

LFG中从词汇结构到功能结构的转换^{*}

冯志伟

(教育部语言文字应用研究所 北京 100010)

[摘要] 本文讨论了在计算语言学中有重要影响的词汇功能语法的理论和方法, 着重介绍了成分结构和功能结构的基本概念, 并通过实例着重分析了从成分结构到功能结构转换的具体过程和实现技术。

[关键词] 词汇功能语法; 成分结构; 功能结构

[中图分类号] H08 [文献标识码] A [文章编号] 1003-5397(2004)04-0105-08

Transfer from C-Structure to F-Structure in LFG

Feng Zhiwei

Abstract: The Lexical Functional Grammar (LFG) is an important formal grammar in the computational linguistics. The author discusses the fundamental conceptions of constituent structure (c-structure) and functional structure (f-structure), shows the concrete procedure of transfer from c-structure to f-structure, and analyzes the approaches of this transfer by the examples.

Key words: lexical functional grammar; c-structure; f-structure

词汇功能语法(Lexical Functional Grammar, 简称LFG)是美国语言学家卡普兰(Ronald Kaplan)和布列斯南(Joan Bresnan)于1982年在《词汇功能语法——一个语法表示的形式系统》(Lexical-functional Grammar: A Formal System for Grammatical Representation)一文中提出的。这种语法为自然语言的描述和语法知识的表达提供了一个有效的模式, 它不仅可以解释幼儿的语言习得的机制, 而且还可以解释人类处理自然语言的行为, 从而满足了自然语言计算机处理的需要, 在机器翻译和自然语言处理中得到了广泛的应用。

词汇功能语法主要由词库、句法和语义解释三部分组成。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)。在构成成分结构时, 还附加有功能等式(annotated functional equation)和限制性等式(con-

[收稿日期] 2004-02-01

[作者简介] 冯志伟, 男, 研究员, 博士生导师, 主要从事应用语言学和计算语言学研究。

* 本文得到国家社会科学基金(项目号: 03BY019)和教育部科研项目(项目号: ZD1105-53A)的资助。

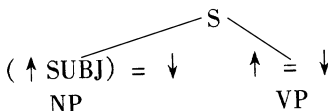
straint equation)。

成分结构是语言的外部结构,它表示句子成分的先后次序,它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,然后进入到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,不同的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,一部分一部分地分别进行描述,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写向功能结构转换,功能描写由一组等式构成,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,便得到了这些语法信息的最终组合结果——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 conditions)。

成分结构是词汇功能语法中句法描写的一个平面。它是由上下文无关的短语结构文法来表示的,它的形式是一般意义上的短语结构树。树形图上的结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:

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

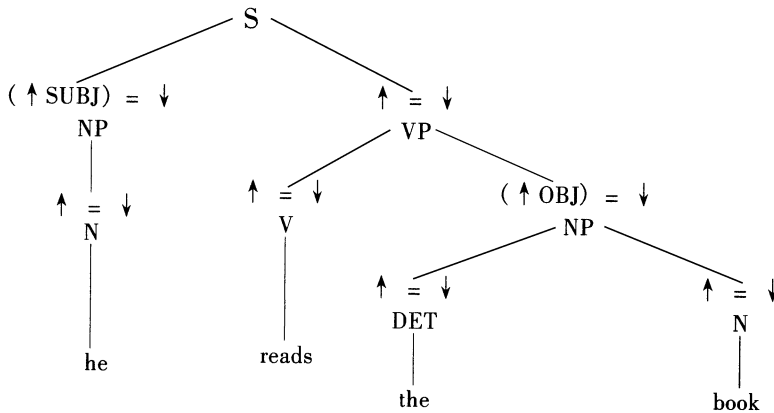
规则中采用了向上单箭标“ \uparrow ”和向下单箭标“ \downarrow ”来表示范畴的支配关系。向上单箭标“ \uparrow ”表示直接支配成分,向下单箭标“ \downarrow ”表示被支配成分,这些箭标可以用在功能等式中,等式左边为限定成分,等式右边为限定值。例如,(\uparrow SUBJ) = \downarrow 读为“直接支配成分的主语等于被直接支配成分的语法功能”, \uparrow = \downarrow 读为“直接支配成分的语法功能等于被直接支配成分的语法功能”。上面的短语结构规则表示句子 S 由 NP 和 VP 组成,VP 前的 NP 是句子的主语。NP 的下方“(\uparrow SUBJ) = \downarrow ”是它的功能等式,表示这个 NP 继承了它的父结点 S 的主语(SUBJ)特征,“(\uparrow 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结点 S 的功能信息。这个短语结构规则也可以用树形图表示如下:



