Elmo双向lstm，基于特征。有对于生词的网络。（如何截取有效成分？若仅是要生词）

Gpt单向transformer

Bert双向transformer，任务是瓦解与nsp，静态mask。

Xlnet：乱序，但代表的其实只是求全概率的顺序不一样，并不代表句子的顺序变了。要求的词总是后几位。这也是种某种意义的和输入不一样。Stateful transformer。

Roberta：动态mask，去nsp，更大语料。

FreeLb-roberta：pgd对抗训练。

Albert：加了一层参数更少的层或者说降低了W的秩。以获得更多参数

T5：所有任务都是翻译任务。