英语语言中的兼类问题及英汉机器翻译系统的对策

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:研制并实现一个实用型的英汉翻译系统，有许多问题要解决。其中，英语语言中的兼类问题是要着重考虑的，因为这个问题涉及到词议的辨识和句法分析，直接影响到译文质量。

The Countermeasures of an English－Chinese Machine Translation System against some Problems of English Grammatical Category Ambiguities

Abstract:This paper mainly summarizes the phenomena of Grmmatical Category Ambiguities in the English language as well as some influences on English－Machine Tranalation System. On the other hand,the paper presents some appropriate countermeasures which deal with some problems of Grammatical Category Ambiguities in rearch and development of an English－Chinese Machine Translation System.

一、英语语言中的兼类问题

兼类词(Grammatical Category Ambiguity)是指具有不只一个词类特征的词。在英语语言中，兼类词是普遍存在的。如果从英语词汇学的角度来考虑这个问题，英语语言中的实词和虚词都存在着兼类问题。有人曾对英语中的兼类词进行过分类统计，英语中大约有二十多种兼类词，但是对一个实用型英汉机器翻译系统来讲，主要要解决的兼类就是实词和虚词的兼类问题。实词的兼类主要有以下几种：第一是名动同形，比如 cage 作为名词是"笼"的意思，作为动词是"装入笼中"的意思。第二是动名同形，比如 desire 作为动词是"想要"的意思，作为名词是"愿望"的意思。第三是形动同型，比如 calm 作为形容词是"平静"的意思，作为动词是"使安静"的意思。

另外还有其它兼类现象如：形名同形(如 bitter,natural)；副形同形(如 pretty)；助动同形(如 can 可用为助动词和动词)等等。

以上是英语实词兼类问题，即同形判别的问题。其中名动同形和动名同形的问题是主要的，也是一个实用型机器翻译系统中最难的问题。

实词的兼类在英语中较为突出，英语中的虚词兼类的问题也是存在的，比如，"when" 这个词(在高立电脑翻译系统中这种词我们称之为功能词)，它可以充当下列词类：

第一：用作疑问副词，如："When did you see Margaret last?";

第二：又可以用作连接副词，如："I would like to know when they will let him out.";

第三：同时它还可以作关系副词，如："This is the hour when the place is always full of women and children."

因为它的词类不同，它连接句子的功能就不一样，所以这种词也属于机器翻译中的同形判别问题。

二、兼类词的处理策略

由于英语中有"名词优于动词"的现象，而名动同形，动名同形等兼类问题又大量存在，功能词的兼类现象也不同程度的存在，所以这会对一个实用型机器翻译系统有一些不利的影响。我们在上机调试词典规则的时候，遇到了大量的这类问题，如果这类同形问题处理不当，会产生岐义结构，也会造成词义辨识和句法分析的失败。为此，我们必需在英汉机器翻译系统中来解决这个问题，这样才能使一个系统更实用化。

2.1 基本策略一：共性与个性相结合

建立抽象的同形区分的共性规则，再由词典规则根据需要随机调用，以实现具体词的同形辨识；极为特殊的，不能在共性规则中解决的，可以尽可能在具体规则中去解决。

以上基本策略的实现取决于系统的设计方法，我们高立电脑翻译系统的设计具有比较扎实的语言学基础，它以具有普遍意义的语言学公理理论和原则作为语言分析器的理论基础，以推理规则运算化的机器词典代替传统的信息参数词典，使句法规则与词的个性规则相结合。这样的设计思想为兼类判别提供了方便的条件。

下面我们以"book"(名动同形)的规则为例来看一下它的同形辨识情况：

(1) book 名词：书 汉语参数：本 形态参数：V# 语义参数：书

(2) book of reference → 参考书

(3) account book → 年刊

(5) bill book → 支票薄

(6) blue book → 蓝皮书

(7) book 排除动词的用法保留名词

(8) 如果 book 作成动词，则 book + NP → 预定

下面对其基本格式做一下简单说明。(1) 中的内容是 book 这个词的基本信息，其中要说明的是"本"是该词的汉语参数，用于源语分析以后的汉语生成，如 A book，汉语应译为"一本书"，"V#"表示该词有动词的用法，这个形态参数为以后做兼类判

