

HNC 理论的语句格式的新进展^{*})

缪建明^{1,2} 张全²

(中国科学院研究生院 北京 100039)¹ (中国科学院声学研究所 北京 100080)²

摘要 为了适应语言中惯常出现的 SVO 表达语序, HNC 理论对原有语句格式代码表(98 版)进行了大幅度的调整, 重点是重新界定基本格式的范围, 扩充基本格式代码, 使所有 SVO 语序的表达都可以用基本格式表示。同时, 为了便于计算机处理和人类专家使用, 新格式代码对代码中的每一数字都做了更加细致的约定, 可以直接从代码得到具体的格式信息, 而不用查阅格式代码表。本文详细说明了新版格式代码的设计思路及其带来的新变化, 指出了新版格式代码与 98 版格式代码表的不同。

关键词 HNC 理论, 语义块, 句类表达式, 语句格式, 语序

The Change of HNC New and Old Sentence Format Code Form

MIAO Jian-Ming^{1,2} ZHANG Quan²

(Graduate School of the Chinese Academy of Science, Beijing 100039)¹

(Institute of Acoustics, Chinese Academy of Science, Beijing 100080)²

Abstract To express the usually word orders in the SVO languages, HNC theory carries on the adjustment by a large margin to old sentence format code form (98 editions). Defining the range of the basic format again, expanding the basic format code, and making all expressions of SVO word order to enable be expressed with the basic form are the focal point of the adjustment. In order to make it convenient for the computer and human experts to deal with, every figure in the new format code make careful agreement. Users can directly receive concrete format information from the code and needn't look for the code from the sentence format code form. Comparing with the 98 editions of sentence format code form, the text explains the way of designing of the new sentence format code and its new following change in detail.

Keywords HNC theory, Semantic chunk, Sentence category (SC) expressions, Sentence format, Word order

1 引言

传统语言学早已发现自然语言有 VO 和 OV 两种基本类型, 统计表明: VO 型占语种总数的 56%, OV 型占 44%^[2]。VO 型语言又以 SVO 型为主, 所占比例接近 63%, 而 VSO 型和 VOS 型仅占 37%^[2]。汉语曾被列为典型的 SVO 语言^[1], 但这是一个误解, 汉语实际上兼具 VO 和 OV 语言的双重特性, S、V、O 之间 6 种可能的排序 SVO、SOV、OVS、VOS、OSV、VSO 汉语都可以采用。而英语可以采用的仅 SVO 和 OVS 两种, 因此 HNC 理论认为单纯从语序的表层划分语言为 VO 和 OV 型也许不是一个合适的描述, 必须配合概念结构才能对语句给出完整合理的描述。

语句的概念结构不仅决定于语义块的类型和个数, 还决定于它们的排序。传统语言学句类研究的范围包括按谓语的结构分类研究和句子的结构方式, 既讨论了句子动词的性质, 又研究了句型成分的组合情况, 两者在一个平面上来研究, 增加了相应的复杂性, 使得在句子谓语特性研究的同时还要考虑句型成分的组合带来的困扰。HNC 理论认为, 应该把语义块的类型和个数(大体相当传统语言学之“句型”)、语义块的排序(大体相当传统语言学之“句式”)放在两个层面上研究, 先研究语义块的类型和个数, 在此基础上结合语义块的排序情况构成相应的句类表示式, 通过两者的组合形式构成概念

结构的完整描述。这样的研究方法把属于语言表达习惯的部分脱离语义块类型研究, 而放在语句格式的研究范围中, 语义块类型研究脱离了语种的特性, 把混杂研究出现的交织型的难度去除, 使得研究起来更加有效。

语句格式(可简称格式)这一术语的引入是 HNC 理论对传统句型、句式术语的继承和发展。格式定位于传统语言学对于语序变化的描述, 这是继承; 但同时格式立足于 HNC 理论的语义块和语句的句类表示式, 这是发展。众所周知, 自然语言中单句存在双 O(双宾)现象, 这时, 如果形式上按 S、V、O1、O2 排序, 则存在 24 种语序。如何对此进行描述呢? 实际上, 自然语言单句不仅存在双宾(双 O)现象, 还存在双 S(双主)和双 V(双谓)现象^[3], 依照传统语言学的 S、V、O 划分, 对此又如何进行描述呢? HNC 理论的处理对策是把 S、V、O 的 6 种可能排序现象和 S、V、O1、O2 的 24 种可能排序都纳入格式代码的范畴进行描述, 而把双 S(双主)和双 V(双谓)现象纳入句类代码的范畴进行描述。语句格式体现的是主语语义块之间的顺序及其组合变化的情况, 具体而言, 体现的是语句的句类表达式的主语语义块的组合顺序及其冗余标记、变格、时态等信息。语句格式应包含各种语言的语言表述习惯, 同时便于区分不同语言的表达习惯和使用者记忆。由于格式概念立足于语句的句类表示式, 而句类表示式的统一数学表示式是参照 SVO 语言的特性表达, 所以语句格式和句

