

文章编号: 1003-0077(2017)01-0230-09

从短语到构式: 构式知识库建设的若干理论问题探析

詹卫东

(北京大学中文系、中国语言学研究中心
北京大学计算语言学教育部重点实验室, 北京 100871)

摘要: 构式语法(construction grammar)在汉语语法学界已引起持续关注,但在自然语言处理领域,将构式语法理论应用到计算机自动句法语义分析中的研究还很少见。该文提出构建现代汉语构式知识库的语言工程任务,讨论了构式与传统语法单位的关系、构式的形式表示、构式的内部小类及主要特征等。

关键词: 构式; 知识库; 组合性; 非递归性

中图分类号: TP391 **文献标识码:** A

On Theoretical Issues in Building A Knowledge Database of Chinese Constructions

ZHAN Weidong

(Chinese Linguistic Research Center, Department of Chinese, Peking University
MOE Key Lab of Computational Linguistics, Peking University, Beijing 100871, China)

Abstract: Although the construction grammar has already drawn much attention in the field of Chinese linguistics, scholars in the field of natural language processing are rarely concerned with this approach to syntactic and semantic parsing of Chinese. This paper proposes a linguistic engineering project of building a knowledge database of Chinese constructions. Some key issues in this task are discussed, including the differences between the construction and the classical grammar units, the formalism of the construction, the types and features of the construction and so on.

Key words: construction; knowledge database; compositionality; non-recursive

1 引言

从注重描写语言事实的结构主义语言学到注重解释语言机制的转换生成语法,20 世纪的主流语言学理论对于语言系统中组合的结构规则做了大量富有成效的研究,无论是在具体语言的句法语义规则体系方面,还是在规则知识的形式化表示方面,都取得了可观的成果。而随着研究的深入,也有不少语言学者认识到,已有的理论和方法对于分析语言系统中的“不规则”现象(或称“边缘”现象),常常显得力有不逮。在这一背景下,自 20 世纪 80 年代认知功能语法学派兴起以来,语言使用中的种种超乎一般规则的现象,受到了语法

学界越来越多的重视,从初期对习语的个案研究,如 Fillmore 等对英语中习语“let alone”的分析^[1],到后来对大量的特殊句法结构的更为系统的研究,如 Paul 等对“What’s X doing Y”的分析^[2],Goldberg 对英语双及物构式、致使-移动构式、动结构式、非及物移动构式、移动构式等所做的系统的论元结构分析^[3],逐渐形成了语言学中的“构式主义视角”,发展出跟基于短语结构的生成语法体系大不相同的基于构式(construction)的整体语法观。这种语法学研究视角很快在国际语言学界产生比较广泛的影响,2013 年出版的《牛津构式语法手册》分五大部分、27 章对构式语法各方面的研究做了全面的介绍^[4]。

鉴于上述发展态势,为推进构式语法理论在汉

收稿日期: 2016-09-20 定稿日期: 2016-10-16

基金项目: 国家重点基础研究发展计划(2014CB340504);教育部人文社科重点研究基地重大项目(13JJD740001, 15JJD740002);国家社科基金(12&ZD175)

语语法领域的研究,并使构式语法理论在对外汉语教学、中文信息处理等领域发挥实际效用,本文提出,应建设一个有一定规模的能覆盖大量现代汉语真实语料的现代汉语构式知识描述数据库(以下简称“构式知识库”),即采用类似词库的方式,将真实语料中实际运用的构式形式逐条收录,并详细描写每个构式的属性。这一语言工程实践的目的,是从外延上初步确定现代汉语构式的大体范围,一方面为汉语构式语法的理论研究提供丰富实例,另一方面也希望通过积累丰富的对汉语构式特征的描述,为计算机自动句法和语义分析,对外汉语语法教学等提供数据支持。

构建现代汉语构式知识库,需要制订详细具体的知识库编写技术规范,为此,首先就要在理论层面,结合计算机分析的需要,对构式的相关基本概念做深入讨论。下文第二节对构式的内涵及外延做学理上的仔细分析,对构式知识库收录的对象加以界定;第三节厘清构式与传统语法单位(特别是短语)之间的关系;第四节讨论知识库中构式的形式表示;第五节讨论构式的内部分类和主要特征;第六节说明构式与修辞之间的关系,交代构式知识库暂时无法处理的语言现象。最后第七节是结语,简要说明构式知识库目前的实施情况。

