

HNC 理论的辅语义块以及语义块标记符

张艳红

中国科学院声学所 声场声信息国家重点实验室 II 部 北京

摘要: 本论文基于 HNC 理论, 做了以下几方面的工作: 阐述辅语义块的类型和特征, 并给出辅语义块的四种形式; 利用北京语言学院“现代汉语语料库”中抽取的 1069 个“v+着+v”式的句子分析了以实词为首标记符的辅语义块, 并从工程的角度, 分析了把这种辅语义块作为 QE 的必要性; 针对俞士汶《现代汉语语法信息词典详解》中的“介词表”, 在对 8872 个标注了介词结构的句子进行人工句类分析的基础上, 得到 32 个主语义块首标记符、56 个辅语义块首标记符、9 个两可语义块首标记符。

关键词: 辅语义块 两可语义块 标记符

HNC theory's secondary semantic-chunk and tags of semantic-chunk

Zhang Yanhong

State key lab section II Acoustic Institute of Chinese academy of Sciences, Beijing

ABSTRACT: This paper does some research based on HNC theory: it illustrates the type and character of the secondary Semantic-Chunk (sSC), and tells about four kinds of structure of sSC, and analyzes mainly sSC with notional word tags on the basis of the Sentence-category analysis to 1069 “v+着+v” sentences from BLCU's “Modern Chinese corpus” and points out the necessity to take this SC as QE from the angle of practice, and enumerates tags of SC.

Keywords SECONDARY SEMANTIC-CHUNK DUAL SEMANTIC-CHUNK TAG

1 HNC 理论的语义块思想

HNC 理论认为语义块是语句中相对完整的基本语义单位, 是构成语句的下一级结构单位。语义块从类型上分为主语义块和辅语义块两大类。语义块的主辅之分与句类思想密切相关。主语义块是句类的函数, 它在一个句子中的有无、个数和具体内涵随句类的不同而不同。辅语义块则具有不确定性, 弱依赖于句类。句类判别是判定该句子的核心谓语反映了作用效应链中的哪一或哪些环节或哪几个环节的链接, 句类分析是要分析出构成句子的一个个语义块。

2 关于辅语义块

HNC 理论中, 辅语义块的类型有: 方式 Ms (Means)、工具 In (Instrument)、途径 Wy (Way)、条件 Cn (Condition)、参照 Re (Refer)、因 Pr (Premise)、果 Rt (Result)。工具包含有生产工具、生活工具和信息工具, 材料、原料、和动力也属于工具的范畴。材料又包含生产材料、生活材料和信息材料; 原料有天然和人工之别; 动力也有天然和人工之别。条件包含时间、环境、社会条件、一般物质条件和前提条件。辅语义块的类型之间有交叉, 如方式和途径之间存在两可的情况, 原因和结果之间也存在两可的情况, 即因果不分。结果也包含着目的。另外, 目的和参照引导的语义块有主辅不分的情况, 我们称之为两可语义块。因果不分和主辅不分都居于辅语义块的分类之下。

辅语义块表示式为 fK , 它弱依赖于句类。一个句子有几个辅语义块是由语用知识与语境决定的, 句类知识无法对它作出确定性的描述。但是, 辅语义块相当于传统句法中的状语。而对于状语, 我们又要排除掉在 HNC 理论中被分析为 QE (E 语义块的附属部分) 的那部分, 如“已经、将来、正在、应该、再、又”等。辅语义块大部分是带有标记符的, 我们称之为辅语义块标记符, 它存在于 HNC 理论的三大语义网络之一——语言逻辑概念网络 I 中, 属结点 I1 行。按照 HNC 理论, 辅语义块的最大特点是, 它的出现带有 I 指示或 I 说明, I 指示指狭义的 I 类概念, 如: “用、按照”等标记符, I 说明包括综合类概念, 如“方式、途径、器材”等 (参见 HNC 理论的综合类概念结点表, 综合类概念用 s 表示)。辅语义块一般位于 E 语义块前面。我们可以这样说, 对于每一个句子, 明确句类后, 找出它所需的主语义块, 剩下的都是辅语义块。