^{*}) 本文承国家 973 项目“自然语音理解的交互引擎研究”(2004CB318104)、中科院声学所知识创新工程项目“HNC 语言知识处理理论及”的资助。缪建明 博士研究生, 研究方向: 自然语言理解处理。张全 研究员, 博士生导师, 研究方向: HNC 自然语言理解处理, 计算机语言学。

类表示式天然具有合理性,并对和句类表示式顺序一致的格式赋予特殊的代码(“! 0”)。

1998年制定的旧版本,在设计上虽然体现了HNC理论把句型和句式放在两个层面上研究、通过两者配合对语言现象进行表述的总体思想,但是编码设计的全面性是靠简单的排列来解决的,没有在全局性的基础上考虑各种语言的语序差异,在设计中明显没能体现出各种语言(特别是典型的几种SVO型自然语言)之间语序的显著特点。同时,语言表层空间的时态、形态、格的变化也没能考虑进去,具体的编码需要通过查表得出,这些成为旧版本存在的问题。随着HNC理论深入到翻译引擎的研究领域,对英语的标注不能不考虑到英语的时态、形态和格变化的问题。如果对这些变化回避处理,对于英汉语言之间差异的体现是明显不够的。而旧版本却由于当时的局限性对此正是采取了回避的策略,这显然不符合新的英语语料标注的发展。为了将新的语言研究成果融入语句格式中,解决旧版本存在的这些问题,进一步配合语句概念结构的研究,新语句格式代码表顺势而生。

2 旧语句格式的设计

HNC语句格式的引入是在语句的句类表示式基础上归纳发展起来的,语句格式的定义要求参照句类表示式的语义块顺序。GBK1大体对应主语S,GBK2和GBK3大体对应宾语O。HNC理论根据特征语义块EK体现作用效应链某一个环节还是某几个环节可把句类分为基本句类和混合句类。基本句类和混合句类的语句,其统一数学表达式为:

$$SCn = GBK_i + (EK) + GBK_m (m = 2 \sim 4)$$

表达式中GBK表示广义对象语义块,EK表示特征语义块,SC表示语句的数学表达式,GBK的数字后缀*i*表示广义对象语义块的序号,而左边符号SC的第一个数字后缀*n*表示语句的主语义块个数。特征语义块EK带有括号表示它可以不存在,这意味着句类中存在无特征语义块的特殊句类。在此基础上,考察语义块排列顺序的各种可能变化后设计出了基本、规范、违例、省略4种格式。

2.1 基本格式

定义为主语义块顺序与语句的句类表示式的顺序相同,不允许广义对象语义块位置任意变动。由于当时还没有考虑时态、形态和格变化等信息,基本格式仅有一种,用! 0格式代表。旧基本格式的设计主要考虑语句格式代码表达式同句类表示式之间语义块顺序的一致性问题,这种设计方便了在句类表示式的基础上考虑语义块的语序和变化问题,便于把格式与句类表示式的顺序结合起来通过对比来研究格式变化。

看下面的例子(使用到的标记符号意义参见附录):

例1 我们党 || 已成为 || (在全国执政 | 五十多年 &、拥有 | 六千四百多万党员 | 的大党)。
YB1 Y02
YB2

Y02J = YB1 + Y02 + YB2 (注:基本格式! 0表示可省略不写)

