

核 心 概 念 篇

词汇功能语法

LEXICAL FUNCTIONAL GRAMMAR

词汇功能语法是 R. M. Kaplan 和 J. Bresnan 于 1982 年提出的一种形式语法 (Kaplan & Bresnan, 1982)。

1. 词汇功能语法的理论构架

词汇功能语法的理论构架可表示如图 1：

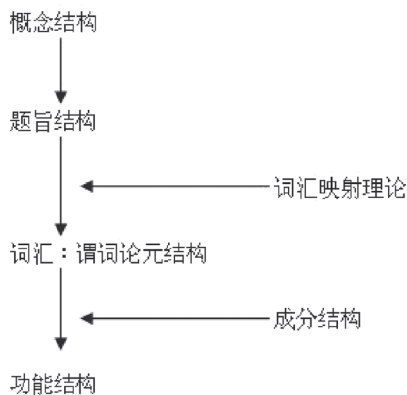


图 1 词汇功能语法的理论构架

概念结构是不同概念在逻辑上的关系，它的表达方式与语言学无关；题旨结构由不同的题旨角色构成，它的表达方式与语言学有关。题旨结构是语言经过筛选之后保存下来的概念结构的骨架，与抽象的句法结构相对应，是赋予每个论元不同语义角色的依据。词汇映射理论用于解释题旨结构与词汇项目的对应关系。题旨结构经过词汇映射得到完整的词汇之后，便进入了句法的范围。

词汇功能语法主要由词库、句法和语义解释三部分组成。表示语义的谓词论元结构首先在词库里通过词汇编码而分配到一个语法功能。这种语义与语法功能之间的联系在词库里是可以通过词汇规则进行转换的。当一个词条最后取得正确的语法功能编码以后，它就可以和其他表示语法意义的词条一起构成词汇输入进入句法部分。

句子在句法部分有两个表达层次：成分结构 (constituent structure, 简称 c-structure) 和功能结构 (functional structure, 简称 f-structure)。

成分结构是语言的外部结构。它表示句子成分的先后次序，是由一组短语结构规则映射而成的树形结构。语法功能通过句法编码进入短语结构规则，然后进入树形结构的相应位置。这一层次代表句子的句法排列和语音表达。

功能结构是语言的内部结构。它表述各语言成分之间的关系，代表句子的语义。一般说来，不同语言的内部结构的表达方式大体上是一致的，因而功能结构具有普遍性；而不同语言的外部结构却有着很大的不同，因而成分结构具有差异性。

功能结构和成分结构是两个具有不同形式的独立体系，词汇功能语法把它们明确区别开来分别进行描述，然后又把它们合在一起，使人们对语言的结构获得一个总体的印象。成分结构可以通过功能描述向功能结构转换。功能描述由一组等式构成，可以很容易地进行计算机编码。成分结构描述了语言的表层结构，成分结构中的单词承载了大多数语法信息。功能等式规定了这些语法信息的组合方法，经过有穷步骤的运算之后，便得到了功能结构。为了确保功能结构的正确性，还要对功能结构的合格性进行判别，为此，词汇功能语法还规定了“合格性条件”（well-formedness condition）。

R. M. Kaplan 和 J. Bresnan 证明，在词汇功能语法中，由成分结构到功能结构的运算在数学上是有定解的（decidable），而且所有的运算都只需要“合一”（unification）这种简单的运算方式。合一就是当信息相冲突时，运算失败；当信息不冲突时，运算成功。合一运算是数理逻辑中“并”运算的一种适于自然语言处理的特殊形式。

词汇功能语法可以更加具体地表示为如图 2 的模式：

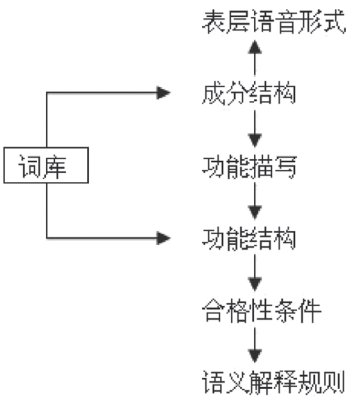


图 2 词汇功能语法的模式

2. 成分结构

成分结构是词汇功能语法中句法描写的一个层面。它是由上下文无关的短语结构语法来表示的，它的形式是一般意义上的短语结构树。树形图上的节点带有句子中的词或短语预示的功能信息。这些信息由语法规则右部的符号所带的“功能等式”来表示。

