第 41 卷 第 6A 期
2014 年 6 月

计 算 机 科 学
Computer Science

Vol. 41 No. 6A
June 2014

面向专利机器翻译的要素句蜕识别和转换研究

张冬梅 晋耀红

(北京师范大学中文信息处理研究所 北京 100875)

摘 要 为了改善专利机器翻译中要素句蜕的翻译效果,提出了一种基于规则的要素句蜕识别和转换方法。通过分析汉语要素句蜕的格式,提取了汉语要素句蜕的描述特征,在此基础上制定了要素句蜕的识别规则。通过对比汉英要素句蜕,总结了两者的差异,在此基础上制定了汉英要素句蜕的转换规则。最后,将识别规则和转换规则应用到一个已有的机器翻译系统中。测试结果表明,这种方法可以有效地实现对要素句蜕的识别和转换,进而提高了专利文本中要素句蜕的机器翻译效果。

关键词 要素句蜕,识别,转换,规则,机器翻译,专利

中图法分类号 TP391.1 **文献标识码** A

Recognition and Transformation for Element Sub-sentences in Patent Machine Translation

ZHANG Dong-mei JIN Yao-hong

(Institute of Chinese Information Processing, Beijing Normal University, Beijing 100875, China)

Abstract This paper proposed a rules-based method for recognizing and transforming the element sub-sentences to improve the translation quality of them in patent machine translation. By analyzing the format of Chinese element sub-sentences, we extracted the description features, and created the recognition rules based on that. By comparing Chinese and English element sub-sentences, we summarized the differences between them, and created the transformation rules based on that. At last, we applied the rules to an existing MT system. Experiment shows that the method can recognize and transform element sub-sentences, and then improve their translation quality in patent machine translation.

Keywords Element sub-sentences, Recognize, Transform, Machine translation, Patent

1 引言

所谓要素句蜕是指一个句子蜕化为语块或语块的一部分,形式上变成一个定中短语结构:其中原句子的一个语块作为描述中心,其他语块作为修饰成分。如例 1(“〈〉”括起来的部分是要素句蜕,下同):

例 1 所述屏幕将〈投影器投影的图像〉分离至各个视场以形成 3D 图像。

参考译文 The screen separates〈an image which is projected from a projector〉 into fields for realizing a 3D image.

在例 1 中,语块“投影器投影的图像”由句子“投影器投影图像”蜕化而来,原句子中的语块“图像”成为了描述中心,其余部分“投影器投影”成为了修饰成分,并且增加了一个“的”字来表示这种修饰关系。我们把形如“投影器投影的图像”这样的由句子蜕化而来的定中短语称为要素句蜕。要素句蜕仍然能传达出原句子所蕴含的语义,只是侧重点可能发生了变化。

在现代汉语中,要素句蜕是一种常见的语言现象,以本文研究的专利语料为例,在笔者随机抽取的 932 个专利句子中,含有要素句蜕的句子有 223 个,约占到句子总数的 24%,要素句蜕的总数是 291 个。可以说,要素句蜕的翻译效果将直接影响到机器翻译的整体效果。

要素句蜕的机器翻译具有较大难度,首先,汉语要素句蜕本身比较复杂,要素句蜕是由句子蜕化而来的定中短语结构,因此除了表示修饰关系,内部还存在主谓或动宾关系,这无疑增加了结构的复杂性;其次,汉英两种语言中的要素差异明显,汉语要素句蜕只有定中短语这一种形式,而英语要素句蜕的形式要丰富得多,例 1 中的英语要素句蜕是一个定语从句形式,这个英语要素句蜕还可以表述成“an image projected from a projector”,这是一个动词过去分词短语的复合形式。

汉英要素句蜕的形式差异,要求在机器翻译时必须对要素句蜕进行转换处理,转换的前提是识别,因此,本文的研究分为要素句蜕的识别和转换两个子任务,并将最终的研究成果应用到一个已有的机器翻译系统上,以提高要素句蜕的机器翻译效果。

