

## 中心语驱动的短语结构语法

冯志伟

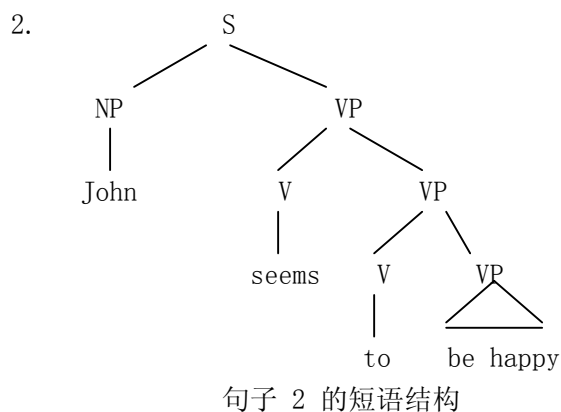
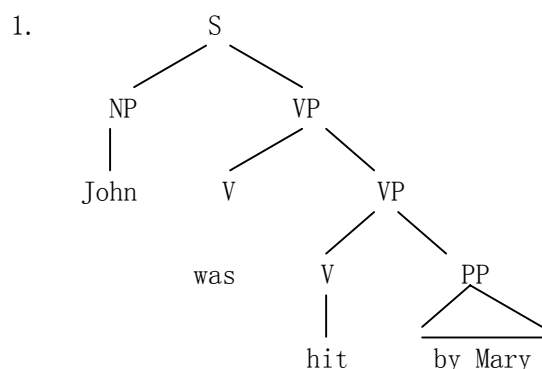
1984年，C. 珀兰德(C. Pollard)和沙格(Sag)在《中心语驱动的短语结构语法分析》(Parsing Head-driven Phrase Structure Grammar)的论文中，提出了中心语驱动的短语结构语法(Head-Driven Phrase Structure Grammar, 简称HPSG)。

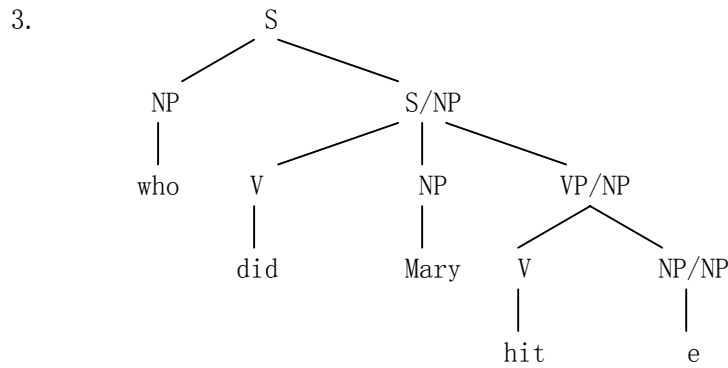
中心语驱动的短语结构语法是在广义短语结构语法的基础上提出的，它基本上继承了广义短语结构语法的原则，并根据自然语言处理的实践进行了重要的改进。这种新的语法理论的突出特点，就是特别强调中心语在语法分析中的作用，使整个语法系统由中心语来驱动。

我们来看下面几个英语句子：

1. John was hit by Mary  
(约翰被玛丽打了)
2. John seems to be happy  
(约翰似乎是幸福的)
3. Who did Mary hit ?  
(玛丽打了谁?)
4. John tries to finish the job  
(约翰试图结束这个工作)

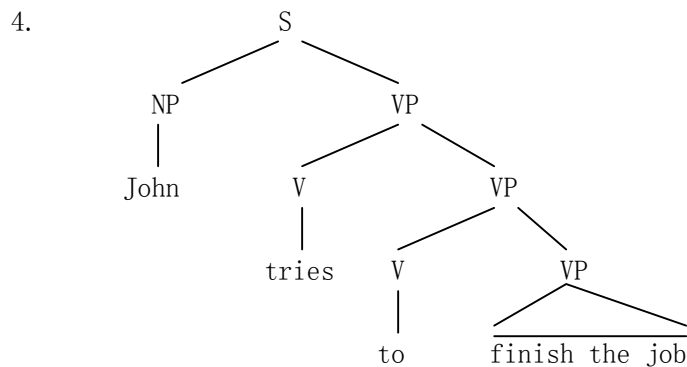
这些句子的短语结构分别如下：





其中的 e 表示 who 的踪迹。

句子 3 的短语结构



句子 4 的短语结构

根据广义短语结构语法，我们可用如下的直接支配规则来生成上述句子：

1.  $S \rightarrow H, NP$
2.  $VP \rightarrow H, VP$
3.  $VP \rightarrow H, NP$

显而易见，对于英语来说，还需要下列规则来处理 VP：

4.  $VP \rightarrow H, NP, NP$
5.  $VP \rightarrow H, NP, PP$
6.  $VP \rightarrow H, NP, VP$
7.  $VP \rightarrow H, NP, S$

在这些规则中，VP 规则的使用是由作为中心语的动词的出现情况决定的。例如，在不定式标志 to 之后，必定出现 VP，这可用规则 2 来表示；当动词 give 有两个宾语时，可用规则 4, 5 来表示，……等等。可见，规则的使用必须考虑中心语的词汇项目的基本性质，也就是必须考虑中心语的次范畴化特征 SUBCAT 的值，从而用中心语来驱动规则的使用。