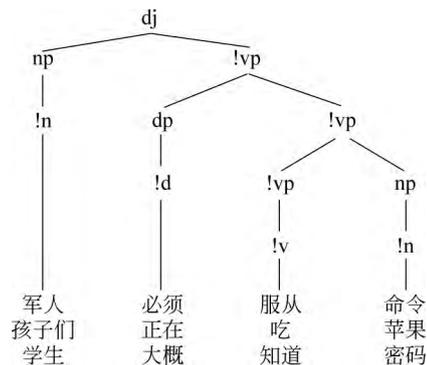
2 构式的界定

汉语语法学界对构式的定义,一般沿用 Goldberg 的说法:“C 是一个构式,当且仅当 C 是一个形式(Fi)和意义(Si)的对应体<Fi, Si>,而无论是形式或意义的某些特征,都不能完全从 C 这个构式的组成成分或另外的先前已有的构式推知。”^[3]

构式最主要的性质在于其形式和意义无法从其构成成分的形式和意义推知。要比较准确地理解这一内涵,不妨从对立的一面来加以认识,从传统的语法研究来说,词组(短语)由词构成,许多词组的形式和意义都可以由其构成成分的形式和意义推知,其特点可以归纳为:(1)内部构造可以而且适合分析为二分支层级树结构;(2)内部成分的功能范畴明确;(3)内部成分的可替换性(可扩展性)强;(4)整个结构的语法功能由中心成分的语法功能决定;(5)整个结构的语义由其组成成分的语义组合得到。

下面是现代汉语中常见的主谓短语结构的构造示例:

例 1



在例 1 所示的短语结构树中,dj 代表小句(内部一般为主谓结构关系),np 代表名词性短语,vp 代表动词性短语,dp 代表副词性短语,n、d、v 分别代表名词、副词、动词。np、vp 这些语言单位都具有递归性(recursive),即 np 和 vp 中都还可以再包含 np 和 vp,从而使得整个结构具有很强的扩展能力,即生成更长的同型结构的能力^①。

跟例 1 所示的作为语言中常规组合的短语结构相对,构式作为超常规的组合,其特点则可以归纳为:(1)构式不适合分析为二分支层级树结构;(2)内部成分的功能范畴不明确;(3)内部成分的可替换性(可扩展性)弱;(4)结构或者无中心成分,或者整体的功能不由中心成分的语法功能决定;(5)整体的语义不能由组成成分的语义组合得到。(6)除上述特点外,典型的构式往往具有独特的交际功能(价值),是比较有特色的表达方式。

一个形式和意义结合体,如果具有上述六项特征中的任何一条,就可以看作是构式,符合的特征越多,则其作为构式的典型程度就越高。下面看两个简单的例子,例中加粗的部分不适合按照传统的短语结构组合来分析,而适合看作构式。

例 2 (眼看)梨树长得有两个人高了(,却总是只开花不结果)。

例 3 天底下长得像的人多了去了。

例 2 中的“有两个人高”,例 3 中的“多了去了”都不大容易按照一般短语结构的组合模式分析,用类似例 1 那样的二叉树的结构来描述例 2、例 3 的内部层次构造有困难。用目前比较主流的自动句法分析工具(无论是短语结构分析器还是依存句法分

^① 例 1 中的“孩子们在吃苹果”就很容易扩展为更长的同型结构:“那些调皮的孩子们 已经在 大口大口地 吃 偷来的苹果”。

析器)来分析这样的句子,得到的结果往往都不如人意^①。显然,要处理这类含有构式表达的句子,需要专门针对构式的特殊性,研究其知识表示和用法规律。

在“现代汉语构式知识库”语言工程的工作框架中,我们把典型的构式看作是无递归性的非平凡的(non-trivial)短语结构。典型的构式作为既含常项又含变项的语言单位,是对常规短语结构语法组合的必要补充。下面进一步分析构式与短语之间的关系。

3 构式与传统语法单位的关系

一般来说,汉语学界通常把语法单位分为语素、词、词组、句子等四个层级。这些单位都是所谓的“音义结合体”,即同时具有语音(物理)形式和语义(心理)内容的语言单位^②。其中语素和词作为基本单位一般认为其长度和数量都是有限的,而词组则是长度无限的(至少是潜在无限),总的数量更是无限的,不适合穷尽性的罗列,需要研究其组合的类型和规则。一般假定,词组的组合类型可能是有限的,组合的规则(制约条件)也是相对有限的。根据有关汉语词组本位语法体系的思想,汉语句子和词组的构造大体相同^[5-7]。因此,把词组的组合规则搞清楚,句子的主要组合规则也就清楚了。

