

“现代汉语构式知识库”填写规范

(征求意见稿)

课题组内部文档、请勿扩散

詹卫东

北京大学中文系

北京大学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

最近更新：2018-10-05

目 录

§0 引言	5
§1 关于构式的基本认识	8
1.1 构式的界定	8
1.2 构式与传统语法单位的关系	10
1.3 构式的内部成分及其构造	14
1.4 构式的分类	18
1.5 构式与修辞	20
§2 构式知识库的设计概要	22
2.1 构式的形式与意义	23
2.1.1 构式形式：常项与变项的概括度及精细度	23
2.1.2 构式“一形一义”与“一形多义”	27
2.1.3 意义相近的构式	29
2.1.4 同形构式	31
2.2 构式的句法功能	32
2.2.1 构式句法功能范畴的确定	32
2.2.2 构式句法功能特征的描写	33
2.3 构式的语境语用信息	33
§3 构式知识库的数据表及字段设置	35
3.1 构式基本信息数据表	35
3.1.1 构式形式	35
3.1.2 构式变体	37
3.1.3 义项	38
3.1.4 构式特征	38
3.1.5 构式类型	39
3.1.6 音节数	40
3.1.7 组块数	40
3.1.8 是否可扩展	40
3.1.9 实例	40
3.1.10 变项数量	40
3.1.11 常项数量	41
3.1.12 释义模板	41
3.1.13 近义构式	41
3.1.14 反义构式	42
3.1.15 上位构式	42
3.1.16 下位构式	43
3.1.17 否定形式	43
3.1.18 疑问形式	43
3.1.19 备注	43
3.2 构式变项信息	43
3.2.1 变项序位	43
3.2.2 变项句法特征	43

3.2.3	变项语义特征	44
3.2.4	变项可替换度	44
3.3	构式常项信息	44
3.3.1	常项序位	44
3.3.2	常项句法特征	44
3.3.3	常项语义特征	44
3.4	构式内部成分约束关系	45
3.4.1	变项间关系	45
3.4.2	常项与变项间关系	45
3.4.3	组块关系	45
3.5	构式整体句法属性	45
3.5.1	构式的句法功能范畴	45
3.5.2	构式的句法功能分布特征	48
3.6	构式整体语义属性	49
3.6.1	字面义	49
3.6.2	交际义	49
3.6.3	语义识解机制	50
3.7	构式整体语用属性	56
3.7.1	感情色彩	56
3.7.2	语体色彩	56
3.7.3	领域限制	56
3.7.4	上下文环境	56
3.8	构式研究文献	56
§4	构式知识库填写工作中常见问题举例	58
4.1	是词还是构式	58
4.2	是短语还是构式	61
4.3	变项的概括度：“a+中+的+a”还是“X+中+的+X”	64
4.4	变项的概括度：“n1+的+n2+v+得+a”还是“np+v+得+X”	65
4.5	变项的概括度：m+q 还是 qp	66
4.6	构式形式是否最简	66
4.7	“构式变体”与“近义构式”	67
4.8	构式实例的形式与意义要求	67
4.9	半凝固型构式中变项不应超过两项	68
	参考文献	69

致 谢

本规范的研究工作得到教育部人文社科重点研究基地重大项目“现代汉语构式知识库建设及其应用研究”（13JJD740001），国家社科基金重大项目“汉语国际教育背景下的汉语意合特征研究与大型知识库和语料库建设”（12&ZD175）和国家社科基金面上项目“语言知识资源的可视化技术研究”（12BYY061），以及国家重点基础研究发展计划“融合三元空间的中文语言知识与世界知识获取和组织”（2014CB340504）资助，特此致谢。

北京大学中文系陆俭明先生、沈阳师范大学文学院夏军老师对本规范文档的起草提供了宝贵意见和建议。中文系多名研究生参与了规范初稿的起草与讨论，他们是：刘洪超、马腾、黄思思、侯人渝、田骏、苗宇晶、李安然、夏雪、赵贤。马腾设计并实现了构式知识库在线填写系统的第一个版本，田骏和李安然进一步完善了该系统。他们也都先后参与了构式知识库具体条目的填写工作。在此一并致谢。

构式知识库网页系统网址：<http://ccl.pku.edu.cn/ccgd>

“语言是规则与不规则的混合体”

Language is a mixture of regularity and idiosyncrasy.

Ronald Langacker, 1987, Foundations of Cognitive Grammar, p.411

—— 题记

§ 0 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性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，常常显得力有不逮。正是在这一背景下，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，就受到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（1988）对英语中习语“let alone”的分析，到后来对大量的特殊句法结构的更为系统的研究，如Fillmore等（1999）对“What's X doing Y”的分析，Goldberg（1995）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，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，同时也很快吸引了中国语言学者的注意。如张伯江（1999）对汉语双及物构式的研究，就是这一学术潮流在国内初起时的典型代表。从那时起到现在，十多年的时间里，有关汉语构式语法的研究已经积累了不少的成果，但是，总体而言，汉语学界在这方面的研究仍属初期阶段，还有几个方面的重要问题有待深入展开：（1）关于构式语法理论的一些基础问题。比如，构式的定义，构式的外延（范围），构式与传统的语法单位（短语结构、词、语素等）的关系，单个构式的多义性，多个构式之间的相互关联及其系统性，等等，都还没有在汉语构式本体研究基础上形成一套相对成熟的理论体系；（2）关于从构式的角度研究汉语所获得的语言知识的表示问题，比如构式的内部成分该如何范畴化，与该构式整体的关系如何处理，构式的语义如何表示等等；（3）构式语法知识的应用问题，包括在面向人的语言教学中如何应用，在面向计算机的自然语言信息处理中如何应用，特别是如何将构式语法的研究成果用于计算机的自动句法分析和语义分析，在当前有关汉语构式的研究中，更是鲜有涉及。

相比之下，国外构式语法理论和应用方面的研究在过去十多年可以说有了很大的发展。2013年出版的《牛津构式语法手册》（Hoffmann & Trousdale, 2013）分5大部分、27章对构式语法各方面的研究做了全面的介绍，包括构式语法的原则和方法、不同分支和流派、不同语法单位的构式性研究，构式的习得与认知心理学研究，构式的变异与演化研究，等等。其中有两方面的研究工作特别值得汉语构式语法研究（特别是面向计算的构式语法研究）重视：一是关于构式知识表示的形式化理论框架的研究工作；二是以

构式语法的视角重新设计计算机自然语言分析和理解的系统架构。前一个方面的研究工作以伯克利构式语法（Berkeley Construction Grammar, BCG）和基于符号的构式语法（Sign-based Construction Grammar, SBCG）为代表；后一个方面的研究工作以体验构式语法（Embodied Construction Grammar, ECG）和流变构式语法（Fluid Construction Grammar, FCG）为代表。这些工作不仅将构式知识形式化表示的理论研究引向深入，同时也推动了语言知识工程以及自动句法分析的实践。比如 Fillmore 在其主持的语言知识工程“框架网”（FrameNet）项目中开辟了专门针对英语构式的语料库标注（FrameNet Construction）研究（Fillmore et al, 2012），伯克利加州大学计算机系的构式语法研究人员开发了基于 ECG 的句法分析器（Bryant, 2004）。

国外构式语法的这些前沿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”）原有的分析模式。比如 SBCG 就是以中心词驱动短语结构语法（HPSG）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的知识描述。

跟上述“以新代旧”的思路不同，我们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子自动句法分析中，我们认为，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。我们的基本认识是，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，这种自底向上（bottom-up）的研究思路，更有可能实现 1+1>2 的效果。

基于上述思路，为推进构式语法理论在汉语语法领域的研究，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，我们将尝试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¹，并详细描写每个构式的内部构成情况、构式整体的语法、语义、语用属性。这一语言工程实

¹ 数据库中收入的构式主要有以下来源：（1）从已有文献（包括期刊论文、学位论文、专著、工具书等）中搜集；（2）通过人工和半自动的方法从大规模真实语料（主要是 CCL 语料库）中搜集。

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。

本文档对构建现代汉语构式知识库涉及的基本概念及具体的填写内容进行详细说明。下文的组织方式为：§1 节阐述我们对构式的基本认识，给出构式的工作定义，说明现代汉语构式知识库收录构式的标准；§2 节阐述构式知识库设计的指导思想及主体框架；§3 节对构式知识库的数据组织方式及各填写项目的要求和规范做详细说明及示例。§4 节针对构式知识库填写中的常见问题，结合实例的分析给出处理方案。

§ 1 关于构式的基本认识

1.1 构式的界定

汉语语法学界对构式的定义，大多沿用 Goldberg (1995) 的说法: "C is a CONSTRUCTION iffdef C is a form-meaning pair <Fi, Si> such that some aspect of Fi or some aspect of Si is not strictly predictable from C's component parts or from some other previously established constructions." (Goldberg 1995: 4) “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。”从这一定义出发，有关构式的认识通常可以展开表述为：

(一) 构式是形式和意义 (包括功能) 的结合体。

(二) 构式本身能表示独特的语法意义，自身有独特的语义配置方式。

(三) 构式的形式或意义，都不能从其组成成分或已有的其他构式推知，因此构式具有“不可预测性” (unpredictability)。

自从构式作为一个语法单位术语被提出以来，学术界关于构式的内涵和外延的争论就一直在持续。Goldberg 本人在其后续研究中对构式的界定也有所扩展: "Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency." (Goldberg 2006: 5)

“任何一个语言学模式，只要其形式或意义 (功能) 中的某些特征不能从其组成成分或已有的其他构式中严格预测出来，该模式就是构式。此外，如果一个语言学模式的使用频率足够高，那么即便该模式是完全可预测的，它也可以被作为构式看待。”可以看到，新的定义在强调构式“形式—意义”的不可预测性 (跟原定义一致的内容) 之外，又增加了可预测而使用频率高的类型。按照后一种定义，构式的范围无疑会更大一些。

我们认为，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，即考虑语言单位中形式和意义可以从其构成成分的形式和意义推知的情况。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。这样的词组，不需要作为 (特殊的) 构式看待。比如：“张三喝了一杯咖啡”“那衣服很漂亮”等等，就属于这样的比较规则的短语构造。这样的结构单位，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；

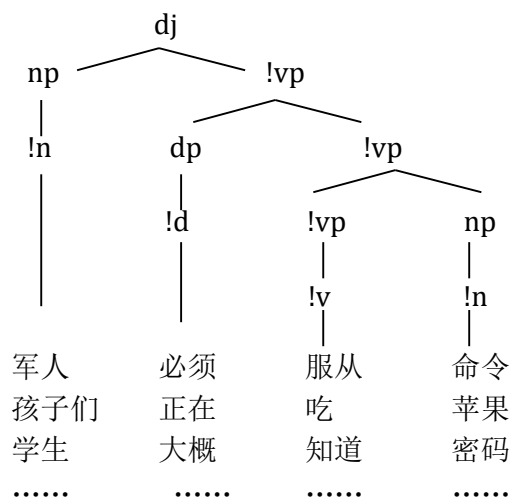
(3) 内部成分的可替换性 (可扩展性) 强；

(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 (dj 代表小句, 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) :

例 1:



在例 1 所示的短语结构树中, 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²。

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 6 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: 他常常为**芝麻绿豆大**的一点小事而长吁短叹。

² 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**”。



例 2 中的“有两个人高”，例 3 中的“芝麻绿豆大”都不大容易按照一般短语结构的组合模式分析，上图中要用二叉树的结构来描述两例的内部层次构造有困难，另外，即便采用二叉树结构来描述，树节点中如何对“高”“两个人高”“芝麻绿豆大”等进行范畴化，也比较困难。而如果从构式的角度来看，“有两个人高”是“有 + np + ap”的模式，其中 np 限定为简单的“m+q+n”（数+量+名）模式，ap 限定为单音节的“高”“大”“深”等少数几个有限的词语（一般词典中标记为形容词 a），或前面加上“那么、这么”修饰的“那么高、这么高”等有限的组合。表达事物的某个方面的量值（如高度、深度、体积等）。“芝麻绿豆大”是“np+大”的组合模式，其中 np 限定为单词（如“绿豆、碗口、米粒”等）或两个双音节名词的并列形式（如“芝麻绿豆”）。表达说话人用比拟夸张的方式描述物体面积、体积或事物重要程度非常小或非常大。

在“现代汉语构式知识库”语言工程的工作框架中，**本文档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**（详见下文 1.3 的说明），**是对常规短语结构语法组合的必要补充。**

1.2 构式与传统语法单位的关系

为了更好地阐述我们对构式的认识，这一小节将构式跟传统语法单位对比，从传统语法单位的角度出发进一步来看构式的性质。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³。宽泛而言，这四级语法单位是从小到大的关系。其中最小的单位——语素——没有内部结构，语素可以组合成词，有不少语素也可以单独成词。通常研究汉语的语法结构规则不大重视语素组合成词的规则（即所谓的构词法），而是以词作为最基本的语法单位，关注词如何组成词组，如何组成句子的规则（即所谓的短语结构规则）。之所以如此，是因为在一般人的心目中，词的长度有限，总的个数（词汇量）也相对有限，可以穷尽性（至少是相对地）罗列，以外延方式给出词的集合（即词典），描述每个词形式和意义之间的对应关系。而词组则是长度无限的（至少是潜在无限），总的个数更是无限的，不适合穷尽性的罗列，需要研究

³ 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）。

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（郭绍虞，1978；朱德熙，1982, 1985），汉语句子和词组的构造大体相同，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。一般把句子按照内部复杂程度的不同，分为单句（简单句/单一句）和复句（复杂句/复合句）两大类。从词组本位语法体系的视角来看，单句和复句都可以再进一步细分，看作是某种类型的词组（比如主谓词组、并列词组或状中词组，等等）。

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 1.1 小节中已做了概要说明，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，但二者作为语法单位，并无截然可分的明确界限，我们更倾向于认为，二者具有连续性。

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
阶梯 教室	沙发 社交
森林 大王	足球 寡妇
水果 大王	胶囊 旅馆
.....

