实验报告：

SimLex-999

Scale Up Event Extraction Learning via Automatic Training Data Generation:

If we

can utilize such CVT information, we can then label sen-

tences without needing to explicitly identifying any triggers.

The first stage identifies the key arguments in a sentence.

If a sentence contains all key arguments of a specific event

type, it will be considered to imply an event mention of this

specified type.

XLNET：　Autorepressive vs Auto-encoding,其实BERT(DAE)的损失函数就是来自与Auto-encoding,

无监督的学习：那张有很多unlabeled data + labeled data

没有上下文

skipGram,

有上下文，

bert, elmo, xlnet

AutoEncoder

Bert 用的denoising AutoEncoder, 加了噪声，

自回归和自编码

Bert存在两个问题，被预测的单词的之间的独立假设；还有预测和

Emlo:不能存在双向

自回归：

缺点: 单向

优点：没有噪音和存在依赖关系

自编码：考虑双向，但是有噪音和独立假设

XLNET的memeory是什么？