心 理 学 报
第2期 ACTA PSYCHOLOGICA SINICA 1990年

汉语句子的多层次分析[1)]

李 荣
中国科学院心理研究所，北京

摘 要

本文试图建立一个中文句子分析的计算机模型。在该模型中，句法、语义信息对中文句子的多层次分析是同等重要的。为此，我们提出了一些语义规则，利用这些规则可将语义关系综合进句法和词典，使该模型能同时进行不同层次的加工。对汉语一些典型句型的应用表明，该模型是行之有效的。

一、前 言

人理解语言是一个复杂的过程，但有一点可以肯定，即人在理解句子过程中是综合使用句法、语义信息和知识的[1]。犹其是对汉语这种缺乏形态变化，词序也比较灵活的语言来说，语义及知识的作用就更突出了。因此，如何对句子中的句法、语义信息进行表达和应用就是一个句法分析程序的关键。

我们认为下面三个层次的描叙是表达一个汉语句子的意义所必不可少的：

（1）成分(constituent)层次：即句子的表层句法结构，由词、短语等范畴及其间关系来表达；

（2）功能(function)层次：即成分间的语法关系，由主语，宾语等范畴来表达[2]；

（3）格(case)层次：即深层的语义关系，由施事，受事等格角色来表示。

这三个层次的描写能比较充分地反映句子的基本意义，为正确理解实际交际形式(如对话或篇章)中的句子提供了基础。

本文试图建立一个汉语句子的分析模型，该模型以句子为输入，输出为上述三个层次的表达。需要指出的是，在以此为目的的分析中，严格意义上的知识(世界知识)只起辅助作用，分析的完成更多地依赖于句法和语义(也是一种知识)，因此，我们将主要讨论句子及词的句法、语义关系的综合(integrative)表达和利用，并采用一种基于预期(expectation)的分析策略，完成多层次的分析。

二、问 题

从成分、功能和格三个层次来考察汉语，我们能发现汉语的一些基本特点，从中获得一些分析句子的线索。

1. 成分和功能的脱离

汉语中同样一种成分可以担任不同的功能，换句话说，一种成分在担任不同功能时形

1） 本文于1989年7月18日收到。

态上是没有变化的。动词就是一个典型的代表，如：

（1） 吃是享受 动词“吃”作主语

（2） 他不想吃 宾语

（3） 他吃饭 谓语

2. 功能和格脱离

汉语中功能和格之间没有严格的对应关系，如动词的宾语往往可以对应不同的格：

（4） 吃饭 宾语＝受事

（5） 吃食堂 处所

（6） 吃大碗 工具

更一般地，功能和格之间有下面的模糊对应关系：

（1）施事主要作主语，偶尔也作宾语；

（2）受事一般作宾语，有时也作主语；

（3）其它格既可作主语亦可作宾语。

3. 介词的灵活性

介词是进行格分析少数几个可以借助的形式标志之一，然而汉语中介词往往可以支配不同的格，如介词“把”：

（7） 把信交了 介词宾语＝受事

（8） 把鞋走破了 工具

（9） 把个大嫂死了 当事

4. 兼类现象严重

汉语中词的兼类现象比较严重，如“在”就兼属三类词：

(10) 他在家 在＝动词

(11) 他在家看电视 介词

(12) 火车在飞奔 副词

从上面的讨论不难看出句法和语义关系是解决问题的关键。具体来说这些关系体现为词序和词之间的句法、语义限制(constraint)[3]，不同功能的词表示不同的关系，反之，要确定一个词的功能也要借助于它所处的关系。

三、句法规则及功能分析

句法规则描叙句子的组成成分及关系，同时表达成分与功能间的对应。我们采用重写式来表达句法规则。下面是程序所使用的一部分规则。成分右边的括号中是该成分所对应的功能，成分的上标表示成分的配价，这一概念将在后面说明。

$$S \rightarrow PP^{*}\left\{\begin{matrix}(NP)\\(VP)\\(S)\end{matrix}\right\}\ (SUBJ)\ ADV^{*}\ (PP)\ NP^{*}\ (SUBJ)\ VP\ (HEAD)\ VP^{*}\ (HEAD)\ (AUX)$$

$$VP \rightarrow ADV^{*}\ V^{-1}\ (HEAD)\ (AUX)$$

$$\rightarrow ADV^{*}\ V^{-2}\ (HEAD)\ (AUX)\left\{\begin{matrix}(NP)\\(VP)\\(S)\end{matrix}\right\}\ (OBJ)\ (AUX)$$