别做准备。

从(2)到第(6)是有关book的一些成语规则。

第(7)是为用共性规则进行兼类判别而做的一条随机调用的规则，系统要从这条个性规则转入共性规则去做兼类判别，如果经过系统分析该词为动词，那么就做下一条规则(8)取"预定"的汉义。

下面我们用一个英文句子来说明一下系统的分析过程是如何把book分析成动词的。如：I book a ticket. 为了节省篇幅，我们只就"book"这个词而论。

查"book"的规则 .....[1]

经分析后跳过"book"的成语规则 .....[2]

转查"book"的调子规则(7).....[3]

见到该条规则后，进入系统的共性规则 .....[4]

经分析后，共性规则将"book"判成动词。这条共性规则是：

如果book的上位词不是形容词性的物主代词或名词的所有

格形式，则可以认为book是动词 .....[5]

如果book分析成动词，则去做(8).....[6]

通过以上的步骤我们可以看出，高立电脑翻译系统的语言处理规则分为个性规则和共性规则两部分。个性规则为共性规则创造兼类判别的条件，共性规则接受到个性规则提供的调子规则信息以后经过上下文环境的分析来完成兼类判别的操作。当然，并不是这么做了，就能解决英语中的一切兼类问题，对比较特殊的情况，应在个性规则下去解决。比如"pretty"这个词是形副同形，如何把它判为副词词义为"很"，我们认为就在这个词下用个性规则来判断就可以了。在"pretty的词上做这样一条规则：pretty+形容词(语义为"性")→相当。这样就把pretty的形副同形的问题解决了。

以上两个例子简要说明了英语中实词兼类判别的一些情况。这些情况说明英语中的兼类问题实质上是一个共性问题，我们可以通过形态分析和句法分析，有时甚至是用语义分析来解决这些问题。这些问题解决的好坏，直接影响到长句子的译文质量。我们知道，英语的长句中复合句偏多，而复合句多数都是由传统语法中所谓的从属连词，关系代词，或连词等来引导的。这些词，尽管是少数，但兼类问题却很普遍，这些问题不解决，一个实用型的机器翻译系统是不可能实用的，也只能在短句子上徘徊不前。下面我们应拿"that"这个词来看一下它的几种兼类情况。

首先它可以作连接代词，来构成宾语从句和表语从句，如：

例一：They drilled wells and hoped that they would find oil to make a profit.

例二：The fact is that China is very rich in natural resources.

其次，它可以当连词引导主语从句，如：

例三：That air is not an element is true.

还有，它可以作关系代词，来连接定语从句，如：

例四：There are also millions of tiny living things that float in the sea.

因为它有不同的语法功能，所以起的作用是不一样的，这种功能词的兼类现象对一个实用型机器翻译系统来说是一个严峻的挑战，处理好这种问题，对系统的进一步开发是比较关键的。我们知道，解决一个兼类词离不开它的上下文的语言环境，只有把这个具

体的词放到具体的语言环境中，才能找到它的支撑点，也只有找到了这个支撑点，我们才能确定它的语法作用，如上文例二这句话，我们可以在 that 的个性规则写这样一条规则：

如果"that"满足下列条件：

(1)BE(且不为 ING 词) / (2)NP(名词或人称代词) / (3)VP(动词是谓语)，那么"that"就为表语从句的用法。这条规则就可以解决"that"这个词引导的表语从句问题。

同样例三里的"that"的规则我们可以这样来描述：

如果"that"满足下列条件：

(1)that 前是句首或一个句子 / (2)that 引导的句子不能是疑问句 / (3)NP(名词或人称代词) / (4)VP1(that 从句的谓语) / (5) VP2 (主句的谓语)那么其结论应为主语从句。这样，我们就确定了"that"这个词引导的主语从句的问题。