英语句子“ He reads the book ”这个句子的成分结构可以表述如下(见下页)。

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



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。

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,而真正的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。

功能结构是词汇功能语法句法描写的另一个平面。它是一个属性值矩阵 (attribute-value matrix),其基本结构可以表示为:

属性	值
A	a
B	b
C	c

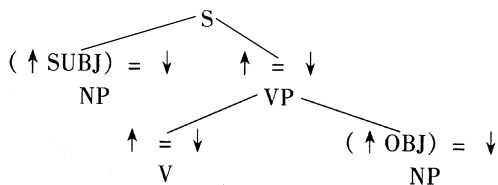
在这个属性矩阵中,第一列 A, B, C 等表示属性,第二列 a, b, c 等表示相应属性所取的值。

英语句子“he reads the book”的功能结构是:

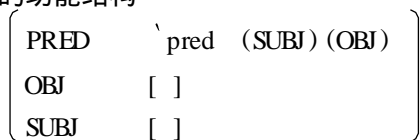
SUBJ	PRED	' he
	ABST	-
	GENDER	MAS
	NUM	SING
	PERS	3
	CASE	NOM
TENSE	PRESENT	
PRED	' read (SUBJ) (OBJ)	
OBJ	SPES	the
	DEF	+
	NUM	SING
	PRED	' book

成分结构和功能结构之间存在着对应关系。这种对应关系,是一个句子的成分结构能够转变为相应的功能结构的基本根据。下面我们来研究成分结构和功能结构的对应关系。

例如,成分结构:



与这个成分结构相对应的功能结构

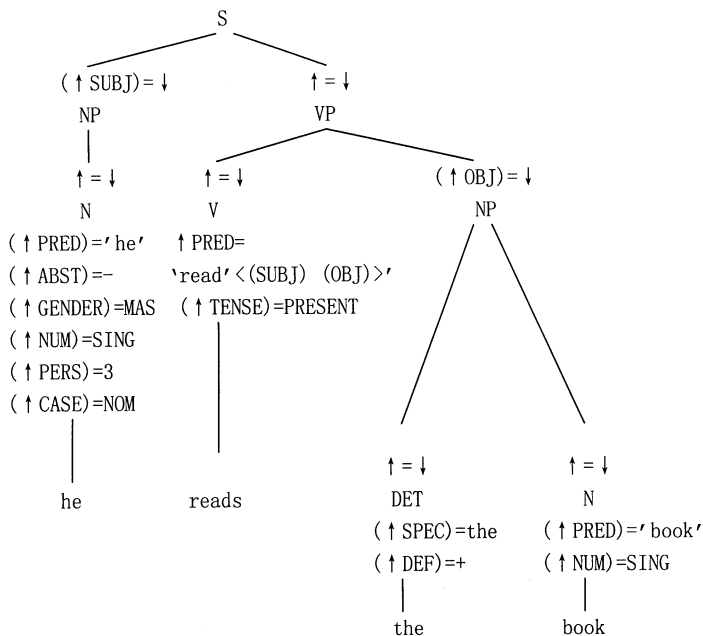


其中 pred 表示某个具体的谓词,如 read 等。

在成分结构与功能结构之间有如下的对应关系 1. 结点 S, VP 和 V, 由于 VP 和 V 都有功能等式 = , 它们是功能中心语, 因而对应于整个的功能结构。2. 带有功能等式 (SUBJ) = 的结点 NP 和带有功能等式 (OBJ) = 的结点 NP, 由于它们不是功能中心语, 它们只能分别对应于上述功能结构中的 SUBJ [] 和 OBJ []。