2 相关研究

要素句蜕的概念最早在文献[1]中提出,文献[2]对要素句蜕的相关内容进行了详细解释和论述,这两篇文献是进行要素句蜕研究的理论基础,本文中提到的很多概念和术语都

本文受国家高技术研究发展计划(863)(2012AA011104)资助。

张冬梅(1982—),女,博士生,主要研究方向为中文信息处理、机器翻译,E-mail:zhangdongmei@mail.bnu.edu.cn;晋耀红(1973—),男,教授,主要研究方向为自然语言处理、机器翻译。

源自其中。

文献[3]提出了句蜕处理的规则及算法，这些算法和规则的提出，标志着要素句蜕从语言描述阶段进入了可计算阶段，但该研究侧重于理论方面的探索，研究结果以战略性为主，制定的规则比较笼统，并且没有对规则进行形式化的描述，因此无法直接将研究成果用于机器翻译系统。文献[4]中描述了要素句蜕的构成，通过汉英对比导出了汉英变换规则。文献[3,4]为本文的研究提供了启发和借鉴，本文在其基础上，进一步总结汉语要素句蜕的构成规律，并制定了具体的形式化的识别和转换规则。除此之外，文献[5]总结了汉语要素句蜕翻译成英语时的几种常见形式及转换规律。此外，还有一些研究从定中短语的角度对要素句蜕这一语言现象进行了研究和论述[6,7]。

上述提到的研究主要是关于要素句蜕的基本理论及要素句蜕机器翻译的基本规律总结。这些研究为面向机器翻译的要素句蜕的识别和转换研究提供了参考和借鉴意义，但没有为要素句蜕机器翻译的实现提供完整的形式化规则体系，因此也都未能将研究成果直接应用到机器翻译中。本研究正是要在这些研究的基础上，面向机器翻译的实际需要，对要素句蜕的识别和转换规律进行形式化的描述，制定出具体的、细化的规则，并将规则应用到一个已有的机器翻译系统中，改善要素句蜕的翻译效果。

3 汉语要素句蜕识别

本节将首先在句子的框架内对汉语要素句蜕进行分析，接下来按类型对汉语要素句蜕的格式进行分析总结，最后，基于规则对汉语要素句蜕进行识别。

3.1 汉语要素句蜕分析

要素句蜕由句子蜕化而来，因此，对它的分析也应该在句子的框架内进行。以下面的句子为例：

例 2 〈由 A/D 转换器转换的电压和电流〉被转换为功耗。

参考译文：〈The voltage and current converted by the A/D converter〉 are converted into the power consumption.

在这个句子中，要素句蜕"由 A/D 转换器转换的电压和电流"由句子"电压和电流由 A/D 转换器转换"蜕化而来，这个句子的格式是"GBK2＋L0＋GBK1＋EK"[1)]，我们对由该句子蜕化而来的要素句蜕也进行类似的分析，得到如下的分析结果（见图 1）。

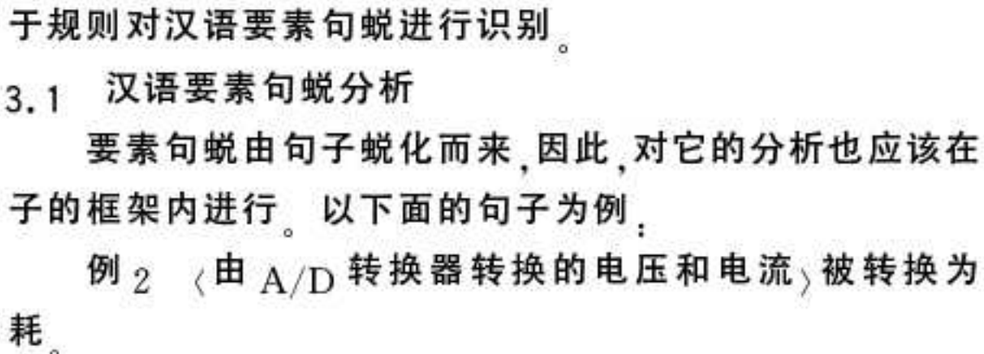

图 1 汉语要素句蜕分析结果

该汉语要素句蜕的格式是"l0＋gbk1＋ek＋mk＋gbk2"，其中 jtmk 表示要素句蜕的标记，在汉语中，jtmk 通常对应"的"字，少数情况也用"之"来表示，其他符号的含义与句子层面相应的符号意义相同，只是为了表示句蜕比句子低一个层级，采用小写字母来表示。