例2 中国人民 || 已拥有 || 一个欣欣向荣的社会主义祖国。
RB1 R611 RB2

$$R611J = RB1 + R611 + RB2$$

例3 这个巨大变化 ||, 是 || (中华民族 | 发展) 的一个历史奇迹 /。
DB jDJ DC

$$jDJ = DB + jDJ + DC$$

三个例子是不同句类(双对象效应句、单向关系句、是否判断句)的句类表示式在! 0格式下形成的句子。

2.2 规范格式

定义为主语义块顺序偏离语句的数学表示式,但在相邻GBK之间一律加上语义块标记的语句格式。规范格式的定义体现了语言表达习惯中多个广义对象语义块之间存在边界模糊、天然要求加上边界分隔标记的语言自然去模糊的特性。

旧规范格式约定形式为! 1mn,其中*m*、*n*的顺序大致按照从小到大的自然编码顺序编码。即:对于3块句,*n*自动消失,*m*的不同取值表示3主语义块的五种排列方式:

! 11 即 *m*=1 表示 GBK1 在第一位,GBK2 在第二位, EK 在第三位。

! 12 即 *m*=2 表示 GBK2 在第一位,GBK1 在第二位, EK 在第三位。

! 13 即 *m*=3 表示 GBK2 在第一位, EK 在第二位, GBK1 在第三位。

! 14 即 *m*=4 表示 EK 在第一位,GBK2 在第二位, GBK1 在第三位。

! 15 即 *m*=5 表示 EK 在第一位,GBK1 在第二位, GBK2 在第三位。

对于4块句,*m*的不同取值表示第一个主语义块GBK的编号,其余主语义块的排列顺序靠*n*来决定:

! 11 即 *m*=1 表示 GBK1 在第一位

! 12 即 *m*=2 表示 GBK2 在第一位

! 13 即 *m*=3 表示 GBK3 在第一位

! 14 即 *m*=4 表示 EK 在第一位

*n*用来补充表示4主块句在*m*确定后其余3主语义块的排列顺序(*n*取值1~6),*n*的不同取值按照从小到大的自然顺序排列其它的6种排列方式^[4]。

看下面的例子:

例4 明天下午的桥牌比赛 || 由张三 || 通知了 || 李四。
T3C TA T3 TB

$$! 133T3J = T3C + TA + T3 + TB$$

例5 张先生 || 送 || 玫瑰花 || 给李小姐。
TA T TC TB

$$! 112T2J = TA + T2 + TC + TB$$

例6 The invasion by Western powers || gradually reduced
A XY02
|| China || to the status of semi-colonial (and semi-feudal society.
YB1 YB2

$$! 111XY02 * 311J = A + XY02 + YB1 + YB2$$

例7 In his War Plans cubbyhole ~ || ~ that afternoon
Cn Cn
~ ||, he || had received || an unexpected word || from on
TB T10 T1C 119
high
TB1

$$! 112T10J = TB + T10 + T1C + TB1$$

2.3 违例格式

定义为主语义块顺序偏离语句的数学表示式同时在相邻GBK之间全部或部分未加语义块边界标记的格式。违例格式体现了自然语言交流过程中交流双方在表达对象清楚的情况下省略标记的语言特性。

旧违例格式的约定形式为! 2mnp,其中的*m*和*n*的约定同规范格式中*m*和*n*的约定^[4]。增加的*p*表示语义块标记

缺省的各种情况:

$p=0$ 语义块标记全部缺省(实际情况下 0 省略不写)

$p=1$ 第一个语义块标记缺省

$p=2$ 第二个语义块标记缺省

看下面的例句:

例 8 红旗 || 漫卷 || 西风。

B X A

! 20XJ=B+X+A

例 9 <打了|李四|的张三>跑了。

<X B A>

! 24XJ= X+B+A (这种句式只存在于 GBK1 要素句蜕中)

例 10 <What | the Chinese people | see > || is || a
<T1C TB1 T1>

prosperous socialist motherland.

要素句蜕中;! 22T1J=T1C+TB1+T1

2.4 省略格式

定义为省去语句数学表示式中某个或某些主语语义块后的格式,它以! 3 开头。省略格式必须和基本、规范与违例三大格式搭配使用,在出现省略主语语义块的时候才会出现,用来标识省略是在哪种格式下发生。省略格式不同于违例格式中边界标记的省略,它是对句类表示式中主语语义块的整体省略,这种情况多见于句群中小句对前面小句中主语语义块省略的情况。

旧省略格式的约定形式为! 3m,m 的不同取值表示省略了不同的语义块(张全、萧国政,2001):