$$\rightarrow ADV^{*}\ V^{-3}\ (HEAD)\ (AUX)\ NP\ (OBJ)\ \begin{Bmatrix}(NP)\ (OBJ_2)\\(VP)\\(S)\ (OBJ_2)\end{Bmatrix}\ (AUX)$$

$$\rightarrow ADV^{*}\ V^{-3}\ (HEAD)\ PP\ VP\ (Vcomp)\ (AUX)$$

$$NP\rightarrow (DET)\ (M)\ \begin{Bmatrix}NP\\N\\PRO\end{Bmatrix}\ (HEAD)$$

$$PP\rightarrow P(HEAD)\ (NP)\ (OBJ)$$

图1　部分句法规则

其中一些符号的含义为：

HEAD——核心　　OBJ_2——第二个宾语

Vcomp——补语　　ADV——副词

AUX——助词　　DET——指示词

M——量词

功能分析仅凭句法规则无法完成，如成分序列V＋NP＋VP中，VP可能是宾语：

(13)　告诉他现在别来

也可能是补语：

(14)　请他现在别来

其原因在于动词“告诉”和“请”虽然能支配相同的成分，但成分的功能不同。这种由词所引起的不同语法关系可用我们后面提出的“词汇网络”来表达。

四、配价（Valence）及格关系

汉语中动词和介词都能和一定数量的成分发生格关系，如动词“吃”要求二个NP分别担任施事和受事格。我们将一个词所要求的格的数目称为词的配价，即如果词W要求n（n≥0）个格，则W的配价表示为：

$$Val(W) = -n \quad 或 \quad W^{-n}$$

如Val(吃) = －2。

根据配价，可将动词分为三类：

V^{-1}类，只要求一个格，如“死”、“笑”、“病”等；

V^{-2}类，要求二个格，如“吃”，“喜欢”、“认为”等；

V^{-3}类，要求3个格，如“告诉”、“教”、“请”等；

有些词具有二种配价，如“死”：

(15)　他死了　　Val(死) = －1

(16)　他死了父亲　　Val(死) = －2

不同配价的动词能支配的成分也不同，这一点反映在图1所示的句法规则中。

配价的主要作用在于表达短语的格特征，为此我们将配价扩展到任一成分：

若XP是某一成分，且使用了句法规则：

$$XP\rightarrow\cdots X_1(HEAD)\cdots X_n(HEAD)\cdots Y_1\cdots Y_m$$

$Y_1,\cdots,Y_m$是充当格角色的成分，则

$$Val(XP) = \sum_{1}^{n} Val(X_i) + m$$

简言之，任一成分的配价也表示该成分所要求格的数目。因此短语“吃了”和“吃了饭”是不同的，因为Val(吃了) = Val(吃) +0 = −2+0 = −2，而 Val(吃了饭) = Val(吃) +1 = −2+1 = −1，不同的配价表示两个短语能和不同数目的成分发生语义关系。一般地，成分XP^{-n}(n≥0)能和n个其它成分发生格关系。

根据配价可以得出下面的语义规则：

SR_1：

如果 使用了句法规则$Y \rightarrow \cdots NP \cdots X^{-n} \cdots, n>0$

则 如果NP和X的格角色R匹配

则记下关系R = X并修正X的配价，使X^{-n}变为X^{-n+1}。

这里“匹配”是指二个对象的句法、语义特征匹配。建立格关系后，X的配价增加1表示它的一个格角色被填充。

SR_1适用于很多句型。下面是应用SR_1分析(17)中一些格关系的过程，[ ]表示成分所要求的格，每当建立一个格关系，相应格角色就被从[ ]中删除(图2，图3)。

图2 短语分析，计算配价

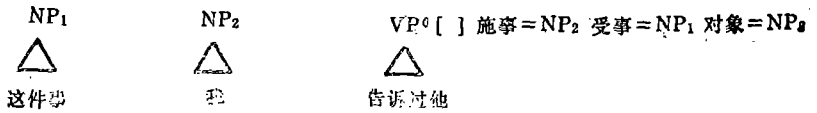

图3 对NP_2及VP应用SR_1，NP_2匹配施事格。
再对NP_1及VP应用SR_1，NP_1匹配受事格

(1)这件事我告诉过他

根据配价，也可将介词分为二类。P^{-1}类如“被”、“在”，它们要求一个格，故其处理和动词一样。P^0类如“把”、“跟”，不要求格角色，其宾语担任的是相关动词的格。这类介词需要特殊的处理。