辅语义块有四种形式: 第一, 以实词为首标记符的辅语义块。如, “有些西方人带着有色眼镜看中国”、“他乘车上学”、“他们打着红旗反红旗”这一类在传统句法中被称为“连动句”的句子, 便是这种情况。按照 HNC 理论, 第一个动词, 也即第一个 E 语义块, 不为主语义块, 而是辅语义块。决定句类的 E 语义块由第二个动词来充当。第一个动词及其宾语或动词短语“带着有色眼镜”“乘车”“打着红旗”等分别表示方式、工具、途径。当然, 传统语法所谓的“连动句”在 HNC 理论中并不都是这种情况。如“我去教室拿书”, “去教室”就不是辅语义块。第二, 以介词为首标记符的辅语义块。它包含两种情况: 一种是只有块首标记符, 如“按照”(表示参照); 一种是前面有块首标记符, 后面有块尾标记符, 即搭配类型的, 如“拿……来说”(表示参照), 但是, 对于这种搭配型的标记, 又要注意真尾和尾标的区别, 比如: “以……为……”的“为”是尾标, “……来说”是真尾。第三, 只有块尾标记符的辅语义块, 如“……以后”等。第四, 没有标记符的辅语义块, 主要是表示时间的条件辅语义块。

3 以实词为首标记符的辅语义块

HNC 理论的复合句类是指一个句子存在两个甚至多个 E 块，而且它们分别含有作用效应链不同环节的信息。它的基本特征是两个述语共用一个或两个语义块。常见的复合句类相当于传统语法中所说的连动句和兼语句。

关于连动句，它的定义和范畴在语法学界存在分歧。“武震脱了鞋走进屋去”是连动句，“他拍着手笑”、“他骑上自行车走了”、“他拉长了脸瞧着我们”是不是连动句呢？在这点上分歧颇大。对于这类句子，王力先生把前一个动词称为“末品谓语形式”，就是说，它具有谓语的形式，却永远不做整句的谓语，只能用来修饰句中的谓词。黎锦熙先生称它为“副词的附加语”，吕叔湘先生称它为“动词的前附加语”。他们都认为是动词做了状语，而不是谓语。《语法讲话》和《现代汉语》认为“连动式就是动词（包括副动词）结构式”，“是几个动词或动词性词组连用的格式”，把这类句子划归连动式的范围。有人甚至主张废除“连动句”这个术语，而更名为“连谓句”。有人干脆不把这一类单立，把它和兼语句、紧缩句等并为一类，称为“复杂谓语”。“连动式”在语法系统中的地位从来都是不稳定的。

基于语义分析的 HNC 理论，进行工程上的句类分析时，对这类句子，首先是作为复合句类来处理。上面提到的例子中，“拍”、“骑”、“拉长”等动词被标记为 E1。但是还需要作第二步的深层分析，废除它的主块地位，把和它一起构成的整个完整的语义块变换为辅块，并判定它的辅块类型。这就是本节所要论述的“以实词为首标记符的辅语义块”。

根据汉语中此类句子的分布特点，笔者着重考察了“v+着+v”式的句子。当然，并不是所有“v+着+v”的句子都是以实词为标记符的辅语义块的情况。如：

“他们面临着重新学习的任务。”(SJ=SB+S+SC)

虽然是“v+着+v”的形式，但是“v+着”不充当辅语义块。相反，“v+着”充当 EK。

利用了从北京语言学院 200 万字的语料中抽出的 1069 个句子，经过仔细的筛选，找出 303 个该类型的句子。为了突出研究的重点，对原语料的一些句子进行了前后删除，只取“着”前后具有 v 且能构成一个句子为限定条件。几个“v+着”并列且作用相同的句子考虑在内；句中具有语音停顿（其标志为 v2 前有逗号）的句子不予考虑；完全相同的句子不再重复收录；中间有“就”、“又”等连接性副词的不考虑；“他脸上露着笑说”等此类有单独小主语的不考虑。该语料属叙述体。

对 303 个句子进行了人工句类分析，以第二个 v 作为主 E 语义块，分析该句子的句类，考察 v1 与它后面的成分所构成的语义块的类型，并着重考察了 v1 的概念结点的分布情况。

句类分布情况如下：

句类	句数	百分比	排序
T3	74	24.42	1
T2b	46	15.18	2
P	43	14.19	3
X	37	12.21	4
S	29	9.57	5
T1	22	7.26	6
X2S	17	5.61	7
Y	9	2.97	8

T2	6	1.98	9
RY	4	1.32	10
R	3	0.99	11
X1S	3	0.99	11
T4	3	0.99	11
X2	2	0.66	12
XT2	1	0.33	13
XT1	1	0.33	13
RS	1	0.33	13
YS	1	0.33	13
X2ST3	1	0.33	13

