

现代汉语复杂句蜕块研究¹²

唐兴全

北京语言大学人文学院 100083

E-mail: tang_xq@263.net

摘要: 语义块是 HNC 理论语句层面表述模式的重要内容, 是句子的语义构成单位。语义块的整体或局部可以由句子蜕化而来, 称为句蜕, 它是语义块的特殊构成。句蜕之间或句蜕与其他成分之间的嵌套或组合以及句蜕内部语义块的复杂变化等形成复杂句蜕, 包含复杂句蜕的语义块称为复杂句蜕块。本文研究了复杂句蜕块的各种类型及其分布, 并探讨了它们的处理策略。

关键词: 复杂句蜕 句蜕 语义块 HNC

Studies on Complex Sentence Decomposition in Modern Chinese

Tang Xingquan

Beijing Language & Culture University, Beijing 100083

E-mail: tang_xq@263.net

ABSTRACT: Semantic chunk is the semantic component of sentences. Sentence decomposition (SD) which means a sentence degenerated to be a semantic chunk or a part of a semantic chunk, is a special type of composition of semantic chunk. Nesting or combination between SD forms complex sentence decomposition (CSD). We Studied on CSD and discussed the processing strategies of each type of CSD in this thesis.

Keyword: CSD, sentence decomposition, semantic chunk, HNC

1 引言

语义块是 HNC 理论为建立语句表述模式而提出的重要概念, 它是句子的语义构成单位, 从对句子语义构成的贡献大小的角度, 它可以分为主语义块和辅语义块两种。主语义块又分为特征语义块、广义对象语义块, 分别用 E、JK 来表示。特征语义块大致相当于句子中的谓语部分。如: “张三 || 打了 || 李四”, “||” 为语义块边界标记, 这个句子有三个语义块, “张三”、“打了” 和 “李四”。语义块的构成可能是相当复杂的, 复杂的表现之一是语义块中含有句蜕。句蜕是指语义块的整体或局部由句子蜕化而来, 它是语义块的特殊构成形式。如 (本文下面的例句中, 句蜕语义块用下划线表示):

美国发表的人权报告指责巴西限制公民权利。

¹ 本文是在笔者硕士学位论文的基础上形成的。

² 本文得到国家重点基础研究发展规划 (973) 项目 G1998030506 资助。

句中的划线部分是全句中的第一个语义块，用 JK1 表示，它是由“美国发表人权报告”这句话蜕化来的。句蜕之间或句蜕与其他成分之间的嵌套或组合形成复杂句蜕，包含复杂句蜕的语义块称为复杂句蜕块。如：

位于伦敦西南部基尤的公共档案局昨天公布的文件引起了极大反响。

划线部分是两个句蜕的嵌套，具体说来，“位于伦敦西南部基尤的公共档案局昨天公布的文件”是一个句蜕，其中又嵌套一个句蜕“位于伦敦西南部基尤的公共档案局”。复杂句蜕块是本文的研究对象。

复杂句蜕块研究将有助于 HNC 句类分析技术的完善，提高句类分析技术对句蜕的处理水平，并且将丰富对于现代汉语句子和短语的认识和研究。

2 复杂句蜕块的定义及范围

句蜕块的复杂程度有所不同，大致可以分为以下四种类型，各类型的复杂程度依次递增：

- 0 语义块=基本句蜕
- 1 语义块=包装句蜕
- 2 语义块=变异句蜕
- 3 语义块=句蜕块嵌套

0 型句蜕块指的是整个语义块由一个简单的基本句蜕构成，基本句蜕包括原型句蜕和要素句蜕两种形式，原型句蜕是指一个句子不做任何变换，直接蜕化为一个语义块。如“中国实行增值税制度有利于与国际市场接轨”，本句的 JK1 就是原型句蜕。要素句蜕是以句子中的一部分（一个或多个语义块）为中心语，其他部分为修饰语的句蜕，它分为 E 要素句蜕和 JK 要素句蜕。如“位于里斯本市拉帕区的帕德雷·巴托洛梅乌小学有着悠久的历史”，句中划线部分是一个 JK 要素句蜕。基本句蜕在处理时只进行简单的还原即可，比较容易分析。1 到 3 型句蜕块则较为复杂，其中涉及的定义我们下面将有解释。我们把包含包装句蜕、变异句蜕或句蜕块嵌套等句蜕复杂形式的语义块称为复杂句蜕块，我们所说的句蜕的复杂形式是指句蜕之间或句蜕与其他成分之间的嵌套或组合，以及句蜕内部语义块的复杂变化。我们的分析是围绕广义对象语义块展开的，包括上述四种类型的后三种。

