

# 汉语句蜕处理\*

晋耀红

知识创新基地语言语音及交互信息技术部

中国科学院声学研究所 100080

E-mail:jinyaohong@sina.com

**摘要:** 汉语句蜕现象丰富多彩, 本文讨论了句蜕现象的各种表现, 从句类的角度给出句蜕的预期, 并根据句类演绎规则, 给出了句蜕的处理策略和解决方案。

**关键词:** 句蜕 句类 句类演绎规则 预期

## The Processing for Chinese Sentence Degeneration into Chunk

Jin yaohong

Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080

Jinyaohong@hotmail.com

**Abstract:** The sentence degeneration into chunk (SD) is rich and colorful in Chinese. The paper discusses the behave of three types of sentence degeneration which are Key-Element SD, Packed SD and Original-Shape SD. And from sentence category the expectation of SD is given. Thorough Sentence Category Deduce Rule(SCDR), the processing strategy for SD is stated.

**Keyword:** sentence degeneration, Key-Element SD, Packed SD and Original-Shape SD, expectation, Sentence Category Deduce Rule

## 1. 前言

我们先看一组句子。

美国发表的人权报告指责巴西限制公民权利。

参与这个计划的俄罗斯企业也遇到了一些严重的问题。

这一切都体现了中共中央对本期理论研讨班的高度重视。

我们开发产品的决心不会改变。

香港接受了一项开创历史先河的光荣任务。

邓小平同志是我党我军我国各族人民公认的享有崇高威望的卓越领导人。

以上这些句子, 就是汉语句蜕的典型代表。

如果把第一个句子翻译成英文, 大致是:

---

\* 本文得到 973 项目 G1998030506 的支持

The report which USA had released criticizes Brail to restrict the human right.

可以看出来，在英语中，使用了一个“which”从句表达同一意思。汉语由于缺乏“which”之类的连词，所以采用了对句子进行适当变形的方式来表达。汉语主要的变形手段就是对句子的整体或局部加“的”。

从语义上来说，句子中下划线的部分虽然也是一个完整的句子，但是它们在整句中是以一个语义块的身份出现的。因此，我们把它叫做句蛻。句蛻是语义块融合性的表现。

句蛻现象并不是汉语所特有的，上面英语的从句就是一种句蛻，英语中还有其他类型的句蛻，如：

Dealing with the attack between these two country is difficult. 只是英语中的句蛻一般都有形态变化（比如这里用动名词），或者附加关系代词、连词等变化来实现。

由于汉语句蛻没有明确的标志信息和形态变化，所以对句蛻的确认存在一系列困难：

1. 句蛻条件的判断；
2. 句蛻类型的确定
3. 句蛻范围的确定；
4. 句蛻与句子整体的语义关系的确定；
5. 句蛻内部的语义关系的确定；等。

本文将重点讨论汉语句蛻的表现，以及 HNC 句类分析系统对句蛻处理的策略与实现。

## 2. 句蛻的表现

句蛻现象表现各异，为了讨论方便，HNC 把句蛻分成三种类型：原型句蛻、要素句蛻、包装句蛻。

原型句蛻就是一个句子不作任何变换，直接蛻化成一个语义块。如：“加拿大采取上述决定会导致它驻在当地的军队遭到塞族武装的报复。”

要素句蛻是以句子中一部分（一个或多个语义块）为中心成分，其它部分为说明成分而发生的句蛻。一般情况下，作为中心成分的部分都移到句子末尾，并加“的”予以指示。如：“参与这个计划的俄罗斯企业也遇到了一些严重的问题”，其中，“俄罗斯企业”为中心成分。如果把这个句蛻恢复成句子，就是“俄罗斯企业参与这个计划”。

包装句蛻是在以上两种句蛻的基础上，前后加上一些包装概念，一起构成一个语义块。包装句蛻的一般形式是：前包装+（要素句蛻或原型句蛻）+后包装。对包装句蛻形成的语义块而言，一般情况下，后包装概念是这个语义块的要素，而前面的句蛻是这个要素的说明成分。如：“香港接受了一项开创历史先河的光荣任务。”其中，“一项”和“光荣任务”分别是前包装和后包装。“开创历史先河”是一个原型句蛻。