次范畴化规则用特征结构表来表示，写为 [SUBCAT]，这实际上也就是次范畴化特征。我们在讲广义短语结构语法时曾经说过，动词的次范畴化特征，就是该动词在形成一个句子时所欠缺的所有范畴的集合，如果是不及物动词，它要形成句子还欠缺一个主语，因此，它的次范畴化特征就是主语；如果是及物动词，它要形成一个句子时还欠缺主语和宾语，因此，它的次范畴化特征就是主语和宾语。单词的次范畴化特征用特征结构表 (list) 来直观地表示。语言单位的远距离联系也可以通过普遍语法的原则来表示。所有合格的语言单位都要用合一的方法来进行运算。

由于中心语驱动的短语结构语法重视词汇（特别是中心语）的作用，根据中心语的次范畴化特征，就可以十分方便地把中心语的语法信息与句子中其它成分的语法信息联系起来，使得整个句子中的信息以中心语为核心而串通起来

珀兰德提出的中心语驱动的短语结构语法，系统地总结了这些语法现象，突出了中心语在语法分析中的地位，并把 SUBCAT 作成成分表 (list) 来取值，逐个地详细描述作为中心语的动词的性质。与上述的广义短语结构语法的 VP 规则相对应，中心语驱动的短语结构语法对于中心语动词的 SUBCAT 作了如下的描述：

1. V[SUBCAT <VP, NP>]

这可描述 seem, do, be, try 等动词。

例如，John seems to be happy (约翰似乎是幸福的)

NP VP

2. V[SUBCAT <NP, NP>]

这可描述 love, hit, kill, read 等动词。

例如，John loves Mary (约翰爱玛丽)

NP NP

3. V[SUBCAT <NP, NP, NP>]

这可描述 give, send, spare 等动词。

例如，John gives Mary a book (约翰给玛丽一本书)

NP NP NP

4. V[SUBCAT <PP, NP, NP>]

这可描述 give, send, buy 等动词。

例如，John gives a book to Mary (约翰把一本书给玛丽)

NP NP PP

5. V[SUBCAT <VP, NP, NP>]

这可描述 persuade, expect 等动词。

例如，John persuades Mary to leave (约翰劝玛丽离开)

NP NP VP

6. V[SUBCAT <S, NP, NP>]

这可描述 expect, believe 等动词。

例如，

John expects every man to do his duty (约翰希望人人尽责)

NP NP S

SUBCAT 的值中，最后的一个 NP 是主语，其余的值是在上面的 VP 规则中出现的补足语。在广义短语结构语法中，SUBCAT 的值的排列顺序在语义上与动词相结合的顺序有关。对于英语来说，SUBCAT 的各个值的排列顺序在大多数情况下与句子中各个成分的逆顺序相对应。

使用这样的 SUBCAT 属性，上述 6 个 NP 规则可表示为如下两个补足语规则和 SUBCAT 属性原则：

— 补足语规则：

1.  $M \rightarrow H C_1$

2.  $M \rightarrow H C_2 C_1$

— SUBCAT 属性原则：

M 的 SUBCAT 的值应该与 H 的 SUBCAT 的值中从左而右地清除了与补足语  $C_1$  和  $C_2$  相一致的部分之后留下的部分相一致，也就是说，在 H 的 SUBCAT 的值中，清除了与补足语  $C_1$  和  $C_2$  相一致的部分之后，留下的部分应与 M 的 SUBCAT 的值相一致。例如，在直接支配规则

$S \rightarrow H, NP$

中, S 相当于 M, H 应是 VP, 因此, S 可表示为  $V[\text{SUBCAT} \langle \rangle]$ , VP 可表示为  $V[\text{SUBCAT} \langle NP \rangle]$ ,  $C_1$  为 NP, 这时, H 的 SUBCAT 的值 NP 与  $C_1$  的值 NP 相一致, 在 H 的 SUBCAT 的值中清除了这个仅有的相一致的部分 NP 之后, 留下的部分为空集, 这样, M 的 SUBCAT 也为空集。可见, 这个直接支配规则符合 SUBCAT 属性原则。这个原则后来发展成“饱和原则”, 我们后面还要进一步讨论。

在中心语驱动的短语结构语法中, SUBCAT 的值是可以改变的。例如, 为了表示被动句, 可以设定如下的词汇规则来改变 SUBCAT 的值:

$$V[\text{SUBCAT} \langle \dots, NP, NP \rangle] \Rightarrow V[\text{PAS}+; \text{SUBCAT} \langle \text{PP}[\text{by}], \dots, NP \rangle]$$

这里, 输入侧最左端的 NP (主语) 与输出侧的 PP 相对应, 输入侧从右数起第二个 NP (宾语) 与输出侧最右端的 NP (主语) 相对应, PAS+ 表示动词为被动式。

由于中心语驱动的短语结构语法特别重视中心语的作用, 根据中心语的次范畴化特征, 就有可能十分方便地把中心语的语法信息与句子中其它成分的语法信息联系起来, 使得整个句子中的信息以中心语为核心而串通起来, 用复杂特征来表示句子的各种信息, 为自然语言的计算机处理提供了方便。这种语法理论已经在一些机器翻译系统得到应用, 具有很强的生命力。