根据配价计算方法，由P^0类介词形成的PP的配价将是一个正数，如Val(把他) = Val(把) +1 = 0+1 = +1。成分的配价为+n(n>0，通常为1)意味着该成分中的n个直接成分(通常为NP)是另一成分的格角色。根据汉语的特点，可以得到一条更简明的语义规则：

SR_2

如果 使用了句法规则 $XP \rightarrow \cdots PP^{+1} \cdots VP^{-n} \cdots, n>0$

则　如果PP的宾语和VP的格角色R匹配

则　记下关系R＝介词宾语，修正PP及VP的配价，

使PP^{+1}变为PP^{0}，VP^{-n}变为VP^{-n+1}。

SR_2　主要用于分析含介词短语的句子。图4，图5是SR_2在分析(7)时的应用。SR使得程序在分析介词短语格关系时，能综合考虑介词，介词宾语及动词三个因素，从而获得正确结果。

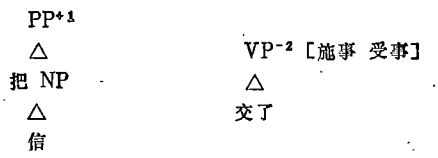

图4　短语分析，计算配价

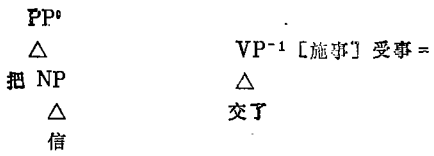

图5　对PP及VP应用SR_2，NP和受事匹配

其它词，如名词，不要求格，故它们的配价都为0。事实上，目前我们认为除动词、介词外的其它词都不具备配价这一语义特征。

五、　词汇网络及词典

心理学的研究[4]表明，人在理解句子时，每读到一个词都会激活一些高层次的认知操作。我们认为这些操作应包括：

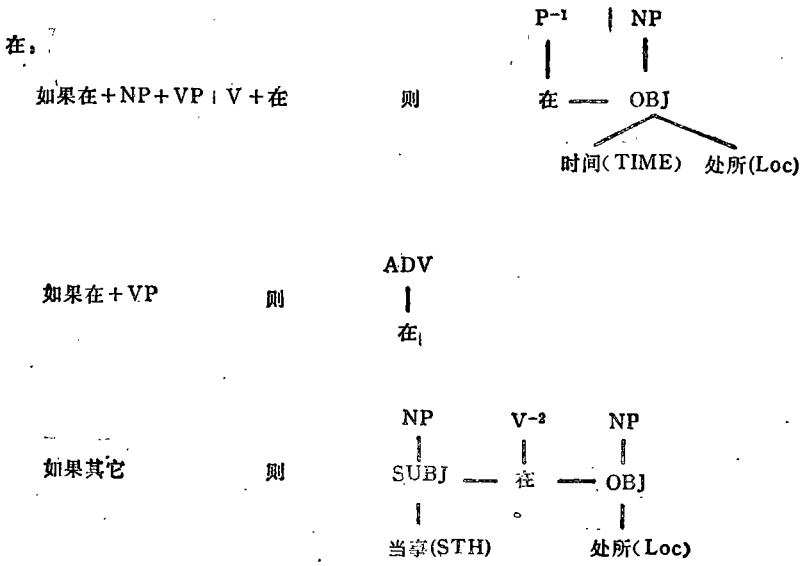

图6　“在”的定义

(1)根据上下文信息确定词的功能；

(2)根据词的功能建立句法和语义关系；

因此，词典中每个词由一组产生式规则来定义，使得机器在遇到一个词时，能完成(1)和(2)的操作。规则的"如果"部分表示词的上下文，"则"部分表示词的功能——它所具有的句法语义关系，这用词汇网络来表示。图6是"在"的定义，"则"部分是词汇网络，网络中横线表示支配关系，竖线或斜线表示对应关系，( )中是格角色的语义特征。

词汇网络中功能用来表示语法关系，它起到了联接成分和格角色的作用。这种显性的综合表达使得词汇网络能方便地用于确定词和短语之间的语义关系，例如要确定"在"后NP担任的格，只需考察词汇网络中与OBJ对应的格角色就可作出正确判断。对于更复杂的情形则需将词汇网络和分析短语之间格关系的语义规则结合起来才行。

六、对汉语一些句型的格分析

在这一部分，我们将应用词汇网络和语义规则对汉语一些典型句型中的格关系进行表达和分析，以证明其有效性。

1. 双NP句

双NP句的第一种情形句型为：

$$NP(SUBJ)\ NP(SUBJ)\ VP^{-2}(HEAD) \quad 其中$$

$$VP^{-2}\rightarrow\cdots V^{-2}(HEAD)\cdots|\cdots V^{-3}(HEAD)\cdots NP(OBJ)\cdots$$