例如，短语结构规则

$$\begin{array}{ccc} S & \rightarrow & NP \qquad \qquad \qquad VP \\ & & (\uparrow \text{SUBJ}) = \downarrow \qquad \qquad \uparrow = \downarrow \end{array}$$

采用了向上单箭标“ $\uparrow$ ”和向下单箭标“ $\downarrow$ ”来表示范畴的支配关系。向上单箭标“ $\uparrow$ ”表示直接支配成分，向下单箭标“ $\downarrow$ ”表示被支配成分。这些箭标可以用在功能等式中，等式左边为限定成分，等式右边为限定值。例如，“ $(\uparrow \text{SUBJ}) = \downarrow$ ”读为“直接支配成分的主语等于被直接支配成分的语法功能”，“ $\uparrow = \downarrow$ ”读为“直接支配成分的语法功能等于被直接支配成分的语法功能”。上面的短语结构规则表示句子 S 由 NP 和 VP 组成，VP 前的 NP 是句子的主语。NP 的下方“ $(\uparrow \text{SUBJ}) = \downarrow$ ”是它的功能等式，表示这个 NP 继承了它的父节点 S 的主语 (SUBJ) 特征，“ $(\uparrow \text{SUBJ})$ ”表示 NP 的全部功能信息就是支配它的 S 的主语功能信息，“ $\downarrow$ ”表示该符号本身（即被 S 直接支配的成分 NP）。VP 下方的功能等式“ $\uparrow = \downarrow$ ”表示 VP 所带的全部功能信息就是支配它的父节点 S 的功能信息。

这个短语结构规则也可以用树形图表示如图 3：

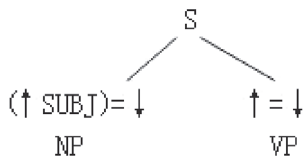


图 3 短语结构规则的树形图表示 (1)

这里要注意三个问题：

- 上箭标和下箭标都是表示功能信息，而不是表示节点；
- 下箭标表示的是本节点的功能信息，而不是本节点之下的子节点的功能信息；
- 等号不仅表示两边的值相等，而且还表示进行合一运算，也就是说，运算前要进行相容性的检查。

又如，短语结构规则

$$\begin{array}{ccc} \text{NP} & \rightarrow & (\text{DET}) \qquad \qquad \qquad \text{N} \\ & & (\uparrow=\downarrow) \qquad \qquad \qquad \uparrow=\downarrow \end{array}$$

表示 NP 由限定词 DET 和名词 N 组成，限定词是可选的成分，N 继承了其父节点 NP 的功能信息。这条语法规则实际上代表了如下两条短语结构规则：

$$\begin{array}{ll} \text{a. NP} & \rightarrow \text{DET} \qquad \qquad \qquad \text{N} \\ & \uparrow=\downarrow \qquad \qquad \qquad \uparrow=\downarrow \\ \text{b. NP} & \rightarrow \text{N} \\ & \uparrow=\downarrow \end{array}$$

它们可以用树形图表示如图 4：

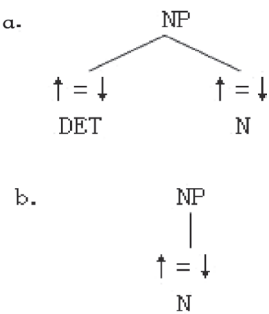


图 4 短语结构规则的树形图表示（2）

再如，短语结构规则

$$\begin{array}{ccc} \text{VP} & \rightarrow & \text{V} \qquad \qquad \qquad \text{NP} \\ & & \uparrow=\downarrow \qquad \qquad \qquad (\uparrow\text{OBJ})=\downarrow \end{array}$$