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，前一个名词（n1）是修饰性成分（定语），后一个名词（n2）是中心成分，后者表示整个名词性词组所代表的事物类别，前者是对后者的属性限定，在后者所代表的事物集合中划分出一个子集。比如“花园洋房”，指的是“洋房”这样一个建筑（房屋）类别，其属性特征是“带有花园”。这是这一类词组的共性。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 组的例子是有较长历史的组合。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）。如果之前没有接触过这些实例，很可能无法从两个 n1 和 n2 的字面意义推导出词组整体表达什么意思。比如“输入法”跟“皮肤”可能存在什么关系呢？为什么用“输入法”来限定“皮肤”呢？“足球寡妇”是“寡妇”吗？“胶囊旅馆”大概是“旅馆”，但“胶囊”所指的事物跟“旅馆”能构成什么关系呢？等等。显然，就 B 组这些实例来说，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，我们很难为 n1+n2 组合找到一个一以贯之的通用的释义模板，也就是说，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 组合因其内部语义关系的多样性（复杂性），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（即符合 n1+n2 模式的 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 1.4 小节有关凝固型构式的说明）。但是也需要注意，从理论上讲，类似例 4B 组的实例可能会层出不穷，很难穷尽性地收录。在实践中，如果要把例 4B 中这样的例子作为构式处理，也只能是酌情在构式库中收入一些实例。

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⁴。下面例 6 中尝试对例 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。

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

例 5B 组实例结构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。“男人中的男人”指某人跟同属男人集合中的其他个体相比，在男人的特性程度上更高、更突出；“天才中的天才”指某人跟同属天才集合中的其他个体相比，在天才的特性程度上更突出、更典型。这个语义无法从结构中的任何一个成分、或者这些成分的语义直接组合得出，即语义上具有不可推导性。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）。

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、跟……比赛”的意思；例 7B 组中的“对”是“对仗、对联、配对”的意思。例 7A 组实例的 X 和 Y 都是 np（名词性短语），X 和 Y

⁴ 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。

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（一般是实词性成分），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”或者处理为动词，或者处理为连词。“对”一般都取光杆形式。如果看作是动词，基本不能像一般动词那样附加时体成分（比如“着、了、过、正在”等）或有对应的反复问形式（如“X 对不对 Y”），X 和 Y 虽有一定的扩展性，但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 4-7 的分析，大致可以从对传统语法单位的认识背景上来概括我们对构式作为一种语法单位的理解。下面归纳的这些基本认识将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**，我们把构式看作是大于传统语法单位中的词的单位，收入构式库中的构式至少由两个以上的词组成。许多构式语法学者把词也看作构式，是强调其“形式——意义”配对的一面。如果仅从“形式——意义”配对这一面来界定构式的话，那么，传统语法单位中的词确实应当算是构式。但是，在构式知识库的工作中，我们的倾向是有意强调构式与传统语法单位中的词的区别（正如我们也有意强调构式与一般短语结构的区别）。这个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，既不同于常规短语结构，也不同于一般的词。对于一个构式，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，内部都包含了若干词项，而整体语义并不是内部这些词项语义的简单相加。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。更多实例可参见下文 1.4 小节关于凝固型构式跟传统语法单位中词的关系的分析说明。

（2）**构式基本无递归性**（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**”，

它们组成一个完整的整体，谁离开了谁，单独都不再成其为构式⁵（至少不再是原先的构式）。

1.3 构式的内部成分及其构造

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（参见例 1 所示的树状图）。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（可参见陆俭明 2009b，苏丹洁 2010，苏丹洁、陆俭明 2010 等）。不过，关于“语块”的理解，学界在不同发展阶段有过不同的看法，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（2013）指出：“在理论探讨和实践分析的过程中，学界就语块作为一种特殊的语言交际单位其内涵在下面这些方面基本达成了共识：就其结构而言，语块是由连续或不连续的词语或其他有义元素整合而成，具有**实体性、模块性**；就形义关系而言，语块作为一个整体，在形义关系上具有共时上的不可分析性或较低的分析性，即**形义整合性**，每个语块都有特定的语用功能；就表达系统而言，语块不是在语言交际中通过语法规则临时生成的，而是早已储存在大脑中的，具有**很高的预制性**；就交际过程的即时加工而言，语块具有**易于提取性**，对学习者和交际过程而言，语块具有鲜明的**整存整取特征**。这是对语块基于交际性质的内涵的最基本的理解。”在薛小芳、施春宏（2013）中，语块被看作是“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。

显然，从薛小芳、施春宏（2013）对“语块”的上述认识出发，构式的内部成分并不适用“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 3 个语块构成，其中每个语块的名称很显然都是指的语义范畴；再如把表达容纳量义的结构表示为“容纳量-容纳方式-被容纳量”（一锅饭 吃了 十个人），也是由 3 个语块构成，其中各语块表达的也是语义范畴（陆俭明，2012）。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（比如词类）。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

⁵ 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 —— 村口一株老槐树”“一锅饭吃十个人 —— 一锅饭十个人”等等。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），因而也可以处理为一个新的构式。

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（参见詹卫东，2005）。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。上面的存在义构式和容纳量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。再回过头来看容易拆解为语块（组块）的构式，如果基于这种内部构成描述方式，只有先识别出构式整体，才能“自上而下（top-down）”地分析出该构式的各内部语义块，而从组装的角度看，当一开始扫描到“一锅饭”时，并不能识解出这个单位就一定对应着表“容纳量”义的语块，即便接下来扫描到“吃了”，也可能再接上某些表数量的后续成分后，形成的整句可能是“一锅饭 吃了 半锅，还没吃饱……”这样的并不表达容纳量义的句子，从“自底向上（bottom-up）”的角度来看构式的分析，用“语义组块”的方式来描述构式的组成，对于识别语句中的构式（至少是计算机的自动识别），并无显著的优势。

基于上面的认识，对于构式的内部成分的描述，我们倾向于不以语块（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。下面的图示可以说明对于构式内部成分的两种可能的描述方式：

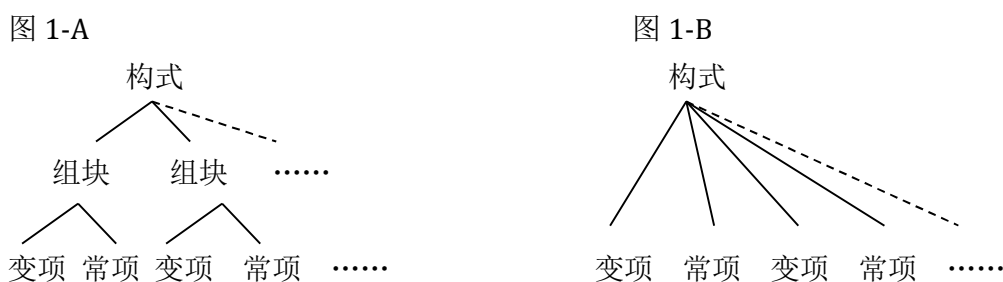


图 1-A 是以“组块（语块）”作为描述构式内部成分的单位，但如果要进一步描述组块的内部构成，仍然涉及到变项和常项等语法单位。图 1-B 直接以变项和常项作为描述构式内部成分的语法单位，充当构式内部变项成分的单位则用传统短语结构语法体系中常用的功能范畴，如 n、v 等词类以及 np、vp 等短语功能类表示。因此，从分层描述语法单位的思路讲，图 1-A 的架构代表了在观念层面表示构式内部组成状况的一种可选的模式，但从工程实践角度讲，大量的构式属于“一个构式对应着一个组块”的情况，内部或者不大容易或者没有必要再拆解为不同的组块。对于那些内部能拆解为不同组块的构式，也主要是从语义范畴的角度为组块定性，要详细描述组块本身的构成，还是从变项和常项成分的角度来界定更为直接，因为这些组块跟短语结构不同，没有递归性，这就意味着设置“组块”这样的中间单位意义不大。

以“你唱你的戏”为例，下面图 2A-C 展示了不同的内部结构表示方式（dj 表示“单句”，一般内部为主谓结构关系）：

图 2-A

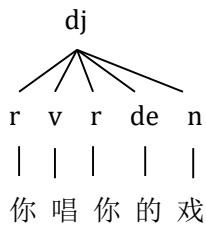


图 2-B

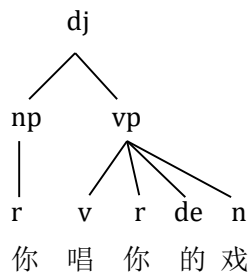


图 2-C

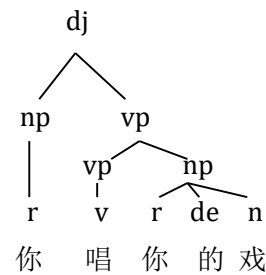
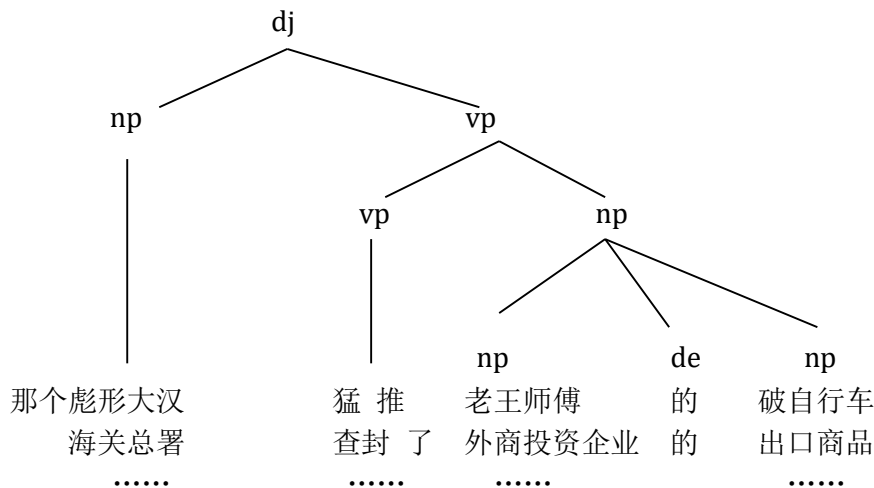


图 2-A 把构式内部成分看作是线性序列组合，由 5 项成分一次组合成型，不需中间结构单位；图 2-B 把“你唱你的戏”看作是先由两个组块构成，其中 vp 组块再由“v+r+de+n”这 4 项组成；图 2-C 是短语结构的层次嵌套组合模式，其中 vp、np 都具有递归性（即组成成分中包含了自身）。图 2-C 代表的结构模式很容易产生出图 2-D 所代表的短语结构组合实例，这些实例已经属于常规短语组合，不再是“你唱你的戏”这样的构式了。

图 2-D



短语结构的内部构造之所以采用树状层级结构，是因为：（1）内部成分之间，有的地方比别的地方更容易插入其他成分；（2）内部成分可以替换为同范畴的更大单位（递归）；（3）内部成分可能改变语序。为了描述这些句法操作造成的变化形式，就有必要对短语结构采取层级划分的方式描述其内部成分之间在发生组合关系时的亲疏差异。而对于构式来说，其内部无法插入其他成分，无法递归，无法改变语序，因此，也就没有必要像短语结构那样采取层级分块的范式描述其内部组成，即不需要假设中间语法单位“组块”（语块），只需要简单地罗列其组成成分的线性序列，就可以表示构式的“一次性压合成型”的构造特点了。

基于上述认识，我们在具体表示构式的内部构成形式时，采用图 2-A 所示的简化模式，即构式由常项和变项的线性组合来表示（关于“构式形式”的格式规范，参见 3.1 节说明）。在构式与其组成项成分之间，不再假设中间层级单位。值得说明的是，大多

数构式就是一个独立的整体，或者说是“单组块构式”。不过也有少数的单组块构式可以复用，形成内部为并列关系的“双组块构式”⁶。即使是对“双组块构式”，在描述其内部构成形式时，也仍然可以采用变项加常项的线性序列表示方式，而不引入“组块”这一中间层次，代价是构式的形式表示可能是一个比较长的符号序列。表 1 中给出了单组块构式和双组块构式的构式形式表示的示例：

表 1:

	单组块构式	双组块构式
构式实例	爱去不去 爱吃不吃 爱听不听	他跳他的舞，你唱你的戏 你唱你的，他写他的 你瘦你的，我胖我的
构式形式 (构式模式)	爱+v+不+v	r1+v1+r1+的+n1 , r2+v2+r2+的+n2 r1+a1+r1+的 , r2+a2+r2+的

上表中“单组块构式”用“爱+v+不+v”表示该构式的组成形式，其中“爱、不”为该构式的常项，前后两个同形的动词 v 是变项。表中双组块构式的两个组块有完全相同的序列，除表中所列的“r+v+r+的+n”和“r+a+r+的”外，还可以有“r+v+r+的”（比如“你唱你的”“你游你的”）形式。这个双组块构式内部两个组块之间就是简单的并列关系，在描述其构式形式时，也跟单组块构式一样，按照线性顺序列出了全部变项和常项，没有涉及组块这一中间层次。

按照上述分析框架，前文举过的表“存在义”的和表“容纳量义”的构式，在构式知识库中就不是采用语义组块的方式来记录其内部构成，而是采用常项-变项模式来记录。比如“存在义”构式可以记录为“sp+v+着+np”（还有变体形式“sp+v+了+np”），其中 sp 代表处所性短语，np 代表名词性短语。“容纳量义”构式可以记录为

“m1+q1+n1+v+m2+q2+n2”（有变体形式 m1+q1+n1+v1+了+m2+q2+n2），其中 m、q、v、n 分别代表数词、量词、动词、名词。这两个构式的内部成分均全部用变项表示，没有常项。对于各个变项，再进一步描述需要满足的约束条件（详见下文 3.1-3.4 节的说明）。另外，有一些构式可能内部成分较多（特别是有一些带有明显修辞色彩的构式），但仍然属于“单组块”构式类型。例如：

- (1) 有人问张瑞敏成功的秘诀是什么？他的回答是“**第一是创新，第二是创新，第三还是创新**”。
- (2) 总书记讲了三大任务，**第一是学习，第二是学习，第三还是学习**。
- (3) **走直线，走直线，走直线，重要的事情说三遍**。

上面例 1-2 对应的构式可以记录为：第一+是+X，第二+是+X，第三+还是+X

⁶ 这只是一个假设。从历时的角度讲，也可能是先出现双组块构式，然后省略其中一个组块，造成单组块构式。构式知识库目前的描写内容仅限于构式的共时表现，没有涉及构式的历时发展（包括来源和形成机制）问题。

例 3 对应的构式可以记录为： X, X, X, 重要+的+事情+说+三+遍

这两个构式具有相近的表达功能，即强调在某个议题中，X 是非常重要的。构式中变项 X 的具体语言形式（语法范畴）不确定，可以由较多功能类充任（比如上面例中的 n、v、vp 等）。虽然这些构式中内部包含的成分比较多，但仍然是整体构成一个组块。组块内的常项和变项是相互依赖的关系，整合在一起，共同表达一个独特的语义。

1.4 构式的分类

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。

Croft & Cruse(2004) §9.4 曾以跟传统语法单位对应的方式给出了构式的类型及示例：

表 1: Croft & Cruse(2004)的构式类型及示例

Construction type	Traditional name	Examples
Complex and (mostly) schematic	syntax	[Sbj be-tns Verb -en by Obl]
Complex, substantive verb	subcategorization frame	[Sbj consume Obj]
Complex and (mostly) substantive	idiom	[kick-tns the bucket]
Complex but bound	morphology	[Noun-s], [Verb-tns]
Atomic and schematic	syntactic category	[Dem], [Adj]
Atomic and substantive	word/lexicon	[this], [green]

Goldberg(2006)也展示了不同构式的复杂度（complexity）和抽象度（abstraction）由低到高的连续变化情况。下面是 Goldberg(2013)对构式单位分级的概括。

表 2: Goldberg (2013) 的构式类型及示例

Construction	Examples
Word	Iran, another, banana
Word (partially filled)	pre-N, V-ing
Idiom (filled)	Going great guns, give the Devil his due
Idiom (partially filled)	Jog <someone's> money, <someone's> for the asking
Idiom (minimally filled)	The Xer the Yer (The more you think about it, the less you understand)
Ditransitive construction	Subj V Obj1 Obj2 (He gave her a fish taco. He baked her a muffin)
Passive construction	Subj aux V Ppp (PPby) (unfilled) (The armadillo was hit by a car)

（上表跟原文不完全一致，引用时根据排版需要形式上略有调整，不影响内容）

上述文献中对构式类型的认识，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。相对而言，我们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，我们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 3：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