3 复杂句蜕块分析

在观察、分析大量语料的基础上，我们对复杂句蜕块的三种主要类型，即语义块=包装句蜕、语义块=变异句蜕块和语义块=句蜕块嵌套，逐一进行了分析研究。

3.1 语义块=包装句蜕

3.1.1 包装句蜕及其分类

基本句蜕的前面或后面如果加上别的成分，就形成包装句蜕。所加的成分称为包装品，有前包装品和后包装品之分，它们可以是词或短语（我们只探讨后包装品，并用\…/标记），包装品包装的句蜕称为包装体。如“我∥看到了∥沙漠变成良田的美景/”，JK2 是包装句

蛻，其中“沙漠变成良田”是包装体，“美景”是包装品。依据包装体的不同，包装句蛻可分为原型包装句蛻和要素包装句蛻，对于后者，汉语中常见的是E要素包装句蛻，如：

经济学家 || 公布了 || 对经济 || 的预测\数据 /。

量子电脑 || 的研制\工作 / || 现在还处于 || 十分原始的阶段。

对原型包装句蛻，汉语一般都用“的”明确指示出包装品。如：

反对派 || 在众议院 || 提出 || 弹劾 || 总统的要求 /。

全球化 || 就是 || 美国 || 搞 || 经济帝国主义的\宣言 /。

但当包装品前有指示代词时，也可以不用“的”来作标记。如：

这些人 || 从未接受过 || 埃斯特拉达 || 担任 || 总统\这一现实 /。

JK要素句蛻包装的包装品前一般都有“的”。

3.1.2 包装品研究

要素包装句蛻的包装品范围广泛，但有一种特殊的类型值得注意，即一些单字充当包装品的情况，如“感、度、量”等，语言学上一般认为这些词缀型的汉字与它前面的词组成新词，因此在分词时，把它们作为一个词切分出来，如“责任感”、“客流量”等。我们从语义理解的角度考虑，认为这些单字往往是对前面所陈述情况的一种概括，所以认为它们也是包装品。例如：

这种材料 || 大大影响了 || 芯片 || 的集成\度 /。

来咨询 || 的客户\量 / || 很大。

这些措施 || 大大增强了 || 干部群众 || 对党和政府 || 的信任\感 /。

原型包装句蛻是定中结构的一种复杂类型，朱德熙、马真（1981）及古川裕（1989）等语言学家曾对此进行过研究。原型包装句蛻里的包装品和此原型句蛻中的Ek不能直接搭配，换句话说，两者之间没有潜在的格或及物性语义关系。原型包装句蛻的包装品一般有较严格的限制，只有具备某种特定的语义特征的名词才有资格充任包装品的语义核心，而不是任何名词都能随便自由地充当这一核心。表达具体概念的词语很少充任原型句蛻的包装品，朱德熙先生曾在《语法讲义》里，从语法结构的角度作过阐明：“值得注意的是这类（即本文所说的原型包装句蛻）偏正结构的中心语总是表示抽象概念的名词。”这种说法是有一定概括力的，原型包装句蛻其包装品的核心名词绝大部分都是表达抽象概念的，它们大体可以分为7个小类：表达时空概念的，如“刹那”“场所”；表达思考内容的，如“决心”“主张”；表达信息转移的，如“消息”“报道”；表达属性的，如“毛病”“作风”；表达感知内容的，如“感觉”“声音”；表达事的，如“事实”“情景”，以及其他一些情况。但我们认为只说“表达抽象概念的名词”还不够全面，因为根据我们考察，还有个别表示具体事物的名词也可以作包装品的中心，例如：

