

论旨网络的描写和 HNC 句类表示的比较分析¹

李千驹

唐兴全

林杏光

北京青年政治学院

北京语言大学人文学院

中国人民大学对外语言文化学院

qianju_li@163.com

tang_xq@263.net

摘要: 动词的语义框架研究是汉语信息处理研究中的一个热点。论旨网络和 HNC (概念层次网络) 理论的句类都是对句子语义框架研究的有益尝试。本文就论旨角色与 HNC 理论中的语义块, 论旨网络描述的主要信息论旨模式 (格框架) 与 HNC 句类的分类标准等问题, 进行了建设性的比较分析。

关键词: 自然语言处理 语义框架 论旨网络 HNC 句类

Comparison and Analysis of the Study of the Case Frame about Chinese Verbs and the Sentence-Category of HNC

Li Qianju

Tang Xingquan

Lin Xingguang

Beijing youth politics college

Beijing language and culture university

Renmin university of China

Abstract: The study of the Semantic Frame about verbs is the focus of Chinese Information Processing. As far as the pattern of Semantic Expression concerned, the Case Frame and the sentence-category of the theory of HNC both are instructive study. In the paper, we compare the Case with the Semantic Chunk, expound the classified standards of the Case Frame and the Sentence-Category of HNC.

Key words: Natural Language Processing, the Semantic Frame, the Case Frame, the Sentence-Category of HNC

1 引言

要让计算机真正理解自然语言的语句仅靠语法知识是不够的, 还必须进行句子语义框架的探索。在语句的语义表述模式上, 有用论旨网络来描写汉语动词的组合框架的, 如格框架; 有通过演绎的方法得到若干语句联想脉络并对此进行表示的, 如 HNC (概念层次网络) 理论的句类。动词格框架和 HNC 句类都是对句子语义框架研究的有益尝试。

论旨网络是述语动词四种句法属性的整合, 包括述语动词的论元属性 (指带几个必有论元)、论旨属性 (论旨角色及其语义约束)、论元的范畴属性 (论旨角色的语法类) 以及论旨角色的句法功能 (主语、宾语、补语等)。论旨网络是一种语义网络, 用来描述述语动词的语义组合关系。述语动词格框架研究就是用论旨网络的方法对述语动词语义组合关系进行描写的。近 10 多年来, 研究动词格框架的不止一家, 林杏光、鲁川二位先生堪称这一方面研究的集大成者。本文中所述的论旨网络的描写就是指中国人民大学的林杏光先生研究得到的汉语动词格框架, 具体成果可见林先生的《词汇语义和计算语言学》一书²。动词

¹本文得到国家社科“九五”重大项目“信息处理用现代汉语词汇研究”之“现代汉语述语动词机器词典的扩展和槽关系研究”以及国家重点基础研究发展规划 (973) 项目 G1998030506 的资助

²由于林杏光先生的《词汇语义和计算语言学》一书中大多采用“格框架”“格”等术语, 我们在本文中出于对原文的尊重仍沿用这些术语以对应于“论旨网络”“论旨角色”等。

格框架还有了工程性描写的重要成果，由清华大学和中国人民大学共同研制的《现代汉语述语动词机器词典》就是运用格框架对 4500 个汉语动词所进行的工程性描写的成果汇编。国内外做动词格框架研究的比较多，但做工程性描写的则少见，因此上述词典的建立具有现实的开拓性。

HNC 理论为自然语言信息处理的语义研究拓展了新的思路，它在自然语言词汇和语句两个层面分别建立了比较完善的表述模式。HNC 句类就是其语句表述模式的最重要组成部分，它所描述的同样是动词与周围名词等其他成分的语义组合关系，是对句子层面概念联想脉络的表述。

本文试对动词格框架和 HNC 句类进行比较，旨在促进二者在相关问题上的交流。

2 格框架和 HNC 句类的定义

格框架研究是在格语法理论的启发下进行的。格语法是美国语言学家菲尔墨在 60 年代中期提出的着重探讨句法结构与语义之间关系的一种语法理论和语义学理论，它主要研究句子的语义平面。格是指在句子中和动词有及物性关系的体词或体词性结构。比如“果子压断了树枝”，其中“果子”是施事格，树枝是受事格。这种及物性语义关系称为格关系。格框架是述语动词同其周围必有名词成分的语义组合模式，是基于格关系对句子作出的语义描述。比如动词“打倒”的格框架是：