作为汉语传统的语法单位中最重要、覆盖面最广的单位,词组(短语)的主要特点在上文中已做了概要说明,这里通过一些例子的分析进一步来讨论词组内部成分之间(语义)关系的多样性,以及成分的高可扩展性(即构成成分的递归性)。在这两点上,典型的构式都具有跟词组(短语)不同的性质。此外,下文的分析同时还想说明,尽管构式跟词组(短语)构造有显著的不同,但二者之间并无截然可分的明确界限。

例4 N1+N2(定中结构名词性词组)

A	B
北大 教师	细胞 词库
体操 明星	输入法 皮肤
花园 洋房	网络 水手

上面例中A、B两组都是名词+名词形成的定中结构关系的词组,但是,A、B具体的实例中n1和n2的语义关系存在明显的多样性,即很难找到一个共同的释义模板(替换表达形式),来对这些实例做统一的释义。而就A、B两组实例比较来说,A组的

例子是容易理解的常见组合。B组的例子则是出现时间不长,比较新的词组,代表的是新的事物(现象)。它们也符合1.1节对构式的界定标准,即结构整体的语义很难从其构成成分推导出来。那么,它们所代表的模式“n1+n2”是不是构式呢?因为n1和n2之间语义关系的多样性,我们很难给出一个“形式——意义”配对,以构式的方式来描述n1+n2组合。从这个角度说,n1+n2组合不适合看作为一个抽象的构式类型(type),而只能作为一般的词组(短语)来处理。不过,n1+n2组合中的某些具体实例(token),特别是例4中B组的例子,则可以作为构式实例来看待,相当于词一级的语法单位(可参见下文有关凝固型构式的说明)。

例5 np1+中+的+np2(定中结构名词性词组)

A	B
生命中的瞬间	男人中的男人
阿Q生命中的六个瞬间	天才中的天才
阿Q短暂生命中的六个惊心动魄的瞬间	奇迹中的奇迹

例5A组的实例np1和np2都可以做比较自由的扩展,这些实例体现的是常规短语组合的性质,即高可扩展性。例5B组的例子中np1和np2由同形的单个名词(n)充任,基本不具有扩展性^③。整体表达一个共同的语义,即n所表示的事物集合中最突出、最典型的元素。从例5A和例5B组实例的对比来看,尽管二者都可以用“np1+中+的+np2”来模式化,但5A是常规的词组;5B是构式。5B可以更精确地模式化为“n+中+的+n”(前后两个n同形)。如果要对例5B的实例进行扩展,生成的例子或者可接受度较差,或者难以接受。如例6所示:

例6 a. ?数学天才中的数学天才

b. * 中国的数学天才中的中国的数学天才

下面再看一个介乎短语和构式之间的语法形式的例子。

^① 读者可以用依存句法分析器(比如<http://www.ltp-cloud.com/demo/>)和短语结构句法分析器(比如<http://nlp.stanford.edu:8080/parser/>)来测试这类含有构式表达的分析结果。

^② 相比之下,音素、音节这样的语言单位就是只有语音形式而无意义内容的纯语音单位(而非语法单位)。

^③ 例5B中有的实例可能存在非常有限的扩展形式,例如:“男人中的顶级男人,天才中的超级天才,超级天才中的超级天才,……”,扩展形式将原式中的隐含义显式地表达出来了。

例 7 X 对 Y

A	B
中国 对 丹麦	天 对 地
林丹 对 陶菲克	大陆 对 长空
中国队的林丹 对 印尼 队的陶菲克	梨儿腹内酸 对 莲子 心中苦

例 7A 组中的“对”是“对阵、对抗”的意思；例 7B 组中的“对”是“对仗”的意思。例 7A 组实例的 X 和 Y 都是 np(名词性短语)，X 和 Y 的长度(音节数)可以不同，但语义都是指人或团体；例 7B 组实例的 X 和 Y 可以是任意的语言成分，X 和 Y 的长度相同，语法范畴、语义范畴、声调特征等也要求对应(即符合对联的形、音、义约束条件)。如果把例 7 格式看作短语(词组)，例 7A 和 B 两组实例都只能看作是主谓结构或并列结构，其中的“对”、X 和 Y 等成分都受到比较强的约束限制。如果把例 7 格式看作构式，从形式上说，跟 X、Y 的可扩展性相悖；从语义上说，“X 对 Y”的整体语义主要由其中的“对”决定，例 7A 和 7B 中的“对”可以处理为两个义项，使得整体语义大致上可以由其组成成分推导得到。我们认为，例 7 反映的现象就是介于短语(词组)和构式之间的一种情况。既有短语的性质，也有构式的性质，或者反过来说，既不是典型的短语，也不是典型的构式。