森喜朗 || 向奥运会马拉松冠军高桥尚子 || 颁发 || “国民荣誉奖”的\照片 / || 被登在《读卖新闻》上。

这些瓷器上 || 都装饰有 || 兰色和白色 || 相间的中式图案 /。

3.2 语义块=变异句蛻

变异句蛻是指伴随语义块分离、句类转换或语义块并合现象的基本句蛻和复杂句蛻。

3.2.1 伴随语义块分离的变异句蜕块

语义块的分离就是同一语义块的不同构成部分之间插入了一个或多个语义块，或语义块的一部分。如“我||被一阵吵闹声||打断了||思路”，被“打断”的应该是“我的思路”，整体作为一个语义块，本句中它们分离了。判定是否分离的依据是 HNC 句类表示式，因为在 HNC 句类表示式中，明确规定了不同句类中语义块的位置和数量。

语义块分离现象可能出现句蜕的各种形式中。原型句蜕中的语义块分离和句子中的语义块分离形式是一样的，如（我们用[...]来表示分离出去的部分）：

大家||都知道||他||被领导||撤消了||[科长职务]。

我||很感激||他||为[我]||安置||工作。

在要素句蜕中，语义块分离主要是要素句蜕中心成分的分离，尤其是量词短语或指代词语与中心成分的分离。如：

公安部门||查封了||[那些]出售||盗版光盘||的书店。

[每一个]以官方身份||访问||北朝鲜||的人||都必须为金日成雕像||献花圈。

其中，“那些”和“书店”分离，“每一个”和“人”分离。

包装句蜕中的分离有两种情况，一是包装体内部的语义块发生了分离；二是包装品发生了分离。下面两个例句分别代表这两种情况：

[那些]反对||这项提议||的议员的如意算盘||落空了。

德国党卫军的一名杀手||于 1939 年||组织了||[一次]袭击||德国边境事件。

其中，“那些”和“议员”分离，“一次”和“事件”分离。值得注意的是要素包装句蜕中的分离，应区分是要素句蜕内部语义块的分离还是包装品的分离。

3.2.2 伴随句类转换的变异句蜕块

句类转换的概念是句类概念的必然产物，它是指一句类的内容换用另一句类来表达。句类转换大致有五种基本类型，下面我们仅以向存在判断句的转换为例进行说明。如：

他早就发现了这种不好的苗头。

他早就对这种不好的苗头有所发现。

句类转换形成的存在判断句有下面三种主要形式（其中的 JK1、JK2、E 分别是转换前句类的 JK1、JK2 和 E；例句中‘=>’后为转换前的形式）：

A: <E+JK2+的>+有+JK1

<E+JK2+的+JK1-1>+有+JK1-2 (JK1-1、JK1-2 是 JK1 的两个部分)

如: <出席||茶话会||的>||有||党和国家领导人。=>党和国家领导人||出席||茶话会。

<参加||聚会||的朋友>||有||王刚和李铁。=>朋友王刚和李铁||参加了||聚会。

B: JK1-1+有+JK1-2+E+JK2

如: 这个村||有||几十个学生考上了大学。=>这个村的几十个学生||考上了||大学。

C: JK1+有+E

如: 他的学习||有了||很大进步。=>他的学习||进步了。

伴随句类转换的变异句蜕块是指还原后的句子包含句类转换的句蜕块。如：

\{历史||是||人民||创造的\}这一观点/||符合||马克思主义的唯物史观。

大家||都知道||\{他||是||以种种谎言||博得||姑娘的芳心的\}。

3.2.3 伴随语义块并合的变异句蛻块

语义块并合是指两个或多个要素句蛻共用一个广义对象语义块的现象。如：

这些孩子 || 都是 || 父母 || 疼爱、爷爷奶奶 || 娇惯 || 的小皇帝。

上句中 JK2 包含两个句蛻，“父母疼爱的小皇帝”和“爷爷奶奶娇惯的小皇帝”，两者共用 JK2 “小皇帝”。两个句蛻还可以并合不同位置的语义块，比如：

请大家 || 选出 || 你 || 最喜爱的、来自 || 内地 || 的歌手。

“歌手”分别充当了前一句蛻的 JK2 和后一句蛻的 JK1。伴随语义块并合的变异句蛻块中也可能出现语义块分离和句类转换现象。

3.3 语义块=句蛻块嵌套

句蛻块嵌套是指在句蛻语义块中又出现句蛻的现象。包含句蛻块嵌套的语义块层次多而复杂，无疑是语句分析的难点。句蛻块嵌套构成的语义块主要有两种基本类型：

一，包装句蛻的包装体是两个或多个句蛻。它们一般为同类型句蛻，如：

每一个以官方身份访问北朝鲜的人 || 都必须为受 || 人 || 爱戴 || 的金日成和
让 || 人 || 好奇 || 的金日成的宝藏 || 增添 || 新的东西。