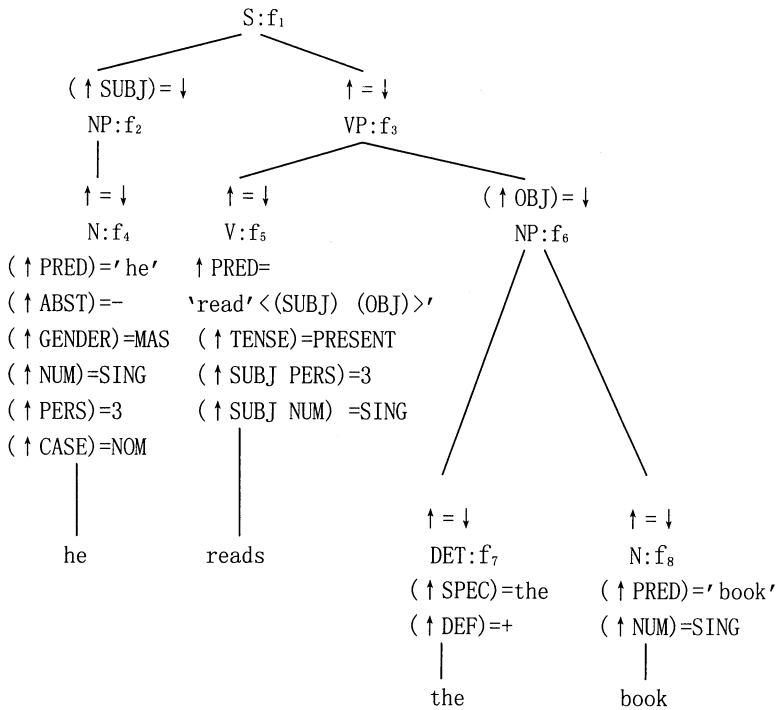
将成分结构转变为功能结构不能直接进行, 而应该通过功能描写 (function description) 这个中介成分来进行。即, 要首先把成分结构转变为功能描述, 再由功能描述转变为功能结构。

从成分结构到功能描写的转变分三个步骤来进行。第一步: 将成分结构进行语法功能编码, 并插入词项。例如, “he reads the book” 这个句子经过第一步后变为:



第二步: 把功能变项 (f₁, f₂, ..., f_{n-1}, f_n) 分配给 S 结点及其他各个附向下向箭标的结点。例如, 上面的树形图中, 对于 S, NP, VP, N, V, NP, DET, N 等结点分别分配功能变项 f₁, f₂, f₃, f₄, f₅, f₆, f₇, f₈ 得到:

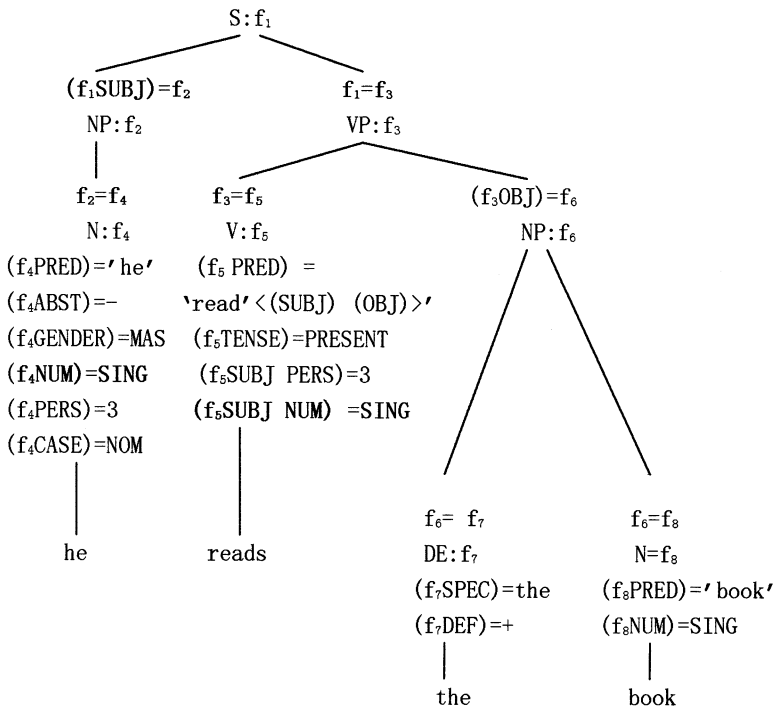
第三步: 用功能变项代替成分结构中的所有的上下箭标, 得到句子的功能描写。例如, 在上面的树形图中, 把各个功能变项分别代入到直接支配它的上箭标以及所有受上箭标支配的下箭标中去。此时, 成分结构中所有功能等式的集合, 就是该句子的功能描写。例如, 在功能