通过上面例 4-例 7 的分析，大致可以归纳出有关构式特性的以下三个要点，用于指导构式库操作层面的具体工作，特别是哪些语法形式应收入构式库。

(1) 构式有组合性，构式应看作是大于传统语法单位中的词的单元，构式与传统语法单位中的词的区别，主要在于组合性。收入词典(词库)中的词作为最基本的语法单位，可以不关心其内部构成。而构式库的定位是介于词库和短语规则库之间的语言知识资源。收入构式库中的构式，可以向内透视到其更为基本的构成成分——词项。而其“特殊之处”在于，这些词项不是按照一般常规的组合方式组合到一起，而是滋生出了原词项没有的意义。例如，“用数据说话、用事实说话、用拳头说话、哪儿的话”等组合，这些语言单位，一般不会收入传统的词典，又不是常规的短语(词组)，就适合收入构式库中。

(2) 构式无递归性(或无扩展性、少扩展性)，这一点主要是把构式跟一般的短语(词组)对比来看。因为无(或少)递归性，带来的后果就是，构式的长度一般较为有限，不大能像普通词组那样扩展得很长。构式内部成分的可替换性一般不高。

(3) 构式是从传统语法单位中因其“形式——

意义”配对的特殊性而抽离出来的语法单位。也正因为如此，构式跟传统的语法单位保持着一定程度上的交叠关系，无法截然分开。从语义角度看，构式的语义往往无法着重落实在其构成成分中的某一个(或几个)上，即没有中心成分和从属成分之分，而是由全部构成成分共同决定的(参见上文关于例 5B 的讨论)。这个特点或许可以进一步概括为“构式中的成分，一个也不能少”。当然，这是就句法组合层面来说的。构式在实际使用中受到语用因素的影响，也可能出现省略其中某项成分的情况，比如“村口有一株老槐树”可以省略“有”成为“村口一株老槐树”；“一锅饭吃十个人”可以省略“吃”成为“一锅饭十个人”等等。不过，从理论上说，省略后的构式跟省略前相比，其“形式——意义”配对必然有所差别(比如省略形式有更多歧义的可能性，“一锅饭十个人”可能理解为并列式名词性短语，因而实际上也对应着另外一个构式。

4 构式的形式表示

上文将构式与短语结构的性质做了对比。短语结构的内部构成一般以 np、vp 等短语功能分类范畴为基础，表示为层层二分的嵌套树状结构。而构式则应该表示为线性的序列结构。但具体如何表达构式的线性序列结构，仍有不同的选择。比如有的学者就建议构式可以由语块(chunk)构成^[8-10]。不过，关于“语块”的理解，语言学界在不同发展阶段也有过不同的看法，相对多数的意见是把“语块”看作介于词和短语之间的语言单位，或称为“词汇性短语”(lexical phrase)。薛小芳、施春宏把语块看作是“比较特殊的构式”。构式的外延更广，既包括由具体的语言成分形成的具有预制性的“实体块”，也包括相对抽象的结构构式(比如双及物构式“np1 + v + np2 + np3”)。前者既是构式，又是语块，而后者则只看作是构式，而不看作是语块^[11]。

本文认为，构式的内部成分并不适合用“语块”来描述。如果要用语块去作为描述构式内部成分的基本单位时，实际上会很自然地把“语块”看作是一个“语义组块”。比如，把汉语中表存在义的结构表示为“存在处所—存在方式—存在物”(桌上 放着一个花瓶)，由三个语块构成，其中每个语块的名称很显然都是指语义范畴。用语义组块表达构式内部成分有利于廓清构式的基本格局，便于人来认识该构式，但从计算机形式处理的角度来说，这些语义块

对应的语法形式单位,却不能一以贯之地用语义范畴加以描述,仍然要求助于句法范畴。这有点类似于在表达短语结构的内部构成时,如果采用“主语+谓语”这种模式来描述,就不如采用“np+vp”这样的可递归嵌套至词一级语法单位的描述模式^[12]。另外,也是更重要的问题是,如果仔细考虑构式的整体性,以及语块的“形式——意义”整合性,一个构式内部通常应该是一个“整体块”,而并不总是可以拆解为承担不同功能的语块。上面的存在义构式似乎

容易拆解一些,但是像“天才中的天才”(n+中+的+n)“多了去了”(a+了+去+了)“漂亮什么漂亮”(a+什么+a)“你唱你的”(r+v+r+的)等等许多构式,其内部成分都很难拆解为不同的功能语块。

