

基本词典与专业词典的关系

刘志杰

刘 倬

北京高立公司语言信息研究所

中国社会科学院语言研究所

「摘要」实用型的机器翻译系统要面向科技文献的翻译，必需要解决好基本词典和专业词典的关系。以往人们设计一个面向科技文献的机器翻译系统时，绝大部分系统是优先去查专业词典，实践证明这种方法不可取，它不仅不能提高译文质量，反而使系统的译文质量下降。本文结合英语词汇学的理论，根据多年的机器翻译的上机实践提出一个解决方法，即以基本词典为主，专业词典为辅的原则，比较理想的处理了两部词典的协调关系，从而在一定程度上提高了系统的译准率。

一、以基本词典为主，设计好基本词典

根据英语词汇学的说法，现代英语中有 60 万个词汇，在这浩如大海的英语词汇中，最常用的是大约 5 千个词汇，这些常用词汇出现的频率极高，结构功能繁多，多义兼类现象严重，搭配能力和派生能力很强，这些词汇通用于各种文体，因此有人说这部分词汇是英语词汇的共核；那么共核之外还有 2 万多个通用词汇，它们的使用频率略低于共核，但也很常用，也通常用于各种文体，英语中称这部分词汇为次共核。一般的中型英语词典这部分词汇大约收录 4 万左右；除了共核词汇和次共核词汇以外，其它大量的英语词汇，可以大致称为非核词汇。

依据这个词汇学的基本理论，我们认为实用型的机器翻译系统中其基本词典的词汇量应该包括英语中共核词汇和次共核词汇，即基本词典中的词汇量至少应该收 4 万左右的词汇。而把一些非共核的英语词汇应该放在各类专业词典中去处理。

在基本词典的设计中，主要设计好动词、介词、关系词和连接词，这些词是基本词典中的主干。

二、以基本词典为主，专业词典为辅

在机器翻译系统中，决定一个英语单词语法功能的词典应该是基本词典，专业词典只能保障专业术语的准确性，不能决定一个英语单词语法功能，所以在协调基本词典与专业词典的关系上，应该是以基本词典为主，专业词典为辅的原则来处理词义的提取及词的基本用法分析。在查完基本词典确立该词的用法以后再决定是否查专业词典提取专业词义，这样做能解决许多两部词典的协调问题，主要方法如下：

- (一)词类相同时，按专业词典优先的原则提取词义
- (二)词类不同时，按兼类判别的结果控制词义的提取
- (三)词形不同时，以词形控制词义的提取
- (四)固定词组，以基本词典优先的原则控制词义的提取
- (五)特殊的动词普通和专业兼用，靠基本词典中给出的专业控制开关查询专业词典并提取词义

总之，基本词典是机器翻译系统的关键，是保证译准率的前提条件，要以基本词典为主的原则去协调两部词典的关系，用基本词典决定基本词汇的词法句法功能，用专业词典去解决科技术语的准确性。