表示 V 继承了其父节点 VP 的功能信息，NP 继承了其父节点 VP 的宾语 (OBJ) 特征。

这条语法规则用树形图表示如图 5：

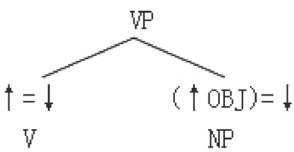


图 5 短语结构规则的树形图表示（3）

这时，由于V继承了VP的功能信息，而VP又继承了S的功能信息，所以，V也就继承了S的功能信息，也就是说，V共享了S的全部信息。上面规则的VP和这个规则的V上都有功能等式 $\uparrow=\downarrow$ ，这个功能等式使得V、VP和S共享了全部的信息，因此可以把带有功能等式 $\uparrow=\downarrow$ 的节点称为“功能中心语”。功能中心语所承载的信息直接并入到它们的父节点所代表的功能信息之中；这些信息非常重要，它们表示了所分析的句子或短语在功能上的基本格局。 $(\uparrow \text{SUBJ})=\downarrow$ 与 $(\uparrow \text{OBJ})=\downarrow$ 表示相应节点所代表的功能信息在父节点中的具体功能。

成分结构代表句子的表层结构中的单词的顺序，它是语言描写的输入部分。成分结构由短语结构规则决定。每一种语言都有其特有的短语结构规则。从理论上说，这些短语结构规则能够生成这一语言中的任何句子。把短语结构规则所表示的各个范畴映射到树形图上，就得到了一个句子的成分结构。

例如，英语句子“He reads the book.”。这个句子的成分结构可以结合上述的语法规则表述如下（图6）：

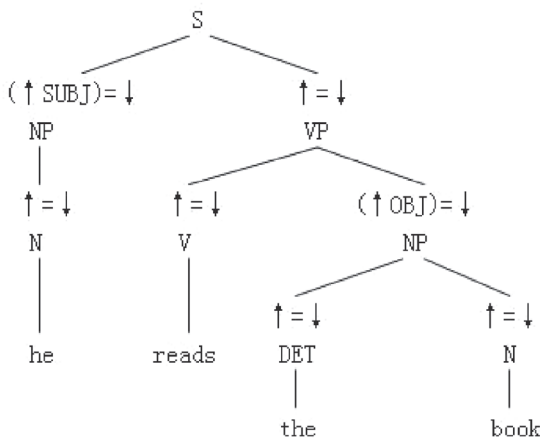


图6 成分结构的树形图表示

成分结构只能由短语结构规则映射而成。除此之外，我们没有其他手段来改变成分结构中各个单词的先后次序。即便是具有结构依附关系句子的成分结构，也要用短语结构规则来生成。

例如，英语句子“I wondered what he read.”就是一个具有结构依附关系句子。在这个句子中，read的宾语就是what，即he read what。图7是构成这个句子的成分结构所要用的短语结构规则：

我们在短语结构规则中使用了双箭标“ $\Downarrow$ ”。双箭标主要用来表示成分结构中范畴之间非直接支配的依赖关系，特别是远距离的支配关系。双箭标必须成对使用，凡是标有“ $\Downarrow$ ”关系的成分必须依附于标有“ $\Uparrow$ ”关系的成分，而“ $\Uparrow$ ”关系的存在要以“ $\Downarrow$ ”关系的存在为前提。

