# 英语词缀与HNC符号的自动变化

张运良

（中国科学院声学研究所，北京 100080）

## 1. 本项研究的意义

英语是当今世界上最重要的语言之一，因此也是自然语言处理的一个重要研究对象。在英语的自动处理当中，必然面临着词语知识库相对处理的文本的不足的问题。因为英语词汇量很大，不可能也没有必要把所有的词汇都收入词语知识库；即使知识库的容量允许把所有的词语都收入知识库，新的词语也会层出不穷。在这种情况下，就呼唤一种自动处理机制。当然，由于派生只是英语新词产生的方法之一，这一研究对于英语自动处理能够起到有限的推动作用。这种机制在某种情况下是对人类猜词能力的一个模拟，如果能够结合句类知识进行排除，那么准确性也会提高，而不仅仅是提供可能的结果。另外，对于英语词缀的考虑，也许反过来会对于英语词语知识库的填写起到一定的指导作用。

## 2. 词缀对HNC符号变化的几种影响

由于目前没有英语的词语知识库，所以研究还是只能停留在演绎上，通过推演，列举其中的可能的情况，而且这种列举可能是不完备的。另外，而文中所举的例子，都是根据汉语相应词语做的一种推测。

### 2.1 词缀与类别符号的变化

这里的类别符号是指五元组符号（v，g，u，z，r）及它们之间的组合（以下简记为A类）和挂靠物概念的概念类别（p，w）及其派生和组合的概念类别（包括和五元组符号的组合，以后简记为B类）。这种情况下，概念的层次符号不发生变化。比如假设write对应的HNC符号为v******，那么writer对应的符号可以推导为p******。（******，仅仅表示两个自语对应的HNC符号相同的层次符号，并不表示具体字母或者数字的个数，下同。具体的符号需要知识库填写人员确定，也许这里******代表9238）。这种变化包括A类内部的变化，B类内部的变化和AB之间的变化三种情况。除了-er外，还有–ment，–ee，-eer，-ism，-hood，-ery等等。

### 2.2 词缀与层次符号的变化

原以为词缀不会改变对应的层次符号的高层，但是英语的词缀是很复杂的，涉及的层面是很多的，所以改变高层符号也是可能的。而且在实际寻找中发现这种情况是实实在在存在的。比如在基本物和物性概念之间转换，cloud（云）属于基本物概念，而cloudage（云量）属于物性概念。其它的象-ize，-ise，-fy，-ify，-able，-ible都可能发生了高层符号的变化

中层符号的变化最为普遍，中层符号主要表达局部联想脉络的对偶性，对比性和包含性特征。英语的词缀在这三个方面都会产生影响。

如arctic的HNC符号为******j21021，而antarctic的符号为******j21022。这就是从对偶性角度来看英语的词缀对HNC符号的影响，这种表示相反的词缀还很多，比如non-，un-，in-，il-，im-，ir-，ig-，dis-，de-，-less，等等。

词缀也可以改变中层对比性符号，比如infrared和ultraviolet就必然改变了*****cnk中的n值。事实上这里还涉及了另一个重要问题：本来n都是定义好了的，k在（1，n）的范围之内，所以在实际的应用中或者允许k超越n的范围，比如k＝0或者n＋1都是许可的，或者预留一部分空间，比如概念节点表的k的范围取自（2，n－1）。此外象super-，mini-，hypo-，over-，pre-等都能够从对比性特征角度改变中层符号。

词缀也可以从包含性概念角度改变中层符号，比如若age的符号为******－，则subage（亚代）的符号为******－0。类似的有hypo-，-age（集合，如baggage）等。

理论上讲，词缀对于词语的HNC符号的底层也会产生相应的影响。

## 3. 自动处理的方略

遇到新词，先分析词缀是什么，然后比较除去词缀部分和词表中的词的关系（因为添加词缀后，词干可能会发生变化，所以比较的结果给出两者相似的权值为宜），主要是词的构成，然后根据构成提出几种可能，利用对词缀的研究，给出几种可能的概念映射符号，带入句类中进行分析和检验。

## 4. 本研究的不足

首先，词缀可能不仅仅表达一个含义，如上面的hypo-，-age就是这样。所以必将为假设HNC符号带来困难。另外目前词缀的判断只能依赖于拼写形式，而有些根本不是词缀，如live和liver一点关系都没有。最重要的是没有权威的相对完备的英语词语知识库提供归纳和验证的材料。这决定了本研究只能停留在目前这一状态，而不能取得更深远的进展。

## 5. 总结

英语词缀纷繁复杂，本文试图通过英语词缀的变化表象得到对应的深层的HNC映射符号的变化规律。当然这项研究充满了重重困难。但是目前的研究已经为英语自动处理提供了解决部分新词识别的一种方法，同时这种考虑也反过来会促使HNC中层对比性符号制定时留有一定的余地，以便与后续的自动处理相配合。

---

作者简介：张运良（1979--），男，吉林九台人。现为中科院声学研究所硕士研究生，主要研究方向为自然语言理解、机器翻译。