基于上面的认识,对于构式的内部成分的描述,本文主张不以语块(组块)作为基本单位来表示构式的组成,而是以常项(constant)和变项(variable)为基本单位,来记录构式的组成。图1以“你唱你的戏”为例,展示了不同的内部结构表示方式。

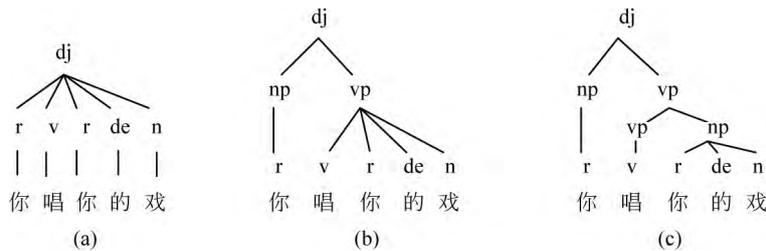


图 1

图1(a)把构式内部成分看作是线性序列组合,由五项成分一次组合成型,不需中间结构单位;图1(b)把“你唱你的戏”看作是先由两个组块构成,其中vp组块再由“v+r+de+n”这四项组成;图1(c)是短语结构的层次嵌套组合模式,其中vp、np都具有递归性(即组成成分中包含了自身)。本文认为构式适合采用图1(a)的形式表示方式。

5 构式的分类及主要特征

上面从跟传统语法单位的对比角度阐释了我们对构式内涵的认识,这一节再进一步来看构式包括

哪些具体的类型,或者说从外延的角度来认识构式。

国外一些构式语法学者对构式的分类基本上涵盖了传统语法的大多数语法单位,是对语法单位整体的性质持一种比较宽泛的构式观^[13-15]。相对而言,本文则持一种相对狭义的构式观,从跟传统语法单位的对应来看,本文把汉语中的构式分为四类:(1)凝固型构式,(2)半凝固型构式,(3)短语型构式,(4)复句型构式。它们之间的差异可以通过下表来呈现。概括而言,从(1)到(4),也是反映了成分复杂度和抽象度的增加,具体体现在组块数量、变项数量、构式长度的增加,以及成分类型多样性程度、可替换性程度的增加。

表 1 构式的分类及特性

类型	组块数	变项数	长度	成分多样性	成分可替换性
凝固型构式	1	0	固定不变(短)	无	无
半凝固型构式	1	1-2	稳定(短)	低	低/中
短语型构式	1	≥1	可变(中)	中	中/高
复句型构式	1-2	≥2	可变(长)	高	高

表2给出各类构式的一些具体实例。

表 2 构式类型及示例

构式类型	构式示例
凝固型构式	用+脚+投票、羡慕+嫉妒+恨、买+面子、打+光棍、必须+的 托福+热、创业+热、留学+热、牢骚+帝、表情+帝、数学+帝 开心+ing、彷徨+ing、7×24
半凝固型构式	v+来+v+去(走来走去、搬来搬去、买来买去、商量来商量去) n+百+出(错误百出、洋相百出、丑态百出、花样百出)

续表

构式类型	构式示例
短语型构式	n+中+的+n (天才中的天才) 不是+n+的+n (不是办法的办法) -+q+比+-+q+a (一天比一天冷) a+不+到+哪里+去 (坏不到哪里去) a+就+a+在+X (妙就妙在这个反差)
复句型构式	别+说+是+X,就是+Y+也+Z 例如：别说是他只有一张嘴,就是有一百张嘴也说不清。 别说是在中国,就是在国外也享有盛誉。 v+也+不是,不+v+也+不是 例如：看也不是,不看也不是 走也不是,不走也不是

凝固型构式大致相当于传统语法单位中的词。不过,跟词不同的是,凝固型构式有组合性,即凝固型构式由多个词组合而成,比如“用脚投票”就是由三个词“用、脚、投票”组合而成的一个凝固型构式。一般词汇中的词语,特别是像纯粹的单纯词,即由一个语素构成的词,仍应按传统观念,看作词级语法单位,不作为构式处理,比如像“仁波切”(指活佛、上师、智者,德高望重的人……)这样的词。

半凝固型构式介于传统语法单位的词和短语之间。跟凝固型构式不同的是,它一般包含一个或两个变项,一旦变项被具体的词语替代,形成的单位就跟凝固型构式很相似。半凝固型构式跟短语型构式

的界限比较模糊,相当多的半凝固型构式一般为四音节(传统上也称为“四字格”固定短语),而且内部成分之间往往有对举并列的关系。短语型构式在音节长度、内部成分的关系等方面更多样一些。