从上述有关句类的具体分析中我们可以发现，“着”前面的 *v* 主要是单音节的词；E 语义块主要是人的行为、活动，所以前面的“*v*+着”也主要描述或表现人的状态、行为，它的概念结点的分布情况以 5 行表示人的形态的 *v*51a 所占的比重最大，其次是 1 行表示运动过程的 *v*1098、2 行的表示接收的 *v*21 和表示信息转移的 *v*23、4 行的表示主宰与从属的 *v*44、7 行的表示情感的 *v*713 和表示态度的 *v*711、1 行表示语言逻辑概念的 *lv*11。在这些词的具体 HNC 符号中，1 行、2 行前面应该加表示“生理及本能活动”的 6 行的“62 或 64 或 65”，标明是人类活动；从语义上讲，“着”是一个很有影响力的词，*v* 加上“着”后可以完全改变它的意义，从语法上讲，汉语虽然是缺少形态变化的语言，但“着”作为一个后缀，它可看作是一个指示符，不同于一般的指示符，把 *v* 虚化成一个 *lv* 概念。

软件工作者遇到这种类型的句子，无论现在是把它当作 E1，还是句蜕，应该根据 E2 形成的句类以及 E1 的概念结点，参考上述结论，进行深层分析，判定它的辅块地位及意义。

从机器翻译的角度来说，我们又可灵活掌握。在分析具体语种的语法形态时，我们发现，以实词为标记符的辅语义块“*v*+着”，在英语中，*v*2 是主动词，*v*1 是以现在分词做状语的形式或以介词短语来表示的。所以在机器翻译中，当英语为目标语时，汉语中的 *v*1 在英语中没有相应的介词短语来翻译时，一定用现在分词来表示，在句中充当状语。所以出于机器翻译的考虑，工程上的需要，我们把这种在理论上分析为辅语义块的“*v*+着”变换为主 E 语义块的“上装”——QE。这样在机器翻译中，QE 加了相关知识，但是，在英汉汉英翻译中，保证了辅语义块的首标记符之间的对译是介词与介词之间的对译。

4 关于标记符

HNC 理论的三大语义网络之一——语言逻辑概念网络（标记为 *l*），是对汉语的介词和连词等的分类。作为语义块标记符的主要是介词。我们针对俞士汶《现代汉语语法信息词典详解》中所列的“介词表”，对介词的这种标记符功能进行分析。

主语义块标记符：

句类格式有标准和非标准的区分，非标准格式包括规范格式。规范格式是汉语的一种特殊形式，指在句类中，标准格式一般是 JK1+EK+JK2，但是借助于介词作为标记符，语义

块的位置进行变动，形成 JK1+^JK2+EK、JK2+^JK1+EK 这种规范格式。

如：“他们把大批工程都包揽下来了。”

该句为主动承受句，结构为：X11J=X1A+^XBC+X11，“把”作为 XBC 的标记符。

辅语义块标记符：

大部分辅语义块是带有标记符的，主要是以介词作为首标记符的。这些介词除了有标记符的作用，自身还有独立的逻辑意义，它能表明辅语义块的类别。如“按照”表示参照辅块、“为了”表示目的辅块。

如：“为了更好地担负起理论工作的历史重任，我们必须严格要求自己。”

两可语义块标记符：

辅语义块的 7 种类型中，途径、工具、方式、条件是纯粹的辅语义块，辅语义块标记符和语义块内部出现的综合类概念或 HNC 符号能保证类型的纯粹性，参照，结果（或称目的）就不是那么纯粹了，除了有相应的综合类概念如“……的标准（原则，方面，高度）、……的目的（结果，目标）”等外，语义块内部会出现对象、内容之类的概念，所以 HNC 理论分析出两可语义块，参照对象、参照内容和目的对象、目的内容。从语义上来说，它们与 EK 所表达的意义也较密切，是 EK 所需的对象或内容，但是 EK 的前后已经存在对象内容语义块并且满足 EK 所需的语义块个数，我们把这种语义块称为两可语义块，或者说，这种语义块主辅不分。参照内容、参照对象和目的对象、目的内容，分别记为 ReC、ReB、RtB、RtC，他们也带有逻辑标记符。

如：“作为一个干部，他明白自己的职责所在。”