在这种语法中, 所有的语言单位都是通过特征结构来表示的。特征结构要描述语音、句法和语义的信息, 把它们分别表示为 [PHON]、[SYNSEM]。再把这些特征值结合起来, 就可以确定语言单位的声音和意义之间在语法上的关系。语法也是以特征结构的方式来表示的, 这些特征结构也就是语言单位的合格性的限制条件。中心语驱动的短语结构语法与广义结构语法的主要区别在于, 在中心语驱动的短语结构语法中, 特别重视词汇的作用, 词汇借助于合一的形式化方法, 构成一个层级结构, 在这个词汇层次结构中的信息可以相互流通和继承, 在全部的句法信息中, 词汇信息占了很大的比重, 而真正的句法信息只占了不多的比例。

在中心语驱动的短语结构语法中, 一个句子的结构可以形式地用表示式 (Sign) 来描述, 最简单的表示式包括 [PHON] 和 [SYNSEM] 两大部分:

$$\left[ \begin{array}{c} \text{PHON} \langle \rangle \\ \text{SYNSEM} \end{array} \right]$$

其中, [PHON] 是句子的语音部分。例如, 句子 “Kim saw the girl” 的语音部分可表示为:

PHON <Kim, saw, the, girl>

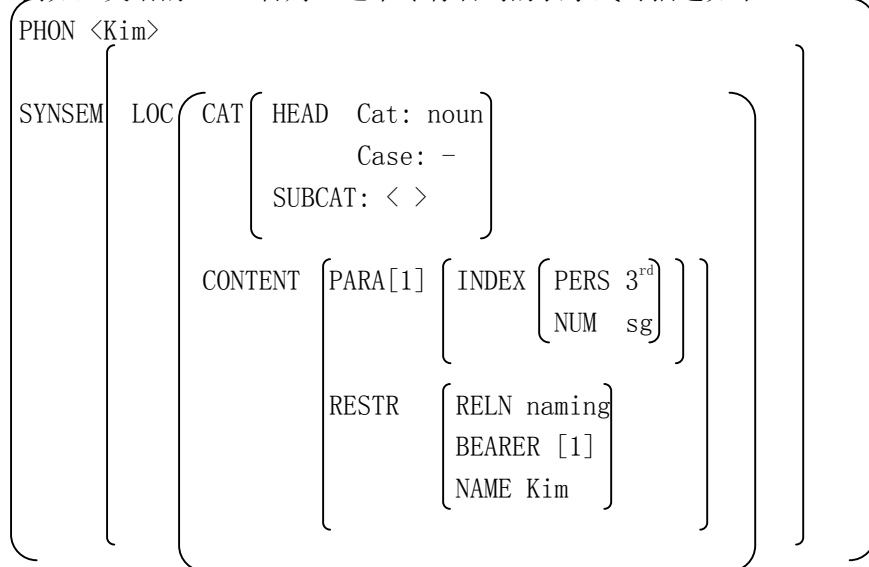
[SYNSEM] 是句子的句法语义部分, 其基本结构又可以用类似的表示式描述如下:

$$\text{SYNSEM} \left[ \begin{array}{c} \text{LOC} \left[ \begin{array}{c} \text{CAT} \left[ \begin{array}{c} \dots \end{array} \right] \\ \text{CONTENT} \left[ \begin{array}{c} \dots \end{array} \right] \end{array} \right] \\ \text{NONLOC} \end{array} \right]$$

在句法语义部分中, LOC 表示实位成分 (local), 用于记录在句子中实际位置的信息,

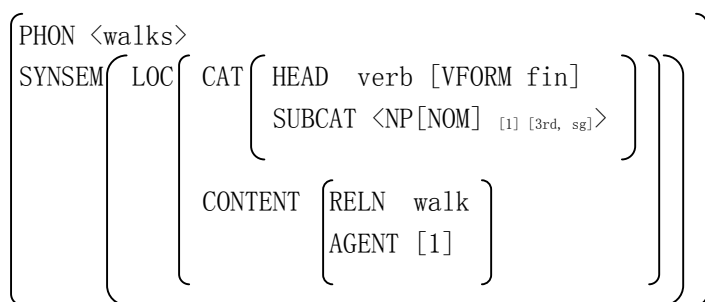
NON-LOC表示空位成分(no local), 用于记录有远距离关系的空位信息。LOC进一步分为CAT和CONTENT, CAT表示范畴(category), 说明句子成分的形态和句法特征, CONTENT表示含义(content), 说明句子成分的语义特征。

例如, 英语的Kim (吉姆) 这个专有名词的表示式可描述如下:



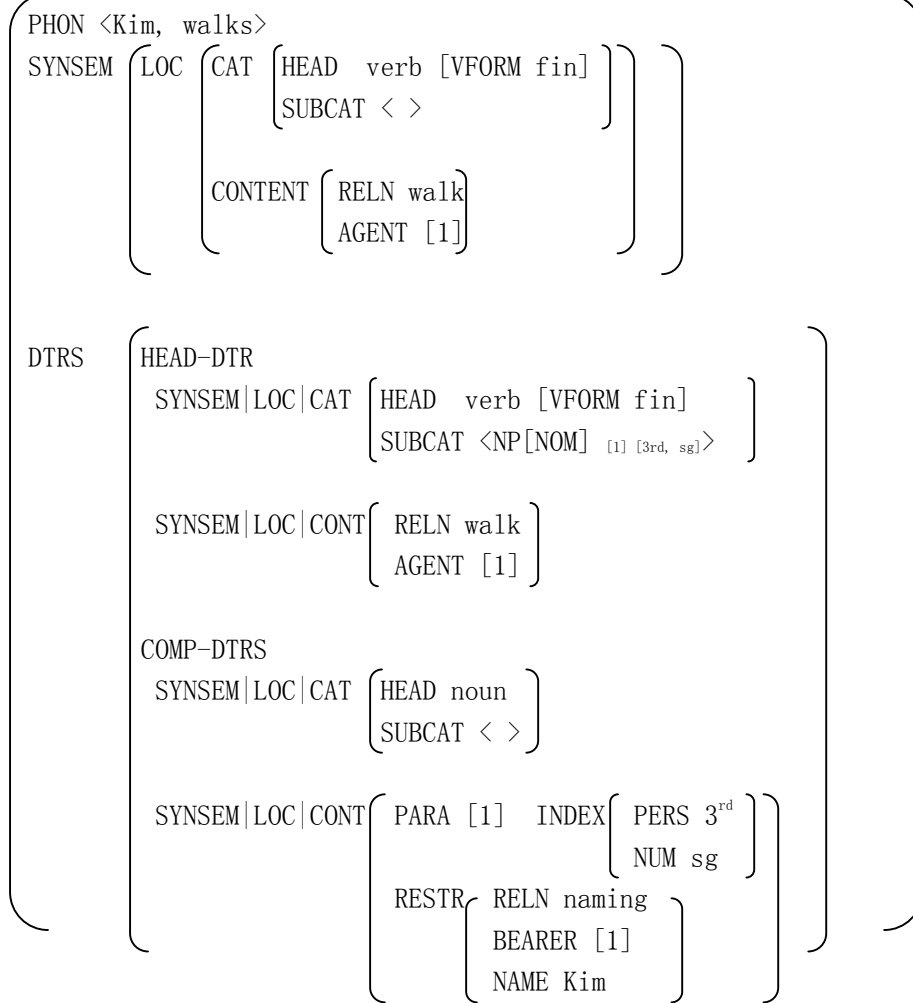
上述表示式中, PHON部分的语音是Kim, SYNSEM部分的句法语义只有实位成分, 没有空位成分, 实位成分的CAT记录了HEAD(中心语)的范畴特征(Cat)为noun(名词), 格特征(Case)为“-”(没有格), SUBCAT(次范畴)特征为<>, 实位成分的CONTENT记录了Kim的含义(即语义特征), PARA表示参数(parameter), 其INDEX(标引)有PERS(人称)和NUM(数)两项, PERS为第三人称(3<sup>rd</sup>), NUM为单数(sg); RESTR表示限制参数(restriction), 共有RELN(relation, 表示关系), BEARER(表示承担者), NAME(表示名字)三项, RELN为naming(命名, 即给人取名字), BEARER后注明[1], 表示它的参数与PARA(1)相同, NAME后的Kim就是给承担者取的名字。这些特征恰当地表达了Kim这个词的语音和句法语义特性。

又如, 英语walks(走路)这个动词的表示式可描述如下:



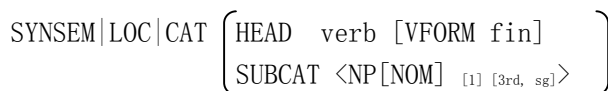
上述表示式中, PHON部分的语音是walks, SYNSEM部分的句法语义只有实位成分, 没有空位成分。实位成分的CAT记录了HEAD特征是verb, 它的动词形式是限定动词[VFORM fin], SUBCAT记录了walk的次范畴为<NP[NOM] [1] [3<sup>rd</sup>, sg]>, 这是一个主格(NOM)、第三人称(3<sup>rd</sup>)、单数(sg)的名词短语NP, 参数取[1]; 实位成分的含义CONTENT记录了其关系(RELN)为walk, 其施事者(AGENT)的参数取[1]。我们知道, 在专有名词Kim中的参数也取[1], 因此, walk的次范畴要取Kim的参数[1]。

短语Kim walks的结构表示式可描述如下：

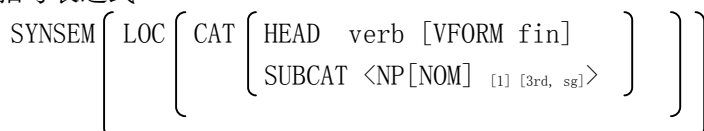


在上述的句子表示式中，PHON部分的语音是Kim walks，SYNSEM部分记录了短语的句法语义；值得注意的是，短语表示式中增加了DRTS (Daughters)部分，用于描述短语的子节点的信息。子节点分为HEAD-DRT和COMP-DRTS两种，HEAD-DRT是中心语子节点，描写这个短语中的walks的特征，COMP-DRTS是补足语子节点，用于描述短语中Kim的特征。可以看出，除了SUBCAT不同之外，短语中HEAD的特征值与中心语子节点的HEAD中的特征值是相同的，它们在结构上是共享的，由于短语和中心语子节点所处的层次不同，它们的次范畴应该不同，在短语这一层，SUBCAT为<>，在中心语子节点这一层，SUBCAT为<NP[NOM]<sub>[1] [3rd, sg]</sub>>，它要求一个主格、单数、第三人称的NP作补足语。这个补足语就在COMP-DRTS中描述。