复句型构式对应传统语法单位的复句,一般由两部分组成。有的复句型构式中的一部分可以独用。像上文举过的“你唱你的,他写他的”,前后两个组块就相对独立,可以单用。

对于知识库中一个具体的构式条目,除了标记其所属类型外,还要描述其主要的形式特征和语义特征。表3、表4是目前知识库中构式特征的设置及其示例。

表3 构式语法特征及其示例

构式特征	示例
复现(同形)	(病情)一天比一天严重 男人中的男人
复现(同构)	左一个报告,右一个请示 丁是丁,卯是卯
省略	糊涂得可以 弹钢琴弹的 两个房间六个人 你一粒我一粒
冗余	因为时间的关系 像做贼似的
异序	老百姓一个 是他说漏嘴的 ^①
语法错配	被和谐 好你个李云龙 那个紧张啊

表4 构式语义特征及其示例

构式特征	示例
论元异常	跑了我一身汗 他高我一个头 你找我什么事
语义错配	他是猪肉馅儿,我是羊肉馅儿
否定	什么一流大学 看什么破电影

^① 异序特征假定了构式在短语结构(常规)组合中有一个对应的常规的语序。比如对于“是他说漏嘴的”,就是假定原始语序是“说漏嘴的是他”,按照一般短语组合规则分析为 np(说漏嘴的) vp(是他),主谓结构。谓语部分“是他”前移到主语 np“说漏嘴的”前面,形成“是他说漏嘴的”(强调句)。

续表

构式特征	示例
周遍	谁也看不懂
主观量	一去就是三年

一个构式的“构式特征”取值允许多选,比如“他是猪肉馅儿”构式既有“语义错配”特征,也有“省略”特征。

6 构式与修辞

典型的构式在“形式——意义”配对时往往有很强的超常错配特点,包括语法范畴、语义范畴、语用范畴等多个层面的错配。而这样的语言现象,正是传统上修辞关注的对象,即字面意义(由形式直接推导的意义)与实际表达意义(形式上无法直接推导的意义)之间的超常联系。从这个角度说,修辞(寻求非常规的表达方式)大概是形成构式的主要驱动力。在语言实际使用中,最新的修辞表达方式都是临时偶发的,而一旦某种修辞表达方式因许多人的跟随

模仿而反复出现,成为一种模式,进而固定下来,就成为语言系统中一种稳定的表达方式,即构式。这样的构式使用时间长了,其修辞意味就很有可能会大大降低,可能不再被人们视为一种“新奇”的表达,又成为“常规”表达方式中的一员了。从动态的观点看某个时间段内语言中的构式,比较突出的,有很强构式感的表达形式,往往就是这段时间内流行的一种修辞表达方式。在当今的互联网媒体时代,因为信息交流的便利,传播速度非常快,一种新奇的表达方式,从出现到普及,从小范围使用到全社会通用,往往可以在很短时间内完成,从一种偶发的修辞语言现象,迅速成为一个语言社团中大家都比较熟悉也比较乐于使用的新兴构式。表5是一些网络语言中常见的带有鲜明修辞色彩的构式例子。

表5 修辞型构式示例

构式(模式)	构式(实例)
(1) 被 X	被自杀 被幸福 被奥数 被没有资格
(2) 各种 X	各种卖萌 各种悲剧 各种睡不着觉
(3) 舌尖上的 X	舌尖上的中国 舌尖上的思考 舌尖上的浪费
(4) 都是 X 惹的祸	都是月亮惹的祸 都是相亲惹的祸
(5) 有一种 X 叫 Y	有一种毒药叫成功 有一种误差叫数据造假
(6) 你的 n1 是 n2 老师教的	你的语文是体育老师教的
(7) n1 v 的不是 n2, 是寂寞	哥吃的不是面,是寂寞
(8) 不想当 n1 的 n2 不是好 n3	不想当将军的厨子不是好裁缝

需要注意的是,有的修辞表达方式着重在两个具体词语之间的搭配超出常规,或者某个具体的词语的语义发生转指,例如,

例8 看见你的声音(“声音”跟“看见”搭配,从听觉现象转为视觉现象)

例9 农民工的声音(“声音”转指其所承载的“看法、观点、诉求”等内容)