下面给出各类构式的一些具体实例：

表 4：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)
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) 一+q+比+一+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，就是在外国也享有盛誉。 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3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4音节（也有人称为“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，如果从传统的结构层次分析的角度来看，内部可以二分，但分出来的两个部分是否各自成为相对独立的组块，则不一定。像上文举过的“你唱你的，他写他的”构式，前后两个组块就相对独立，可以单用。但表4给出的两个复句型构式的例子，内部两个部分之间有很强的相互依赖关系，因而不宜看作独立的组块。像这样的复句型构式，整体应看作一个组块。

1.5 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。下面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 1:

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，比如：

(1) **看见你的声音** （“声音”跟“看见”搭配，从听觉现象转为视觉现象）

(2) **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：

(3) **基层干部很忙很辛苦，工作是 5+2，白加黑，**礼拜六保证不休息，礼拜天休息不保证。****

(4) **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。**

(5) **重庆老汉一根扁担“挑出”3个大学生**

(6) **中学生踢球“踢”进北大 全国仅两人**

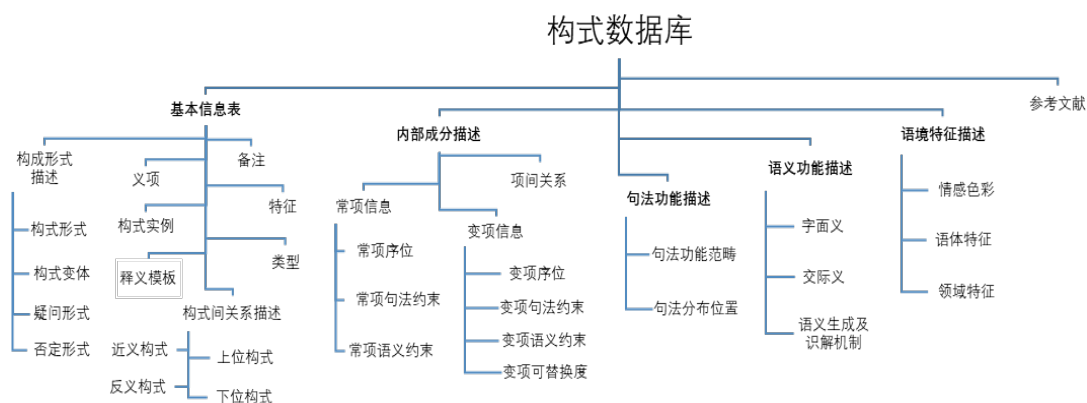
(7) **华罗庚说过，读书有两个过程，先要**把书读厚**，这是消化的过程，然后再**把书读薄**，这才是理解的象征。**

像上面例 3-7 这样的表达形式，在目前我们的构式知识库实践中，只能暂时看做纯粹的修辞现象，不大适合作为构式来加以描写。

§ 2 构式知识库的设计概要

上一节从几个方面阐述了我们对于构式的认识，基于这些认识，可以形成收录构式条目的操作标准：（1）构式内部成分有组合性（即可分解出更小的语法单位）；（2）构式内部成分无递归性（即可采用线性组合方式描述）；（3）构式“形式——意义”超常规配对；（4）有的构式具有独特交际价值（特别是带有修辞色彩）。按照以上标准将满足条件的语法单位（作为构式）录入数据库后，需要对构式进行尽量全面的描述。对个体构式而言，是深入认识该构式的意义和用法；对语言系统而言，则是希望像短语结构系统那样，能够在一定程度上把离散的构式相互关联起来，形成一些构式系统。

下面是构式数据库的结构框架示意图。



构式知识库中对每条构式的描述信息主要可以分为四个方面：

（1）关于一个构式的基本信息，包括该构式的形式方面和意义方面的各种特征。

（2）一个构式的相关构式，包括一个构式的变体形式，一个构式的同义构式、反义构式，等等，这部分相当于是关于一个构式的“聚合”（跟聚类相关的）特征。

（3）构式在使用时表现的特点，主要是构式作为一个整体，在参与组合时，在句中可能承担的语法功能⁷。这部分相当于是关于一个构式入句时的“组合”特征。

（4）构式的语用、语境信息，包括对交际参与者（说话人、听话人）的约束、语体约束、句类约束、领域约束等等，即一个构式在交际中依赖的“环境”特征。

下面分别说明各部分内容涉及到的主要项目。下文第3节将进一步对每个具体项目在数据库中对应的字段设置及填值要求等做详细说明。

⁷ 对于句法上相当于体词性成分的构式（如“不是办法的办法”），可能会像一般的 np 短语那样，在句中充当谓词的论元角色。对于这样的“动态”语义功能的描写，可以在构式专项语料库中标注。在构式知识库中仅描写关于一个构式的相对静态（稳定）的信息。

2.1 构式的形式与意义

2.1.1 构式形式：常项与变项的概括度及精细度

构式库中记录构式，首先碰到的问题就是如何表示一个构式。对于凝固型构式，因为无变项成分，可以像词典中收录词语一样，直接以构式实例本身来记录。而对于半凝固型构式、短语型构式、复句型构式，涉及到变项成分，如何表示变项成分，需要遵循一定的规范。为保持跟短语结构语法分析的最大兼容性，构式中的变项成分尽量采用传统的词类范畴（如 **n** 表示名词、**v** 表示动词等）和短语功能类范畴（如 **np** 表示名词性短语、**vp** 表示动词性短语等）来表示。如果一个变项成分可以由词组（短语）充任，则选用短语功能类范畴表示变项，如果一个变项成分只能由词充任，则选用词类范畴表示变项。对于无法确定语法范畴的变项成分，可以采用 **X**、**Y** 等通配符号表示（注意 **X**、**Y** 均采用大写字母，以跟采用小写字母的词性标记和短语标记相区别），然后再具体描述 **X**、**Y** 可以选择的语法范畴集合。一般来说，用 **X**、**Y** 等表示的变项成分，既可以是单个词充当，也可以是短语充当，而且所属的语法类型通常有多种情况。在构式形式中，相同的变项符号，表示构式实例中采取的是同形的语言成分，如果是不同形但同类的成分，可以在变项成分符号后加数字 1、2 等加以区别，如 **X1**，**X2**；如果不同形又不同类，则采用不同的符号加以区别，如 **X**，**Y**。参见 3.1.1 的说明。

下面是一些构式形式的示例。

表 1:

构式形式	构式实例	说明
a+着+呢	高着呢 早着呢 好玩儿着呢	这个构式只有一个变项成分，且由形容词充任，用 a 标记
v1+了+v2+ v2+了+v1	吃了喝喝了吃	这一构式有 4 个变项成分，由两个同形的 v 充任，分别记作 v1 和 v2
都+是+X+惹+的+祸	都是月亮惹的祸 都是相亲惹的祸 都是我乱说话惹的祸	这一构式的变项成分 X 可以是 n 、 v 等词语，也可以是像“我乱说话”这样的主谓结构，因此用 X 标记
左+一+个+X，右+一+个+X 左+一+个+X1，右+一+个+X2	左一个电话，右一个电话 左一个主义，右一个流派 左一个对不起，右一个很抱歉	这一构式中两个变项成分可以同形，也可以不同形，且既可以是词，也可以是短语，因此用通配符 X 标记

值得注意的是，有的构式中的变项成分是引语用法⁸，即重复上文中已经“说”过的成分，这样的变项成分往往可以由不同范畴的语言成分来充任，因而只能用 X 来标记，比如上表中的“左+一+个+X1，右+一+个+X2”，该构式的实例中“左一个对不起，右一个很抱歉”中的“对不起”“很抱歉”就是这种情况，是重复上文（或语境）中已经出现过“对不起”“很抱歉”。表中另一个实例“左一个电话、右一个电话”则是歧义的，既可以指不断有人打电话来，也可以指有人不停在言谈中提到“电话”这个词。后一种情况，在书面上，应该写作：左一个“电话”，右一个“电话”。但因为汉语书面语中标点符号使用方面比较灵活，有可能书写者在引语用法时把引号略去了，造成同形歧义。具有引语用法变项的构式往往也有同形的非引语用法变项，除这里举的“左一个 X，右一个 X”例子外，不妨再举两个例子：

- (1) v 来 v 去（跑来跑去） X 来 X 去（你来你去、啊来啊去）
 (2) v 呀 v 的（眨呀眨的） X 呀 X 的（你呀你的、死呀死的）

上面给出的是构式形式的正例，下面再看一些错例，其中少部分例子是构式形式表示有问题的情形，大部分例子属于不合作为构式收入构式库的情形。

表 2:

构式形式	构式实例	错误情况说明
np+v+着+vp	国产冰箱用着放心 臭豆腐闻着臭 题看着很多	构式实例中“v+着”后的成分并不是 vp，而是 ap
np+可+餐	秀色可餐 美色可餐 秋色可餐	构式变项成分并不是 np，而是有限的几个 n
mq+比+mp+a	一部比一部气人	构式中表数量的部分应该是“一+q”，而不应该记作 mp。因为构式中的数词仅限于“一”，应作为常项成分处理，不应看作变项成分的一部分
打+死+我+也+不+vp	打死我也不回去上学	构式实例中“我”可以替换成其他代词（如“他”）或名词（如“小编”），因此“打死”后的成分应处理为变项，用 n、r 等分别标记
vp+的	必须的 绝对的	“必须的、绝对的”可以处理为凝固型构式，不含变

⁸ 具有这种用法特点的构式数量有多少，在变项成分的约束方面是否存在某些共性，需要做更全面系统的考察。

		项。其中的“必须”“绝对”在常规短语组合中都是副词（并非 vp）。“必须的、绝对的”应分别处理，二者并无统一的释义模式，不应看作一个构式的实例
看+不+得+vp	看不得老人这样辛苦	实例中“老人这样辛苦”是主谓结构（单句），不是 vp，构式形式中用 vp 标记变项不准确，应该用 dj 标记
v+r+个+np	告诉你个秘密	“告诉你个秘密”是双宾语构式的实例，如果用“v+r+个+np”来表示双宾语构式，概括面不够。双宾语构式的语义抽象度过高，暂不收入构式库
v1+np+v2	找粥喝	“找粥喝”属于普通的连谓结构，不应处理为构式
vp+有木有	三个月一次有木有 通宵看书有木有	实例中“三个月一次”不是 vp，构式形式中用 vp 标记变项不准确。“有木有”属于网络语言中的用法，可以作为新词收入词库，也可以处理为凝固型构式
X+灾+Y+祸	幸灾乐祸 天灾人祸 躲灾避祸	这些构式实例并无统一的意义，尽管形式上有“X+灾+Y+祸”的“表面共性”，但缺乏一个统一的整体构式义，整个结构也不属于同一个语法类（“天灾人祸”是 np，“幸灾乐祸”是 vp）因而不应看作一个构式。这些实例可以看作是独立的词项收入词库
a+X+a+Y	美轮美奂 呆头呆脑 古色古香	错误情况同上
n1+多+n2+少	僧多粥少 凶多吉少 苦多乐少	构式实例中的变项实例并不都是名词，因而构式形式

		中的变项不能仅用 n 标记。再进一步，整个格式并没有比组成成分简单相加多出的语义，不应作为构式收入
np+np+np+np+什么的	文艺青年美少男传奇人物伟人什么的	实例中确实有 4 个 np,但理论上这里 np 的个数可以是 1、2、3、4、5、……多种情况，用“np+np+np+np+什么的”表示构式不适合概括这种有不定个数 np 连续出现的情况（相当于有嵌套）。这种情况适合把“什么的”处理为助词（类似“等、等等”），“np+什么的”处理为一般的短语结构，而不是构式
n1+只是 X ， n2+才是+Y	璀璨只是一瞬，幻灭才是永恒	格式难以概括出统一的语义，不应看作构式
要么+n1+a1 ， 要么+n2+a2	要么物质强大，要么精神强大	“要么……要么……”是已经定型的关联成分，可以按照一般普通的复句关系分析，不应收入构式库

通过上述示例，可以看到，构式形式的表示主要反映了人对构式变项和常项的概括认识。对于语言单位（成分）的概括，总是面临概括过度（*overgeneralization*）和概括不足（*undergeneralization*）两个问题。在具体实践中，如果二者需要权衡，那么宁愿倾向于概括不足，即为了追求精确性，可以丧失一定的覆盖率。比如，构式变项的设置一定要考虑如何进一步对其进行约束（具体的约束项目设置见下文 3.2 小节的说明）。如果一个变项缺乏有效的约束手段，则要谨慎考虑是否设置该变项。这种情况下，变通的做法就是增加常项的数目，或者完全将变项改为常项，以丧失覆盖率为代价，换得对一个构式的精确表示（将原先的非凝固型构式改为凝固型构式了）⁹。

⁹ 这里的处理策略有两方面的考虑：（1）构式库收入构式的首要目的是为计算机信息处理服务，希望将来能够将真实语料中的“构式”跟一般的常规“短语”的分析结合起来。对计算机的应用来说，追求匹配精度，比追求概括的广度更有实用价值。如果是为了语言教学服务的目的，则会适当地向追求概括度倾斜；（2）构式形式表示从精确的形式向概括的形式泛化，更容易一些。相反，如果要从概括的表达形式向精确的形式细化，在操作上要困难得多。因此，在构式库建设的初期阶段，优先考虑构式形式表示的精确度，丧失一定的概括度，从工程角度，是更为合理的选择。

2.1.2 构式“一形一义”与“一形多义”

下面再来讨论构式的意义描述问题，特别是构式的多义性问题。

构式的语义表示大致可以分为简单和复杂两类情况。简单的情况指一个构式的语义是单一确定的，可用一个对应的释义模板（基于构式形式的变换表达式）来表达。例如：

表 3:

构式形式	构式实例	构式释义
a+着+呢	高着呢 早着呢 好玩儿着呢	a 的程度很高 很 a
a+不+到+哪里+去	高不到哪里去 好玩儿不到哪里去	a 的程度不高 不怎么 a
v1+了+v2+ v2+了+v1	吃了喝喝了吃	不停地 v1v2（没完没了）
零+v	零污染 零排放 零容忍	完全不 v 绝对不 v

复杂的情况指一个构式的语义包含的内容较多，甚至可以看作是有多个义项的，无法用一个简单的释义模板加以概括。语义内容较多的构式一般除了构式字面触发的意义解释，还涉及到该构式的使用语境以及社会常识等，才能对该构式的完整意义做出较为详细的说明。例如：

表 4:

构式形式	构式实例	构式释义
被+X	被自杀、被吸烟、被下岗、被怀孕 被奥数、被间谍、被“高铁” 被正态、被平均、被幸福、 被“满意” 被“67%”、被第一 被没有资格、被涨工资	句子形式：S+被+X 句子语义： X 对应的事件记作 E(x)； E(x) 的隐性施为者记作 A； 说话人记作 B； (1) A 断言 S 是 E(x) 的主动行为者； (2) B 认为 S 没有 E(x) 的意愿，甚至 S 的实际状况为 E(x) 的对立面事件，比如“（断言）自杀（实为）他杀” “（断言）没有资格（实为）有资格”，因而 A 的断言并非真实情况； (3) B 的交际意图在于揭示：A 之所以如此断言，有某种趋利避害的动机，该动机导致的这种断言行为使 A 获利，使 S 受损。

构式的释义可能会涉及到字面义 (literal meaning) 与交际义 (real meaning) 两个层面。其中“字面义”指一个构式的实例仍然可能按照普通短语结构组合方式解读时表达的意义；“交际义”指一个构式的实例在使用时通常具有的特殊语义解读。比如“a+哭+了”构式 (美哭了, 帅哭了, 萌哭了), 字面义是按照一般述补结构的模式来释义, 即“因 a 而哭” (比如: 因为见到美的事物而哭), 实际的交际语义则是“太 a 了”, 即“美哭了”是“太美了、美的程度极高”的意思¹⁰。也有的构式, 可能因为用法比较固化, 绝大多数实例只能有交际义解读, 不再能还原出字面义解读, 如“a+着+呢”构式, 一般用例都是解读为“很 a, a 的程度很高”的意思。只有极少数的用例 (如“汤还热着呢”) 还有可能还原出字面义解读, 即“a 所表达的性状在持续”这一语义。

构式的语义还有概括度的区别, 一个意义可以细分为几个有关联的下位意义。从这个角度讲, 构式可以有多义性。例如: “什么 X”构式就有两种语义解读 (或者说是两种用法)¹¹:

(1) 元语否定

例 a 甲: 你昨天在电影院怎么搂着个女人?