为了节省空间，DRTS中采用了简洁的表示方法，在SYNSEM和LOC这两层都没有使用括号，而是使用“|”来代替括号。



代表了括号表达式



从Kim walks这个短语的结构表示式中可以看出，在有中心语的短语中，短语的HEAD的特征值同中心语子结点的HEAD的特征值在结构上共享。这是中心语驱动的短语结构语法词汇信息流通的最重要的原则，叫做“中心语特征原则” (Head Feature Principle)。

因为短语所在的层次也就是父结点所在的层次，因此，我们可以换一种方式来表达这个“中心语特征原则”：

在有中心语的短语中，父结点的HEAD的特征值同中心语子结点的HEAD的特征值在结构上共享。

中心语驱动短语结构语法中的词汇信息流通的原则还有：

—奉献原则(Contribution Principle)：在有中心语的短语中，父结点的CONTENT的特征值与中心语子结点的CONTENT的特征值等同。

这意味着，父结点的CONTENT的特征值来自中心语子结点，也就是说，中心语子结点要向它的父结点奉献特征值。

例如，短语Kim walks表示式中，父结点的CONTENT的特征值

$$\text{CONTENT} \left[ \begin{array}{l} \text{RELN walk} \\ \text{AGENT [1]} \end{array} \right]$$

就是由它的中心语子结点中的CONTENT的特征值奉献的。

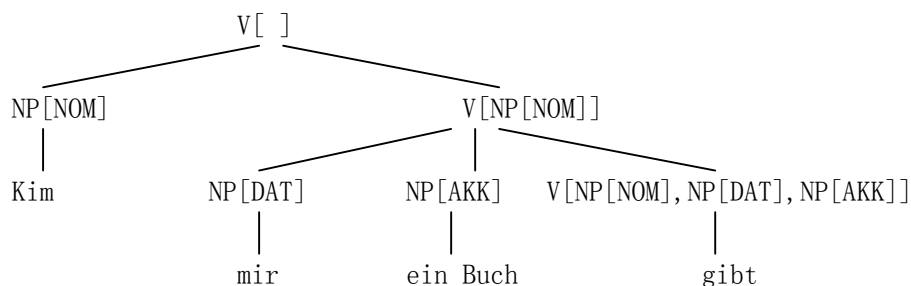
奉献原则可用结构表示式描述如下：

$$\left[ \begin{array}{l} \text{PHON} \langle \rangle \\ \text{SYNSEM|LOC|CAT...} \\ \text{SYNSEM|LOC|CONT[1]} \\ \text{DRTS} \left[ \begin{array}{l} \text{HEAD-DRT} \\ \text{SYNSEM|LOC|CAT...} \\ \text{SYNSEM|LOC|CONT[1]} \\ \text{COMP-DRTS...} \end{array} \right] \end{array} \right]$$

父结点的CONT[1]来自中心语子结点的CONT[1]。

—饱和原则(Saturation Principle)：在有中心语的短语中，父结点的SUBCAT之值等于其中中心语子结点上的SUBCAT之值减去补足语子结点上的有关特征值。

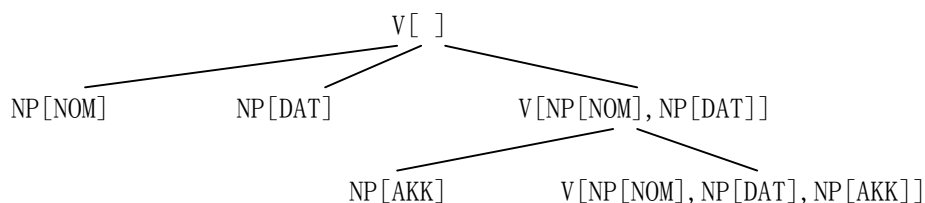
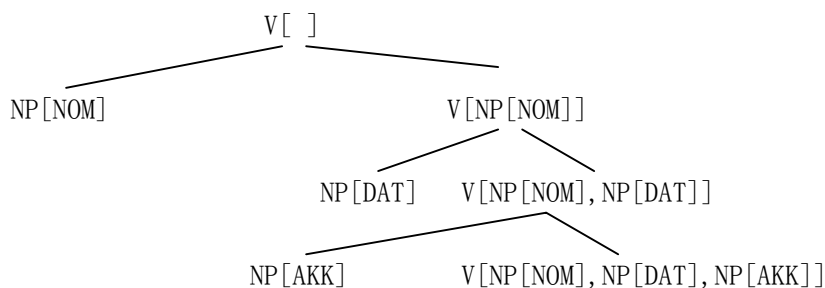
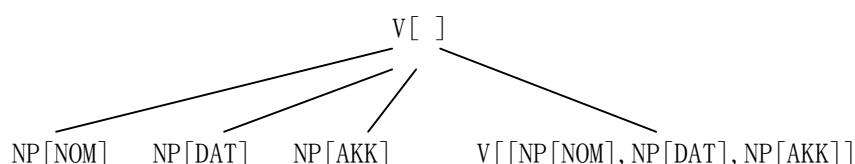
例如，德语短语“Kim mir ein Buch gibt” (吉姆给我一本书) 的树形结构图如下：