像上面这样的修辞表达形式,如果要看作构式,则只能处理为“凝固型构式”。因为这类表达形式只用实例化的形式,没有进一步模式化。

还有不少具有鲜明的修辞色彩的表达形式,尽

管其形式和意义之间的配对也符合构式的特征要求,但从形式化表达的角度看,很难表述为常项加变项的模板。例如,

例10 基层干部……工作是5+2,白加黑,礼拜六保证不休息,礼拜天休息不保证。

例11 袁君是大连一名电视台记者,过着有选题忙死,没有选题死忙的高压锅生活。

例12 重庆老汉一根扁担“挑出”3个大学生

例13 中学生踢球“踢”进北大 全国仅两人

例14 读书有两个过程,先要把书读厚,这是消化的过程,然后再把书读薄,这才是理解的象征。

像上面例 10-例 14,只能看做纯粹的修辞现象,很难作为构式来加以形式化描写。

7 结语

近年来国外构式语法的研究有一个鲜明的共性,无论是在形式化表示体系的理论建构方面,还是在基于构式语法的计算框架设计方面,研究者的目标都是从“构式主义”视角出发来搭建新的语法分析框架,视图替代(或“收编”)原有的语法分析模式。比如 SBCG(基于符号的构式语法)就是以 HPSG(中心词驱动短语结构语法)的形式化方法为底本,以构式来整合传统的短语结构语法体系中的各类语法单位,将以往对短语组合规则的描述全盘移植并扩充升级为对各类构式^①的知识描述^[16]。跟这种“以新代旧”的思路不同,本文对汉语构式的研究更倾向于走“融合”之路,特别是在面向计算的汉语句自动句法分析中,本文认为,探索如何把构式的分析融合到传统的短语结构语法体系中,是更为务实的策略。正如著名的认知语言学家 Ronald Langacker 所指出的那样:“语言是规则与不规则的混合体”^[17],本文的基本认识是,传统的短语结构语法可以处理大量的语言常规组合现象,而构式则适用于描述非常规的组合。二者可以结合。尤其是在目前阶段,对汉语具体构式的个案研究较多,且研究兴趣大多集中在构式的语义描写、构式的语用功能特色、构式的认知机制等等方面,对于汉语构式知识的形式化表示,以及在计算机自动分析中的应用等,都还很少涉及,更谈不上深入系统的研究。如果完全照搬英语构式语法的形式化体系(如 BCG, SBCG 等)来描写汉语的构式,替代传统的短语结构语法形式化体系,这种自顶向下(top-down)的设计理念遭遇“水土不服”的可能性很大。相反,如果以相对稳定且比较成熟的短语结构语法形式化(范畴)体系为基础框架,将汉语构式知识的新发现结合进来,在缺乏“形合”(hypotactic mechanism)约束的汉语大环境中,注重更多地发掘“意合”(paratactic mechanism)约束的条件,则稳妥得多。

在上述思想指导下,我们已经在现代汉语构式知识库中收录了 800 余条构式,其中短语型构式 407 条(50.62%)、半凝固性构式 265 条(32.96%)、复句型构式 114 条(14.18%)、凝固型构式 18 条(2.24%)。关于构式常项、变项数目的分布统计,构式不同特征的比例等信息,可访问 <http://ccl.pku.edu.cn/ccgd/> 查询详情。

对这些构式,如何进行释义,如何描写其常项、变项成分的约束条件,以及构式在使用中的整体约束条件,为自动句法语义分析提供帮助,是我们下一步工作的重点。

致谢:本文写作中得到北大中文系陆俭明教授、沈阳师范大学文学院夏军副教授提出许多宝贵意见和修改建议。北大中文系研究生马腾、刘洪超、侯人渝、黄思思、田骏、苗宇晶、李安然、夏雪、赵贤参加了课题大量工作,特此致谢!

参考文献

- [1] Fillmore C J, Kay P, M C O'Connor, Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone[J]. *Language* 1988,64,(3): 501-538.
- [2] Kay, Paul, Fillmore, Charles J. Grammatical Constructions and Linguistic Generalizations: the What's X doing Y? Construction[J], *Language* 1999,25(1): 1-33.
- [3] Goldberg A E. A Construction Grammar Approach to Argument Structure[M], The University of Chicago Press. 1995.
- [4] Hoffmann T., Trousdale G., (eds.), The Oxford Handbook of Construction Grammar, Oxford University Press. 2013.
- [5] 郭绍虞,汉语词组对汉语语法研究的重要性[J],复旦大学学报,1978,1.
- [6] 朱德熙,语法讲义[M],商务印书馆,1982.
- [7] 朱德熙,语法答问[M],商务印书馆,1985.
- [8] 陆俭明,多维视野下的汉语教学——第七届国际汉语教学学术研讨会论文集[R],广西师范大学出版社,2009: 3-8.
- [9] 苏丹洁,试析“构式-语块”教学法——以存现句教学实验为例[J],汉语学习,2010: 2,83-90.
- [10] 苏丹洁,陆俭明.“构式-语块”句法分析法和教学法,世界汉语教学[J]. 2010: 4,557-567.
- [11] 薛小芳,施春宏.语块的性质及汉语语块系统的层级关系[J],当代修辞学,2013: 3,32-46.
- [12] 詹卫东,以“计算”的眼光看汉语语法研究的“本位”问题[J],汉语学报,2005: 1,64-73.
- [13] Croft W., Cruse D A. Cognitive linguistics. Cambridge[M]. Cambridge University Press. 2004.
- [14] Goldberg A E. Construction at work: The Nature of generalization in language[M]. Oxford University Press. 2006.