$m=0$ 省略了特征语义块

$m=1$ 省略了 GBK1

$m=2$ 省略了 GBK2

$m=3$ 省略了 GBK3

看下面的例子:

例 11 {没有|共产党} || , {就没有|新中国}。

PBC1 PBC2

! 30P21J=PBC1+(P21)+PBC2 (省略特征语义块 P21)

例 12 敏锐地把握 || \<我国社会生产力|的发展>趋势

D01 DBC

Δ 和要求/。

! 31D01J=(DA)+D01+DBC(省略作用者语义块 DA)

例 13 Marx, Engels, Lenin Δand Comrades [Mao Zedong Δ and Deng Xiaoping] || were all || \ shining examples of { integrating | theory | with practice } Δand of { making |
{R110 RB1 RB2}
theoretical innovations }/。

原型句蜕中;! 31111XR110 * 311J=(A)+R110J+RB1+RB2 (省略作用者语义块 A)

3 新格式的设计

在参考了旧版语句格式表之后,在考虑了 SVO 型语言的语种特性和语言“时态、形态、格”变化的表达习惯后,本着语言现象描述的便利性、基本格式与 SVO 语言的联系、SVO 语言的天然合理性、格式的数学穷尽性与物理实在性(即语种个性)、便于记忆这 5 个具体的设计原则,在旧版基础上,进行了新格式的设计。

3.1 基本格式

定义为语句数学表示式中特征语义块 EK 处于第二位的语句格式,它以字母! 0 开头。新定义允许广义对象语义块

位置任意变动,变动最大之处在于将原有的基本格式的要求放宽到“特征语义块 EK 处于第二位”,允许广义对象语义块位置变动。新基本格式把作用句的效应表达这一汉语语言习惯表达从省略格式中脱离出来,用“! 02”和“! 03”格式表示,这种表达实际是对作用者 GBK1 一种隐含的表达,不能简单地用省略来替代处理,应作为特殊的一类基本格式。

新基本格式编码规则可归纳为以下三条:

1) 由于! 01 格式的含义与! 0 格式相同,因此取消了! 01 格式。! 011 表示 4 块句特征语义块后的 GBK 位置进行了交换,即形成“GBK3+GBK2”的顺序。

2) ! 02 除表示 GBK2 位于特征语义块左侧外,还表示特征语义块做了变形(如英语中采用的被动形式),并且约定 GBK1 不出现,即无需恢复 GBK1(对于省略,在分析时需要恢复 GBK1)。! 021 表示 GBK2 和 GBK3 向左移位,GBK2 移到 EK 前,GBK1 循环成为最右侧的 GBK。! 022 只对 4 块句有效,表示特征语义块右侧的 GBK 位置进行了交换,即形成“GBK1+GBK3”的顺序。

3) ! 03 的约定与! 02 的约定类似,除表示 GBK3 位于特征语义块左侧外,还表示特征语义块做了变形,同时 GBK1 不出现。! 031 表示 GBK2 循环成为最右侧的 GBK。! 032 只对 4 块句有效,表示特征语义块右侧的 GBK 位置进行了交换,即形成“GBK2+GBK1”的顺序。

完成主语语义块位置的描述之后,需要根据冗余标记出现的情况和 GBK 变格的情况,添加相应的冗余代码和变格代码。

4) 主语语义块的冗余标记信息分别用 7、8、9 表示 GBK1、GBK2 和 GBK3 前有冗余信息;

5) 主语语义块的变格信息,分别用 01、02、03 表示 GBK1、GBK2 和 GBK3 变格。

6) 反格式符号“~”放在格式符号“!”之前,表示两块句和 3 块句中语义块的顺序正好相反,四块句中不存在反格式符号。看下面例句(注:为了对比说明方便,例句同旧格式的例句):

例 1 我们党 || 已成为 || <在全国执政|五十多年 &、拥有|六千四百多万党员|的大党>。

Y02J=YB1+Y02+YB2

例 2 中国人民 || 已拥有 || 一个欣欣向荣的社会主义祖国。

R611J=RB1+R611+RB2

例 3 这个巨大变化 || , 是 || \<中华民族|发展>的一个历史奇迹/。

jDJ=DB+jDJ+DC

例 4 张先生 || 送 || 玫瑰花 || 给 李小姐。

! 0118T2J=TA+T2+TC+TB

例 5 The invasion by Western powers || gradually reduced ||
A XY02
China || to the status of semicolonial (and semi-feudal society.
YB1 YB2

! 09XY02 * 311J=A+XY02+YB1+YB2

例 6 In his War Plans cubbyhole ~ || ~ that afternoon ~ || , he || had received || an unexpected word || from on high.