在这种句型中，NP和格角色之间有多对多关系：

NP	NP	例 句
施事	受事	(18) 我饭吃了
受事	施事	(19) 饭我吃了
		(20) 那本书我已经给他了

这种句子可以看作某个格角色从宾语"移位"到主语形成的，这意味着某些格角色可以对应多个功能。如动词"吃"的受事可以作宾语或者主语，当受事和施事同时作主语时，就形成了(18)、(19)那样的句子。这一关系用词汇网络很容易表示(图7)。这类句子可用SR_1进行格分析，具体过程可参看图2，图3。

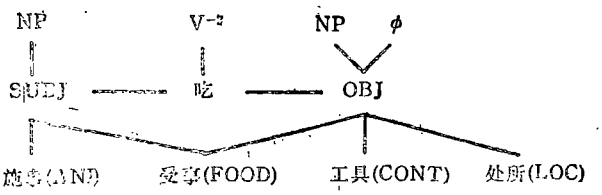

图7 "吃"的词汇网络(φ表示空成分，意味着"吃"的宾语可以省略)

第二种情型的句型为：

$$NP(SUBJ)\ NP(SUBJ)\ VP^{-1}(HEAD) \quad 其中$$

$$VP^{-1}\rightarrow\cdots V^{-1}(HEAD)\cdots$$

如：

(21) 他身体挺棒

这类句子的特点是第一个NP和VP只有间接语义关系，我们把这种NP统称为系事格。为此，我们得出下面的语义规则：

SR_3：

如果　使用了句法规则　$X\rightarrow\cdots NP\cdots VP^0\cdots$

　　则　记下关系　系事=NP

使用SR_1和SR_3可以对这类句子作出正确分析(图8、图9)。

2.兼语句

句型为：NP(SUBJ) VP^{-1}(HEAD)　其中

$$VP^{-1}\rightarrow V^{-3}(HEAD)\ NP(OBJ)\ VP^{-n}(Vcomp),\ n>0$$

图8　短语分析，计算配价

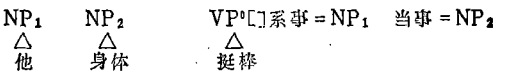

图9　对NP_2及VP应用SR_1，NP_2匹配当事；
对NP_1及VP应用SR_3，确定系事

成分和格角色之间也有多对多关系，如

NP(OBJ)	VP(Vcomp)	例句
受事(施事)	目的	(22) 张老师叫你们就去呢
受事(施事)	原因	(23) 我喜欢这孩子懂事

这类句子的特点是NP(OBJ)兼任V(HEAD)及VP(Vcomp)的格角色。由于NP(OBJ)和VP(Vcomp)间的语义关系已通过VP(Vcomp)的配价反映出来，故词汇网络只需表达这类动词本身所具有的关系就行了(图10)。

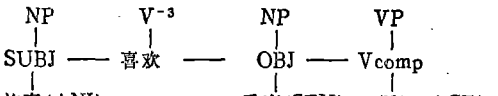

图10　"喜欢"的词汇网络

利用词汇网络和SR_1可以确定NP(OBJ)兼任的格，图11,12是分析(23)的一个片断。

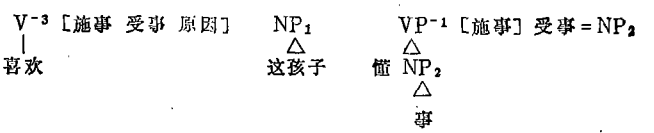

图11　短语分析，计算配价

图12　利用词汇网络，确定V的受事、原因；
对NP_1及VP应用SR_1，VP的施事匹配NP_1，

3.被动句

这里被动句指含有"被"字短语的句子,句型为:

NP(SUBJ) PP VP^{-n}(HEAD) 其中

PP→P(HEAD) (NP) (OBJ) P→被

这种句型的特点是由于"被"字的出现,NP(SUBJ)被强制置为受事格,如:

(24) 他被石头砸了脚 他=受事

(25) 敌人被打败了 敌人=受事

换句话说,"被"消除了NP(SUBJ)和VP(HEAD)之间其它语义关系的可能性,这意味着"被"使VP^{-n}变为VP^0。综合其它关系,可以得到"被"的词汇网络(图13)。

