmo处理OOV问题的能力，我们使用它嵌入参数。（采取elmo词向量进行交换。）

a是argument，E是Elmo嵌入。我们选择前10%的最相似的论点作为候选，并使用Softmax运算对它们的相似性来分配概率。

We choose the top 10 percent most similar argu

ments as candidates, and use softmax operation on

their similarity to allocate probability.

选了十个相似的argument，用softmax决定替换概率。而这种度量embeding（不是训练embeding）使用elmo的平均值得到的。

替换adjunct word被称为数据增强中的语义平滑。

当增强结果与真实情况很不符时会反而造成过拟合的现象。

语义对抗学习也可以当成一种数据增强，例如生成很像狗的又符合人的判断基准的猫。

为了不彻底改变意思，我们采用相似性作为选择的标准新的参数。 它基于以下两个注意事项：一是发挥两个论点同样的角色在语义上可能会有很大的不同； 另一个是论点扮演的角色在很大程度上取决于它的上下文。 因此，我们应该选择语义上合适的参数与上下文相似且连贯。我们在嵌入之间使用余弦相似度衡量两个论点的相似性。 和由于ELMO能够处理OOV问题，因此我们将其用于嵌入参数：

采用了一个argument的不同出现的平均值。

***X***ʹ Theory**** (pronounced ‘*X*-bar Theory’) states that each constituent consists of four basic components: *head* (the core meaning), *complement* (compulsory element), *adjunct* (omissible element), and *specifier* (an omissible phrase marker)

短语由指示语、核心词和补语组成

短语结构语法（PSG）来概括成分的模式。每个成分由两个部分组成：头部（核心含义）和补语（构成核心含义的其余成分）

在动词短语“ [ate icecream ravenously]”中，动词“ ate”需要补语“ icecream”，而补语“ ravenously”则可以省略。

*ead* (the core meaning), *complement* (compulsory element), *adjunct* (omissible element), and *specifier* (an omissible phrase marker)

*complement* 是不可以省略的非核心成分。

一个是对于短语结构意义上的不可以缺失，一个是

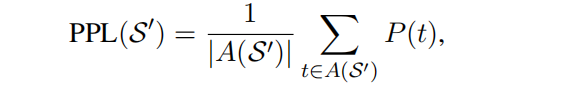
数据增强。

1. NP+VP

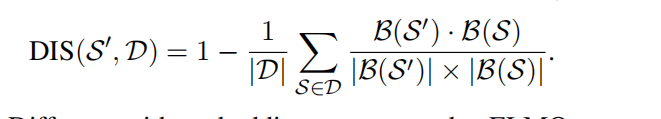
最后依靠perplexity+bert词向量打分。

# 重要：需要一种显式的结构来学习jointly

复杂性（PPL）与对数版本的蒙面困惑（Devlin等人，2018年）不同，我们取那些被重写为困惑的附加标记的平均概率。产生的句子S0：



最后有个打分。上面那个是对邻接词替换的打分。下面则是bert的打分。大概是个句子的总体映像打分。



用的【cls】的向量。