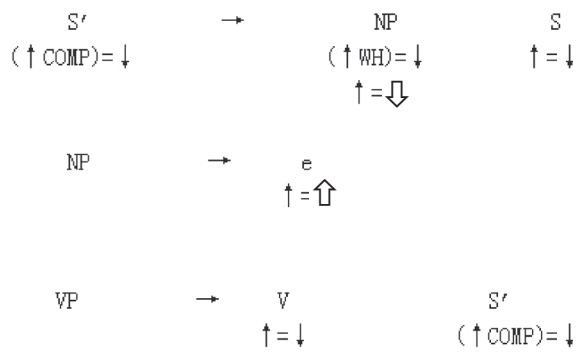


图 7 带双箭标的短语结构规则

在规则中，e 表示空位，通过上述规则的映射，我们可得到 “I wondered what he read.” 这个句子的成分结构（如图 8 所示）：

因为双箭标必须成对使用，所以，e 就代表了 what，二者是相通的。可见，成分结构也可以表示远距离的支配关系。

短语结构规则是句法规则，此外还有词法规则。词法规则是由词典信息提供的，它带有语法功能的预示信息，在词汇功能语法中占有重要地位。例如，下面的词汇项表：

he:	N,	(↑PRED)='he'
		(↑ABST)=-
		(↑GENDER)=MAS
		(↑NUM)=SING
		(↑PERS)=3
		(↑CASE)=NOM
read:	V,	(↑PRED)='read <(SUBJ) (OBJ)>'
		(↑TENSE)=PRESENT
the:	DET,	(↑SPEC)=the
		(↑DEF)=+

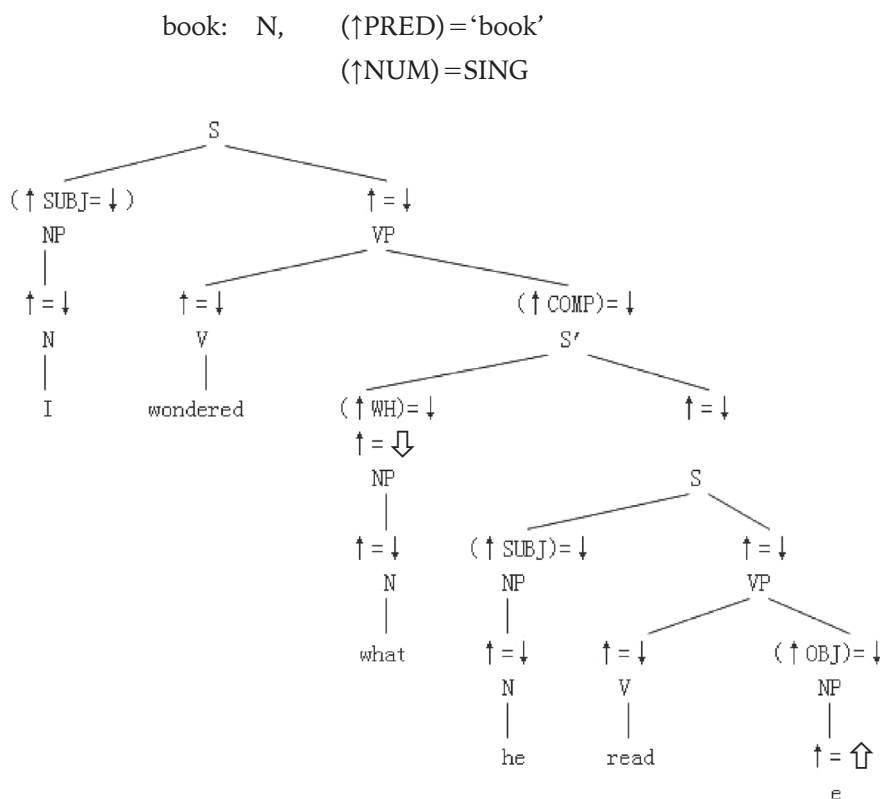


图 8 句子的成分结构

词汇功能语法把词汇按词的不同意义立项，词汇项所含的信息有语法范畴和功能等式。功能等式的形式与短语结构规则中的功能等式完全一致，便于用统一的方法来处理语言信息。

在上述词汇项中，he 这个词汇项的“(↑PRED)='he'”表示它的父节点具有功能 PRED（谓词），其具体的信息为‘he’；“(↑ABST)=-”表示它的父节点具有功能 ABST（抽象），其具体的信息为“-”（不抽象），因此，he 是个具体的物；“(↑GENDER)=MAS”表示它的父节点具有功能 GENDER（语法性），其具体信息为 MAS（阳性），“(↑NUM)=SING”表示它的父节点具有功能 NUM（数），其具体信息为 SING（单数）；“(↑PERS)=3”表示它的父节点具有功能 PERS（人称），其具体信息为 3（第三人称）；“(↑CASE)=NOM”表示它的父节点具有功能 CASE（格），其具体的信息为 NOM（主格）。the 这个词汇项的“DET (↑SPEC)=the”