乙: 什么“女人”! 他是个男的!

例 b 甲: 你昨天在电影院怎么搂着个女人?

乙: 什么“女人”! 那是我老婆!

(2) 负面评价

例 c 这什么女人! 脾气这么大!

构式“什么 X”的“元语否定”用法是指出说话人断言中的 X 表述不恰当, 应该换成其他更得体的说法; “负面评价”用法则是指言谈中所指的某个具体的 X 很糟糕、很差劲。前一种用法下的 X 是引语性质的语法单位, 不限于 np 成分, 可以是其他范畴的成分, 例如:

例 d 什么漂亮, 她太难看了

什么漂亮, 她简直是倾国倾城

什么带薪休假, 全是假的。

后一种用法下的 X 仅限于 np 成分, 不妨再多看两个例子:

¹⁰ 比较: (1) 小汤姆把我萌哭了。(2) 小汤姆真是萌哭了。例 1 中“萌哭了”可以按照字面义来解读, 即因为小汤姆太萌, 因此我 (开心/激动) 哭了。例 2 中“萌哭了”则表示“太萌了”的意思。小汤姆并没有“哭”。

¹¹ 多义词的多个义项之间一般有本义、引申义、基础义、常用义、罕用义等等不同。在词典释义中, 往往通过义项顺序安排 (编号不同), 来体现不同义项之间的地位差异。在更为深入的词义分析文献中, 除列举一个多义词静态的义项区别外, 还需要解释不同义项的关联、来源。在构式知识库的具体描述层面, 构式多义性的呈现方式, 也与多义词类似, 采用“义项”来表示 (参见 3.1.3), 对于义项之间的关联, 在“构式语义识解机制”字段中描述 (参见 3.6.3)。这里所举的“什么+X”构式的两个义项, 有可能都来自“什么”的疑问用法 (比如表“负面评价”义) 的用法, 可能是由“疑问”到“质疑”再到“批评”的语义发展脉络)。这两个义项本身, 不一定有直接的引申关系。

- 例 e 什么破电脑，电源都没有！
什么乱七八糟的，全部重写！

2.1.3 意义相近的构式

有一些构式形式上不完全相同，但表达的语义比较相近，如果把这些构式放在一起考虑，也涉及到如何确定构式形式的问题，即在构式形式的概括度和精细度之间，需要做一些权衡。下面先看一组例子：

- 例 1) 我看**都没看**怎么写意见？
 2) 他**连看都没看**，就同意盖上县政府的大印
 3) 老师**看都没看**一眼，拿过来，顺手甩到讲桌里
 4) 这种国产新型装甲车，梁永葆过去**连看都没看到**过，不要说指挥了
 5) 林珠根本**看都不看**秘书一眼
 6) 这世界上吃穿用的东西可真多，**看都看不过来**
 7) 非有博士的水平不行，我是**看都看不懂**，别说往上填了
 8) 这个姓马的老儿，弄了一大车珍宝都据为己有，我们**连看都看着**
 9) 大楼挡住了他的视线，**看也看不远**
 10) 这几天他回来很晚，**我看也看不见**他，我有啥办法！”
 11) 别人别说想要，**连看也看着**。
 12) 他们**看也不看**望远镜，就肯定“这玩意儿并不十分可靠”

观察上述例子不难发现，例中加粗部分形式上不完全相同，但表达的语义非常相近，都有否定发生某种行为或实现某种行为结果的意思，同时蕴含了一个字面外的意思，就是该行为或行为结果按常理是很容易或者应该发生（实现）的，但由于某种原因，实际上却没有发生（实现）。例中实例对应的构式形式可能的表述包括（但不限于）：

表 5:

v+都+没+v	v+都+不+v	v+都+没+vp	v+都+不+vp
v+也+没+v	v+也+不+v	v+也+没+vp	v+也+不+vp
连+v+都+没+v	连+v+都+不+v	连+v+都+没+vp	连+v+都+不+vp
连+v+也+没+v	连+v+也+不+v	连+v+也+没+vp	连+v+也+不+vp

如果要强调上述构式的联系，那么上述构式形式可能可以抽象出一个更为概括的模板：**连+vp1+d1+d2+vp2**（记作 C1）。在这个模板中，构式的形式可以进一步细化描述为：第 1 项“连”可以省略，第 2 项 vp1 均由单个动词充任，第 3 项 d1 由副词“都”或“也”充任；第 4 项 d2 由副词“没”或“不”充任，第 5 项 vp2 可以是单个动词，也可以是述宾、述补等动词性词组。不过，这样描述模板中各项目的变化形式，仍未能概括以上例 1-12 的所有情况，比如，d2 表否定的“没”和“不”可以并入 vp2 中，形成的构式形式为：**连+vp1+d+vp2**（记作 C2），对应的实例如例 8、11。此外，上述

例 1-12 也只是反映了这类语义相近构式的部分情况，还有一些形式没有体现出来，比如：

- 13) 他的心就像铁打的，不要的就是不要，**连看一眼都不看**
- 14) 湖南卫视所有节目都没看过！**一眼都不看**
- 15) 这么好的木材，您怎么**一眼都看不上**呢？
- 16) 这里的男人们，却连看都没有看过她，甚至偷偷的**看一眼都不敢**
- 17) 那种地方，探出头去**看一眼都够人受的**
- 18) 华山栈道**看一眼都脚软**

将例 13-18 跟上面的例 1-12 联系在一起考虑的话，如果想突出这些实例之间的语义共性，就可能把它们所对应的构式形式进一步概括为：**vp1+d+vp2**（记作 C3）或者 **X+d+vp**（记作 C4）。其中的 d 副词限于“都”“也”。构式形式的表示从 C1 到 C4，是不断泛化的一个过程，覆盖的同类实例越来越多，但与此同时，基于构式形式能够生成的构式实例也越来越可能无法再适用同一个语义解释。C1 构式的组成成分最多，因而相对比较容易限定其中 vp1 跟 vp2 的中心动词相同，并且通过指定 d2 为“没、不”，vp2 为肯定形式等约束条件，将 C1 构式能够生成的实例限制在比较有限的范围内。到了 C4，X 中可以包含“连”，也可以不含“连”，X 的语法范畴可以是 vp，也可以是 qp（比如例 14、15 中的“一眼”），或者不在上面例句中的 np（比如“一部电影都没看”），甚至 X 可以对应两个范畴（比如“一个字他都不认识”）；构式中 d 后的 vp 可以是否定形式也可以是肯定形式（比如例 17、18），如果按照 C4 的方式来描述这些构式实例，那么 X 和 vp 之间的相互约束就很难表示了。

上面例子展示了从实例到构式形式的一个可能发生（但并不恰当）的概括过程。如果强调从意义的角度把表义相近的构式关联起来，可能导致的一个后果是能够系联的构式形式越来越多，原先概括的意义虽然能够在这些构式中或多或少找到¹²，但在语义细节方面可能很难完全覆盖所有这些构式。因此，还是应该优先保证从形式出发，尽量做到单个构式形式和意义之间的精确配对，然后再考虑多个构式形式因语义相近而存在的联系（参见 3.1.2 小节对“构式变体”和 3.1.11 小节对“近义构式”字段的说明）。这仍然是体现了强调精确度优先于概括度的原则。对于上面的例子而言，试图用 C4 来作为构式形式概括所有这些实例是不可取的。描述这些实例的构式形式应该采用表 5 所示的比较精确的表达形式，同时也需要 C1、C2 等相对概括一些的表达形式，这样的做法

¹² 如果要把这一小节所举的例 1-18 对应的构式形式概括为 C4，甚至更为抽象的“**X+都+Y**”（C5）。这一“构式”的抽象语义可能可以表述为：**X+Y** 对应的命题记作 P，P 可以激活一个命题集合（记作 G）。在 G 中的各个命题可以按照发生可能性大小排成一个级差序列，其中，P 所对应的事实是最不可能或最不应该发生（出现）的，但现实情况是 P 所对应的事实发生了。因此，说话人用这个构式强调了发生 P 所对应的事实具有“超乎常情（常理）”的性质。这样概括的抽象语义或许可以覆盖构式所对应的各个具体实例，但是，这样处理的代价（缺点）也是很明显的，即 X 和 Y 之间难以描述其约束条件，X+Y 对应的命题 P 过于抽象，很难从具体实例的表层符号形式出发转换为具有可操作性的形式语义表达。换句话说，这么高抽象度的构式形式概括，对人的认知理解，可能是有效的，但对于计算机处理相关的构式实例，则过于抽象，不易操作。

是兼顾了构式形式表示的精确性和概括性，能够相对比较容易并且比较准确地描述构式内部成分的约束条件。

2.1.4 同形构式

跟构式的多义性相关的现象是同形构式的问题，即两个构式形式上相同，但语义上并无关系。例如：“不+v1+不+v2”构式，就有两种语义，“不 v1”和“不 v2”之间可以是并列关系（表 6 示例），也可以是条件关系（表 7 示例）。而且即使同为并列关系，语义解读时也还有不同的情况。具体的构式实例的语义要依赖变项 v1 和 v2 的关系。比如表 6 例中的“增”和“减”有中间状态；而“死”跟“活”客观上无中间状态，就滋生出更多的字面外的意思。

表 6:

构式实例	构式语义
(1) 不增不减	保持原量不变
(2) 不死不活	虽然活着，但跟死了也差不多，状况不如人意
(3) 不哭不闹	没有哭闹（发出噪声），很安静
(4) 不知不觉 不折不扣	没有觉察（意识）到 不打折扣、完完全全

条件关系的例子如下表。

表 7:

构式实例	构式语义
(1) 不破不立	如果不先破旧，就不能立新
(2) 不敲不响	如果不敲，就不会响

“不…不…”表示条件关系的构式实例中变项位置可以是像上面这样的单音节动词（这个跟表示并列关系的情况一样），还可以是形容词或动词短语，例如下表。

表 8:

构式实例	构式语义
(1) 不去不 礼貌	如果不去，会（让人觉得）不礼貌
(2) 不看不是人	如果不看，就不是人

从上述示例中可以看到，构式的“形式表示”和“意义解释”可以说是一个硬币的两面，无法截然分开，必须联系在一起加以考虑。《现代汉语词典》（第 6 版）有“不…不…”条目¹³（相当于一个词），并分立了 3 个义项来释义¹⁴，但同时也以单立词条形

¹³ 有学者把《现代汉语词典》中的这类条目称为“待嵌格式”。（参见周荐，2001，现代汉语词典中的待嵌格式，载《中国语文》2001 年第 6 期；孟祥英，2014，《汉语待嵌格式研究》，齐鲁书社）

式收录了“不哼不哈、不郎不秀、不伦不类、不蔓不枝、不偏不倚、不三不四、不痛不痒、不温不火、不瘟不火、不闻不问、不折不扣、不知不觉”等十余个符合“不…不…”格式的实例。含有变项的构式形式可以反映构式一定程度上的能产性，但对释义的概括性提出了较高的要求。不含变项的凝固型构式，则可以像一般的词语释义一样，做到比较准确。在构式知识库收录一个构式条目时，首先就要考虑如何表示该构式的形式，是采用常项加变项的方式，还是直接以全部常项的方式来表示。如果是包含变项的构式，则一定要联系如何对该构式进行释义来考虑变项的设置，包括变项范畴的表达，以及后续对变项范畴如何进行尽可能准确的约束等。

2.2 构式的句法功能

汉语学界对构式的研究，关注构式的形成（来源）机制、内部构成以及语义解释相对较多，而从传统的描写语法角度，说明构式在使用中表现的对外句法功能则相对较少。如果一个句子中包含构式，要分析清楚整个具体的结构（按照短语结构语法的框架），那就必然会像对短语进行功能分析一样，要说明一个构式属于什么样的功能范畴（是 np 还是 vp 等等），充当什么句法角色，承担什么样的语义功能（角色）等。在构式知识库中针对每个具体构式，要回答这些问题，就需要在传统的短语结构语法体系中，逐项来检验一个构式所能和不能实现的句法功能（参见詹卫东，2000）。下面通过几个简单的示例来扼要说明描述构式的句法功能需要注意的问题。

2.2.1 构式句法功能范畴的确定

请看一些“零+X”构式的例子：

- 1) 成人**零基础**学英语
- 2) 上海坚持**零起点**教育
- 3) 企业**零库存**管理的实现途径和方法
- 4) 日本积极迈向“**零排放**”时代
- 5) 在加州销售的所有汽车中 2 % 必须是“**零排放**”汽车
- 6) 莱州黄金冶炼企业实现废水**零排放**
- 7) 网民就“**污染物零排放**”进行了网上辩论

“零+X”构式中变项成分 X 可以是双音节抽象名词，也可以是双音节动词。从使用中表现的分布情况来看，“零+X”构式的整体功能范畴更接近名词性短语（np）。上面例 1-5 中的“零+X”构式实例可以分析为处于修饰语句法位置，其中例 1 的“零基础”可以看作是比较特别的名词性成分作状语（相当于“以零基础的状态学英语”）；例

14 《现代汉语词典》（第 6 版）为“不…不…”格式设置的 3 个义项是：①用在意思相同或相近的词或语素的前面，表示否定（稍含强调意）②用在同类而意思相对的词或词素前面，表示“既不……也不……”③用在同类而意思相对的词或词素的前面，表示“如果不……就不……”。

2-5 的“零+X”构式实例是处于定语位置。例 6-7 两例是“零+X”构式处于 np 短语的中心语位置。

2.2.2 构式句法功能特征的描写

有的构式整体所属的句法功能范畴比较清楚，但具体到构式实例，仍有一些特殊的分布特点需要精细描写。比如：构式“v+来+v+去”和“v+进+v+出”“左+v+右+v”整体功能类都属于动词性短语（vp），不过，具体到构式实例的实际分布情况，就可能仍有不同表现。比如：

1) 他在卡尔身边**绕来绕去地**游着，而且还游得灵活自如。（在“地”前做状语）

2) 树冠被风吹得**摇来摇去**，象喝醉了酒一样。（在“得”后做补语）

构式知识库中应描述“v+来+v+去”可以分布在上面两例所示的“状语、补语”位置。相比之下，“左+v+右+v”（左看右看、左等右等）一般不在“地”前做状语；“v+进+v+出”（飞进飞出、走进走出）一般没有在“得”后做补语的功能。