施事+动词+受事

格关系工程研究课题组通过对 3000 个动词的考察与分析，把 3000 个动词的格框架总结归纳为 58 小类，即在占有语言材料的基础上认为现代汉语中有 58 个述语动词格框架。

HNC 理论在语句层面的表述模式主要包括语义块和句类两部分。语义块是句子的语义构成单位，是句子的下一级构成单位。句类是句子的语义分类，是指句子层面概念联想脉络的描述，而不是语言学中所说的句型。句类可能由句子中的谓语动词决定，但不尽然。语义块与语义块的组合构成句类表示式。比如同样是动词“打倒”，“打倒”做谓语动词，形成基本作用句，句类表示式是：

$XJ=A+X+B$

其中 A 表示作用者，B 表示作用对象。HNC 理论研究得到了现代汉语中的 57 个基本句类。

可以看出，格框架和句类都是对自然语言语句深层结构的描述，都是设法把句子深层的语义平面变成可以解读的显性的表示。格框架和句类的下一级构成单位都是语义单位：语义格和语义块，二者研究得出的格框架和句类的数量（格框架 58，句类 57）也差不多。那格框架和句类是不是就是差不多甚至是重复的研究呢？答案是否定的。

3 格与语义块

格与述语动词的组合形成格框架，语义块之间的组合形成句类表示式，因此，格与语义块分别是格框架和句类的基本构成单位，二者的不同意味着格框架和句类在语句理解方式上的差异。

3.1 格框架中的“格”

格是指和句子中的述语动词有及物性语义关系的体词或体词性结构，是句子的下一级语义构成单位，相当于论旨网格中的论旨角色。根据格与动词的关系，格可分为三种类型：必须格、可选格、禁用格。描述某个动词的格关系特征必不可少的格叫做必须格，相当于论旨网格中的必有论元，必须格的有无影响句子的语义自足性。如：“我削了个苹果”，施事“我”和受事“苹果”是必须格。可选格是那些虽然能与动词搭配，但缺少时并不影响语义自足性的格。如“我用刀子削了个苹果”，工具格“刀子”是可选格。某个动词的禁用格是指不能和该动词组合的格。

格语法的这种格是一切语言的普遍现象。一经菲尔墨提出，就引起全世界不同语种不同国别人们的广泛兴趣，但随后人们就发现，不管用格语法分析哪种语言，关于格，有两个关键问题没有解决：一，应该分多少格？二，如何划分必须格和可选格？针对这两个问题，林、鲁进行过建设性的探讨，指出：“重要的不是说定有多少个格，而是要找到一个划分格的正确方法，我们在这个问题上取得了新的进展，这就是提出了层层控制和层层推进的划分方法……使用者可以根据自己的研究目的选取适合自己需要的层次”。现代汉语格系统分三个层次，共有 22 个格，找到了划分格的正确途径。对如何确定必须格，林、鲁根据前人和时贤的研究成果并结合自己的研究实践总结出了三种方法：主体客体法、加“的”转指法、最小自足主谓结构法。这种探讨把汉语格关系的研究推进了一大步。

3.2 句类表示式中的语义块

对于格语法无法解决的上述两个问题，HNC 理论提出了语义块的概念试图加以解决。

3.2.1 什么是语义块

语义块是句子的语义构成单位，也是理解的基本单位。从对句子的语义构成的贡献的角度，语义块可分为主语义块和辅语义块两种类型。主语义块是构成完整句义的必不可少的成分，辅语义块是句义中可有可无的成分。主语义块分为特征语义块、作用者语义块、对象语义块和内容语义块，分别用 E、A、B、C 表示，后三者称为广义对象语义块，简记为 JK。比如：“我用刀子削了个苹果”，“我”、“个苹果”是广义对象语义块，“用刀子”是工具辅块，“削了”是特征语义块。语义块与语义块的组合构成句类表示式。

3.2.2 语义块是句类的函数

“语义块是句类的函数”，是 HNC 理论关于语义块和句类的基本论点。其基本含义是：特征语义块 E 和广义对象语义块 JK 在一个句子中的有无、个数和具体内涵随句类的不同而不同。一个句类可以含有两个语义块，也可以含有三个语义块、四个语义块，在两个语义块构成的句类中，有不含特征语义块而含两个广义对象语义块的情况。用 JK1、JK2、JK3 分别表示句子中第一、第二和第三个广义对象语义块 JK，两主块句、三主块句和四主块句分别表示为：