3.2 汉语要素句蜕类型和格式

句子蜕化成要素句蜕时，原来的一个语块成为了描述中心，我们对要素句蜕的分类就是基于其描述中心。以汉语最常见的三块句为例[2)]，可以蜕化成 GBK1、GBK2 和 EK 3 种要素句蜕，具体每种类型的要素句蜕构成格式如下：

(1)GBK1 要素句蜕的两种格式是：ek＋gbk2＋jtmk＋gbk1、l0＋gbk2＋ek＋jtmk＋gbk1；

例 3 〈构成//ek 该破坏层//gbk2 的//jtmk 树脂//gbk1〉为脂环式聚烯烃树脂。

例 4 类似于上述的实施例，〈把//l0 树、灌木丛或柱//gbk2 从土地移除//ek 的//jtmk 动力源 312//gbk1〉提供了足够的动力。

例 3 和例 4 的要素句蜕分别对应 GBK1 要素句蜕的两种格式。

(2)GBK2 要素句蜕的两种格式是：gbk1＋ek＋jtmk＋gbk2、l0＋gbk1＋ek＋jtmk＋gbk2；

例 5 在本文所述的工艺方法之前，大部分的 MDI 罐是采用〈专有商业//gbk1 供应//ek 的//jtmk 烃类和乳化剂混合物//gbk2〉进行溶剂清洗。

例 6 〈由//l0 接头 10、20 或 30//gbk1 形成//ek 的//jtmk 密封效果//gbk2〉在很大程度上决定了管子的压力等级。

例 5 和例 6 中的要素句蜕分别属于 GBK2 要素句蜕的两种格式。

(3)EK 要素句蜕的两种格式是：gbk1＋l0＋gbk2＋jtmk＋ek、gbk2＋l0＋gbk1＋jtmk＋ek。

例 7 〈移动终端//gbk1 对//l0 信号能量//gbk2 的//jtmk 探测//ek〉有助于减低复杂性并有助于移动终端在通信系统中被同步化及有助于移动终端跟踪下行链路信号。

例 7 中的 EK 要素句蜕属于第 1 种格式，EK 要素句蜕的第 2 种格式基本上只在理论上存在，在研究过程中所分析的语料中并未见到。

从上述分析可以看到定中短语是汉语要素的唯一形式。因此，从整体上看，要素句蜕具有名词属性和语法功能。从内部构成来看，GBKm(m=1,2,3)要素句蜕的描述中心一定是名词或名词词组；就修饰部分而言，如果不含 l0，通常是动宾或主谓短语，如果含有 l0，则是"介词短语＋动词（或动词词组）"的形式。EK 要素句蜕的描述中心一定是动词，但这个动词已经名词化了；EK 要素句蜕的修饰部分通常会含有 l0，并且是"名词（或名词词组）＋l0＋名词（或名词词组）"的形式。

需要说明的是这里列举的要素句蜕格式，是指要素句蜕基本格式，不包括语块省略、语块分离以及带有辅块的情况，在识别时，对于这些情况我们都予以了考虑。

3.3 汉语要素句蜕识别

要素句蜕的识别需要分两个步骤完成，首先是规则匹配，

---

1) GBK1、GBK2 分别表示广义对象语块 1、2，EK 表示特征语块；L0 表示主语块的标记，在汉语中通常是介词，如"对"、"把"、"由"、"将"、"被"，等等。详细内容请参考文献[1,2]。

2) 三块句的 3 个语块分别是 GBK1、GBK2 和 EK。

其次是节点生成。

1. 规则匹配

根据对要素句蜕格式的分析和总结，我们制定了要素句蜕的识别规则。识别规则实际上是对要素句蜕的格式进行符号化描述。从上面的格式分析可以看到：要素句蜕的构成要素通常包括 gbk、ek、l0 及作为 jtmk“的”。理想的情况是采用这四者来描述要素句蜕的格式，但是因为 gbk 本身的构成比较复杂，不易描述，我们选取了其余三者作为描述特征，也就是依赖点[9]，因为这些特征本身容易描述和辨识，并且能准确地概括出某一种类型要素句蜕的特点。