等式 $(\uparrow \text{SUBJ}) = \downarrow$ 中,用上箭标的功能变项 f_1 来代替 \downarrow ,用下箭标的功能变项 f_2 来代替 \uparrow ,得到 $(f_1 \text{SUBJ}) = f_2$;在结点 VP 的功能等式 $\uparrow = \downarrow$ 中,用上箭标的功能变项 f_1 来代替 \downarrow ,用下箭标的功能变项 f_3 来代替 \uparrow ,得到 $f_1 = f_3$,如此等等。这样,我们便可以得到上面树形图的功能描述如下:

- | | |
|---|--|
| a. $(f_1 \text{SUBJ}) = f_2$ | l. $(f_5 \text{TENSE}) = \text{PRESENT}$ |
| b. $f_1 = f_3$ | m. $(f_5 \text{SUBJ PERS}) = 3$ |
| c. $f_2 = f_4$ | n. $(f_5 \text{SUBJ NUM}) = \text{SING}$ |
| d. $(f_4 \text{PRED}) = \text{he}$ | o. $(f_3 \text{OBJ}) = f_6$ |
| e. $(f_4 \text{ABST}) = -$ | p. $f_6 = f_7$ |
| f. $(f_4 \text{GENDER}) = \text{MAS}$ | q. $f_6 = f_8$ |
| g. $(f_4 \text{NUM}) = \text{SING}$ | r. $(f_7 \text{SPEC}) = \text{the}$ |
| h. $(f_4 \text{PERS}) = 3$ | s. $(f_7 \text{DEF}) = +$ |
| i. $(f_4 \text{CASE}) = \text{NOM}$ | t. $(f_8 \text{PRED}) = \text{book}$ |
| j. $f_3 = f_5$ | u. $(f_8 \text{NUM}) = \text{SING}$ |
| k. $(f_5 \text{PRED} = \text{read}) (\text{SUBJ}) (\text{OBJ})$ | |

这时,树形图如下:



功能描写的形式化特点为计算机处理提供了方便。一个句子如果不合语法,就可以很快地通过计算机的运算查出来。比如说,如果我们把树形图中的 he 改成 they,句子就不合乎语法了。因为如果 they 为复数,则它的词项输入的相应内容也应该是复数,也就是应该有 (f₄NUM) = PLUR,这里,PLUR 表示复数。这样,在功能描写中,功能等式 g 的“(f₄NUM) = SING”应该改为“(f₄NUM) = PLUR”。但是,这样一来,就与功能等式 n 的“(f₅SUBJ NUM) = SING”发生了矛盾。因为由 f₁ = f₃和 f₃ = f₅ 可以得出 f₁ = f₅,将 f₅ 替换(f₁ SUBJ) = f₂ 中的 f₁,可以得到 (f₅ SUBJ) = f₂,用 f₂ 来替换“(f₅ SUBJ NUM) = SING”中的 f₅ SUBJ,得到 (f₂ NUM) = SING,而 f₂ = f₄,所以得 (f₄ NUM) = SING,但这时我们已经有了 (f₄ NUM) = PLUR,因此,计算机根据功能一致性的原则,可以判断出将 he 改为 they 是不合乎语法的,很容易就检查出语法错误。这个事实说明,将词汇功能语法作为语言信息处理的一种基本理论是会受到欢迎的。

功能描写既然是成分结构和功能结构之间的中介成分,所以,它可以向功能结构转变。由功能描写到功能结构的转变主要是由定位 (locate) 和合并 (merge) 两个算子 (operator) 来完成的。整个转变过程是一个综合分析的过程。

定位算子首先定出功能描写中功能等式两边的名称在功能结构中所处的位置,然后由合并算子按照功能结构的格式将功能描写等式两边的内容进行横向排列。这样的过程需要不断重复,直到功能描写中的所有的功能等式都计算完毕,并得到功能结构才告结束。