例四是一个定语从句，可以用下面的规则解决：

如果(1)"that"前是个名词 / (2) "that"本身成分为空，那么其结论应为定语从句。

上文的例二，例三，以及例四中的"that"都是在该词的个性规则中去处理的，这么去做，我们认为是比较理想的，但是如例一，本句中的"that"引导的是宾语从句，所谓宾语从句，肯定是做及物动词的宾语，这样我们完全可以把"that"的这类规则放在其前的动词身上去解决，如动词"hear"有这么一条规则：

让"hear"去查共性规则，判断其是否带宾语从句。

当系统在分析到这条规则的时候，就去转查共性规则。这条规则是：

如果"hear"后面是"that"且"that"后有 VP(谓语)，那么这个"that"就是宾语从句的用法，这样就把 that 引导宾语从句的问题解决了。

这里我们想说明一下兼类问题放在个性规则中处理还是放在共性规则处理的基本标准。因为英语中名动同形和动名同形的现象较多，又因为英语中名词多于其它词类而大量存在，名词转化为动词的随机性很大，所以把这类问题最好放在共性规则中去处理。而其它词的兼类问题最好放在个性规则下去解决。但是无论放在那里，都要考虑到个性规则和共性规则的结合。这么去处理，我们认为有两个好处：

第一，能提高系统的处理速度。如果把兼类问题都归于共性规则中去判断，这样在分析的过程中势必要反复调用共性规则，这样处理的速度就要慢下来，不利于系统的优化。

第二，能使共性规则的条数减少一些，对于个性规则来讲，抽象性降低一些，这有利于规则的维护，规则的抽象性越高，维护起来就越困难，容易造成动一牵百的可能。所以不能将英语中任何语言现象都归于共性中去，尽可能在个性规则上下工夫。英语中越是个性的东西，用的就越普遍，用的越普遍，变化就越大，变化越大，对系统的影响也就越严重。所以这些个性的东西仅靠共性规则去解决，有时会走向反面，甚至会影响全局，就词论词，就在当前词条下去处理，是最好的方法，如果系统局部出了问题只是这个词的问题，甚至是一条规则的问题，而不是整个系统的问题。我们高立电脑翻译系统中兼类问题都是这么去做的，既要考虑到共性问题，又要考虑到个性问题。使共性和个性分开处理，是所有翻译系统采取的措施，而这里更重要的是如何使二者有机地结合起来。所谓有机结合，指的是在个性规则中可以调用共性规则；反之，共性规则也可以转回个性规则。

2.2 基本策略二：使用排除法．

过去，人们在处理兼类判别时使用的是分块独立分析。分块独立分析说到底就是把名动同形，动名同形，形动同形等等的问题各自集中在一块，从形式上看，这样显得规则分块明显，但是英语中的词有的是身兼几类，如果用这种方法去写规则，这势必要造成规则的大量重复，同时也大大增加了规则量，不利用规则的维护，如果控制不好，极容易造成釜底抽薪的恶运。

高立电脑翻译系统在这方面用的是排除法，所谓排除法是指在兼类词中排除一种出现可能性小的词类而去保留常见的词类。我们系统的共性规则就是以这个思想去做的，在共性规则中，我们分出了兼类判别的几种模块，它们包括判断名词的模块，动词的模块，形容词的模块，副词的模块等等。在这些模块中，我们根据英语的特点，或者保留名词而排除动词，或者排除名词而保留动词，等等。这里需要指出的是我们分出的几种模块同以往的分块独立分析中的模块是不同的，从形式上讲，我们分出的模块似乎是独立的，实际上这些模块间是相互联系的，它们之间可以实现相互调用，这样做是要把兼类问题中的一些共性问题集中起来，使这些问题形成一个有机的整体，为源语的分析打下基础；这样做还有一个好处就是减少共性规则的重复。拿实词来讲，不管同形兼类有多少种，只有名、动、形、副四个模块就足以了，这样就减少了规则的数量，对维护规则是非常有利的。