(1)“的”

“的”是汉语要素句蜕的标记[1]，也是要素句蜕重要的提示信息。

(2)ek

ek 是构成要素句蜕的必要条件，没有 ek，要素句蜕就不存在。汉语要素句蜕的 ek 通常是 v 类概念及其修饰和附加成分，它的识别相对容易和简单。

(3)l0

l0 是要素句蜕的另一个重要提示信息。汉语里的 l0 数量非常有限，可以穷举，通常是一些介词，如“由”、“对”、“将”、“被”和“把”，等等。

至于 gbk，虽然可能因其构成复杂而难于描述，但可以采用一种“模糊策略”来处理。从 3.2 节的格式分析中可以看到：gbk 或者定位于 ek，l0 以及“的”三者中的任意两者之间，或者位于块首/块尾，比如 l0(“由”)与 ek 之间的部分就作为 gbk1。

建立规则时还要考虑位置因素。从要素句蜕的格式分析可以看到：我们所选的 3 个描素特征(或其中的两者)之间的相对位置不同，要素句蜕的类型和格式也就不同。再有就是语块的开始和结束两个绝对位置，也是识别时需要考虑及可以借助的因素。

经过上面的描述特征选取和位置分析后，针对例 2 中的要素句蜕，制定了如下的识别规则：

(b){(-3)LC-CHK[L0]&CHN[由]&GBK-B%}+(b){(-2)LC-GCC[W,G,P]}+(-1){LC-CHK[E]&LC-E-SCORE[EL]}+(0)CHN[的]}+(f){(1)GBK-E%&LC-GCC[W,G,P]}=>PUT(-3,JTL)+PUT(-2,JTL)+PUT(-1,JTL)&PUT(-1,LC-E-SCORE,EL-HIGH)+LC-TREE(JTMK,0,0)&PUT(fp,LC-EXP,YSJT-GBK1)&PUT(fp,JTL)+PUT(1,JTL)$

下面对规则的含义进行说明：“=>”的左边是规则的条件部分，右边是操作部分。

条件部分的“{}”内部是对终结符或非终结符的特征的描述，“&”表示“并”的关系。其中用 LC-CHK[x]表示语义属性，如 LC-CHK[E]表示该节点的语义属性是 EK，并且用 LC-E-SCORE[EL]表示它是一个局部(要素句蜕中)的 EK，也就是说这个位置上是要素句蜕的 ek；LC-GCC[x]表示广义概念类别，如 LC-GCC[W,G,P]表示这个节点可以是具体物概念(W)、抽象物(G)，还可以是人(P)；CHN[x]表示汉语的词形或字形，如 CHN[的]表示这个节点对应的是汉语的“的”字，等等。

规则条件部分还包含了位置信息：“( )”的数字表示的是相对位置；其中(0)表示规则的切入点，也就是从该节点起向左右匹配规则，(b)和(f)则分别表示向左和向右找到满足条件的节点，如“(0)CHN[的]}+(f){(1)LC-GCC[W,G,P]}”表示从(0)号位置“CHN[的]”向右能找到“LC-GCC[W,G,P]”。如果节点前没有(b)或(f)，则表示该节点与前/后一节点必须是紧邻的。GBK-B%和 GBK-E%是两个绝对位置，分别表示语块的开始和结束。

操作部分的“( )”内的数字所指位置与条件部分所指位置相同；LC-TREE(CHK-VALUE,start,end)表示生成一个非终结符，也就是一个 TREENODE，CHK-VALUE 值为 LC-TREE 的取值，start,end 是规则左边的编号，程序自动对应到原始串中的绝对位置，比如 LC-TREE(JTMK,0,0)表示在 0 号位置生成一个值为 JTMK(其含义为“要素句蜕标记”)的 TREENODE；PUT(fp,LC-EXP,value)表示给生成的节点赋属性值，如(fp,LC-EXP,YSJT-GBK1)表示给新生成的节点附上属性值 YSJT-GBK1(其含义为“GBK1 要素句蜕”)。

