* **词的定义**

词是指具有一定的整体意义和固定的语音形式、能够独立运用的最小的语言单位。词是语言的基本符号，它是语言中最小的自由的内容和形式相结合的实体单位。词不仅是基本的词汇单位，而且也成为语法的基本单位，语法中无论句法和词法，都以词为必要的基础。语言系统内的各种单位里面，词是居于中心地位的，作用最为突出。

* **词作为翻译单位**

1. 第一代机器翻译系统

第一代机器翻译系统使用的是word-level的翻译单元。这些系统是使用所谓的直接翻译策略设计的，将源文本中的单词直接映射到目标语言中的单词。根据特定的语法规则重新排序以适应源语言和目标语言的结构差别。

1. 统计翻译系统

1980年代末IBM首次开展统计机器翻译研究，IBM 的统计机器翻译方法的基本思想是，把机器翻译看成一个信息传输的过程，用一种信源信道模型对机器翻译进行解释。假设一个目标语言句子 e，经过某一噪声信道后变成了源语言句子 f，也就是假设源语言句子 f通过某种编码后得到了目标语言句子 e，而翻译的目的就是要将 f 还原成 e。并提出了统计机器翻译的基本公式。

1993年IBM提出五种基于词的统计翻译模型IBM Model 1-5

Mode 1将句子切分成词然后进行统计，只考虑单词之间的对译概率，不考虑语序。

Mode 2 考虑了翻译过程单词的位置变化

Mode 3 考虑了一个单词翻译成多个单词的情况，引入了产生概率。

Mode 4 在对齐时不仅仅考虑词的位置变化，同时考虑了该位置上单词的词类

Mode 5 修正模型4，消除了模型4中的缺陷，避免了对不可能出现对齐给出非零的概率

* **缺点：**

1. 第一代翻译系统

从大量翻译实践和理论可知把词作为翻译单位的逐词翻译 ,其操作的层次偏低 , 或歪曲原意 ,或行文不畅 ,往往造成死译或形式主义的翻译。

在不同语言的词汇中，意义能一一对应的词主要是一些专业术语。在大多数情况下，某语言中一个多义词甲的意义与另一语言中相应的乙词的意义往往是互相错综交叉的，它们可能在某个或某些义项上是相同的，而在其他几个义项上却不同，如汉语的“运动”和英语的“movement”，它们有相同的义项，如“表物体运动、社会运动等”，但是也有不同的义项，如汉语可指“体育运动”，英语可指“粪便”等。也可能甲词的全部义项包含于乙词之中，只和乙词的部分义项相同，如汉语的“猫”和英语的“cat”，后者包含前者的全部义项，而“心地恶毒的女人”这个义项是前者所不具有的；也可能某词或某个义项在其他语言的词汇中根本找不到对应的词或义项，例如一部分音译词就是在这种情况下产生的；另外，词语的附加意义或附属色彩，语言之间往往也难以直接对应。

因此，词对词的直接翻译策略很难提供高质量的译文。

1. IBM 词对齐模型

IBM提出的统计机器翻译基本方程式对于机器翻译领域具有开创性的意义，但从理论上讲，IBM 的几个模型只考虑了词与词之间的线性关系，没有考虑句子的结构。对于语序差别比较大的语言对而言，效果可能会不大好。而后基于短语的统计机器翻译方法对此进行了改进。

参考文献：

Martha Thunes. (2017).The concept of ‘translation unit’ revisited*. Bergen Language and Linguistics Studies,8*(1),241-259

Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, Robert L. Mercer (1993) The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational linguistics - Association for Computational Linguistics. 19*(2),263-311.

张卫萍&杨萍.(2002).翻译单位略论.*华北水利水电学院学报,社会科学版,*(2). 93 – 94