下面简要说明句子“he reads the book”的成分结构到功能结构的转变过程。

第一步:由功能等式 (f₁ SUBJ) = f₂和 f₁ = f₃ 可以推出 :f₁f₃ [SUBJ f₂ _____]

这时, f₁ 与 f₃ 合并, f₂ 定位为 f₁ 和 f₃ 的子功能描述,处于 f₁f₃ 下位。f₂ 后面划线的部分表示有待填入的未知数,也就是限定值,这个值可以通过与 f₂ 和 f₄ 有关的功能等式来确定。

第二步:由功能等式 f₂ = f₄,用 f₄ 的值来填充 f₂ 中的划线部分, f₂ 与 f₄ 合并。f₄ 的这些值

是:

(f₄ PRED) = he (f₄ ABST) = - (f₄ GENDER) = MAS

(f₄ NUM) = SING (f₄ PERS) = 3 (f₄ CASE) = NOM

得到:

$$f_1 \left(\begin{array}{l} \text{SUBJ} \left(\begin{array}{l} \text{PRED} \quad \text{'he'} \\ \text{ABST} \quad - \\ \text{GENDER} \quad \text{MAS} \\ \text{NUM} \quad \text{SING} \\ \text{PERS} \quad 3 \end{array} \right) \\ f_2 \\ f_3 \quad f_4 \quad \text{CASE} \quad \text{NOM} \end{array} \right)$$

式中 f₁ 和 f₃ 合并了, f₂ 和 f₄ 合并了, 故 f₁ 和 f₃ 定位于同一位置, f₂ 和 f₄ 定位于同一位置。

第三步: 由功能等式 f₃ = f₅, 把 f₅ 的值填入 f₃ 中, 这些值由功能等式

(f₅ PRED) = read (SUBJ) (OBJ) (f₅ TENSE) = PRESENT

来确定, 分别定位为 f₃ 的 PRED 和 TENSE。

由功能等式 (f₅ SUBJ PERS) = 3 和 (f₅ SUBJ NUM) = SING, 可把 f₅ 的 SUBJ 的 PERS 和 NUM 的值填入 f₃ (也就是 f₁, 因为它们合并了) 的 SUBJ 的子功能结构中, 而这个 SUBJ 的子功能结构中已经有 PERS 和 NUM, 并且它们的值完全一致, 故可取已有的 PERS 和 NUM 的值来代之。得到如下的功能结构:

$$f_1 \left(\begin{array}{l} \text{TENSE} \quad \text{PRESENT} \\ f_3 \left(\begin{array}{l} \text{SUBJ} \left(\begin{array}{l} \text{PRED} \quad \text{'he'} \\ \text{ABST} \quad - \\ \text{GENDER} \quad \text{MAS} \\ \text{NUM} \quad \text{SING} \\ f_2 \quad \text{PERS} \quad 3 \\ f_4 \quad \text{CASE} \quad \text{NOM} \end{array} \right) \end{array} \right) \\ f_5 \left(\text{PRED} \quad \text{'READ (SUBJ) (OBJ)'} \right) \end{array} \right)$$

第四步: 由功能等式 (f₃ OBJ) = f₆, 把 f₆ 的值填入 f₃ 中, 定位为 f₃ 的 OBJ, 由功能等式 f₆ = f₇ 和 f₆ = f₈, 把 f₇ 和 f₈ 的值填入 f₆ 中, 充当 f₃ 的 OBJ 中的值, 它们由下列功能等式提供:

(f₇ SPEC) = the (f₇ DEF) = + (f₈ PRED) = book (f₈ NUM) = SING,

得到:

$$f_3 \left(\begin{array}{l} \text{OBJ} \left(\begin{array}{l} \text{SPES} \quad \text{the} \\ f_6 \quad \text{DEF} \quad + \\ f_7 \quad \text{NUM} \quad \text{SING} \\ f_8 \quad \text{PRED} \quad \text{'book'} \end{array} \right) \end{array} \right)$$

第五步: 把 OBJ 的值填入 f₃ 中, 得到整个句子的功能结构如下:

	SUBJ	<table style="border-collapse: collapse;"> <tr><td style="padding-right: 10px;">PRED</td><td style="padding-right: 10px;">\</td><td>he</td></tr> <tr><td style="padding-right: 10px;">ABST</td><td style="padding-right: 10px;"></td><td>-</td></tr> <tr><td style="padding-right: 10px;">GENDER</td><td style="padding-right: 10px;"></td><td>MAS</td></tr> <tr><td style="padding-right: 10px;">NUM</td><td style="padding-right: 10px;"></td><td>SING</td></tr> <tr><td style="padding-right: 10px;">f₂ PERS</td><td style="padding-right: 10px;"></td><td>3</td></tr> <tr><td style="padding-right: 10px;">f₄ CASE</td><td style="padding-right: 10px;"></td><td>NOM</td></tr> </table>	PRED	\	he	ABST		-	GENDER		MAS	NUM		SING	f ₂ PERS		3	f ₄ CASE		NOM
PRED	\	he																		
ABST		-																		
GENDER		MAS																		
NUM		SING																		
f ₂ PERS		3																		
f ₄ CASE		NOM																		
TENSE PRESENT																				
PRED \ READ (SUBJ) (OBJ)																				
	OBJ	<table style="border-collapse: collapse;"> <tr><td style="padding-right: 10px;">SPES</td><td style="padding-right: 10px;"></td><td>the</td></tr> <tr><td style="padding-right: 10px;">f₆ DEF</td><td style="padding-right: 10px;"></td><td>+</td></tr> <tr><td style="padding-right: 10px;">f₇ NUM</td><td style="padding-right: 10px;"></td><td>SING</td></tr> <tr><td style="padding-right: 10px;">f₈ PRED</td><td style="padding-right: 10px;"></td><td>\ book</td></tr> </table>	SPES		the	f ₆ DEF		+	f ₇ NUM		SING	f ₈ PRED		\ book						
SPES		the																		
f ₆ DEF		+																		
f ₇ NUM		SING																		
f ₈ PRED		\ book																		
f ₁																				
f ₃																				
f ₅																				

词汇功能语法克服了短语结构语法的生成能力过强而分析能力不足的缺陷,在当代语言学中独树一帜,这是语言信息处理的重要成果之一。

[参考文献]

- [1] Kaplan, R. M., & Bresnan, J. Lexical Functional Grammar: A formal system for grammatical representation[A]. In Bresnan, J. (Ed.), *The Mental Representation of Grammatical Relations*[C], pp. 173—281, MIT Press, Cambridge, MA, USA, 1981.
- [2] 冯志伟. 计算语言学探索[M]. 哈尔滨:黑龙江教育出版社, 2002.
- [3] 冯志伟. 中心语驱动的短语结构语法[A]. 语言学问题集刊(第一辑)[C]. 吉林人民出版社, 2001.
- [4] 俞如珍, 金顺德. 当代西方语法理论[M]. 上海外语教育出版社, 1997.

第二届中国文字学国际学术研讨会在荆门举行

第二届中国文字学国际学术研讨会于6月28日至7月3日在湖北荆门举行。会议由南开大学、湖北大学主办,荆门市政府、荆门市社联、荆门市外国语学校协办。南开大学向光忠教授、湖北大学舒怀教授共同主持了这次学术研讨会。来自海内外的文字学学者近80人与会。

与会学者向大会提交了文字学学术论文80余篇(有的学者因故缺席也提交了论文),涉及到文字学研究的诸多领域,集中展示了近年来文字学研究的新成果。这次学术研讨会,与会学者没有门户之见,相互切磋质疑,展示了朴实的会风。在会中,资深学者鼓励中青年学者发表创见,中青年学者向资深学者虚心学习,表现出了空前团结的局面。会议期间,学者们还参观了荆门博物馆、荆州博物馆、随州博物馆,实地考察了古文物、古文献、古文字,收获颇丰。台湾、香港和国外的学者们说这次研讨会是真正的学术会议,与会学者一致认为这次到荆门参加学术研讨会不虚此行。这次学术研讨会对推动文字学的研究和文字教学具有重要的意义。

在会议的开幕式上,荆门市政府、人大、政协、市委宣传部的领导到会表示了祝贺,并在会议期间多次看望了与会学者。

(赵 铮)