在上面的树形图中，每个结点的SUBCAT之值标在它后面的方括号中，NOM表示主格，DAT表示

给格，AKK表示宾格，每个易于看出，如果V[ ]为父结点，那么，它的中心语子结点V[NP[NOM]]的SUBCAT之值[NP[NOM]]减去它的补足语NP[NOM]的值[NOM]为[ ]，这恰恰是父结点的SUBCAT之值；如果V[NP[NOM]]为父结点，那么，它的中心语子结点V[NP[NOM], NP[DAT], NP[AKK]]的SUBCAT之值[NP[NOM], NP[DAT], NP[AKK]]减去它的补足语NP[DAT]和NP[AKK]的值[DAT]和[AKK]之后，恰恰等于父结点V[NP[NOM]]的SUBCAT之值[NP[NOM]]。

如果不管树形图的叶子结点，那么，上面的树形图还可以改变为如下一些形式：



不论怎样改变树形图的组合方式，中心语子结点的 SUBCAT之值减去补足语子结点的相关值，一定等于父结点的SUBCAT之值。这说明，中心语子结点上的SUBCAT之值是饱和的，它不可能再增加了，它已经处于饱和状态了。

饱和原则可用结构表示式描述如下：

$$\left( \begin{array}{l} \text{SYNSEM|LOC|CAT} \left[ \begin{array}{l} \text{HEAD}[n] \\ \text{SUBCAT} \langle [1] \dots [m] \rangle \end{array} \right] \\ \text{DTRS} \left[ \begin{array}{l} \text{HEAD-DRT} \quad | \text{SYNSEM|LOC|CAT|HEAD}[n] \\ \quad \quad \quad | \text{SYNSEM|LOC|CAT|SUBCAT} \langle [1] \dots [m], [n] \rangle \end{array} \right] \\ \text{COMP-DRTS} \langle n \rangle \end{array} \right)$$





CONTENT [INDEX \*1\* [PERS 3<sup>rd</sup> NUM sg]]>]  
 CONTENT [RELN walk WALKER \*1\*]]

从中我们了解到walks是一个关于走路的行为，它是一个限定动词，要求一个主语做走路的行为者，但不要求宾语。

■ 我们有如下的语法规则，这条规则只适用于VP为不及物动词的场合。规则用我们的表示式可写为：

[SYNSEM [CAT [HEAD \*1\* SUBCAT <\*2\*>]  
 CONTENT \*4\*]  
 DRTS [HEAD-DRT [SYNSEM [CAT [HEAD \*1\* SUBCAT <\*2\*>]  
 CONTENT \*4\*]  
 PHON \*3\*]  
 COMP-DRTS < >]  
 PHON \*3\*]

这个规则中的\*1\*, \*2\*, \*3\*, \*4\*是一些临时性的参数，用于简化规则的写法，含义如下：

\*1\* = [VFORM fin]  
 \*2\* = [CAT [HEAD noun SUBCAT < >] CONTENT [INDEX \*1\* [PERS 3<sup>rd</sup> NUM sg]]]  
 \*3\* = (1 2 walks)  
 \*4\* = [RELN walk WALKER \*1\*]

把④与这个规则中的中心语子结点HEAD-DRT合一，得到：

⑤ [SYNSEM [CAT [HEAD [VFORM fin  
 SUBCAT < [CAT [HEAD noun SUBCAT < >]  
 CONTENT [INDEX \*1\* [PERS 3<sup>rd</sup> NUM sg]]>]  
 CONTENT [RELN walk WALKER \*1\*]  
 DRTS [HEAD-DRT [SYNSEM [CAT [HEAD [VFORM fin] SUBCAT <...>]  
 CONTENT [...]  
 PHON <(1 2 walks)>]  
 COMP-DRTS < >]  
 PHON <(1 2 walks)>]

现在我们得到了一个以不及物动词walks为中心语的VP。

■ 我们还有如下的语法规则，这个规则表示：一个饱和的短语可以包含一个中心语短语和一个在这个中心语短语前面的补足语。因为这里的VP只要求一个主语，所以正好使用这条规则。规则用我们的表示式可写为：

⑥ [SYNSEM [CAT [HEAD \*1\* SUBCAT < >]  
 CONTENT \*4\*]  
 DRTS [HEAD-DRT [SYNSEM [CAT [HEAD \*1\* SUBCAT <\*2\*>]  
 CONTENT \*4\*]  
 PHON \*3\*]  
 COMP-DRTS <[PHON \*5\*  
 SYNSEM \*2\*>]  
 PHON (\*5\* < \*3\*)]

这个规则中的\*5\* = (0 1 Kim).