JK2 是一个包装句蛻，包装体是“受人爱戴的金日成”和“让人好奇的金日成”两个句蛻。

二，句蛻语义块中的某个语义块又包含句蛻，这是句蛻块嵌套的最常见类型。如：

穷人 || 送给 || 即将入学 || 的孩子 || 的喇叭筒 || 都是 || 很便宜的。

JK1 是一个要素句蛻“穷人送给即将入学的孩子的喇叭筒”，此要素句蛻中的 JK2 又是一个要素句蛻“即将入学的孩子”。句蛻块嵌套现象按照句蛻的类型还可以细分为若干子类。

4 复杂句蛻块分布

句蛻块和复杂句蛻块的分布比例是我们得到的主要数据。这一统计结果有助于获得对句蛻块和复杂句蛻块现象更明确的认识，并为句类分析系统制定适当的句蛻处理策略提供依据。我们的统计所依据的语料是从人民日报、参考消息、光明日报等报纸杂志中抽取出来的真实文本，约 125800 字，共有 6967 个语串（语串的定义是：书面语中由断句符号分隔出来的语言片断，断句符号包括句号、问号、叹号、逗号、分号和冒号），主要文体涉及新闻、散文等。其中部分语料是译文。

我们的语料中，由于对句蛻进行了明确的形式标记，因此可以根据这些标记统计句蛻的有关信息。我们统计了含句蛻语串和含句蛻语义块在所选语料中的出现情况，结果见表 1：

表 1 含句蛻语串和含句蛻语义块出现情况统计

| | 语串 | 语义块 |
|-----------------|-------|-------|
| 总数（个） | 6967 | 15273 |
| 含句蛻数量（个） | 2207 | 4336 |
| 含句蛻语串或语义块百分比（%） | 31.68 | 28.39 |

要素句蛻和原型句蛻是句蛻的两种基本类型，包装句蛻是在要素句蛻和原型句蛻基础上形成的复杂句蛻。经过统计，我们得到了含有这三种形式句蛻的语义块的字数占文本总

字数的比例，见表 2:

表 2 三种句蜕语义块的字数占文本总字数的比例

| | 统计文本 | 句蜕块 | 要素句蜕块 | 原型句蜕块 | 包装句蜕块 |
|---------|--------|-------|-------|-------|-------|
| 字数(个) | 125800 | 31496 | 12975 | 8673 | 9848 |
| 所占比例(%) | 100 | 25.04 | 10.31 | 6.89 | 7.84 |

对句蜕块及复杂句蜕块分布情况的统计有助于为句类分析系统制定适当的句蜕处理策略提供依据，经统计，我们得到的结果如下:。

表 3 句蜕块及复杂句蜕块分布情况统计³

| 类型 项目 | 基本句蜕 | | 复杂句蜕块 | | | | 含句蜕语义块 |
|----------|-------|-------|--------|--------|-------------|-------|--------|
| | 要素句蜕 | 原型句蜕 | 包装句蜕 | | 伴随语义块分离的句蜕块 | 句蜕块嵌套 | |
| | | | 要素包装句蜕 | 原型包装句蜕 | | | |
| 数量(个) | 1478 | 1276 | 146 | 887 | 189 | 360 | 4336 |
| 所占比例(%) | 34.09 | 29.43 | 3.37 | 20.46 | 4.35 | 8.30 | 100 |