值得特别说明的是，构式的句法功能特征描写是一个系统的问题，并不仅仅局限于观察一个构式自身。对于一些形式上相近有联系的构式，需要把它们句法分布情况放到一起来加以考虑。比如上面提到的构式例子“v+都+不+v”（下面例 1）和“v+都+不+vp”（下面例 2），二者形式上相近，从模式匹配的角度讲，后者可以涵盖前者，即如果不加特别约束的话，则前者可以看作是后者 vp 为单个动词 v 时的一个特例。

1) 我把信递给她，没想到她**看都不看**（v+都+不+v）

2) 跟老板打招呼，他**看都不看我一眼**（v+都+不+vp）

在描写“v+都+不+v”构式的句法分布特征时，就需要特别强调，该构式整体作为一个 vp，但其不能再接宾语成分（包括像“一眼”这样的数量宾语）。如果没有这样一条句法功能限制，则对于上面例 2 的分析就会出现歧解。一种分析方式是例 2 为构式“v+都+不+vp”的实例，另一种方式则是可以分析为“v+都+不+v”带宾语“一眼”。后一种分析方式不仅错误，而且纯属多余，需要加以避免。

2.3 构式的语境语用信息

上文 1.5 “构式与修辞”一节已经提到，有不少构式带有浓厚的修辞色彩，因而对于该构式的解读需要在很大程度上依赖语境。反过来说，如果文本中出现了这样构式，也可能可以提示当前语境具有什么样的特征。比如有的构式使用时前后常带引号，提示该构式是引语用法（如“死呀活的”“啊来啊去”等构式）；有的构式常用于标题（如“被 X”“有一种 X 叫 Y”“舌尖上的新年”等构式）；还有的构式是典型的口语用法，可以提示当前是在对话语体中（如“什么 X”“哪儿跟哪儿”“谁跟谁”“你看看你”等构式）。此外，有一些构式可以传递说话人的态度、情感等信息。比如上面举过的“什么+X”构式，就是表达驳斥上文观点或做出负面评价等态度信息的典型例子。下面再多看一些例子：

- 1) 这孩子**早不回来晚不回来**，偏这时候回来。
- 2) 他**要钱没钱、要权没权**，能拿得出什么。
- 3) 国企以前的“政治思想工作”有时带一种强迫性，你**接受也得接受，不接受也得接受**。
- 4) 他发表的小小说数量不算很多，但**写一篇是一篇**。
- 5) 他**要名有名，要钱有钱**，比你强多了。

上面例中加粗部分是构式实例。例 1-3 表达了负面态度（情绪），例 4-5 表达了正面态度。所谓负面（消极）态度，指的是对于句中所陈述的客观事件，说话人或事件主体的心理状态是不满意、无奈等。比如例 1，对于“这孩子这个时候回来”这个事件，说话人的态度是不满。所谓正面（积极）态度，是指对于句中所陈述的客观事件，说话人或事件主体的心理状态是肯定或赞许。比如例 4 陈述的客观事件是“他写的小小说数量不多”，但说话人的主观评价是“这些小小说的质量（水平）不错。”说话人对此持肯定和赞许的态度。

像上面例 1-5 中的构式，类似于词汇中的褒义词和贬义词，可以表达正面或负面评价的主观信息。构式知识库中应对类似上面这样的具有特定的语体色彩、褒贬色彩的构式加以标记。

§ 3 构式知识库的数据表及字段设置

构式知识库主要描述单个构式的特点以及构式之间的关系（比如同义构式、反义构式等）。描述单个构式又有两种角度：一是向内看，描述常项、变项的特征以及项与项之间的关系；二是向外看，描述构式的句法、语义、语用特征。所有字段的描述均服务于构式的计算机处理以及进一步的语言学研究。上面两节对构式的形式和意义、句法功能、语用等方面做了分析，这一节对构式知识数据库描述的具体字段设置做详细说明。

3.1 构式基本信息数据表

3.1.1 构式形式

取值类型：文本（symbol unlimited）

说明：构式形式由常项和变项组成。一个构式可以只包含常项，也可以只包含变项，但更为典型的情况是构式既含常项又含变项。各项之间用“+”分隔。

构式形式的常项成分中除汉字外，还允许含英文字母或数字。另外，构式形式中也可以包含中文逗号。逗号起到分隔作用。原则上，一个构式形式中必需包含至少一个“+”分隔号，以体现构式的“组合性”。如果构式形式中没有“+”号，则该条目在理论上更适合收入“词典”，而不是“构式库”。换句话说，**构式形式的表示不允许只有一个常项或只有一个变项的情况¹⁵**。

在构式库中，“构式形式”是一个构式的唯一形式表征。“构式形式”字段值和“义项”字段合在一起起到主关键字（primary key）的作用，即决定数据库中一条唯一的记录。

关于“构式形式”的填写，要注意以下几项操作原则：

（1）构式形式中的常项成分一般为传统语法单位中的词，应尽量避免把多个词合并为一个单位作为构式的常项成分。比如“一个”“那个”都不适宜作为常项成分，应分开为“一+个”“那+个”。这样在将“构式形式”跟经过切词处理的文本语料匹配时更方便处理。再举一个例子：如果在构式形式中用“v+不胜+v”来概括像“数不胜数、

¹⁵ 举例来说，如果认为“一动不动”是一个凝固型构式，考虑收入构式库，则构式形式应该描述为：一+动+不+动，而不应该是“一动不动”，后者意味着整个单位被当做一个词来看待。前一种形式意味着这个凝固型构式由四个常项“一”“动”“不”“动”组成。此外还有一些个别的特殊现象值得一提，比如“7×24”（表示一周七天一天二十四小时全天候做某事），内部也具有一定的组合性。但如果要作为凝固型构式收入构式库，就需要表示为“7+×+24”，这样的形式太过特殊，我们认为更适合作为一般词汇看待，不宜作为凝固型构式收入构式库。“7×24”的实际用例：“电子团员证”是上海共青团首次推出的全功能电子团员证管理平台，是纸质团员证在互联网时代形态变化和功能提升的一种尝试。它不仅完整保留了纸质团员证的所有功能，更打破了时间和空间的限制，实现团务工作7×24小时在线办理、证件信息移动展示等。

防不胜防、举不胜举”这样的例子，就犯了常项表示的错误，构式形式中“不胜”应分开为两个词，构式形式的正确表示应为“v+不+胜+v”。有一些语言单位长期共现合用，已经重新组合为一个相当于词的单位，这样的情况也可以作为一个单位分析为构式的常项，比如“不是”“就是”等。

(2) 构式形式的变项可以用 a、n、v、m、q……等词类标记或 ap、np、vp、mp、qp……等短语功能类标记表示，也可以用 X、Y、Z……等通配符标记表示。如果一个构式形式中有多项同类标记，则用数字下标 1、2、3……区别。对于完全同形的变项成分，用相同的符号标记。**变项范畴如果是明确的，应优先选择词类标记，其次是短语功能类标记，应尽量避免使用 X、Y、Z 等通配符标记。**只有在一个变项成分可以由词、短语、句子等长度不一的语言单位充任时，才考虑用通配符标记。通配符标记配合数字下标表示两个变项在构式实例中总是要求同属一个范畴，例如“X1、X2”和“Y1、Y2”表示前两个变项同属一个范畴，且跟后两个变项属不同范畴。若两个通配符标记表示的变项没有同属一个范畴的要求，则采用 X、Y 分别标记。

(3) 构式形式应尽可能简短，除非必要，勿增项目（常项和变项）。比如：

甲	n+也+a+不+到+哪里+去	阿 Q 也好不到哪里去
乙	a+不+到+哪里+去	好不到哪里去

甲例的形式中包含了乙例中的“构式形式”，但甲中的 n 和“也”这两项并不是构成一个构式的必要成分。相反，乙例中的每个成分，对于整个构式，都是必需的。因此构式库中应收录构式形式乙作为一个构式条目，而不应该收录甲形式。甲应该看作是构式形式乙的一个应用实例。

(4) 构式条目不避部分重复。即使是构成成分相近的构式形式，也可以同时将这多个相近的构式收入构式库。这样处理有利于将构式形式跟分词和词性标注语料去做模式匹配。例如：

- A: 你走你的阳关道，我走我的独木桥 (r1+v1+r1+的+n1, r2+v2+r2+的+n2)
 B: 你走你的 (r+v+r+的)
 C: 走你的 (v+r+的)

形式上，A 包含了 B，B 包含了 C。如果按照上面第 (3) 条操作原则，就只需要将 C 形式收入构式库，而不需要收录 A 和 B。但是，构式库中可以同时收录这三条。这样有利于计算机处理，省去了像短语结构那样层层组合的麻烦，可以提高匹配的准确率。这里 A、B、C 三个构式之间的关系，跟上面阐述第 (3) 条原则时举的例子的情况不同。第 (3) 条中的甲例并不是一个独立的构式，其中“n+也”是附加在后面的构式上，构式“a+不+到+哪里+去”是一个谓词性结构（可归入形容词性短语 ap），“也”作为状语修饰其后的 ap，然后充当谓语，甲例整体是一个主谓结构短语。第 (4) 条原则中的例 A 虽然包含 B，B 包含 C，但三个形式都可以看作是独立的构式¹⁶。如果只收入 B 构式

¹⁶ C 构式可以看作是 B 构式省略“主语”r 造成的。C 构式一般只用在对话语境，作为祈使句使用。因此构式中的 r 仅限于“你”（不能是“我、他”）。对于 C 构式的变项成分的约束及语境约束，需要在构式

形式，那么，碰到 A 形式时，其内部构成就会分解为“B+n”两个部分，但这样就无法描述“你走你的”（B）跟“阳关道”（n）之间的结构关系，显然，这样分解是不合适的。只有把 A 也作为一条独立的构式形式收入构式库，才能避免这个问题。

从表面形式上看，上面第（4）条原则似乎跟第（3）条原则的精神相悖，但实则不然。这两条原则实际上强调了收录构式应遵循的共同标准：第 3 条是说，没有达到构式的标准，就不该收录。第 4 条是说，即便有相近的形式已经收入构式库，也不影响再增加一条符合标准的构式条目入库。第 3 条原则和第 4 条原则共同作用，可以使构式库中收录的构式条目能最大限度地覆盖语言系统中的构式实例，同时又不至于鱼目混珠、掺和凑数。

（5）“构式形式”字段的取值跟下文将谈到的“构式实例”字段有对应关系。下面表中所示的“构式形式”跟“构式实例”不严格对应，是错误的：

构式形式	构式实例
看+np+vp+的	看你这么说的 看他给闹的 看把你急的

从“构式实例”取值来看，对应“构式形式”中“np+vp”变项部分的有“你这么说的”“他给闹”“把你急”。其中“把你急”不符合“np+vp”模式的要求。“给闹”也不符合典型的 vp 范畴。从这里所举实例的情况来看，“构式形式”应该用 3 个模式（即在构式库中立 3 条记录）去对应这里所举的 3 个实例：看+np+vp+的；看+np+给+vp+的；看+把+np+vp+的。

3.1.2 构式变体

取值类型：文本（symbol unlimited）

构式变体是“构式形式”的变异形式。在形式上，“构式变体”与“构式形式”大致相同，一般只是在某个常项成分上有细微差异。例如，“a+不+到+哪儿+去”就是“a+不+到+哪里+去”的变异形式，两者唯一的差别就在于常项中一个是“哪里”一个是“哪儿”。

“构式变体”字段填写的格式规范与“构式形式”一致，并且允许有多个构式变体，多个构式变体之间用“|”分隔。

如果一个构式没有“构式变体”，则可以不填（值为空）。“构式变体”字段缺省值为空。

出于条目精简考虑，若“构式形式”A 的“构式变体”字段取值中包含构式 B，则构式 B 不需要再作为一个独立的条目出现在构式数据库中，即数据库中不应录入一个“构式形式”为 B 的条目。

库中详细描述。另外，计算机在寻找文本中的构式形式实例时，遵循“最大匹配原则”，如果能够匹配成功 A 构式，则不会再匹配 B 构式；如果能够匹配成功 B 构式，则不会再匹配 C 构式。

3.1.3 义项

取值类型：整数（number limited）

说明：填写“0、1、2、3”等整数，指一个构式的义项编号。对于只有一个义项的构式，“义项”字段值为0（缺省情况）。如果一个构式有多个义项，则分别用1、2、3等表示第1个义项，第2个义项，第3个义项，等等。

上文2.1节讨论构式的多义性时已经指出，一般来说构式应遵循“一形一义”原则。不过，构式的形式概括和意义概括具有相对性。形式越具体，意义也会相对越具体越确定，形式如果相对抽象（能产性强、覆盖面广），则意义也就会相对抽象、笼统，从而包含多个下位义项。

一个构式的多个义项的顺序没有强制性要求。原则上，尽量按照各义项的出现频率（常用义、非常用义）、出现先后（如本义、引申义）等标准排序。

在词典编纂中，一个词的义项划分有一定的相对性，因而也往往容易出现不同意见。构式的“义项”划分也还需要在积累一定的个案分析的基础上提炼标准和原则。上文2.1.2小节举了“什么+X”构式可以分立两个义项的例子，其中“元语否定”义可以作为第1个义项；“负面评价”义可以作为第2个义项，前者更为常用，X的类型更多，整体分布范围更广；后者相对低频，X的类型不多。

3.1.4 构式特征

取值类型：文本（symbol unlimited）

取值范围：复现、省略、冗余、异序、语法错配（语法形式特征）

含否定成分、含疑问成分（语法形式特征）

论元异常、否定义、负面评价、周遍、主观大量、

主观小量、语义错配（语义特征）

修辞、网络用语（其他特征）

构式语法特征示例

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 ¹⁷
语法错配	被和谐 好你个李云龙 那个紧张啊

¹⁷ 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）。

含否定成分	照顾 不过来 不 写 不知道
含疑问成分	读 哪门子 书

构式语义特征示例

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
否定义	什么一流大学 看什么破电影 怕个毛
负面评价	这 叫 什么 画家 啊
周遍	谁也看不懂
主观大量	一去就是三年
主观小量	没 几 天 算 你 赢

说明：一个构式的“构式特征”允许多选，即可以选择一个特征或同时选择多个特征。比如“他是猪肉馅儿”构式既有“语义错配”特征，也有“省略”特征。

语义特征中的“论元异常”指的是构式中出现的动词通常具有的论元结构跟它在构式中实际共现的论元结构不一致。比如“跑了我一身汗”，在动词“跑”后面出现了两个名词性论元成分“我”“一身汗”。而“跑”通常是不及物动词，其论元结构为“跑<施事:[有生]>”，即只跟一个有生名词性成分（充当施事角色）共现。在“跑了我一身汗”中，“我”是通常的施事论元，“一身汗”是超常出现的结果论元。还可以说“我跑了一身汗”，不及物动词“跑”出现在只有及物动词才能出现的SVO句法配置中。

构式特征的取值具有开放性（unlimited），允许编辑人员在确定构式特征时，在现有的特征之外，自行给出一个新的特征标签。一般来说，有两种情形：

- (1) 当前构式的特点不适合用已有的特征来加以描述；
- (2) 可能需要对已有的特征进行细化，做更精细的描写，比如“主观量”根据需要，进一步细分为“主观大量”或“主观小量”。

