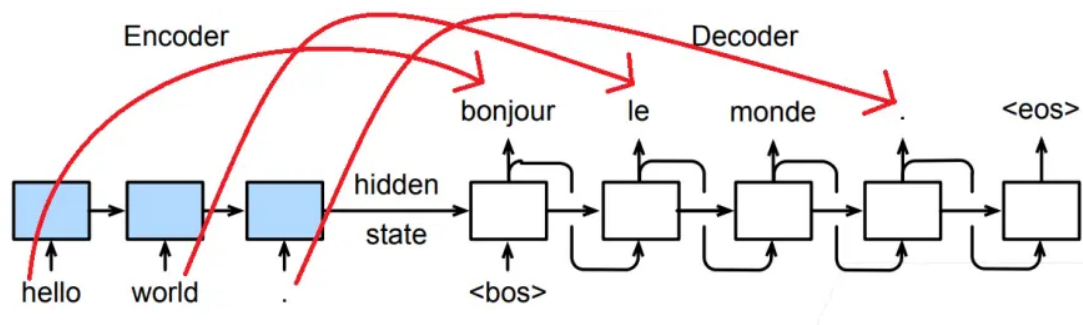


使用注意力机制的 Seq2Seq

- 动机

- 机器翻译中，每个生成的词可能相关于原句中不同的词



- 原有的 Seq2Seq 模型不能直接对此建模（解码器接收到的是编码器最后的隐藏状态，看不到之前的）
- 源句子中的所有信息虽然都包含在这个隐藏状态中，但是要想在翻译某个词的时候，每个解码步骤使用编码相同的上下文变量，但是并非所有输入词元都对解码某个词元有用。将注意力关注在源句子中的对应位置，这也是将注意力机制应用在 Seq2Seq 模型中的动机

- 加入注意力

- 编码器对每个词的输出(h)作为 key 和 value（它们相同）
- 解码器 RNN 对上一个词输出(h)是 query，即预测下个词时用当前预测输出和编码器各状态进行 attention 计算
- 注意力的输出和下一个词的 embedding 合并进入