把前面得到的⑤与这个规则中的HEAD-DTR合一，得到：

⑦ [SYNSEM [CAT [HEAD [VFORM fin SUBCAT < >]

```

          CONTENT [RELN walk WALKER *1*]]
DRTS [HEAD-DRT [SYNSEM [CAT [HEAD [VFORM fin]
          SUBCAT <[CAT [HEAD noun SUBCAT < >]
          CONTENT [INDEX *1*
          [PERS 3rd NUM
sg]]]]>]

```

```

          CONTENT [RELN walk WALKER *1*]]
          PHON <(1 2 walks)>]
COMP-DRTS <[PHON *5*
          SYNSEM [CAT [HEAD noun SUBCAT < >]
          CONTENT [INDEX *1*]]>]

```

```

          PHON <*5* < (1 2 walks)>]
    这里, *1* = [PERS 3rd NUM sg]

```

现在我们得到了一个句子（其次范畴为零），但是这个句子的主语（在表示式中是补足语）还没有发音信息。

■ 把③与⑦中的COMP-DRTS合一，得到：

```

    ③ [SYNSEM [CAT [HEAD [VFORM fin SUBCAT < >]]
          CONTENT [RELN walk WALKER [PERS 3rd NUM sg]]]
DRTS [HEAD-DRT [SYNSEM [CAT [HEAD [VFORM fin]
          SUBCAT <[CAT [HEAD noun SUBCAT < >]
          CONTENT [INDEX *1*
          [PERS 3rd NUM
sg]]]]>]

```

```

          CONTENT [RELN walk WALKER
          [PERS 3rd NUM sg]]]
          PHON <(1 2 walks)>]
COMP-DRTS <[PHON <(0 1 Kim)>
          SYNSEM [CAT [HEAD noun SUBCAT < >]
          CONTENT [INDEX [PERS 3rd NUM sg]]]]>]

```

```

          PHON <(0 1 Kim) < (1 2 walks)>]

```

至此，句子的主语有了发音信息，整个句子分析完毕。

上面的分析过程阅读起来比较困难，请读者仔细阅读，是可以理解的。为了帮助读者更清楚地了解中心语驱动的短语结构语法对于语言信息处理的作用，我们再用更加简洁的方式来描述用中心语驱动的短语结构语法来分析自然语言句子的过程。

我们不再写出完整的特征表示式，单词和短语的范畴一律用诸如“NP:”这样的方式来表示。

我们来分析英语句子“Sue the guy with the mustache”（控告那个留小胡子的家伙）。其中，Sue有歧义，作名词用时是人名，作动词用时是“控告”。

自底向上分析过程如下：

① Sue可为名词，也可为动词，首先把它当作名词，得

N: Sue

② 名词可以作NP的中心语，得到：

NP:[HEAD [N: Sue]]

- ③ NP可以做句子S的主语补足语COMP, 得到:

```
S: [HEAD ?  
    COMP1 [NP: [HEAD [N: Sue]]]  
    MOOD declarative]  
    Declarative表示陈述句。
```

- ④ 对于Sue, 不能再继续分析下去, 因此, 回到输入句子中, 取下一个词the, 这是一个限定词, 而限定词必定是一个DetP的中心语HEAD, 并且一个DetP必定是一个NP的补足语COMP. “(MOD)”表示NP有一个可选的定语(modifier).

```
Det: the  
DetP: [HEAD [Det: the]]  
NP: [HEAD ?  
     COMP [DetP: [HEAD [Det: the]]]  
     (MOD)]
```

- ⑤ 继续往前走, 我们必须回到输入句, 找下一个词Guy. Guy必定是一个名词, 它是NP的中心语HEAD, 而且要求一个补足语(COMP) DetP.

```
N: Guy  
NP: [HEAD [N: guy]  
     COMP [DetP: [HEAD [Det: the]]]  
     (MOD)]
```

- ⑥ 这个NP不足以构成句子, 我们回到输入句, 继续分析后面的with, the, mustache等单词. 它们分别是介词、限定词、名词, 构成一个介词短语PP.

```
Prep: with  
PP: [HEAD [Prep: with  
        COMP ?]
```

```
Det: the  
DetP: [HEAD [Det: the]]
```

```
NP: [HEAD ?  
     COMP [DetP: [HEAD [Det: the]]]]
```

```
N: mustache
```

```
NP: [HEAD [N: mustache]  
     COMP [DetP: [HEAD [Det: the]]]]
```

```
PP: [HEAD [Prep: with]  
     COMP [NP: [HEAD [N: mustache]  
                COMP [DetP: [HEAD [Det: the]]]]]]
```

- ⑦ 可以把PP作为前面的NP的修饰语, 得到:

```
NP: [HEAD [N: guy]  
     COMP [DetP: [HEAD [Det: the]]]  
     MOD [PP: [HEAD [Prep: with  
                COMP [NP: [HEAD [N: mustach]
```

COMP [DetP: [HEAD [Det: the]]]]]]]]]]

这时，这个复杂的NP处理完毕。继续往前分析。

⑧ 单词已经全部走完了，可是，S还没有找到它的HEAD，所以，我们只好回溯到第一个词 Sue，这次我们把Sue作为动词，这样一来，这个句子成了命令句(imperative)，并且隐含着句子的主语是you。

V: sue

VP: [HEAD [V:sue]  
COMP ?  
(MOD)]

S: [HEAD [VP: [HEAD [V: sue]  
COMP ?  
(MOD)]]]  
COMP [HEAD [N: you]  
PHON < >]  
MOOD imperative]