该句为效应句，结构为：Y0J=^ReB+YB+Y+YC，“作为”是参照对象的标记符。

在对 8872 个标注了介词结构的句子进行人工句类分析（对 8872 个句子中没有出现的表中所列的介词，则从本年度《人民日报》上寻找例句），得到了分析结果。如下表：

主语义块标记符表

主 语 义 块 标 记	纯粹 主 语 义 块 标 记 符	把、被、为（阳平）、对、对于、连、往、跟、同、和、与、就、让、叫、替、管、将	
	兼 类 主 语 义 块 标 记 符	兼 辅	打、由、向（着）、朝（着）、以、于、为（去声）、通过、从、经过、经
		兼 两 可	给、至于 为（去声）、从

辅语义块标记符表

辅语义块标记符	纯粹语义块标记符	按、按照、本着、根据、基于、鉴于、据、像……一样、离、拿、依、依据、依照、照、照着、遵照、趁、趁着、当、当着、及至、借、临、每当、凭、凭借、顺着、随着、为了、为着、沿、沿着、因、因为、用、由于、在、正如……那样、只限、至、待、自、自从	
	兼类语义块标记符	兼主	打、由、向(着)、朝(着)、以、于、为(去声)、通过、从、经过、经
		兼两可	为(去声)、从

两可语义块标记符表

两可语义块标记符	纯粹两可语义块标记符	作为、关于、针对、就……而言、拿……来说	
	兼类两可语义块标记符	兼主	给、至于为(去声)、从
		兼辅	为(去声)、从

《现代汉语语法信息词典详解》的“介词表”。表中所列介词的总数为 85 个，笔者选取了 76 个（被去掉的是“比、乘、除、除了、尽、连同、随、望、为（去声）），并外加 4 个（分别是“拿、拿……来说、随着、就……而言”），总计 80 个。其中，主语义块标记符有 32 个，辅语义块标记符有 56 个，两可语义块标记符有 9 个。在 80 个介词中，纯粹作为主语义块标记符的有 17 个，占总数的 21%；占主语义块标记符总数的 53%；纯粹作为辅语义块标记符的有 43 个，占总数的 54%；占辅语义块标记符总数的 77%；纯粹作为两可语义块标记符的有 5 个，占总数的 6%；占两可语义块标记符总数的 56%。既可作主语义块标记符又可作辅语义块标记符的有 13 个，占总数的 16%；占主语义块总数的 40%；占辅语义块总数的 23%。既可作主语义块标记符又可作两可语义块标记符的有 4 个，占总数的 5%；占主语义块标记符总数的 13%；占两可语义块标记符总数的 44%。既可作辅语义块标记符又可作两可语义块标记符的有 2 个，占总数的 3%；占辅语义块标记符总数的 4%；占两可语义块标记符总数的 22%。三种语义块标记符均可作的有 2 个，占总数的 3%；占主语义块标记符总数的 6%；占辅语义块标记符总数的 4%；占两可语义块标记符总数的 22%。

HNC 理论工程上采用“三部曲”的策略——语义块感知、句类假设及检验、语义块构成。语义块感知阶段要借助于语言逻辑概念去辨识语义块，而作为标记符的介词为此提供了极强的信息。上述统计数据表明：纯粹作为主、辅、两可语义块首标记符的介词占总数的比例分别为：21%、54%、6%，三者相加共占 81%，也就是说：有 81% 大约可以在程序判定语义块类型时，直接从这些介词得到依据。对于介词充当兼类标记符的类型判定，则可以从句类入手，优先考虑主语义块的可能。

参考文献

- [1] 黄曾阳,《HNC(概念层次网络理论)——计算机理解语言的新思路》,清华大学出版社,1998年。
- [2] 俞士汶,《现代汉语语法信息词典详解》,清华大学出版社,1998年。
- [3] 刘月华,《实用现代汉语语法》,外语教学与研究出版社,1983年。
- [4] 晋耀红,《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,1998年。
- [5] 惠湛源,“对‘连动式’的意见”,《中国语文》,1954年3月。
- [6] 史振华,“谈‘连动式’”,《中国语文》,1960年1月。
- [7] 王福庭,“‘连动式’还是‘连谓式’”,《中国语文》,1960年6月。
- [8] 马忠,“试论复杂谓语句”,《中国语文》,1961年7月。
- [9] 吴云芳,《现代汉语介词结构的自动标注》,北京语言文化大学95级硕士学位论文。
- [10] 张艳红,《HNC理论的主辅语义块及其相互变换》,北京语言文化大学96级硕士学位论文。