使用这条规则，不仅可以实别出例 2 中的要素句蜕，还为要素句蜕的构成要素打上了各种标记。

2. 节点生成

节点生成是基于规则匹配的结果。首先，如果一个语块成功匹配了要素句蜕规则，就将该语块生成要素句蜕(YSJT)的父节点。其次，生成构成这个要素句蜕的各个子节点，子节点的生成分成两种情况：对于已经明确打上 JTL 标记的 ek、l0 和“的”，直接生成相应的节点；对于 gbk，则是将 ek、l0、“的”以及 GBK-B%和 GBK-E%任意两者之间的非空部分生成 gbk。

至此，要素句蜕的识别完成，得到如图 1 所示的结果。

4 汉英要素句蜕转换

本节首先分析英语要素句蜕的格式，然后进行汉英要素的对比，最后根据对比结果制定要素句蜕的转换规则。

4.1 汉英要素句蜕对比

为了与例 2 的汉语要素句蜕进行对比，我们对与之对应的英语要素句蜕“The voltage and current converted by the A/D converter”也进行了相同的格式分析，分析结果如图 2 所示。

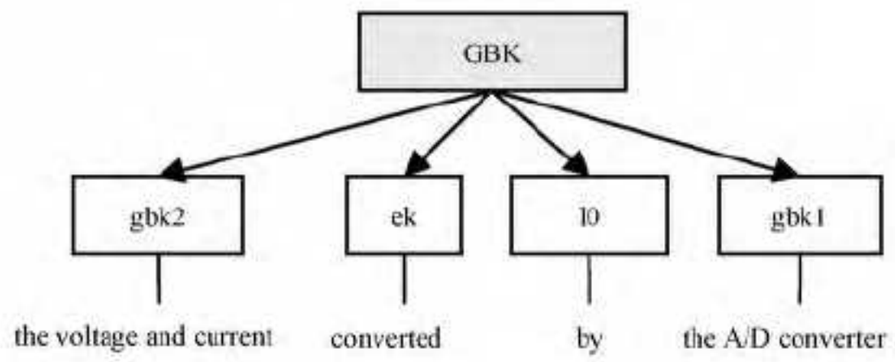

图 2 英语要素句蜕分析结果

该英语要素句蜕的格式是“gbk2+ek+l0+gbk1”(“ek”是非谓语动词形式)，这与对应的汉语要素句蜕有比较明显的差异。

通过分析双语对齐语料，我们总结出汉英要素句蜕的差

1) 要素句蜕中也有省略“的”的情况，但是极少见，在本研究中不考虑这种情况。

异主要体现在形式和格式两个方面。

1. 形式差异

汉语要素句蜕只有定中短语一种形式，上文所列举的汉语要素句蜕都是定中短语。英语的要素句蜕主要有3种形式：

(1)先行词+定语从句

例1中参考译文中的英语要素句蜕就是这种形式。这是英语要素句蜕最常见的形式之一。

(2)非谓语动词的复合形式

这是英语要素句蜕的另一种常见形式。如“The voltage and current converted by the A/D converter”是动词过去分词的复合形式，而要素句蜕“The resin constituting the destructive layer”则是动词现在分词的复合形式。

(3)多个逻辑单元的组合[1)]

这通常是英语EK要素句蜕的常见形式，如下面的例子中的“measurement of desorption of volatile organic compounds with gas chromatography/mass spectrometry”。

例8 其温度和在此温度下的时间是通过〈气相色谱/质谱法对挥发性有机化合物解吸的测定〉来选择的。

参考译文：The temperature and time at temperature are selected based on 〈measurement of desorption of volatile organic compounds with gas chromatography/mass spectrometry〉

2. 格式差异

格式差异主要指构成汉英要素句蜕的各个成分之间的差异。

(1)位置差异

除了描述中心的位置差异，其他成分之间的排列顺序也不尽相同，这在上文所举的例子都有体现。

(2)形态差异

主要是指ek的形态差异，汉语要素句蜕的ek在形态上与句子层面的ek没有差异，英语因为受形态约束，要素句蜕的ek往往采用非谓语形态，比如现在分词或过去分词。

(3)成分有无的差异

汉英句蜕的gbk，ek有良好的对应关系，尽管可能存在位置、形态上的差异。但对于jtmk和l0来说，二者之间却不一定有对应关系，汉语要素句蜕一定以“的”作为jtmk，但英语的要素句蜕可以没有jtmk，而以ek的形态变化来代替；汉语中的一些l0在英语中也不存在对应者，比如下面的例子。