! 0118T10J=TB+T10+T1C+TB1

例 7 红旗 || 漫卷 || 西风。

! 0XJ= B+X+A

3.2 规范格式

定义为特征语义块不处于第二位,但在相邻 GBK 之间一律加上语义块标记的语句格式,它以字母! 1 开头。旧版本定义为主语义块顺序偏离语句的数学表示式,但在相邻 GBK 之间一律加上语义块标记的语句格式。

新规范格式中约定为! $1mnp$,其中 m 表示特征语义块从第二个语义块位置移动的情况;向右移动直接写出移动的位数,向左移动用“4”表示; n 表示第一个出现的广义对象语义块是第 n 个广义对象语义块; p 只出现于 4 块句中,1 表示第二、三个出现的广义对象语义块是按照自然顺序,即 GBK 脚标小的语义块在前,2 表示第二、三个出现的广义对象语义块交换了位置。新规范格式完成后,在汉语中广义效应句不会出现规范格式,同时规范格式更便于使用者记忆。

看下面例句:

例 4 明天下午的桥牌比赛 || 由 张三 || 通知了 || 李四。
! 1131T3J=T3C+TA+T3+TB

3.3 违例格式

定义为特征语义块不处于第二位,而且在相邻 GBK 之间一律未加语义块边界标记的格式,它以! 2 开头。新违例格式除了新基本格式和新规范格式外,在语义块不省略的情况下,其余的格式就是违例格式了。违例格式的特征是特征语义块不位于第二个语义块的位置,同时广义对象语义块之间的语义块标记信息不全或没有。

新规范格式约定为! $2mnp$, m 、 n 、 p 的含义与新规范格式中的相同。另外在违例格式中,对带有语义块标记信息的广义对象语义块,按照新基本格式中冗余和变格信息编码的约定相应给出。

看下面例句:

例 9 〈打了|李四|的张三〉跑了。
! 242XJ= X+B+A(这种句式只存在于 GBK1 要素句蛻中)

例 10 〈What | the Chinese people | see 〉 || is || a prosperous socialist motherland.

要素句蛻中;! 212T1J=T1C+TB1+T1

3.4 省略格式

新旧省略格式的定义和编码都没有改动,定义为省去了语句数学表达式中某个或某些应有的主语义块后的格式,以! $3m$ 开头,新旧省略格式的不同在于搭配其后的另外 3 种格式的编码不同。

看下面例句:

例 11 {没有|共产党 } || ,{就没有|新中国}。

! 30P21J=PBC1+(P21)+PBC2(省略特征语义块 P21)

例 12 敏锐地把握 || 〈我国社会生产力|的发展〉趋势 & 和要求/。

! 31D01J=(DA)+D01+DBC(省略作用者语义块 DA)

例 13 Marx, Engels, Lenin A and Comrades [Mao Zedong A and Deng Xiaoping] || were all || \ shining examples of { integrating | theory | with practice } & and of { making | theoretical innovations } /。

原型句蛻中;! 31! 09XR110 * 311J=(A)+R110J+RB1+RB2(省略作用者语义块 A)

4 新旧格式定义的不同

旧格式设计的编码主要在自然编码顺序的基础上进行,

格式的具体编码通过死记硬背后习惯后熟悉,同时对英语语法范畴“格、时、体、态、式”的各种习惯表达置之不理,未予考虑。改进后的新语句格式对代码选取具有严格的要求,含有确定的设计原则。首先在数字串的结构划分上可严格分为基本位和延伸位,基本位数字串最多 4 位,表示主语义块之间的位置信息,延伸位又分为冗余标记延伸位和变格延伸位,冗余标记延伸位表示基本格式和违例格式中对广义对象语义块前的冗余标记的信息,变格延伸位表示广义对象语义块的变格信息(注:冗余标记延伸位不用于规范格式,因为规范格式要求两个 GBK 之间必须加入标记,但也存在冗余情况,例如对第一个语义块加上标记)。其次新语句格式的编码更体现便于记忆的宗旨,使用者不需死记硬背,只需要记住编码设计的规则,即可以写出语句的相应格式代码,这一编码设计更符合人的记忆特点,克服了旧版本编码自然排序、难于记忆的缺点。以下将说明 4 种格式新旧版本定义和编码的不同。