① 在 SBCG 的框架中,构式涵盖了以往的常规语法单位,如词和短语结构(核心语法现象)和大量的非常规特殊结构,即狭义的构式(边缘语法现象)。

- [15] Goldberg A E., The Oxford Handbook of Construction Grammar[M], Oxford University Press, 2013.
- [16] Boas H C, Sag I A. Sign-based Construction Grammar[M], CLSI Publications, Stanford University.

- 2012.
- [17] Langacker, Ronald W, Foundations of Cognitive Grammar: Theoretical Prerequisites [M]. Stanford University Press. 1987.



詹卫东(1972—),教授,主要研究领域为现代汉语形式语法,中文信息处理,汉语语言知识工程。
E-mail: zwd@pku.edu.cn

中国中文信息学会第八次全国会员代表大会暨学会成立35周年学术会议在京成功举办

2016年12月23—24日,中国中文信息学会第八次全国会员代表大会暨学会成立35周年学术会议在北京中国科技馆隆重举行,会上颁发了“钱伟长中文信息处理科学技术奖”,中国中文信息学会“终身成就奖”,“青年创新奖”,以及中国中文信息学会“优秀博士学位论文奖”。同时在本次会议也发布了2016年《中文信息处理发展报告》。会议邀请了五位专家进行了学术报告。来自中国科协、教育部国家语委等部委领导、学会支撑单位领导和中文信息处理领域的专家学者400余人参加了本次会议。大会开幕式由中国中文信息学会副理事长兼秘书长、中国科学院软件研究所孙乐研究员主持。中国中文信息学会理事长李生教授致欢迎词,中国科协魏军锋副部长、教育部语言文字信息管理司田立新司长和学会支撑单位、中国科学院软件研究所赵琛所长做了重要讲话,讲话肯定了学会工作所取得的成绩,同时从不同的角度分析了中文信息处理的应用需求与战略发展。

颁发“钱伟长中文信息处理科学技术奖”

钱伟长中文信息处理科学技术奖是经科技部批准设立的中文信息处理领域的最高科学技术奖,主要授予该领域在基本方法或关键技术上有原始创新或重大突破,对推动我国中文信息处理事业或行业进步起到重要作用,创造出较大经济效益或社会效益的项目或个人。2016年评选产生了“钱伟长中文信息处理科学技术奖”一等奖两项,分别授予了厦门大学史晓东、陈毅东等完成的“多目标智能化简繁文本双向转换平台”和东北大学朱靖波等完成的“东北大学机器翻译系统 NiuTrans”。

颁发“汉王青年创新奖”

“汉王青年创新奖”设立于2010年,主要授予在中文信息处理领域做出突出贡献的青年学者。2016年该奖项分别授予了清华大学计算机系的刘知远助理教授、复旦大学的张奇副教授、中国科学院软件研究所的韩先培副研究员和苏州大学的熊德意教授。

颁发中国中文信息学会“优秀博士学位论文奖”

中国中文信息学会“优秀博士学位论文奖”由北京拓尔思信息技术股份有限公司捐资设立,旨在鼓励中文信息处理领域的博士研究生在读博期间面向前沿方向、立足原始创新、开拓进取,勇创世界领先的研究成果。2016年优秀博士学位论文奖分别授予清华大学计算机系王超同学和哈尔滨工业大学的唐都钰同学,王超的论文题目是《搜索引擎点击模型构建》,唐都钰的论文题目是《基于表示学习的文本情感分析研究》。

颁发中国中文信息学会“终身成就奖”

本次会议还为哈尔滨工业大学李生教授和新疆大学吾守尔·斯拉木院士颁发了中国中文信息学会“终身成就奖”。中国中文信息学会“终身成就奖”的设立是为在中文信息处理领域做出杰出贡献的老科学家致敬。首届中国中文信息学会“终身成就奖”的得主是唐亚伟先生、董振东教授、俞士汶教授和倪光南院士。