“修辞”是指具有修辞意味的构式。比如：都是月亮惹的祸、哥抽的不是烟，是寂寞。

“网络用语”是指在网络媒体环境中兴起、常用的构式。比如：爽到爆、帅到没有朋友。

3.1.5 构式类型

取值类型：文本（symbol limited）

取值范围：凝固型、半凝固型、短语型、复句型

说明：参见上文1.4节对构式的类型划分所做的说明。

“构式类型”只能单选。

3.1.6 音节数

取值类型：整数 (number unlimited)

说明：填写构式的音节数。对于音节数确定的构式直接填写构式的音节数。例如，构式“n+帝”中的n一定是双音节（如“表情帝、牢骚帝”等），所以构式的音节数是3。对于音节数可变的构式，可以填写构式音节数的范围。例如，构式“a+什么+a”中的a可以是单音节（例如“高什么高”）也可以是双音节（例如“漂亮什么漂亮”），所以构式的音节数是4-6。两个整数之间用短横线（-）连接。如果构式变项成分的可能性较多，不易确定音节长度，可以填写音节数范围的最小值，最大值空缺，形如“n-”。

3.1.7 组块数

取值类型：文本 (symbol limited)

说明：填写“单”或“双”。大多数构式为一个组块，填值为“单”。像“你写你的，他唱他的”这类构式，由两个同型结构组成，可以看作是双组块构式，填值为“双”。

3.1.8 是否可扩展

取值类型：布尔 (Boolean)

说明：有的构式可以并立扩展，如“麻烦就麻烦在担保上，麻烦在还款手续上”，是“a+就+a+在+X”构式的扩展形式。对于这类构式，“是否可扩展”字段取值为“可”。有的构式没有并列扩展形式，只能单用，如“一个不留神”这类“一+个+vp”构式，没有扩展形式。对于这类构式，“是否可扩展”字段取值为“否”。

3.1.9 实例

取值类型：文本 (symbol unlimited)

说明：凝固型构式的“实例”字段不需要填写。其他类型的构式，“实例”字段中需填写具体的构式实例，每个构式至少3例。

实例与实例之间用字符“|”隔开。

实例应该尽量覆盖变项的不同类型，如“那+个+X+啊”（释义为“很X”或“真X”），其中X的取值可以是“形容词、心理动词、动宾词组”等等，那么这几种情况的例子就应该各举一个。

构式实例应该尽量取自真实语料，不宜自造。如若没有对应实例或对应实例较少，则应重新考虑构式知识库中是否应该收录该构式。可以调查他人的语感，征求更多意见后再做处理。

3.1.10 变项数量

取值类型：整数 (number limited)

说明：变项是指构式中可以替换的成分。该字段填写一个构式所包含的变项的数量，数值为大于等于 0 的整数。

凝固型构式的变项数量为 0。其他类型构式的变项数量大于 0。

一个构式的变项不宜太多，一般 2-4 个为常见情况，如果太多则应重新考虑对构式形式的概括工作是不是可以改进。

3.1.11 常项数量

取值类型：整数 (number limited)

说明：常项是指构式中保持不变的固定成分。该字段填写一个构式所包含的常项的数量，数值为大于等于 0 的整数。

注意：常项一般为传统语法单位中的词。在“可合可分”的情况下（比如“不是”可以合为一个单位，也可以分为“不”和“是”两个单位），原则上倾向从分。但如果一般常用词典中将该单位收录为词（比如《现代汉语语法信息词典》《现代汉语词典》等），则也可以不分（比如“一方面”“另一方面”等都可以作为一个单位）。

注意：原则上，常项数量与变项数量之和应大于 1。（参见 3.1.1 对“构式形式”字段的格式规范说明）

3.1.12 释义模板

取值类型：文本 (symbol unlimited)

说明：这个字段填写构式的直接释义。所谓释义模板，指的是包含构式变项和/或常项成分的一个字符串，该字符串的格式要求跟“构式形式”遵循相同原则，各项之间以“+”隔开。语义上跟“构式形式”所表达的语义相同（或相近）。一个构式至少有一个释义模板（简单情况），也可以对应若干个释义模板组成的集合（复杂情况）。多个释义模板之间用西文字符“|”分隔。具体示例见 2.1 节。

释义模板实质上是用一种线性符号串来转写（释义）原来的构式。对构式的语义分析除“释义模板”外，还可以用框式结构（由各类框架元素组成）来结构化地释义。见下文 3.6 节的说明。

构式的释义目前还难以做到形式化。我们采用“释义模板”加“框架结构”的释义方式，力图做到构式语义分析的结构化表示。

如果一个构式的语义难以变换为线性符号格式的模板化释义，可以采用说明该构式用法、解释其使用环境或语义的方式来表述。区别于模板化的释义，解释性的释义用 # 在释义文字首尾加以标记。形如 # <构式意义说明> #。

3.1.13 近义构式

取值类型：文本 (symbol unlimited)

说明：填写与该构式的构式义相同或相近的构式。这一字段的形式规范与“构式变体”字段一样。

近义词义是对称关系。A 构式的“近义词式”字段填值若为 B，则 B 构式的“近义词式”字段也要填 A 构式。换言之，近义词式字段中填入的构式应该在构式库中是独立的条目。（下面的反义构式、上下位构式等字段也都要满足这个要求，即所填构式应在构式库中也是一个独立的条目）

A 构式的近义词式 B 构式往往在更概括的层次上可以看作同一个构式，只是其中的变项范畴有所不同，因而细分为两个形式上不同的构式。比如“a+就+a+在+X”构式（例如“难就难在持之以恒”）的“近义词式”字段可以填写“v+就+v+在+X”（例如“输就输在大意”）。这两个构式的主要语义特征都有“解释缘由”这个意思。这种情况是“近义词式”字段跟“构式变体”的主要区别所在。后者主要是常项成分有细微差异（参见 3.1.2 对“构式变体”字段的说明）。

除上面的“近义词式”外，还有这样的情况：两个构式的实例具有近义词义，但作为抽象的构式模式，并不能保证二者的实例总是构成近义词义。比如“问东问西”“问这问那”“问来问去”分别是“v+东+v+西”“v+这+v+那”“v+来+v+去”构式的实例。作为实例，这三个构式都有“反复义”，可以看作是构式近义词义的实例。不过，并不是所有的 v 都可以自由地进入这三个构式模式产生出具有近义词义的实例，比如“争来争去”“争这争那”可以形成近义词义，但“争东争西”就不是合格的实例。像这样的情况，这些构式可能并不适合在一个更高的层次上概括为同一个构式，但确实又有一些实例的语义比较接近，而且构式模式能抽象出一个共同的语义范畴（语义特征），比如“反复义”，这样的构式，也应标记它们构成了近义词义。同时在“备注”字段做一些必要的说明。

3.1.14 反义构式

取值类型：文本（symbol unlimited）

说明：填写与该构式的构式义相反或相对的构式。这一字段的形式规范与“构式变体”字段一样。

一个构式可能没有“反义构式”，则“反义构式”字段为空（缺省取值）；也可能有一个甚至多个“反义构式”。不同的反义构式之间用符号“|”分隔。

构式“n+帝”（如“表情帝”）的“反义构式”字段填值为空。

构式“a+着呢”（高着呢）的反义构式为“a+不+到+哪里+去”|“能+a+到+哪里+去”等。

3.1.15 上位构式

取值类型：文本（symbol unlimited）

说明：填写该构式的上位构式。这一字段的形式规范与“构式变体”字段一样。

一个构式可能没有“上位构式”，这时候“上位构式”字段为空（缺省取值）；也可能有一个甚至多个“上位构式”。不同的上位构式之间用符号“|”分隔。

3.1.16 下位构式

取值类型：文本（symbol unlimited）

说明：填写该构式的下位构式。这一字段的形式规范与“构式变体”字段一样。

一个构式可能没有“下位构式”，这时候“下位构式”字段为空（缺省取值）；也可能有一个甚至多个“下位构式”。不同的下位构式之间用符号“|”分隔。

3.1.17 否定形式

取值类型：文本（symbol unlimited）

说明：说明一个构式是否有否定形式的用法。如果没有，填写“无”；如果有，列举具体的形式。

3.1.18 疑问形式

取值类型：文本（symbol unlimited）

说明：说明一个构式是否有疑问形式的用法。如果没有，填写“无”；如果有，列举具体的形式。

3.1.19 备注

取值类型：文本（symbol unlimited）

说明：填写除上述信息之外的关于一个构式的其他必要的基本信息，或者对上述字段的内容做补充说明。

3.2 构式变项信息

3.2.1 变项序位

取值类型：整数（number limited）

说明：填写变项所占据的位置序号。序位的顺序是从左往右，从1开始，变项与常项都参与排序。如“v1+一+q1, v2+一+q2”构式，一共由6项成分组成（其中逗号是分隔符，不计在内），变项“v1、q1、v2、q2”的序位分别为“第1项、第3项、第4项、第6项”；两个常项“一”分列第2项和第5项。

注意“构式形式”中的逗号起分隔作用，不在变项和常项序位中参与计数。

3.2.2 变项句法特征

取值类型：文本 (symbol unlimited)

说明：填写某一个序位上的变项成分的句法信息，主要包括音节数、语素类、词类 (ccat)、短语类 (cpcat) 等，具体句法特征字段名可参考《现代汉语语法信息词典》；也可以直接列举出变项可取的所有值；填写形式为“特征名:特征值”，如“音节数:2”。多个句法特征之间用逗号隔开，形式为“特征名:特征值,特征名:特征值,…”；多个句法特征信息之间是合取 (AND) 逻辑关系。

变项句法特征中特别设置一个“引语”特征，缺省取值为“否”。如果一个变项可以为引语，则标记“引语”特征取值为“是”。(参见 2.1.1 小节)

3.2.3 变项语义特征

取值类型：文本 (symbol unlimited)

说明：填写某一个序位上的变项成分的语义信息，主要包括语义类、主体角色约束、客体角色约束等，具体语义特征字段可以参考《现代汉语语义信息词典》。填写形式跟“变项句法特征”字段的填写形式相同。

3.2.4 变项可替换度

取值类型：整数 (number unlimited)

说明：构式中变项成分有两个统计指标，一个是 type frequency (语类频次)，一个是 token frequency (语例频次)。变项可替换度用前者描述。由计算机根据构式在语料库中的实际表现情况自动统计得到，不需要手工填写。

3.3 构式常项信息

3.3.1 常项序位

取值类型：整数 (number limited)

说明：填写常项所占据的位置序号。序位的顺序是从左往右，从 1 开始，变项与常项都参与排序。如“v1+一+q1, v2+一+q2”中，常项“一”分别占据了“2、5”两个位置。

3.3.2 常项句法特征

取值类型：文本 (symbol unlimited)

说明：填写常项的句法特征信息，参照“变项句法特征”。

3.3.3 常项语义特征

取值类型：文本 (symbol unlimited)

说明：填写常项的语义特征信息，参照“变项语义特征”。

3.4 构式内部成分约束关系

3.4.1 变项间关系

取值类型：文本 (symbol unlimited)

说明：填写变项之间的约束关系。

填写形式为“变项序号 i. 特征 m==变项序号 j. 特征 n”。约束关系之间用逗号“,”分隔。

3.4.2 常项与变项间关系

取值类型：文本 (symbol unlimited)

说明：填写构式中常项与变项之间的句法语义约束关系。

填写形式为“变项序号 i. 特征 m==常项序号 j. 特征 n”。约束关系之间用逗号“,”分隔。

3.4.3 组块关系

取值类型：文本 (symbol unlimited)

说明：描述构式内部是否由两项组块构成。若构式为单组块，则不填(缺省值为空)；若构式可以分析为由两个组块构成，则用<>标记组块的边界，组块之间的关系默认为并列。如果有其他关系，则在组块最后用()标记。

填写形式为“<组块 1><组块 2>(组块关系)”。

3.5 构式整体句法属性

3.5.1 构式的句法功能范畴

取值类型：文本 (Symbol limited)

取值范围：ap、dp、mp、np、pp、qp、sp、tp、vp、dj、fj

以上是传统短语结构语法中对汉语短语功能范畴的一个基本类别划分。可以参照短语如何判定其功能范畴归属的做法，来对一个构式的功能范畴进行确定（参见詹卫东，2000）。

各功能标记的含义如下表所示：

标记	含义	标记	含义
dj	单句	fj	复句
ap	形容词性短语	qp	数量短语
dp	副词性短语	sp	处所性短语
mp	数词性短语	tp	时间性短语
np	名词性短语	vp	动词性短语
pp	介词性短语		

根据分布分析的原则，句法功能范畴的不同类别，对应着在给定的句法结构关系体系中的不同句法结构位置。下表 1 列出了汉语的基础句法结构关系，也即句法结构位置的一个总体描述；表 2 根据一个语法单位出现的典型句法位置的不同，给出了一个句法功能分类的体系。

表 1：汉语句法结构的组合类型

序号	结构关系	句法结构位置		实 例
1	主谓结构	主语	谓语	树叶黄了;小明喜欢看电视;感冒传染
2	述宾结构	述语 1	宾语	喝了三杯酒;学了三年;企图逃跑;送他香烟
3	述补结构	述语 2	补语	洗干净;做得非常好;好得很;吃得完;拿出来
4	定中结构	定语	中心语 1	一斤白菜;老师的眼泪;大红灯笼;削梨的刀
5	状中结构	状语	中心语 2	快跑;认真地学习;把饭吃完;明天见;屋里坐
6	连谓结构	前项	后项	开着窗户睡觉;打电话请医生;派助手办理;请他来
7	联合结构	前项	后项	小说和戏剧;又高兴又难过;批评教育
8	附加结构	中心语 3	附加语	红着;吃了;砍光了;努力奋斗过
9	的字结构	中心语 4	附加语	买菜的;老师表扬了的;冰凉的;慢性的
10	所字结构	附加语	中心语 5	所知道;所了解

表 2: 汉语句法单位的功能范畴¹⁸

序号	标记	功能类名称	句法功能 (位置)													
			a	b	c	d	e	f	g	h	i	j	k	l	m	n
1	dj	单句型短语		+		+										+
2	np	名词性短语	+			+				+	+				+	+
3	vp	动词性短语		+	+		+						+	+	+	+
4	ap	形容词性短语		+			+	+	+		+	+		+	+	+
5	dp	副词性短语									+					
6	pp	介词性短语									+		+			
7	sp	处所词性短语				+										+
8	tp	时间词性短语				+										+
9	qp	数量短语								+						+
10	mp	数词短语								+						+

表中 a-n 各项代表的句法功能为:

a:作主语; b:作谓语; c:作述语₁; d:作宾语; e:作述语₂;

f:作补语; g:作定语; h:作中心语₁; i:作状语; j:作中心语₂;

k:作连谓结构前后项¹⁹; l:作联合结构前后项; m:作中心语₃; n:作中心语₄;

下面给出一些构式的句法功能范畴标注的示例:

构式形式	构式实例	句法功能范畴
a+着+呢	高着呢 早着呢 好玩儿着呢	ap
a+不+到+哪里+去	高不到哪里去 好玩儿不到哪里去	ap
v1+了+v2+ v2+了+v1	吃了喝喝了吃	vp

¹⁸ 表中没有列出复句fj的功能特征,对于fj这样“大尺寸”的语言单位来说,实际是从内部结构特征上来跟dj(单句)相区分而得到的语法单位类别,并不是从句法功能(分布位置)特征角度来分出的类。

¹⁹ 这里为表格的简单起见,把连谓结构前项位置和后项位置并作一个区别特征加以看待了。实际上这两个位置是有差异的,比如pp和ap短语都允许出现在连谓结构后项位置,但不能出现在前项位置。