4.1 基本格式

旧基本格式就是各主语义块严格遵循物理表达式约定顺序的格式。该格式是对语义块顺序信息的充分利用,其基本特征是不对任何语义块给出任何标记信息,因为它认为位置就包含了所需要的全部信息,代表了各语义块的语义角色。对于 4 块句 GBK1+EK+GBK2+GBK3 中 GBK2 和 GBK3 之间、两块句 GBK1+GBK2 之间存在的边界模糊,旧基本格式则采取“蛮不讲理”的态度,对此置之不理。对于印欧语系格、时态的问题,基本格式也采取“置之不理”的态度,例如,英语对“p400e31-0”有 I 和 me 之分,而汉语用一个“我”全代表了,甚至还扩展到代表 my 和 our,这些在旧基本格式中都未予考虑。

新基本格式定义为特征语义块 EK 处于第二位的语句格式,它以字母! 0 开头。新定义允许广义对象语义块位置任意变动,变动最大之处在于将原有的基本格式的要求放宽到“特征语义块 EK 处于第二位”,允许广义对象语义块位置变动。新定义又充分考虑了印欧语言的时态、变格的语言习惯,把涉及到语言表层结构的语言习惯留在格式的设计中,克服了原版本简略处理这一语言习惯的不足,更利于提高计算机自动处理的能力。

从例句中可看出除了原基本格式的例 1、例 2、例 3 保留之外,新基本格式吸收了原有规范格式(如例 5、例 6 和例 7)和违例格式(如例 8)的部分结构。对“全局特征语义块 EK 必须位于第二位”的严格要求使得汉英两种语言在格式中存在很大的差别,英语的语句一级的格式只会出现基本格式,不出现规范格式,这一语序差别格式设计上的体现有利于计算机的自动识别。同时在格式的设计中添加了“7、8、9”的格变化的信息,更丰富了英语“时态、形态和格变化”的表达习惯。

4.2 规范格式

按照传统语言学的观点,汉语表达中在 S、V、O 可能的 6 种排序在汉语中都能出现,英语却只能出现 SVO 语序和 OVS 语序。两种语言这一截然不同的不同来源于汉语偏重逻辑关系,英语则比较强调语言形式。把特征语义块位于第二位的格式从旧规范格式中脱离出去以后,纯净了汉语语种的特性,使得英语语句一级的格式不会出现规范格式,而汉语中广义效应句不会出现规范格式,便于从格式上区分两种语言的不同表达习惯,同时规范格式按照新的编码原则,更便于使用者记忆。

从例句中可以看出,两个英语的例子都进入了新的基本

格式中,旧版本中的部分格式调整到了新基本格式中,规范格式的覆盖范围更小了,但是语种特性却更加突出了。

4.3 违例格式

违例格式的设计体现了语言交流过程中交流双方在相邻 GBK 边界清楚的情况下省略标记的语言特性。虽然这种语言特性体现了自然语言天然抽象表达的特性,对人的交流不会存在困难,但是对于计算机理解自然语言来说则是一个大的困扰,好在标记的省略只特定于部分的特征语义块或者广义对象语义块中,只要根据新定义的违例格式知识结合相关的句类知识后就有效地解决这一困扰。新版本违例格式的编码在基本格式定义变化后进行缩减,采用了编码更易于记忆的设计原则,本质定义上同旧版本违例格式未有太大的变化。

从例句中可以看出,这种结构更多见于古汉语的习惯表达和句脱一级的局部表达习惯中,英汉都存在,但语句一级的表达不常见。

4.4 省略格式

省略格式不同于违例格式中边界标记的省略,它是对数学表示式中主语义块的整体省略,这种情况多见于句群中小句对前面小句中主语义块省略的情况。新旧版本省略格式的定义没有发生改变,具体的格式编码随着前 3 种格式的变化而相应变化。

例句中没有发生变化,只是在具体编码过程中,由于例 13 原是在规范格式基础上省略变为了在基本格式基础上变化,代码发生了相应的变化(! 111->! 09)。

5 新格式存在的问题

由于旧格式对语义块“时态、形态和格”变化采取置之不理的态度,可以很容易列出 2 块句、3 块句甚至 4 块句的排列中统计出格式的完备集合。新格式充分考虑语义块“时态、形态和格”变化,造成数学排列列出格式完备集合总数巨大,并且这也违背了自然语言格式分布不均的使用习惯特点,所以