以上三种类型是句蛻的基本类型，它们的再分类或多次组合，造成了句蛻现象的复杂性。由于原型句蛻就是一个句子充当语义块，所以对它的处理就是对句子处理的再循环，因此，下面重点介绍要素句蛻和包装句蛻。

### 2.1 要素句蛻

理论上说，要素句蛻“的”后的部分可以是多个语义块，但是目前我们只考虑一个语义块的情况。虽然，语句中任何一个语义块都可以做要素句蛻中“的”后的部分。但是一般情况下，要素句蛻加“的”不能影响语句的格式。所以下面以句类格式为指导讨论要素

句蜕。

### 1. 非省略格式

非省略格式格式（包括基本格式、规范格式）的语句中，产生要素句蜕可以有两种手段：

1) 最后一个语义块（包括特征语义块和广义对象语义块），只要它不带语义块切分标志，都可以在其前加“的”形成要素句蜕。

2) 第一个语义块如果是广义对象语义块，则可以把这个块移到末尾，前面加“的”指示。

三主块句的句蜕可能形式有（如果加“的”后，出现重复，则只取其一）：

1) !0J=JK1+E+ 的+ JK2 （无语义块移动）

!11J=JK1+^JK2+的+ E      !12J=JK2+^JK1+的+E

!20J=JK2+E+的+JK1      !21J=JK1+JK2+的+E

!22J=JK2+JK1+的+E

2) !0J= E+JK2 +的+ JK1 （有语义块移动）

!11J= ^JK2+E+的+ JK1      !12J=^JK1+E+的+ JK2

如：美国发表的人权报告指责巴西限制公民权利。（JK1+E+的+JK2）

体现了中共中央对本期理论研讨班的高度重视。（JK1+^JK2+的+E）

参与这个计划的俄罗斯企业也遇到了一些严重的问题。（E+JK2+的+JK1）

### 2. 省略格式

省略格式只可能省略广义对象语义块 JK（E 块省略，只能是几个特殊的句类，其他句类不允许省略 E），而且是不带语义块切分标志的。省略格式下的句蜕一般不考虑语义块的移动。

因此可以演绎得出，三主块句省略格式下的句蜕可能形式有：

1) !0J= E+的+JK2 （省略第一个 JK）

!01J= E+的+JK1      !11J= ^JK2+的+E

!12J= ^JK1+的+E

2) !0J=JK1+的+E （省略第二个 JK）

!22J=JK2+的+E

如：发达国家利用技术和资金的控制力实现对全球利益的瓜分。（^JK2+的+E）

这个问题的解释已经很清楚了。（JK2+的+E）

以上只对三主块句的要素句蜕作了演绎推理，四主块句可以根据句类格式作同样的演绎。这里可以看出，汉语的要素句蜕只需要加“的”就可实现，而不用像英语一样要使用动名词、不定时、或者从句来实现，可见，汉语是多么的简明啊。

### 2. 2 包装句蜕

从上面对包装句蜕的定义可知，包装句蜕有两种：原型句蜕的包装、要素句蜕的包装。

对原型句蜕的包装，汉语一般都用“的”明确指示出包装概念。对要素句蜕的包装，理论上，所有的要素句蜕都可以包装，但是汉语中一般常见的是特征语义块要素句蜕的包装。特征语义块要素句蜕的包装可以不加“的”。

如：经济学家降低了对明年经济指数的的预测数据。

遭受战争灾难的南人民的生活状况引起国际社会的关注。

### 3. 句蜕的处理策略

句蜕处理的对象是绝对排除序列、相对排除的  $E\bar{L}$  序列中的  $E\bar{L}$ ，还有被多动词处理判断为  $Eg-E\bar{L}$  关系的  $E\bar{L}$  [晋耀红 1, 2001]。

句蜕处理的任务是确定句蜕的范围，给出  $E\bar{L}$  以及与他相关成分的语义角色。

下面对句蜕处理前提条件、处理的基本思路做详细说明。

#### 3. 1 句蜕的条件

因为句蜕是语义块复杂构成的表现之一，而 HNC 语义块基元有四个：E、A、B、C，所以下面的论述将以四个语义块基元为单位而展开。又由于语义块是句类的函数，而句类是广义作用句、广义效应句和两可句类为集团的，所以以下的说明将以句类集合为主线。