例9 在上述外周面与上述对置面上，设有〈将上述流路形成为弯曲的形状的凸凹部〉。

参考译文：On the outer circumference and the opposed face，〈a convex part and a concave part that form the passage into a bent shape〉 are formed.

例9中的中文要素句蜕格式是l0+gbk2+ek+gbk3+mk[2)]+gbk1，与之对应的英文句蜕格式是gbk1+mk+ek+gbk2+gbk3。与前者相比，后者不包含l0。

4.2 影响汉英要素句蜕转换的因素

(1)类型

要素句蜕的类型是影响转换的首要因素，不同类型的要素句蜕通常转换成不同的英语形式。汉语EK要素句蜕通常转换成英语的多个逻辑单元的组合形式，而汉语GBKm(m=1,2,3)要素句蜕则通常转化为英语的定语从句或非谓语动词的复合形式。汉语GBK1要素句蜕和GBK2要素句蜕都可以转换为非谓语动词的复合形式，但是前者通常转换为动词现在分词的复合形式，而后者通常转换成动词过去分词的复合形式。

(2)格式

同种类型的汉语要素句蜕，格式不同，在翻译成英语时也会有差异。表达同一语义的要素句蜕可以呈现不同的格式，格式不同意味着强调的重点会有差异，因此在转换时，也应该保留这种差异，所以不同的格式的要素句蜕在转换成英语时，也应有所差异。

4.3 汉英要素句蜕转换

通过4.1节中汉英要素句蜕的对比，我们已经明确了两者的差异，在此基础上，我们总结出汉语要素句蜕转换成英语时的转换要点：(1)整体形式的改变，如由定中短语转换成定语从句；(2)调序，对节点重新排序；(3)增删节点；(4)改变节点的形式，这一点通常是就对ek而言。

仍以例2的要素句蜕为例，需要对其进行的转换包括：(1)调整各个节点的顺序；(2)删除jtmk(“的”)所对应的节点；(3)将ek(“转换”)变成过去分词形式。

基于上面的分析，我们制定了如下的规则，实现对例2中的要素句蜕转换：

(-3){LC-€HK[L0]&CHN[由]}+(-2)LC-€HK[GBK]+(-1)LC-€HK[EK]+(0){LC-€HK[JTMK]&CHN[的]}+(1)LC-€HK[GBK]=>(1)+(-1){VOI=P}+(-3)+(-2)+DEL-NODE(0)$

规则的左边是根据3.3节的识别规则得到的汉语要素句蜕识别结果，具体含义是：汉语要素句蜕的格式是l0+gbk1+ek+jtmk+gbk2，并且其中的l0是“由”，各部分对应的节点从(-3)到(1)，其中(0)号节点是要素句蜕的标记“的”(jtmk)。右边部分则是进行的转换操作，具体包括：对节点的顺序进行了调整，调整后的顺序为(1)、(-1)、(-3)、(-2)；将(-1)号节点变成过去分词形式(VOI=P)；删除(0)号节点(DEL-NODE(0))。