⑨ 现在，我们回到了输入的句子，得到该句子的一个分析结果：

S: [HEAD [VP: [HEAD [V: sue]  
COMP [NP [HEAD [N: guy]  
COMP [DetP: [Det: the]]]  
MOD [PP: [HEAD [Prep: with]  
COMP [NP: [HEAD [N: mustache]  
COMP [DetP: [HEAD [Det: the]]]]]]]]]  
(MOD)]]]  
COMP [NP: [HEAD [N: you]  
PHON < >]]]  
COMP2  
MOOD imperative]

⑩ 如果我们想得到一切可能的分析结果，那么，上面的结果还忽略了一种结果，这就是PP修饰动词词组VP的情况，这时，NP只含有the guy，是一个简单的名词词组。显然，这样的结果在句法上是成立的，但是在语义上却不合情理。

S: [HEAD [VP: [HEAD [V: sue]  
COMP [NP [HEAD [N: guy]  
COMP [DetP: [Det: the]]]]]  
MOD [PP: [HEAD [Prep: with]  
COMP [NP: [HEAD [N: mustache]  
COMP [DetP: [HEAD [Det: the]]]]]]]]]  
COMP [NP: [HEAD [N: you]  
PHON < >]]]  
MOOD imperative]

自顶向下分析过程如下：

- ① 我们只有两条关于S的规则，一条规则描述陈述句(declarative)，另一条规则描述命令句(imperative)：

S: [HEAD ?  
COMP ?  
MOOD declarative]

S: [HEAD ?  
COMP [NP: [HEAD [N: you]  
PHON < >]]  
MOOD imperative]

- ② 从树形图的顶部S继续往下走一步，只存在两个非空的NP以及两个非空的VP。  
两个NP如下：

S: [HEAD ?  
COMP [NP: HEAD [N]]  
MOOD [declarative]

S: [HEAD ?  
COMP [NP: [HEAD [N]  
COMP [DetP]]]  
MOOD declarative]

两个VP如下：

S: [HEAD [VP: [HEAD [V]]  
COMP [NP: [HEAD [N: you]  
PHON < >]]  
MOOD imperative]

S: [HEAD [VP: [HEAD [V]  
COMP [NP]]]  
COMP [NP: [HEAD [N: you]  
PHON < >]]  
MOOD imperative]

- ③ 顺着树形图的树枝继续往下走，可得到如下结果。

S: [HEAD ?  
COMP [NP: [HEAD [N: sue]]  
MOOD declarative]

S: [HEAD ?  
COMP [NP: [HEAD [N]  
COMP [DetP [HEAD [Det]]]]]  
MOOD declarative]

S: [HEAD [VP: [HEAD [sue]  
COMP [NP]]]

COMP [NP: [HEAD [N: you]  
PHON < >]]  
MOOD imperative]

继续进行自顶向下的分析，便可得到分析结果，限于篇幅，兹不赘述，有兴趣的读者可以自己试着分析。

近年来，国内外语言信息处理界对于中心语驱动的短语结构语法(HPSG)的研究非常热烈。1994年以来，先后在丹麦的哥本哈根、英国的爱丁堡、德国的图宾根、法国的马赛、波兰的波兹南召开过中心语驱动的短语结构语法(HPSG)的专题国际研讨会，1998年8月，在德国的萨尔布吕肯召开了FHCG国际会议，专门讨论中心语驱动的短语结构语法和范畴语法问题。可以说，中心语驱动的短语结构语法是当前国际语言信息处理研究的一个热点。

这种语法为什么会引起国内外这么多的学者的注意？我们认为，这是由于中心语驱动的短语结构语法反映了当前语言信息处理的一些重要的思想，符合世界学术发展潮流的大趋势。当前，中心语驱动的短语结构语法正在如下一些问题上继续进行探讨：

- 第一，强调语法的限制性，力图把人类的语言模型描述成一个特征结构限制的系统，实行严格的词汇主义(lexicalism)，使得词汇的结构和短语的结构都由一些独立的原则来支配。
- 第二，着重于描述具体的、面向表层的结构，尽力避免那些抽象的结构（如空范畴、功能映射等），把成分结构的描述放在重要位置。
- 第三，语法的组织方式也要反映句子几何结构，按层次关系组织起来的语言信息有助于预测某些在结构上绝对不可能存在的语言现象。
- 第四，力图做到中心语选择的局部化。词汇中心语的选择只局限于SYNSEM部分的主语、补足语和修饰语。范畴的选择、中心语的一致关系等由局部化的选择特征的控制。
- 第五，语法中的词汇特征丰富，这些特征不仅仅是简单地列举出来的，而要通过树形图中特征的继承和传递而获得的，所有的词汇特征被组织成一个有继承关系的层级系统。
- 第六，短语的类型也利用继承关系的层级方法来处理，把不同类型的结构的处理方法统一起来，并在这样的基础之上，建立“构造语法”(construction grammar)的一般性概念。

## 参考文献

- [1] Pollard C. Generalized Context-free Grammar, Head Grammar and Natural Language, PhD thesis, Stanford University, 1984.
- [2] Sergei Nirenburg, Machine Translation – Theoretical and Methodological Issues, Cambridge Uni. Press, 1987.
- [3] 冯志伟，生成语法的公理化方法，《生成语法讨论会文集》，1984年。