JK+E; JK1+JK2 (两块句)

JK1+E+JK2 (三块句)

其中,特征语义块 E 决定句类,广义对象语义块 JK 的内涵随句类的不同而不同。如 A 在基本作用句中是产生作用者,在转移句中则是转移的发出者。可见,ECBA 的确切内涵要根据具体句类来确定。针对某一具体的句类,从语义上考虑,它所必需的主语义块的数量及内涵是确定的,比如对转移句而言,除了转移的动作本身(E)外,转移的发出者(转移的起点 A)、接收者(转移的终点 B)、途径(B')以及转移的内容(C),都是必需的。辅语义块与句类的函数关系较弱,因此不进入句类表示式。因此,语义块的数量就和句类表示式联系在一起,所以,每个语义块的具体内涵只要联系到具体的句类就比较清楚了。格的数量就不成为问题了,而确定哪些语义块是必须的,哪些是可选的,只要和句类表示式相比较就可以了。

3.3 格与语义块的构成

格的构成分析和槽关系的研究是分不开的。语块是组成句子的元素,如“人民大学的教师广泛开展教书育人的活动”,这个句子有三个语块:“广泛开展”“人民大学的教师”“教书育人的活动”。可见,语块分为两种:一类是谓词性语块,如“广泛开展”;一类是体词性语块。两类语块之间是格关系,而语块内部的语义关系是槽关系,如“广泛”和“开展”之间的语义关系。体词性语块内部各成分间的槽关系是研究的重点,同时也可以看作对格的构成的研究。槽关系研究的重点是槽类(以谓词类别和受谓词联系的名词类别来划分)、槽序(槽类的排序)和槽系(一个名词所能搭配的槽类数量)。值得一提的是,在鲁川、陈群秀等专家的努力下,槽关系的研究无论从理论到工程都取得了令人欣喜的成果。

语义块构成与格的构成不同的是语义块具有复合构成。语义块复合构成意味着块内各部分之间可以存在分离、成为子句等各种复杂情况。特征语义块的构成有高低搭配、动静搭配等形式,不同部分间可以分离。这里我们重点看一下广义对象语义块的构成。广义对象语义块的构成有良性构成和非良性构成之分。良性构成指 JK 中对象与内容的形式上的有序分解,如作用句中的对象 B 语义块可分解为作用对象 XB、效应对象 YB 和效应内容 YC,表示式为 $B=XB+YB+YC$ 。非良性构成指构成中对象 B 与内容 C 的融合,即 $JK=BC$ 。广义对象语义块也存在分离现象,如“李四被张三打断了腿”一句,本应同属于对象 B 的“李四”与“腿”发生了分离。广义对象语义块分离出去的部分还可以融合到别的语义块中,如“联合国对伊拉克进行了经济制裁”一句中,JK2(“伊拉克的经济”)的一部分“经济”分离出去,融合到了特征语义块“进行了制裁”的内部。这种处理方式是从语义理解的角度进行考虑的。

广义对象语义块可以是一个语句,或者是一个句子的变体。HNC 提出了块扩与句蜕的概念。块扩表现为一个小句,无说明部分与核心部分之分。HNC 理论定义了块扩现象只和特定的句类相联系,比如“希望”做特征语义块时后面的 JK 必须是一个句子,也就是必须块扩。这种处理弥补了格框架中无法用格来命名子句的不足。句蜕则是指那些由句子经过各种变化形成的语义块,如“克林顿对中国的访问将会推动中美关系的发展”,其中的划线部分我们看作是句子“克林顿访问中国”变化得到的,这也是从语义的角度考虑的。

总之,格和语义块同是句子的下一级语义构成单位,都是句子语义框架中必不可少的成分。但二者也有不同:

第一，格是不可分解的，不存在基元与复合之分，而语义块可以分解，有单纯构成和复合构成的区别。

第二，格仅涉及名词短语，或者更具体地说，仅涉及“主谓状”的语义分类。而语义块既包括体词或体词性短语，也包括动词或动词短语，以及其他一些成分。

第三，语义块的概念是在对句子及句子内部各部分间概念上的联系的基础上提出来的，它的构成分析处处考虑到了句子内部各部分之间的语义关系。

4 格框架与 HNC 句类的分类标准

4.1 格框架的分类

必须格和动词一起构成基本句式，简称基本式，基本式即格框架。

我们把最早做的 3000 个动词的格框架总结归纳为 3 大类、9 中类、58 小类，3 大类，是第一个层次。它的分类标准是动词的论元属性，即论旨角色与动词之间的配价关系，也就是看动词能带几个必须格。能带一个必须格的是一价动词，由一价动词组成的格框架即为一价格框架，另外还有二价格框架、三价格框架。9 中类，是第二个层次，它是根据动词的六个次类划分出来的。动词的六个次类是：他动词、自动词、外动词、内动词、领属动词、系属动词。58 小类，是第三个层次，它的分类标准是动词的论旨属性。所谓论旨属性是指某个动词所搭配的必有论元是什么论旨角色以及这些论旨角色在语义上有什么约束。述语动词对有关论旨角色所施加的语义制约的研究是格框架研究的重要环节。

由于动词格框架是建立在有选择地描写的一些语言现象基础上的，并不是在全部语言事实之上得出的。58 类格框架描写的均是有且只有一个述语动词的简单句子，而无法描述由一些特殊的动词或动词的组合等特殊情况形成的句子。张潮生在《语义表达的一些性质》中指出：汉语的一些流水句、无动句、连动、紧缩、动补、省略等结构，无法或不必用一个动词统率一个句子的模式来描述。其中，连动句和兼语句尤为突出。比如“他把他的论文提交大会讨论”，这里用格框架的方法并不能找出一个统率全句的动词，是“提交”还是“讨论”？因此无法说明这个句子属于哪个格框架。这是用格框架描述汉语句子的局限。

4.2 HNC 句类的分类

HNC 句类是根据特征语义块的意义和性质来划分的。特征语义块 E 的核心部分一般由动词充当，它蕴涵了一个句子的基本的语义信息。这种基本的语义信息可以用 HNC 的作用效应链思想来概括。什么是作用效应链？“作用效应链反映一切事物的最大共性，作用存在于一切事物的内部和相互之间，作用必然产生某种效应，在达到最终效应之前，必然伴随着某种过程或转移，在达到最终效应之后，必然出现新的关系或状态。过程、转移、关系和状态也是效应的一种表现形式。新的效应又会诱发新的作用，如此循环往复，以至无穷，这就是宇宙间一切事物存在和发展的基本法则，也是语言表达和概念推理的基本法则。”

（见《HNC 理论》第 29 页）作用效应链的六个环节是表述事物的基本侧面，自然语言的一个句子总是对作用效应链的某个或某些环节进行表达，所谓一个句子的基本的语义信息就是指它所表达的是关于作用效应链的哪个或哪些环节的信息。作用效应链是表达客观事

物的。判断是表达主观思维的，因此，作用效应链加判断形成了七大句类，分别是：作用句 XJ，过程句 PJ，转移句 TJ，效应句 YJ，关系句 RJ，状态句 RJ，判断句 DJ。因此可以说，HNC 划分句类的标准是“作用效应链+判断”，

在七大句类下，又划分出 57 种子类，这是对动词（特征语义块）内涵细分类的结果，称为基本句类。比如转移句下面又分信息转移句、物转移句等。基本句类是自然语言语义结构的基元类型，它们可以混合和复合，形成混合句类和复合句类。

那 HNC 的句类怎样描述上面提到过的格框架不能描述的那些情况的呢？我们以同样的句子为例：他把他的论文提交大会讨论。这里“提交”和“讨论”对句子具有相同的语义贡献，首先是“他把他的论文提交大会”，然后是“大会讨论他的论文”，因此形成复合句类，只不过，两个动词共用两个语义块“大会”和“论文”。

综上所述，我们不难得出，格框架的分类结果是通过归纳的方法得出的，它来源于大量真实语料，具有可操作性，易于为人所理解和接受，但不能共同尽举地覆盖所有的语言现象，还有待在语言工程中总结归纳新的动词格框架。而 HNC 句类是通过演绎的方法得到的，从理论上讲，它具有普遍的适用性，但它的可行性还需进一步用真实语料加以检验。对汉语句子进行有利于信息处理的分类，我们期待着上述两种分类方法相互补足。