通过这条转换转换规则，对3.1节实现了如图3所示的转换。

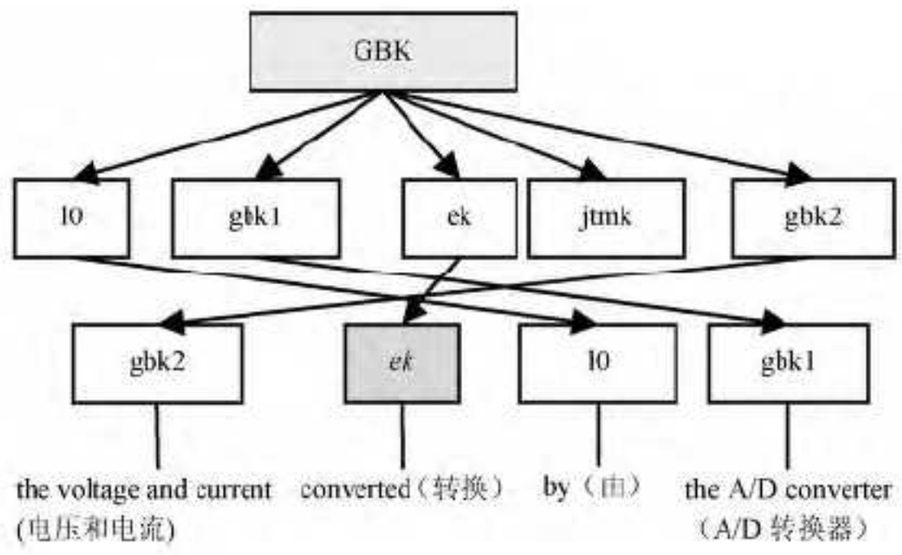

图3 汉英要素句蜕转换的实现

1) 这一概念来自HNC理论，指的是语言逻辑连接词(如汉语中的“的”，英语中的“of”)所联结的两个以上的组合单元(通常是名词性的)。

2) 英语要素句蜕中的mk通常对应的是定语从句中的引导词，如that、which，等等，在这个例子里对应的是“that”。

5 要素句蜕处理在机器翻译中的应用

我们将制定好的要素句蜕识别和转换规则应用到现有的一个面向专利文本的机器翻译系统中[9]，这个机器翻译系统是基于 HNC 理论的规则系统，该系统通过分析、转换、生成 3 个阶段实现对专利文本的汉英翻译，其中分析和转换都分为句子和语块两个层面。要素句蜕的识别和转换处理分别属于语块层面分析和语块层面转换(见图 4)。

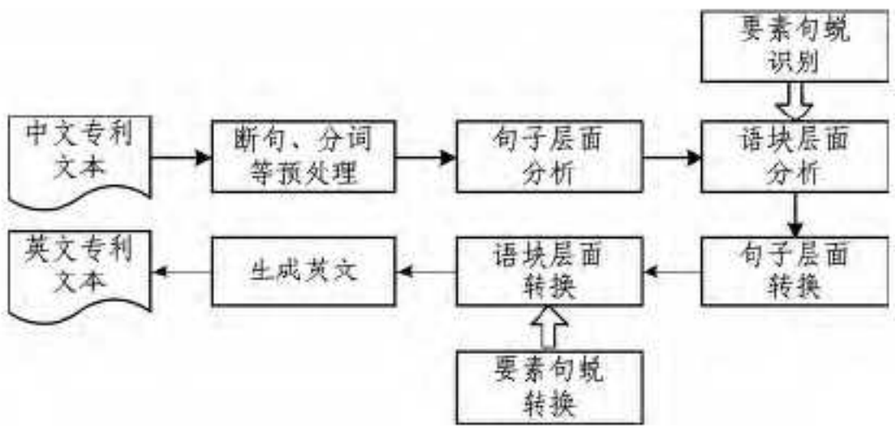

图 4 要素句蜕处理在已有机器翻译系统中的应用

为了验证规则的效果，我们选取语料进行了测试，测试语料来自中国专利信息中心专利检索系统的双语对齐语料，从中选取了 1000 个句子进行测试，得到的具体测试结果如表 1 所列。

表 1 要素句蜕识别和转换结果

要素句蜕总数	识别总数	正确识别	转换总数	正确转换
242	217	163	214	158

根据统计数据，得到要素句蜕识别的正确率和召回率分别是 75.1%和 67.6%，要素句蜕转换的正确率和召回率分别是 73.8%和 65.3%。

为了检验要素句蜕识别处理是否改进了机器翻译效果，我们针对包含要素句蜕的句子，分别测试了其在基线系统和添加了要素句蜕处理后的系统中的 BLUE 值，如表 2 所列。

表 2 添加要素句蜕处理前后的 BLUE 对比

BLUE	1-gram	2-gram	3-gram	4-gram
基线系统	0.4452	0.2522	0.1713	0.1350
添加句蜕处理后	0.4466	0.3595	0.2702	0.2144