表示它的父节点具有功能 SPEC (指示), 其具体的信息为 the (定冠词), “(↑DEF)=+” 表示它的父节点具有功能 DEF (定指), 其具体的信息为 + (是定指)。读者由这种表达方式不难理解其他词汇项的含义。

这里值得注意的是动词的 PRED。在词汇功能语法中, 采用“谓词论元结构”(predicate argument structure) 来表示谓词所带论元的多少以及每个论元所表示的逻辑语义。谓词论元结构的各个论元列在尖括号“<>”之中。例如, read 的谓词论元结构为“(↑PRED)=‘READ<(SUBJ)(OBJ)>’”, 这表示 read 的论元分别是其父节点的主语 (SUBJ) 和宾语 (OBJ)。

词汇功能语法中的语法信息终究来自词汇。功能结构的作用只是检查信息的结构是否合理, 成分结构的作用只是规定信息组合的方式, 而真正带有实质意义的信息全部都来自词汇。因此, 词汇在词汇功能语法中起着决定性的作用, 是我们必须认真加以对待的。

词汇功能语法中的词汇所记录的词汇项目都是形态完全的。词汇中的信息以“定义性功能等式”和“限制性功能等式”的形式来记录。例如, 英语 persuades (劝说) 这个词汇可记录如下:

persuades V. (↑PRED)=‘persuades <(SUBJ) (OBJ) (XCOMP)>’
 (↑OBJ)=(↑XCOMP SUBJ)
 (↑SUBJ PER)=c 3
 (↑SUBJ NUM)=c SING

其中, 前两个等式是一般的“功能等式”, 它们都是定义性的, 因而是“定义性功能等式”; 后两个等式是“限制性功能等式”, 限制性功能等式中有符号 c, 表示“限制”(constraint)。第一个定义性功能等式规定了该动词的谓词-论元关系 (predicate-argument relationship), 该动词要带主语、宾语和宾语补足语 (XCOMP) 三个论元; 第二个定义性功能等式规定了论元之间的控制关系 (control relationship), persuades 这个动词的宾语补足语 XCOMP 没有主语, 它在逻辑上的主语被它上层谓词的宾语所控制 (例如, 在句子 “John persuades Mary to study computational linguistics.” 中, 宾语补足语 to study computational linguistics 在逻辑上的主语受到上层谓词 persuades 的宾语 Mary 的控制); 第三个等式是限制性功能等式, 表示该动词的主语应该是单数第三人称; 第四个等式也是限制性功能等式, 表示该动词的主语的数应该是单数。后两个等式的信息来自词尾 -s, 其余的信息来自 persuade 这个动词词根。

3. 功能结构

功能结构是词汇功能语法句法描写的另一个层面。它是一个属性-值矩阵，其基本结构可以表示为图 9：

属性	值
A	a
B	b
C	c

图 9 属性值矩阵

在这个属性矩阵中，第一列 A、B、C 等表示属性，第二列 a、b、c 等表示相应属性所取的值。

具体地说，功能结构用方括号形式表示语法中各个有意义成分之间的等级关系，从而为在句法层次上做出语义解释提供必要的信息。图 9 方括号内左列首纵行所列的是语法功能或语法特殊标记，笼统地把它叫作“限定成分”。与限定成分在水平方向上相对应的是它们各自的限定值。限定值有三种形式：简单符号、语义形式和子功能结构。子功能结构有自己的限定成分和限定值，而它的限定值还可以有子功能结构，这样就构成了功能结构的递归性质。从理论上说，功能结构可以容纳任何长度的句子。

例如，英语句子 “He reads the book.” 的功能结构如图 10 所示：

SUBJ	PRED	'he'
	ABST	-
	GENDER	MAS
	NUM	SING
	PERS	3
	CASE	NOM
TENSE	PRESENT	
PRED	'read <(SUBJ) (OBJ)>'	
OBJ	SPES	the
	DEF	+
	NUM	SING
	PRED	'book'

图 10 “He reads the book.” 的功能结构