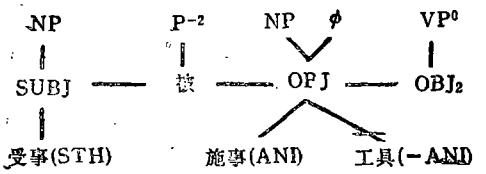

图13 "被"的词汇网络

被动句的分析也需借助词汇网络和语义规则(图14,15)。

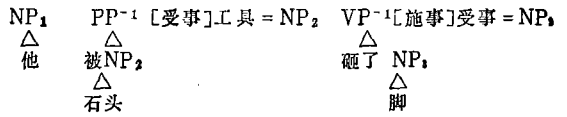

图14 短语分析、计算配价

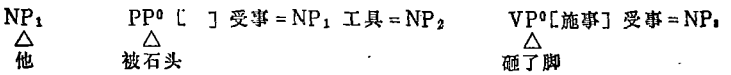

图15 根据词汇网络,置VP的配价为0;
对NP_1和PP应用SR_1,NP_1和受事匹配

由于VP的配价被置为0,故避免了将SR_1用到NP_1和VP上得到错误的结果。

语义规则和词汇网络还可用于分析许多其它的句型,如及物和不及物动词句,连动句、双宾句、把字句,存在句[5]等,配价的思想还可用于分析一些其它的短语,如"的"字结构,的语义关系,相关的讨论见[6]。由于篇幅所限,对这些我们就不作进一步讨论了。

七、分析策略及实例

心理语言学研究表明人在理解语言时能根据已知信息对未知信息作出预期,并以此建立信息之间的联系[7]。对于句子来说,预期的内容就是句法、语义关系,而词成为激活预期的最小单位。因此,一个基于预期的分析模型可以表示为(图16):

这一分析策略能动态地综合利用各种信息完成对句子的多层次分析。

下面是机器分析(26)的运行结果。成分结构由语法树来表示,功能结构和格关系则表示为树节点的特征,特征集中特征的次序为范畴、功能、语义特征或格关系,C_{nn}表示节点编号。

图16　预期分析模型

(26)　他把这个消息告诉了我

C_{74}=〔S (ACT) ((施事 C_{73}) (受事 C_{68}) (对象 C_{61}))〕

C_{73}=〔NP SUBJ (ANI)〕

C_{72}=〔PRO HEAD〕

他

C_{71}=〔PP〕

C_{70}=〔P HEAD〕

把

C_{68}=〔NP OBJ (STH)〕

C_{69}=〔DET〕

这

C_{67}=〔M〕

个

C_{66}=〔NP HEAD〕

C_{65}=〔N HEAD〕

消息

C_{64}=〔VP HEAD〕

C_{63}=〔V HEAD〕

告诉

C_{62}=〔AUX〕

了

C_{61}=〔NP OBJ (ANI)〕

C_{60}=〔PRO HEAD〕

我

八、结束语

本文以人理解语言的特点为基础，针对汉语的问题提出了一个句子分析模型，该模型综合利用句法、语义信息完成对句子的多层次分析，对汉语一些典型句型的应用证明该模型是可行、有效的。汉语是一种十分灵活的语言，从认知和语言两方面作进一步探讨是十分有益的。我们今后的方向是将知识引入模型，由于词汇网络已包含了语义信息，因此很容易将词典扩展为一个知识库，这无疑将会增强模型的分析能力。

参 考 文 献

[1] Jay L.Garfield (ed), Modularity in Knowledge Representation and Natural-Language Understanding, Lawrence Erlbaum Associates Pub., Hillsdale, New Jersey, 1987.
[2] Bresnan J. (ed), The Mental Representation of Grammatical Relations, the MIT Press, 1985.
[3] Jacobson, Pauline & Pullum, Geoffrey K. (ed), The Nature of Syntactic Representation, Dordrecht, D. Reidel Pub. Co., 1982.
[4] Senders J. W., Eye Movement and the Higher Psychology Functions, Lawrence Erlbaum Associates, Pub., Hillsdale, New Jersey, 1978.
[5] 吕叔湘主编，现代汉语八百词，商务印书馆，1984.
[6] 朱德熙，现代汉语语法研究，125—150，1985.
[7] Michael S. K., Psychology of Language, the MIT Press, Cambridge, 1983.

MULTIPLE LEVEL ANALYSIS OF CHINESE SENTENCES

Li Li
Institute of Psychology, Academia Sinica

Abstract

The goal of this paper is to construct a expectation-based computer model in which syntactic and semantic components coordinate their information to facilitate the multiple level analysis of Chinese sentences. For this purpose, we propose some formalism and relevant semantic rules with which semantic relations can be integrated into formal syntax and lexicon. This enables the model to accomplish simultaneous process on different levels. An application to some sentence patterns has shown that the mechanism is feasible.