##### 1、广义作用句

句类表示式中，只含 A 或 B 的语义块，如：T2bJ=TA+T2b+TB2，其中的 TA，TB2 都是只含 A，B 的语义块，这些语义块一般都是以具体概念为中心的，所以他们一般只能允许广义对象语义块的要素句蜕，而且这个要素句蜕的中心是具体概念。

对显含内容 C 的语义块，除了个别（如物转移句的 T2C）不能句蜕以外，其它都可以句蜕，而且是三种句蜕都可以。这主要是由于 C 语义块的融合性造成的。

##### 2、广义效应句

素描句 P11J、换位状态句 S02J，S03J 的 B、C 语义块都有自己明确的定义，不允许句蜕。

其他广义效应句句类的 B、C 语义块一般都可以句蜕。

3、两可句类介乎前面两种情况之间，需要根据具体情况而定。

以上这些知识是 HNC 概念层面的句类知识，他们是句蜕处理的基本原则，如果这些条件满足，就必须进入句蜕处理。

#### 3. 2 句蜕处理

HNC 句类分析处理是以语义块为单位的，因此句蜕处理面临的问题就是从一些可能的语义块序列中选择出最优，句蜕处理就是消解这些语义块序列的可能歧义。这些歧义的消解是建立在句类知识的基础上的，而句类演绎规则是对句类知识的形式化，因此句蜕处理主要是根据句类演绎规则而展开的。

由于汉语“的”在句蜕中的特殊作用，我们对它也给予特殊关注。根据句蜕类型，结合“的”的位置，我们对句蜕的处理分以下几部分。

##### 3. 2. 1 特征语义块要素句蜕

非省略格式下的特征语义块要素句蜕，一般不发生语义块搬移，只是在句尾的 E 块前加“的”予以指示。因此对它的处理，只要去掉“的”，进行语句处理“三部曲”即可。

省略格式下的特征语义块要素句蜕有以下四种情况：

!11J=^JK2+的+E      !12J=^JK1+的+E

!0J=JK1+的+E      !22J=JK2+的+E

对前两种情况，由于是规范格式！31 的省略，而！31 省略在“三部曲”中处理得比较好，因此，这里同样进行“三部曲”的小循环。

违例格式下的省略 JK2+的+E, 是句蜕中比较常见的, 他容易和 JK1+的+E 混淆, 对它们的判断规则如下:

首先形式化为: K+的+E

句类演绎规则:

E-K 满足 E-JK2 预期

E-K 不满足 E-JK1 同行

----->JK2+的+E

E-K 不满足 E-JK2 预期

E-K 满足 E-JK1 同行

----->JK1+的+E

E-K 满足 E-JK2 预期

E-K 满足 E-JK1 同行

----->优先 JK2+的+E

需要说明的是, 对 JK1+的+E, 如果这个 E 有 g 或 r 的非 v 类概念, 也可以判断它为非 v, 它和 K 构成偏正关系。这是一个两可判断, 在处理中, 依据具体情况而定。

如: 这个问题的解决将推动整个事件的发展。 (JK2+的+E)

他的表演博得了阵阵掌声。 (JK1+的+E)

### 3. 2. 2 广义对象语义块的要素句蜕

对没有发生语义块搬移的句蜕, 可以去掉“的”按照语句处理。如:

JK1+E+ 的+ JK2

E+的+JK2 (省略第一个 JK1)

美国发表的人权报告指责巴西限制公民权利。 (JK1+E+ 的+ JK2)

对发生搬移的句蜕, 可以把“的”后的语义块搬到句首, 然后按照语句“三部曲”处理。如: !11J= ^JK2+E+的+ JK1

!12J=^JK1+E+的+ JK2

对伊拉克进行空袭的美国空军已经安全返回。 (^JK2+E+的+ JK1)

如果句蜕都象以上这样经过简单的变化, 就可以和语句一样处理, 那句蜕处理就不存在任何难点了, 麻烦在于在一种结构下, 可能有多个句蜕判断, 比如: E+K1+的+K2, 它是要素句蜕: E+JK2+的+JK1, 还是省略格式的原型句蜕 E+ (K1+的+K2)? 结构的歧义选一处理才是句蜕的难点所在。