如表 2 所列，从两个结果对比可以看到添加了要素句蜕处理系统的 BLUE 值要好于基线系统的 BLUE 值，尤其是前者的 2-gram、3-gram 和 4-gram BLUE 值明显好于后者。这说明要素句蜕处理可以明显改善含有要素句蜕的句子的机器翻译效果。

通过对识别错误的句子进行分析，我们发现引起要素句蜕识别错误或未能识别的主要原因有两点：

一是句子的语块切分错误，要素句蜕识别是在语块中进行的，如果语块切分得不对，要素句蜕很难识别正确，如例 10。

例 10 这些计数器//对//这些数据输入/输出装置发出的总线分配请求数//进行计数。

系统对这个句子的切分结果如下：

这些计数器//对//这些数据输入/输出装置发出的总线//分配//请求数进行计数。

在这个切分结果中，要素句蜕“这些数据输入/输出装置发出的总线分配请求数”所在的语块被从中间切开了，因此要素句蜕也就不可能识别正确。

二是结构嵌套，即要素要蜕又包含另一个要素句蜕，如例 11。

例 11 本发明具体应用于这样的复合产品，它包括〈由〈金属制成的主体〉形成的聚合部件〉。

这个例子中，含有两个要素句蜕，其中“金属制成的主体”这一要素句蜕嵌套在上一级要素句蜕“由金属制成的主体形成的聚合部件”中，嵌套就意味着存在两个 ek、两个“的”，甚至两个 10，这无疑会增加规则匹配的难度。

结束语 本文提出了一种面向专利机器翻译的要素句蜕识别和转换的处理方法。通过分析汉英双语对齐的专利语料，总结了汉语要素句蜕的类型及每一种类型的格式表示，在此基础上总结了要素句蜕识别的普通规律，从而制定了要素句蜕的识别规则；通过对比汉英要素句蜕的差异，总结了汉英要素句蜕的转换规律，并在此基础上制定了汉英要素句蜕的转换规则；在一个已有的面向专利文本的机器翻译系统上，对识别和转换规则进行了验证。实验结果表明：本文提出的方法可以有效地实现要素句蜕的识别和转换，从而改善了专利机器翻译中要素句蜕的翻译效果。但是由于要素句蜕本身的复杂性，以及受其他处理结果的影响，对一些要素句蜕的识别和转换处理没有达到预期的效果。

在下一步的工作，我们将通过分析更多的语料，深化对要素句蜕识别和转换规律的总结，并进一步修改和完善规则，同时加强对与要素句蜕相关的其他内容的处理，以提高要素句识别和转换的准确率、召回率，进一步改善要素句蜕的机器翻译效果。

参 考 文 献

[1] 黄曾阳. HNC(概念层次网络)理论——计算机理解语言研究的新思路[M]. 北京：清华大学出版社，1998

[2] 苗传江. HNC(概念层次网络)理论导论[M]. 北京：清华大学出版社，2005

[3] 李颖. 句蜕构成及汉英变换处理[D]. 北京：中国科学院声学研究所，2004

[4] 李颖，池毓焕. 汉英机器翻译中要素句蜕变换初探[M]//黄河燕. 机器翻译研究进展——2002 年全国机器翻译研讨会论文集. 北京：电子工业出版社，2002：162-171

[5] 刘智颖，晋耀红，池毓焕. 汉英专利机器翻译中的要素句蜕研究[C]//IALP 2011：2011 International Conference on ALP. 2011：193-196

[6] 刘丹青. 汉语关系从句标记类型初探[J]. 中国语文，2005，1：3-15

[7] 朱德熙. 说“的”[J]. 中国语文，1961，12：1-15

[8] Zhu Xiao-jian，Jin Yao-hong. Hierarchical Semantic-Category Tree Model for Chinese-English Translation[J]. China Communications，2012，12：80-92

[9] Zhu Yun，Jin Yao-hong. A Chinese-English Patent Machine Translation System Based on the Theory of Hierarchical Network of Concepts[J]. The Journal of China Universities and Telecommunications，2012，19：140-146