2.3. 基本策略三：利用函数的递归调用．

上面提到在共性规则中，把一些兼类判别的共性问题放在一起，分别形成几个模块，实际上这些模块就是一些函数体，这么去做规则是便于系统在进行源语分析时的递归调用。所谓"递归调用"是指自己调用自己，这为解决上下文中的两个兼类词(如名动同形)连用，可提供一定的方便，这样做有利于兼类词的上下文分析，分析结果的正确性也能得到保障。过去的机器翻译系统使用的是过程加工，这样做也许有一定的道理，但是灵活性差了一些，上下文中的分析功能差了一些，分析出的结果的正确性低了一些。因为英语句子中兼类词分布的随机性很大，在处理这个问题时，实施随调随用即函数的递归调用是比较理想的方法。

2.4. 基本策略四：概率的运用．

所谓"概率"，在高立电脑翻译系统中是指一个兼类词中出现的频繁最高的词类。即把一个英文单词的出现的最大可能的词类保留下下，而排除其它出现可能性较小的词类。比如 machine 这个词，名词出现的频率最高，那么我们就可以排除它出现的可能性较小的动词而保留其名词的词类，即名词优先；又如 design 这个词，因为它动词出现的频率最高，所以保留其动词而排除名词，即动词优先。

这个方法非常有利于个性规则的制作，利用这种方法去写个性规则能减轻共性规则的一些负担，同时又能减少规则的误差。这样使个性词典显得直观一些，对维护个性词典很有益处，同时为共性规则提供了较为准确的判断前题，为共性词典的成功分析兼类问题打下坚实的基础。

上文我们简单介绍了一下高立电脑翻译系统在兼类判别方面的一些处理手段，但是语言是千变万化的，仅仅靠有限的规则去概括复杂多变的语言现象是不可能的，也是有局限的，这样，有的语言现象就有可能解决得不太理想。请看下面的一句话：

As the airplane takes off and lands, you may feel your ears pop. 本句中的 pop 一词是一个兼类词（动形同形），而 feel 这类动词又有这样的用法：

FEEL＋NP＋VP（不带 TO 的不定式作宾语补足语）;FEEL＋NP＋AP（形容作宾语补足语）。

pop 是兼类词，feel 又有两个句型，则 feel your ears pop 是一个岐义结构。对于这类问题，我们认为可以保留岐义支翻译，当然这是没有办法的办法了。

三、小　　结

自从 1992 年以来，我一直在从事高立电脑翻译系统上机调试规则的具体工作，所以上文所讨论的问题是我本人在开发高立电脑翻译系统中遇到的一些具体问，以及解决这些问题的一些办法，整个文章实践性的东西较多。几年来的工作，我深刻体会到一个实用型机器翻译系统到底有多长的生命力，主要还是体现在系统的语言设计方面。一个系统若能够为语言工作者提供较为理想的语言处理环境，则这个系统是有能力进一步发展的。众所周知，机器翻译是一门跨学科的边缘科学，它的形式和方法是一门计算技术，而它的内容则反映了一定的语言学理论和思想。由此，我个人认为，一个机器翻译系统是否有发展前途，主要取决于系统的语言学的功底是否牢靠，而不是程序的不断修补。本文所提出的一些问题我衷心希望能得到先驱者和同行们的指教。

参 考 文 献

[1] 刘　倬：开放型机器翻译系统的设计和实现《第三届中文信息处理国际会议论文集》

[2] 刘倬、傅爱平、李维：基于词专家的机器翻译系统《中文信息学报》1989.4

[3] 李　维：机器翻译词义辨识对策《中文信息学报》1990.1

[4] 冯志伟：迈向实用化和商品化的机译研究《语文建设》1994.7

[5] 孙茂松：汉语中的兼类词，同形词类组及其处理策略《中文信息学报》1989

[6] 张韵斐：《现代英语词汇学概论》北京师范大学出版社　1987.4