下面对句蜕处理中常见的歧义结构作详细分析。

### 3. 2. 3 常见句蜕结构的歧义判断

#### 3. 2. 3. 1 E1+K1+的+K2

参与这个计划的俄罗斯企业也遇到了一些严重的问题。

国家为维持产品的低价格而压低生产资料的价格。

大家对制定这个计划的背景都很清楚。

对这个结构, 黄曾阳先生在《句类分析 20 项难点》中已经作了详细的分析[黄 1999, 第一章 1.2 节], 黄先生的分析是把 Eg、E<sub>l</sub>放到一起考虑的, 我们这里假定已经判断这个 E 为 E<sub>l</sub>[普耀红 1, 2001], 而对它内部句蜕的结构进行判断。

对这个句蜕结构的演绎规则如下:

$E\bar{L}K1$  满足  $E\text{-}JK2$  预期

$E\bar{L}K2$  满足  $E\text{-}JK1$  同行

----- $\rightarrow E\bar{L}JK2+$ 的 $+JK1$  ( $JK1$  要素句蜕)

$E\bar{L}K1$  不满足  $E\text{-}JK2$  预期

$E\bar{L}K2$  满足  $E\text{-}JK2$  预期

----- $\rightarrow E\bar{L}[K1+$ 的 $+K2]$  (省略格式的原型句蜕,  $K1+$ 的 $+K2$  一起充当  $JK2$ )

$E\bar{L}K1$  满足  $E\text{-}JK2$  预期

$E\bar{L}K2$  不满足  $E\text{-}JK2$  预期

$E\bar{L}K2$  不满足  $E\text{-}JK1$  同行

----- $\times E\bar{L}K1+$ 的 $+K2$  (包装句蜕)

### 3. 2. 3. 2.....+的 $+E1 +K1$

经济学家降低了对明年经济指数的预测数据。

对“的”后的  $E\bar{L}$  它天然就是句蜕的下界(我们暂时不考虑两个或两个以上的语义块做句蜕的中心,这种情况虽然理论上会出现,但是实际情况下很少。),那么这里的句蜕只有一种选择,就是作特征语义块要素句蜕,也就是说,这里的  $K1$  只能考虑它是后包装概念,如果不是,那么它就否定前面的  $E1$ ,这里不做句蜕处理。如果是,则把  $E1$  和其前面的部分按照 3. 2. 1 节的策略处理。

### 3. 2. 3. 3 小结

以上对常见的句蜕结构进行了歧义消解,消解采用的是句类演绎规则。句类规则消解策略是开放的,具有良好的可扩充性和可移植性。因为,在句类分析系统中,句类演绎规则是以规则文本的形式存在的,是完全开放的,如果我们发现汉语中新的句蜕结构,或者处理英语的句蜕,只需要更新句类演绎规则文本即可。

## 4. 结束语

汉语的句蜕现象无处不在,但是从句类的角度来看,它的出现又有一定的可预期性,本文根据这种预期性,给出了汉语句蜕处理的解决策略。

句蜕处理依据的是句类演绎规则,而演绎规则的基本操作是句类检验,主要是语义块之间的预期和同行检验,因此句蜕处理的效果取决于预期同行检验的水平,我们下一步的工作将集中于提高检验水平,提高分析的“自知之明”的能力,进而不断提高句蜕处理的水平。

## 参考文献

- [1]黄曾阳,1997 “HNC理论概要”, 中文信息学报, Vol.11, No.4,
- [2]黄曾阳,1998 《HNC(概念层次网络)理论》 清华大学出版社
- [3]黄曾阳, 1999, “HNC理论与自然语言语句的理解”, 中国基础科学, 第一卷, 第一期
- [4]晋耀红 1, 2001, “汉语的多动词难点处理”, HNC与语言学研究研讨会
- [5]晋耀红, 1998, “基于HNC理论的句类分析系统的设计与实现” 中国科学院声学所硕士学位论文
- [6]张全, 1996, “基于HNC理论的语义块感知处理” 中国科学院声学所博士学位论文
- [7]苗传江, 1998, “HNC理论的句类”, “98中文信息处理国际会议”论文集