我们从小学就开始学习的造句方法大多是基于语法规则来完成的。人类需要记忆很多的语法规则，然后使用某种模糊逻辑的方法将这些规则灵活运用，并能进行一定程度的创新。但是目前看来，计算机还不可能拥有我们人脑这样的能力来进行任何程度上的创新。此外，没有人或者书籍可以囊括人类语言的所有语法现象，所以计算机暂时还没有可能按照人类学习语言的方法来处理自然语言。比如向计算机输入一个中文短句“我们决定去郊游”，计算机并不把句子拆分为主谓宾来分析，取而代之的是分析中文词或字共现的概念。对这个句子，计算机会分析在“我”字之后出现“们”字的概率，然后分析在“我们”之后紧接“决”的概率，以此类推，最后对每个句子，计算机都能给出一个总体的概率来表示这个句子正确的可能性，也就是：

自然语言处理就是研究如何让计算机理解并  
生成人们日常生活中所使用的自然语言 , 从而建  
立起人与计算机之间的密切联系, 使其能高效地  
进行信息传递和认知活动 .自然语言处理时经常  
遇到的问题有分词、词性标注、语法分析、句法分  
析和语义分析等 ,这些自然语言中的问题都可以  
使用一些基于规则的语言分析方法进行处理 .但  
对基于规则的系统来说 , 需要将专家的领域知识  
融入各种规则中 ,并且该方法随着规则库的增大  
效率明显下降.直到目前还没能出现一种比较完  
善的表示自然语言的规则系统 .  
随着计算机技术及 Internet 的迅速发展,大量  
联机语料库随之出现 , 计算机处理能力也大幅度  
提高 ,人们自然地想到利用统计方法对这些语料  
及新生成的语言进行分析处理 .由于语料库具有  
信息量大 、领域广 、真实及实时性强等特点 , 我们  
可通过对语料库进行深层加工 、统计和学习, 获取  
大规模真实语料中的语言知识 , 这就是所谓的基  
于统计的自然语言处理 .基于统计的自然语言处  
理模型使用分布函数来表示词 、词组及句子等自  
然语言基本单位 ,它描述了自然语言的基于统计  
的生成和处理规则