零+v	零污染 零排放 零容忍	np
-----	-------------	----

3.5.2 构式的句法功能分布特征

取值类型：文本 (Symbol limited)

说明：描写整个构式可以充当的句法成分（即能够占据的分布位置）。

取值范围：包括以下 20 种句法功能（位置）。

- (1) 作主语
- (2) 作宾语
- (3) 作谓语
- (4) 作定语（指直接作定语，不需要附加“的”）
- (5) 作状语（指直接作状语，不需要附加“地”）
- (6) 作补语（指直接作补语，不需要附加“得”）
- (7) 带宾语
- (8) 带补语
- (9) 被定语修饰
- (10) 被状语修饰
- (11) 联合结构前项
- (12) 联合结构后项
- (13) 连谓结构前项
- (14) 连谓结构后项
- (15) 作介宾
- (16) 带‘的’后作定语
- (17) 带‘地’后作状语
- (18) 在“得”后作补语
- (19) 形成“的”字结构²⁰
- (20) 独立成句

数据库中这一字段填值时可以采用 20 位字符来标记一个构式的分布特征。每 1 位字符有“1”和“0”两个取值，比如第 7 位的取值为“1”，则表示该构式具有“带宾语”分布特征，取值为“0”则表示该构式没有“带宾语”的功能。

²⁰ 例如表示比例关系的构式“m1+q1+m2+q2”（十块三斤），就可以形成“的”字结构：在“十块三斤的是秋茄子”这个句子中，“十块三斤的”为“十块三斤+的”形成的“的”字结构。

3.6 构式整体语义属性

构式的语义除了以“释义模板”方式做线性变换来表示外，还可以分出更多层次，以框架结构的方式来描述构式丰富的语义信息。有的构式可能涉及到“字面义”和“交际义”两个层次（参见 2.1.2），前者指一个构式可以按照一般短语结构的常规组合方式去理解其语义，后者则是以一种整体的方式（非组合）来理解语义。如果二者能够同时成立，就可能造成“双关”的修辞效果。比如：给你点颜色看看。这个表达式一般是取构式义用法，即表达“要教训、惩罚某人，通过展示说话人的实力的方式让对方感到害怕，从而迫使对方服从、认输”（对应的英语表达方式也是一个构式：**I will teach you a lesson**）。但仍然可以用一般的短语结构组合方式去理解，即表达“要向某人呈现某些颜色，请某人观看这些颜色”。这种用法在现代商业（广告）文本中较为常见。比如某公司发布新产品，发布会的主题是“给你点颜色看看”。这一事件的新闻稿正文第一段是：“如果不出什么意外的话，这次发布会的主角应该是 XX 公司旗下一款新的超级手机，可能主打外观和**配色**，但目前还没有一个比较明确的说法。”通过一个具有鲜明构式义的标题来吸引眼球，然后再对这个构式做出一般常规短语义的“重新解释”，达到“双关”的修辞效果。在构式知识库中设置“字面义”和“交际义”这两个字段来描写构式的语义，有助于刻画这类语言现象。此外，构式的语义识解机制也很重要，在构式语义属性描述中，应尽可能深入挖掘构式语义的产生机制。

值得特别说明的是，理论上，对于构式的整体语义来说，构式是“形式——意义”的特定配对，因而不需要（也不应该）有所谓的“构式字面义”和“构式交际义”两个“意义（构式义）”。我们之所以在构式知识库中设置“字面义”和“交际义”两个字段来刻画构式的语义内容，完全是出于实用的考虑，为兼容“短语和构式同形”时的语义分析（特别是像潜在的“双关”修辞）做准备。

另外需要注意的是，语义特征，语用特征（参见 3.7 节）填写时遵循如下规范：

先概括说明，内容尽量简短、格式化。

如有需要，可以在概括说明之后，补充详细的解释以及示例说明内容。用 #...# 加以标记。（参见 3.1.10 “释义模板”的填写规范）

3.6.1 字面义

取值类型：文本（symbol unlimited）

说明：填写构式各项成分直接组合而得到的构式字面意义，即一个构式在仍然可以看作是普通短语结构时表达的语义。

3.6.2 交际义

取值类型：文本（symbol unlimited）

说明：填写构式在实际使用时表达的真实语义。一个构式在交际中表达的语义，肯定是其组成部分的字面意义之外的意义，即不是由各部分意义和句法关系意义简单组合

而成的意义。构式的交际义，是一个构式的所有实例共享的，能够实现特定的交际意图，具有特定的表达效果的意义。在抽象的构式层面，构式的交际义由构式整体（而不是由某个组成成分）确定。在实例化的构式层面，构式的交际义由每个使用中的构式在语境中不断固化（entrenchment）定型。

这里不妨看一个简单的例子：

构式：n+第二

示例：姚明第二 苹果第二 莫扎特第二

字面义：n 在某方面/领域/活动中排名第二

交际义：某人（物）像 n 一样优秀、出色

实例：贝多芬十一岁时，就已经显露了他的音乐天才，被认为是莫扎特第二。

3.6.3 语义识解机制

取值类型：超链接（URL/Link）

说明：填写构式交际义是如何产生的，包括揭示交际义跟字面义之间的联系。在数据库中填入一个超链接，指向详细说明该构式语义的识解机制（文件类型：html/ pdf /word 文件）。在数据库编辑页面，通过上传一个本地文件到服务器实现向数据库中这一字段填入一个超链接（字符串）信息。

下面通过两个更为复杂一些的具体构式例子来说明构式知识库中对于构式字面义、交际义、语义识解机制的描写方式。例 1 只分析了字面义和交际义（语义识解机制不清楚，暂缺）。例 2 涵盖了上述 3 个字段的内容。

例 1：X+不是+Y+吗

先看一组例子：

1) 这句话是个很简单的句子，语法也没有什么复杂，**这要翻译成英文不是很容易吗？**

2) 如果读者像我一样采取拒绝的态度，**扩版岂不是适得其反了吗？**

3) **北大自己不是有编辑出版专业吗**，为什么出版社不招“专业对口”的毕业生呢？

4) **你们不是老来找大毒枭吗**，他就是大毒枭的儿子！

5) …笑着说：“一看你就是底眼儿（外行），哪有问这个的，**说了不是砸自己饭碗吗。**”

6) 我最心爱的东西是邮票，还有和一些苏联朋友的通信。**六十年代初不是鼓励和苏联联系吗。**

7) 人的感性，就是孟子所说的耳目之官。**孟子不是说吗？**「耳目之官不思，而蔽于物，物交物，则引之而已矣」。

8) 没错, 有一位天文学家不是说过吗, 宇宙万物终将灭绝或爆炸。

9) “您怎么知道我爱喝酒?” “您测的是个酉字, 酉字加上三点水不是酒字吗, 所以说您好喝酒……”

上面例 1-9 中的“X+不是+Y+吗”格式可以分为两组: 例 1-3 是一组(下称“甲组”); 例 4-9 是一组(下称“乙组”) 21。

甲组中的“X+不是+Y+吗”是一般的所谓“反问句”, 即句中包含了疑问词“吗”, 但跟一般的真性疑问不同。句子形式上为“疑问句”形式, 但侧重的是“传递说话者的看法信息”, 而不是向听话人索取信息(因“疑”而“问”)。甲组例子的语义可以描述如下:

1') 这句话是个很简单的句子, 文法也没有什么复杂, 这要翻译成英文(应该)是很容易的。

2') 如果读者像我一样采取拒绝的态度, 扩版就是适得其反了。

3') (既然) 北大自己是有编辑出版专业的, 为什么出版社不招“专业对口”的毕业生呢?

甲组“X+不是+Y+吗”的字面义可以进一步概括为: X+(应该/肯定)+是+Y

甲组“X+不是+Y+吗”的交际义可以概括为:

(1) 说话人对“X+(应该/肯定)+是+Y”的命题义(记作 S)持肯定态度;

(2) 说话人认为, 听话人不一定认同 S, 因而用“疑问形式”来传递肯定信息, 以强调 S 的真实性。

(3) 说话人用“疑问形式”传递的“疑问”可以解读为“难道你不认为 S 是事实吗?”

(4) 说话人希望通过这种方式得到听话人对自己观点(S)的认同。

乙组中的例子已经不宜看作是一般的“反问句”, 因为句子已经完全没有传递疑问信息的功能。相比之下, 甲组之所以看作是“反问句”, 是因为甲组例子尽管主要不是“问”, 但仍然传递了部分“疑”的信息。乙组例子中说话人用“X+不是+Y+吗”表达的是陈述功能, 即“X+Y”所表达的命题义。

乙组“X+不是+Y+吗”不需要再如甲组例子那样做字面义解读。乙组构式例子的交际义可以概括为:

(1) 表达“X+Y”命题义(记作 P);

(2) 在语境中存在一个命题(记作 Q), 说话人需要解释(或加强) Q 的真实性, 或者为何要表达 Q, P 可以用来实现这个目的。

(3) 基于上述交际目的, “X+不是+Y+吗”的交际义释义模板可以进一步扩展为“因为 X+Y, 所以 Q”。从这个角度讲, “X+不是+Y+吗”构成的句子可以称之为“释因陈述句”(不同于真性问句、假性问句/反问句)。

²¹ 甲组例子和乙组例子的语调有一定的差异。如果要对具体差异做准确的语音学描写, 需要借助实验手段来实现。

下面根据上面的分析模式对乙组例子的语义解释做具体说明：

4') (因为) 你们老来找大毒梟, (那么我就告诉你) 他就是大毒梟的儿子!

5') …笑着说: “一看你就是底眼儿(外行), 哪有问这个的, (因为) 说了砸自己饭碗。”

6') 我最心爱的东西是邮票, 还有和一些苏联朋友的通信。(因为) 六十年代初鼓励和苏联联系。

7') 人的感性, 就是孟子所说的耳目之官。因为孟子说「耳目之官不思, 而蔽于物, 物交物, 则引之而已矣」。

8') 没错, 因为有一位天文学家说过, 宇宙万物终将灭绝或爆炸。

9') “您怎么知道我爱喝酒?” “您测的是个酉字, 因为酉字加上三点水是酒字²², 所以说您好喝酒……”

上面通过观察语料实例中“X+不是+Y+吗”的细微差异, 把这一构式的表义功能分出了两种情况。从初步的语料调查看, 作为反问句使用的构式实例可能占更多数。CCL语料库中查询“X+不是+Y+吗”后接不同标点符号? 。, ……! 的频次分别为 9163, 169, 755, 25, 325 (查询表达式为“不是\$10 吗”后接不同的标点符号)。可以看到该构式后接问号的用例仍然占绝对多数。这在一定程度上可以推测该构式的甲组语义更为基本, 乙组语义(或用法)可能是后来演化来的。从甲组构式义发展到乙组构式义, 应该在很大程度上受到这一构式所处的上下文语境影响, 当然也受到 X、Y、X+Y 等构式组成成分不同的句法语义性质的影响。但具体是如何影响的, 需要在调查历时语料的基础上做进一步分析。这里对于“X+不是+Y+吗”构式乙组例子的语义识解机制, 暂还无法做详细的讨论。

例 2: 有+一+种+X+叫+Y

先看一组常规短语组合的例子:

- 1) 有一种毒药叫砒霜
- 2) 有一种饮料叫 “非常可乐”
- 3) 有一种中药材叫麻黄草
- 4) 有一种说法叫 “经济要大上, 教育顾不上”
- 5) 有一种婚姻叫事实婚姻
- 6) (某地) 有一种珍贵动物叫羚羊

²² 例 9 中的“X+不是+Y+吗”中的 Y 是体词性成分 (“酒字”), 跟例 4-8 中 Y 为谓词性成分的情况有所不同。为了将构式中的 X 和 Y 连接为一个合法的句子表达命题义, 例 9 在释义时需要在 X 和 Y 之间保留“是”。此外, 值得说明的是, Y 为体词性成分时, “X+不是+Y+吗”构式若脱离上下文语境, 语义会有更多歧解的可能性。比如例 9 “酉字加上三点水不是酒字吗”可能有三种解读: (1) 真性问, 其中“酉字加上三点水不是酒字”如果是上文的一个陈述, 说话人引述这个陈述, 加上疑问词“吗”, 整句是一个普通的是非疑问句 (Yes-no question)。说话人的目的是进一步向听话人求得证实; (2) 反问句, 相当于“你怎么会认为酉字加上三点水不是酒字呢”; (3) 释因陈述句, “(因为) 酉字加上三点水是酒字, ……”

7) (任何物体) 都有一种属性叫做惯性

“有+一+种+X+叫+Y”格式的功能是引入一个新的话题性成分 Y, Y 对听话人(读者)来说, 是一个新的概念(信息)。X 是 Y 的上位概念, 是听话人(读者)比较熟悉的或者更为宽泛的范畴(类别)。这个格式的语义解释模板可以表达为:

Y+是+一+种+X / 有+一+种+叫+Y+的+X

上面例 1-7 的释义如下:

- 1') 砒霜是一种毒药 / 有一种叫砒霜的毒药
- 2') “非常可乐”是一种饮料 / 有一种叫“非常可乐”的饮料
- 3') 麻黄草是一种中药材 / 有一种叫麻黄草的中药
- 4') “经济要大上, 教育顾不上”是一种说法 / 有一种叫“经济…”的说法
- 5') 事实婚姻是一种婚姻 / 有一种叫事实婚姻的婚姻
- 6') 羚羊是(某地的)一种珍贵动物 / 有一种叫羚羊的珍贵动物
- 7') 惯性是(物体的)一种属性 / 有一种叫惯性的属性

近年来, “有+一+种+X+叫+Y”格式出现了一种新的用法, 不能再用上面的短语结构组合释义模式来解释, 而是一种修辞色彩非常鲜明的用法, 需要看作构式来做整体释义。下面看一些例子:

- 8) 有一种力量叫平静 8') ? 平静是一种力量
- 9) 有一种爱叫永不放弃 9') ? 永不放弃是一种爱
- 10) 有一种爱叫放手 10') ? 放手是一种爱
- 11) 有一种毒药叫成功 11') ? 成功是一种毒药
- 12) 有一种从容叫范冰冰 12') ? 范冰冰是一种从容
- 13) 有一种误差叫数据造假 13') * 数据造假是一种误差
- 14) 有一种倒下叫站起 14') * 站起是一种倒下

例 8-14 的释义如果用 8'-14' 的方式, 都是不合适的, 甚至是错误的。这些例子的释义可以表述如下:

- 8'') 平静让人感觉到不心虚、不胆怯, 有力量
- 9'') 永不放弃保护对方, 这是爱的一种方式
- 10'') 放手让对方获得幸福。这是爱的一种方式
- 11'') 成功(有时)对人有害, 如同毒药
- 12'') 范冰冰(格外)从容

13'') 数据造假(非真实数据)当然不是误差, 但跟误差在非真实性这一点上有“表面相似性”。

14'') 站起当然不是倒下, 二者是相反的动作过程。但站起跟倒下往往是连续的两个动作。在某个静态的瞬间, 从身体形态上看, 站起跟倒下也具有“表面相似性”。有的人(悲观者)看到这个动作, 以为是“倒下”; 有的人(乐观者)看到“同样的动作”, 以为是“站起”。