5 HNC 句类转换与格框架的语义制约

HNC 句类转换和格框架的语义制约研究，是我们认为二者应相互借鉴的重要方面。

5.1 HNC 句类转换

HNC 句类转换是指一个句类的内容换用另一个句类来表达。比如，汉语中很多句子，不管句类如何，均可变换成是否判断句，如：

美国以财富分配失调加重换取较快的经济增长。

=> 美国是以财富分配失调换取经济的快速增长的。

前一句句类是一般效应句，句类表示式为 $Y0J=YB+Y+YC$ ，“以财富分配失调”是辅语义块，而后一句句类是是否判断句，句类表示式为 $jDJ=DB+jD+DC$ 。下面是基本作用句 $XJ=A+X+B$ 向被动承受句 $X12J=X1B+X1+XAC$ 的转换的例子：

敌人袭击了我们的一支小分队。

=> 我们的一支小分队遭到了敌人的袭击。

转换前后句类之间的语义块对应关系是：基本作用句的 B（作用对象）转换为被动承受句的 X1B（作用承受者），基本作用句的 A（作用者）和 X（作用）转换为承受句的 XAC（承受的引发者及其表现）。对于句类转换现象，只有对转换句类进行反转换处理，还原出转换前的句子，正确分析出语义块之间的隐含的语义关系，才能得到语句分析处理的正确结果。

HNC 提出的句类转换和语义块变换，体现了现实生活中语言运用的灵活性，说明了自然语言的深层理解和表层运用存在巨大差异。格关系工程性研究中论旨模式也描写了基本式（格框架）和变换式，但它所描写的变换式只是主体、客体乃至邻体的简单移位，不同变换式的对应的格框架是唯一的。对于不同格框架间的转换没有进行研究，可以吸收和借鉴 HNC 这方面的研究成果。

5.2 格框架的语义制约

在格框架中对框架轴心的谓词进行分类,描述必须格名词的语义属性,是研究聚合和组合相互制约的重要方面。格关系的工程研究一改义类词典对语义的静态分类办法,用论旨网格描写汉语动词的组合框架,在动词的组合框架中研究名词性概念的聚合类别,反过来,又以名词的聚合语义类别划分动词的次类。这样,就从论旨角色语义约束的角度,建立起了汉语名词性概念的分类体系和基于格关系的现代汉语述语动词分类系统,进而确定了在述语动词组合框架中的动词和名词的语义相互制约关系。

格关系工程研究课题组根据对 3002 个动词所搭配的常用名词进行考察分析,最终得到了一个包含 212 类名词的现代汉语名词性概念分类体系。而 HNC 句类研究虽然对于某些特定句类的特定语义块,作出了语义上的约束,但论述比较零散,尚未形成知识体系,相对于格框架中的语义制约研究还有不足。

6 总结

格框架是句子中的述语动词同其周围名词性成分的语义组合模式。格框架研究从语言表层到语言深层,从句法层面到语义层面,通过对由述语动词形成的句子模式的归纳,根据三个不同层次的划分标准,得到现代汉语中 58 个汉语动词格框架,并且在格框架中研究动词和名词的语义制约关系,得到了现代汉语的名词性概念分类体系。

HNC 句类是句子的全局联想脉络,为了建立这个全局联想脉络,HNC 提出了以主语语义块构成句类表示式的概念,对语义块提出了特征语义块 E、广义对象语义块 JK 和辅语义块 fK 三分类的概念,对句类提出了基本句类、混合句类和复合句类的三分类标准,对基本句类提出了按作用效应链划分的基本标准,从而得到了 7 大基本句类,57 个基本句类一级子类的重要结论。并且指出,广义对象语义块是句类的函数,句类可由特征语义块唯一确定(但可能存在多个句类代码),这样,最终建立起了整个 HNC 句类体系。

总之,格框架和 HNC 句类研究都为自然语言处理提供了语句表述模式,都是对句子语义框架研究的有益探索。随着这两个体系的不断发展和完善,它们必将在人机对话的领域内发挥更大的作用。

参考文献

- [1] 陈群秀:“信息处理用现代汉语语义分类体系的设计思想”,《计算机时代的汉语和汉语研究》,清华大学出版社,1996年11月。
- [2] 黄曾阳:《HNC(概念层次网络)理论》,清华大学出版社,1998年。
- [3] 林杏光:《词汇语义和计算语言学》,语文出版社,1999年。
- [4] 陈群秀:“现代汉语述语动词机器词典的扩充和槽关系研究”,《语言文字应用》,2001年第四期。
- [5] 鲁川 林杏光:“现代汉语语法格关系”,《汉语学习》,1989年第五期。
- [6] 张庆旭:“汉语述语动词框架分类及其语义限制”,《汉语学习》,1996年第三期。