5 复杂句蜕块处理策略讨论

对复杂句蜕块需依据 HNC 句类知识，利用 HNC 句类分析技术进行分析处理。

HNC 对句子的分析以语义块为单位，按照句类分析的三步曲进行，即语义块感知和句类假设、句类检验、语义块构成分析，“句类”是整个分析系统的灵魂。句蜕处理的任务是确定句蜕的范围，给出句蜕中的 E 块以及与其相关成分的语义角色。因此我们把句蜕语义块同样作为句子来分析和理解，进行句类分析“三部曲”的小循环，而不是仅仅看作定中之类的短语结构。句蜕处理面临的问题首先是句蜕结构的歧义判断，从一些可能的语义块序列中选择出最优，而歧义的消解是建立在 HNC 句类知识的基础上的。我们可以根据句类知识演绎出句蜕处理的规则，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。下面我们以结构“K1+E+K2+的+K3”为例进行说明（其中 K 为名词短语）:

K1+E+K2+的+K3

E-K1 满足 E-JK1 预期

E-K2+的+K3 满足 E-JK2 预期

-----→原型句蜕，即 (K1+E+ (K2+的+K3))

E-K1 满足 E-JK1 预期

E-K2 满足 E-JK2 预期

³说明：由于在语料标注中，没有为伴随句类转换的句蜕块和伴随语义块并合的句蜕块定义明确的形式标记，因此我们对这两类情况占有句蜕块的比例未加统计，但这并不对句蜕块总数的统计造成影响。另外，句蜕块三层以上嵌套由于数量极少，我们也未考虑在内。因此，复杂句蜕块的实际存在数量应该比我们表中所得到的数字要稍多一些。

E-K3 不满足 E-JK3 预期

-----→原型包装句蜕, 即 ((K1+的+K2) +的 K3)

E-K1 满足 E-JK 预期

E-K2 和 E-K3 不满足 E-JK 预期

-----→双层包装句蜕, 即 ((K1+E) +K2+的+K3)

对复杂句蜕块的处理同样要从基本的句蜕结构的歧义判断开始, 比如对包装句蜕的处理首先应是正确判断出包装句蜕, 然后把包装体还原成句子的形式以得到包装体内部的概念联想脉络; 对伴随语义块分离的变异句蜕块的处理关键是语义块分离部分的正确合并; 伴随句类转换的变异句蜕块的处理关键是对句类转换进行反转换处理, 把句子改造成转换前的句子; 伴随块并合的变异句蜕块的处理主要是要把被并合的语义块还原到句蜕的正确位置; 句蜕块嵌套形成的复杂句蜕块的处理需要进行的是句蜕处理的小循环, 即从句蜕块嵌套的最深一层开始作还原处理, 逐级还原, 直至整个复杂句蜕块的还原。

6 总结

本文着重对复杂句蜕块的各种不同类型进行了分析, 对各种类型在真实文本中的分布进行了统计, 并讨论了复杂句蜕块的处理策略。对复杂句蜕块各种类型的更详细归纳与更深入探索, 以及对复杂句蜕块处理策略的深入研究, 将是本文后续工作的主要内容。

参考文献

- [1] 黄曾阳:《HNC·(概念层次网络)理论》,清华大学出版社,1998年。
- [2] 黄曾阳:《句类分析的20项难点》,中科院声学所内部资料,2000年。
- [3] 古川裕:“‘的’字结构及其所能修饰的名词”,《语言教学与研究》,1989年第一期。
- [4] 晋耀红:“汉语句蜕处理”,《自然语言理解与机器翻译》,清华大学出版社,2001年8月。
- [5] 晋耀红:“基于HNC理论的句类分析系统的设计与实现”,中科院声学所硕士学位论文,1998年。
- [6] 苗传江:“HNC句类知识研究”,中科院声学所博士学位论文,2001年。
- [7] 马真:《简明实用汉语语法》,北京大学出版社,1981年第一版。
- [8] 李千驹:“基于HNC理论的汉语定中结构研究”,广西师范大学硕士学位论文,2001年。
- [9] 陆俭明:“‘的’字结构和‘所’字结构”,《语法研究和探索(2)》,北京大学出版社,1983年。
- [10] 张普:“信息处理用现代汉语语义分析的理论和方法”,《中文信息学报》,1991年第三期。
- [11] 薛侃:“现代汉语的块扩及句蜕研究”,中国人民大学硕士学位论文,1999年。