新格式一般不列出代码表的完备集合,仅列出了部分常见的编码,具体代码由使用者在规则基础上自行完成。

新格式代码在旧基本格式基础上融入了大量旧规范和违例格式,虽然从语言表达习惯上强烈体现了汉英的语种特性,但是显然减弱了旧基本格式与句类表示式的天然联系性,导致基本格式! 0 作为英汉两种语言最习惯采用的格式的突出性明显减弱,这不能不说是由“鱼与熊掌不可兼得”所带来的矛盾。

结语 HNC 理论是一个描述语言概念空间的符号体系,这个符号体系的一部分就是句类基元符号体系,可大体对应着自然语言中的语句,HNC 理论通过格式代码和句类代码相互配合能对语言空间中的各种复杂语言现象进行解释。语句格式用来表述语句中主语义块在句中的出现顺序。由于研究的深入,进一步加深了对句类和格式的理解,格式定义和编码也相应地发生了新的变化。变化后的新格式突出体现了以汉英为代表的语言的习惯表达、语法范畴“格、时、体、态、式”的综合运用。本文就语句格式的变化进行了系统的说明,详细阐述了新旧格式之间定义和编码的差异,从而有利于新格式的全面推广。不当之处还请读者不吝指出。

附录 相关的标记符号:

|| 语义块边界标记符;| 句脱中语义块的间隔符号;
{ } 原型句脱标记符;< > 要素句脱标记符;\ / 包装句脱标记符;^ 语义块指示符;& 并联连接符。

参考文献

- 1 胡树鲜. 现代汉语语法理论初探. 中国人民大学出版社,1990
- 2 黄曾阳. HNC(概念层次网络)理论. 北京:清华大学出版社,1998
- 3 黄曾阳. 语言概念空间的基本定理和数学物理表达式. 北京:海洋出版社,2004
- 4 张全,萧国政. HNC 与语言学研究. 武汉:武汉理工大学出版社,2001

(上接第 161 页)

验证性质 $V \diamond (S6, S9)$,先转换为公式 $V(\text{true } u(S6, S9))$,根据模型检测算法,性质成立。

结论 把 UML-Statecharts 用于工作流的建模和验证已经成为当今的趋势,目前已经提出了它的基于 Statemate 的操作语义和基于步(Step)的操作语义,成为工作流验证的基础。本文得出了用 UML-Statecharts 建立的工作流模型是完全的当且仅当根据步语义生成的全局可达状态迁移图是完全的结论,这样简化了工作流完全性的验证。并且在工作流语义相关性验证中,使用了模型检测理论,它通过模型检测算法自动地验证时序性质,使验证过程有效而快速。但是 UML-Statecharts 的分析技术还不够成熟和完善,需要我们不断改进和发展,使其更有效地用于工作流的建模和验证。

参考文献

- 1 Hollingsworth D. Workflow Management Coalition-The Workflow Reference Model [M]. Workflow Management Coalition,

- 1995
- 2 Verbeek H M W, et al. Diagnosing Workflow Process Using Woflan [J]. Computer Journal,2001,44(4):246~279
- 3 van der Aalst W M P. Workflow Verification; Finding Control-Flow Errors Using Petri-Net-Based Techniques. In: van der Aalst W, et al, eds. Business Process Management, LNCS1806, Springer-Verlag,2000. 161~183
- 4 Oren E, Haller A. Formal Frameworks for Workflow Modelling. In:DERI-Digital Enterprise Research Institute [R],2005. 1~20
- 5 Bhaduri P, Remash S. Model Checking of Statechart Models-Survey and Research Directions [J]. CoRR cs. SE/0407038,2004
- 6 Latella D, et al. Automatic Verification of a Behavioural Subset of UML Statechart Digrams Using the SPIN Model-checker. Formal Aspects of Computing,1999,11:637~664
- 7 Dong W, et al. Model Checking of UML State Machines [J]. Computer Engineering and Science,2001,23(6):7~11
- 8 Alur R,Courcoubetis C, Dill D L. Model-checking in dense real-time [J]. Information and Computation,1995,104:2~34