例 8-14 “有+一+种+X+叫+Y”的构式用例可以分为两组，例 8-12 是一组（下称“甲组”），例 13-14 是一组（下称“乙组”）。跟前面例 1-7 的短语结构用法不同，前者是表达一种“客观性的知识”（下称“短语义”），后者则是表达一种“主观性的认识”（下称“构式义”）。其中甲组构式例子的语义解释可以进一步概括为：Y 有 [特别] X 的特征 / Y 是 [特别] 的 X；乙组构式例子的语义解释可以概括为：表象是 X，真相（事实）是 Y

“有+一+种+X+叫+Y”的短语义到构式义的发展符合一般的主观化的认知机制模式，可以图示如下：

	X	Y
图式 1 (集合)	集合 (类)	元素 (成员/特例)
图式 2 (聚焦)	远看 (表象)	近看 (实质)

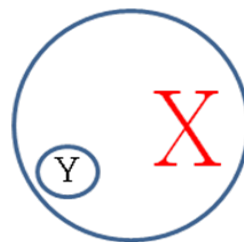


图 1：“有+一+种+X+叫+Y”图式

“有+一+种+X+叫+Y”格式在 X 和 Y 之间建立了一种“集合”和“元素”的关系，X 是集合（上位类），Y 是元素（下位范畴）。这种关系一般是客观的（如例 1-7）。但是，如图 1 所示的这种关系中，Y 和 X 的关系也可以不完全是一种客观知识，而更倾向于说话者的主观认知。这种主观认知可能跟听话人（或社群中多数人）的主观认知比较一致（如例 8-9），更可能的情形则是，这种主观认知是说话人刻意想标新立异提出的一种新看法，往往不是听话人（或社群）一般的认知状态（如例 10-14）。从这个角度讲，“有+一+种+X+叫+Y”构式的表达功能在于表达一种“主观新知”，即发表一种新看法（而非新知识），以显著地区别于“陈说”。也正因为如此，这一构式大多用于文章的标题。从认知角度讲，对于同一个图式，人们主观上可能做出不同的解读，对于上面的图 1，就存在渐进的三种解读方式：

- (1) 成员-集合关系解读：Y 是 X 的一个元素
- (2) 特例-集合关系解读：Y 是 X 的一个特殊元素
- (3) 由表（远）及里（近）聚焦解读：远看是 X，近看是 Y

X 是表象，Y 是实质

其中第一种解读方式对应“有+一+种+X+叫+Y”的短语义。后两种解读方式对应这一格式的两构式义。

通过以上简要分析，我们可以进一步把“有+一+种+X+叫+Y”构式的语义功能表述如下：

- (1) 字面义：将 X 中的某个特定成员“命名”为 Y，通过在 X 和 Y 之间建立客观上的概念上下位关系的方式，引入新的言谈对象 Y。

(2) 交际义：人们以往在对 X 的认识中没有看到 Y 的重要性。而实际上，Y 对于认识 X 是非常重要的。

具体实例的释义又可细分为两类情况：

- 甲类：Y 有 [特别] X 的特征（主观大量）
Y 是 X 集合中的“特例”成员（非典型，但极富特色）
- 乙类：听话人（或一般人）以为 X 是事实（表面现象）
但实际上，Y 才是事实（本质）

下面以表格形式给出更多的“有+一+种+X+叫+Y”的实例及其对应的构式义解读。

实例	X 和 Y 的关系 [构式的语义解读]
让无数老总感觉忧伤的是： 有一种策划叫别人家的策划	X = 策划 Y = 别人家的策划 Y 是 X 集合中的特例 [别人家的策划特别好]
有一种心疼叫做“随便你”， 有一种失望叫做“算了”	X = 心疼（时安慰的方法） Y = 说“随便你” Y 是 X 集合中的特例 [用“随便你”安慰对方实属无奈，其实特别心疼]
有一种忧伤叫“我回你是秒回， 你回我是轮回”	X = 忧伤（的原因） Y = “我回你是秒回……” Y 是 X 集合中的特例 [我对你这么好，你对我这么差，太令人伤心了]
有一种幸福叫转	X = 幸福（的原因） Y = 转（帖/文） Y 是 X 集合中的特例 [文章被转发，感觉特别幸福啊]
有一种春运叫孩子看病	X = 春运（天下难事） Y = 孩子看病 Y 是 X 集合中的特例 [孩子看病是大难事]
有一种友谊叫何炅和谢娜	X = 友谊 Y = 何炅和谢娜（的友谊方式） Y 是 X 集合中的特例 [何炅和谢娜之间的异性友谊非常特别]
有一种胜利叫撤退，有一种 失败叫占领。	X = 胜利 Y = 撤退 Y 是 X 集合中的特例 [撤退是一种特殊的胜利（不合常识）] X = 失败 Y = 占领 [占领是一种特殊的失败（不合常识）]
有一种强拆叫公平	X = 强拆 Y = 公平 X 是表象，Y 是实质 [虽然说是（对某人）强拆，但其实是（为了大家的）]

	公平]
--	------

3.7 构式整体语用属性

3.7.1 感情色彩

取值类型：文本 (symbol limited)

取值范围：褒、贬、中

说明：填写构式的感情色彩，主要包括褒（褒义）、贬（贬义）、中（中性），缺省值为“中”。三个取值中只能选择一个。

3.7.2 语体色彩

取值类型：文本 (symbol limited)

取值范围：口语、书面语、通用

说明：填写构式的语体色彩，主要包括口语、书面语、通用。缺省值为“通用”。三个取值中只能选择一个。

3.7.3 领域限制

取值类型：文本 (symbol unlimited)

说明：填写构式的领域特征，比如“法律用语”“学术领域”“方言”等等。

3.7.4 上下文环境

取值类型：文本 (symbol unlimited)

说明：填写构式的上下文环境的典型特征，比如“用于提示后续句子带有负面评价义”。

3.8 构式研究文献

取值类型：文本 (symbol unlimited)

说明：填写针对该构式进行研究的参考文献。参考文献应填写题目、作者、类型（包括论文、专著、工具书）、发表或出版时间、来源（包括期刊、出版机构或论文集）。

例如，构式“那+个+a+啊”对应的参考文献。题目：《“指+量+啊”句的功能扩展》；作者：许剑宇；类型：论文；发表或出版时间：2012年；来源：《杭州师范大学学报》。

§ 4 构式知识库填写工作中常见问题举例

4.1 是词还是构式

在考虑一个语法单位是否作为构式收入构式库时，常见的一个问题是该单位是看作词还是凝固型构式。跟在传统短语结构语法体系中词（特别是复合词）和词组之间的界限难以严格划分类似，词跟凝固型构式之间的界限也存在模糊性。不过，一方面，我们要尽量在理论层面厘清凝固型构式跟词之间的性质差异；另一方面，在构式知识库的语言工程实践中，即便在构式库中收录了一些“类词”的语法单位，也不是坏事。毕竟，从工程角度讲，区分词和凝固型构式并非目的，描述这些语法单位的用法（句法、语义）特点，才是构建知识库的主要目的。

下表中的例子在不同人处理时，可能有人会作为构式收入构式库，有人会把其中的一些常项看作词，把一些常项跟变项的组合看作一般的短语结构组合而非构式²³。

例号	构式形式	构式实例
1.	铆劲儿+vp	铆劲儿补活儿 铆劲儿吃 铆劲儿干完
2.	可劲儿+vp	可劲儿夸你 可劲儿窜红 可劲儿摔打
3.	可算+vp	可算看到你的笑脸了 可算到了家了 可算顶了天了
4.	百般+v	百般抵赖 百般狡辩 百般讨好
5.	闷头+vp	闷头往前走 闷头游着 闷头抽烟
6.	狂+v	狂吠 狂跌 狂吹
7.	奇+a	（双手）奇瘦 （效果）奇强 （速度）奇快
8.	没准+vp	没准要来查户口 没准吓个半死 没准是他
9.	明摆着+vp	明摆着吃亏 明摆着想赖账 明摆着找麻烦
10.	没成想+X	没成想他竟然在这个时候回来了
11.	没曾想+X	没曾想你是这样的人
12.	哪曾想+X	哪曾想就这样当了演员
13.	X+何苦来	劳民伤财何苦来 你何苦来呢 这又何苦来
14.	你没看见+X	你没看见我都忙坏了 你没看见大家都争着往前冲 你没看见他当时可尴尬了
15.	哪里+X	哪里，还是您有学问 哪里哪里，我还要多学习 哪里，我就是个普通人
16.	那可不+X	那可不，他们可不是一般的有钱 那可不，我答应的事情哪一次不兑现的 那可不，要不怎么说他仗义呢

²³ 在构建构式库的实践工作中，确实也发生过这样的情况。

17.	哪儿的话+X	哪儿的话，没什么不好意思 哪儿的话，这是我应该做的 哪儿的话，您别多想
-----	--------	-------------------------------------

例 1-7 属于一类情况，“铆劲儿、可劲儿、可算、百般、闷头、狂、奇”基本都可以分析为副词，在组合时的功能主要是做状语，其句法分布特点跟一般的副词相同。这些成分，收入词库就可以，不需要连带其修饰的 vp 成分一起收入构式库，因为例 1-7 的构式形式所表达的组合，也都可以看作是常规的短语组合，结构整体并无明显的语义增殖。不过，这种处理方式也不是绝对的。以其中例 4 的“百般”来说，可能有人感觉“百般”不能独用，总要跟一个动词性成分 (vp) 组合才能使用，从这个角度来说，就倾向于把“百般+vp”看作一个整体，作为构式收入构式库。而主张将“百般”看作普通词语的人可能会说，“百般”只是一个黏着程度比较高的副词，在使用中也并不总是需要跟双音节动词组合。语料库里也可以查到像“百般地讨好母亲”以及“百般武艺，样样精通”这样的用例，因此，将“百般”看作一个副词（或副词兼区别词），“百般+(地)+vp”看作常规的状中式短语结构，更为合理。我们的认识是，对于这类争议问题，并无截然可分的界限。如果构式收录从严的话，倾向于不收录“百般+vp”这样的构式形式；而如果从宽的话，则收录这一构式形式也可以接受。前文 1.4 节在介绍半凝固型构式时举过“n+百+出”这样的例子，跟“百般+vp”的情况也比较接近。不过，二者相比，“n+百+出”的构式感更强一些，因为这一构式形式中间很难插入其他成分（不像“百般”跟 v 之间还可以插入助词“地”），而且对结构中 n 的约束比较强，n 只能是双音节，通常多为表达贬义的一些名词（如丑态、错误、洋相等），即便是中性名词加上“百+出”后，也往往用在表达贬义的场合（比如“偷税抗税、逃税骗税、欠税漏税、非法减免税，**花样百出**、无所不用其极”）。跟“n+百+出”相比，“百般+vp”的约束相对要弱一些，比如这里的变项实际上并不限于 v，也可以是 vp，“百般”表示量大、程度高，对后面的 vp 的语义并无特别强的限制，可以是贬义（如“抵赖”）、褒义（如“体贴”）或中性义词（如“恳求”）。可见，一个表达形式是否看作构式，有些时候是个程度问题。从工程角度来说，既难以做到一刀切，也没有必要苛求非此即彼。在构式知识库的建设工作中，重点是准确地描写一个表达形式的句法语义特征，对变项的约束条件等。对于什么样的表达形式应该收录入库，什么样的表达形式又应该排除在外，可以定原则，但操作时也允许一定的灵活度。

例 8-13 又属于一类情况，其中“没准、明摆着、没成想、没曾想、哪曾想、何苦来”从内部构成成分的性质以及分布特点来看，基本可以分析为动词，属于情态义动词，不能后接时体助词（不说“*没准了、*没准着”），主要表达主观情态义，反映了说话人对某个命题所代表事件的态度（可能性大小、是否出乎意料、是否值得做等等）。例 8、9 的“构式形式”中变项用 vp 表示，其实也可以是 dj，比如“没准他还会来的”。例 8-13 这一组跟上面一组例 1-7 的不同主要有两点：（1）例 8-13 这一组中的常项成分“没准”等分布情况要复杂一些，不像“铆劲儿”等副词那么单纯；（2）整体的功能范畴和内部结构关系也要复杂一些，可以是 vp（述宾结构、连谓结构等），也可以是 dj（主谓结构）。例 1-7 整体的功能范畴由变项决定，结构上都属于状中结构。不过，

跟例 1-7 类似，对例 8-13 来说，也是既有可能把其中的常项成分当做词语（动词）收入词库处理，也有可能像表中“构式形式”所表示的那样，整体作为一个构式来分析。

例 14-17 的情况不同于上面例 1-13。例 14-17 中的“你没看见、哪里、那可不、哪儿的话”很难归入现有的词类体系，或者勉强归入某个词类，对于计算机分析的帮助并不大，比如把例 15 中的“哪里”看作疑问代词（像表达问处所的“哪儿”），并不能对说明例 15 中“哪里”的用法提供多大的帮助。因此，例 14-17 中的这些成分作为构式处理更好一些。不过，上表中的“构式形式”表达不准确，应做修改：

例号	构式形式	构式实例
14.	你+没+看见+X	你没看见我都忙坏了 你没看见大家都争着往前冲 你没看见他当时可尴尬了
15.	哪里, X 哪里+哪里, X	哪里, 还是您有学问 哪里, 我就是个普通人 哪里哪里, 我还要多学习
16.	那+可+不, X	那可不, 他们可不是一般的有钱 那可不, 我答应 的事情哪一次不兑现的 那可不, 要不怎么说他仗 义呢
17.	哪儿+的+话, X	哪儿的话, 没什么不好意思 哪儿的话, 这是我应 该做的 哪儿的话, 您别多想

常项部分应该以词为基本单位，词与词之间以“+”分隔开。“哪里”“那可不”“哪儿的话”在使用中，书面上后面都有逗号与其后小句分隔，在“构式形式”中应包含这个逗号（中文字符）。构式中的变项成分多数情况应该是 dj，但也可能因为省略成分，在表层形式上有 vp、ap 等，因此构式形式中变项用 X 通配符标记。如果强调范畴更为明确的话，可以将 X 具体化为 dj、vp、ap 等。

下面再看一些例子，这些例子大部分作为构式并不恰当，其中的“构式实例”大多应作为独立的词项收入词库更为合理。

例号	构式形式	构式实例
18.	耳+X+目+Y	耳濡目染 耳聪目明 耳闻目睹
19.	X+枝+Y+叶	金枝玉叶 粗枝大叶 开枝散叶
20.	风+X+浪+Y	风平浪静 风口浪尖 风吹浪打
21.	风+X+日+Y	风和日丽 风吹日晒 风和日美
22.	风+X+雨+Y	风吹雨打 风鬟雨鬓 风行雨散
23.	X 街 Y 巷	穷街僻巷 大街小巷 花街柳巷 走街串巷 穿街过巷 充街盈巷
24.	风+雨+X	风雨交加 风雨飘摇 风雨如晦
25.	飞+X+走+Y	飞檐走壁 飞沙走石 飞禽走兽
26.	n+里+n+外	话里话外 圈里圈外 戏里戏外

例 18-23 中的构式实例对应的变项范畴内部有多种情况，比如例 19 中三个实例的 X、Y 分别为（金、玉）（粗、大）（开、散），分属名、形、动三个语法范畴。整体上，这三个实例也不属于同一范畴，换言之，并没有因为“X+枝+Y+叶”这个整体（看作构式）的作用，把内部成分的不一致性给一致化了。由此也可以反推：将“X+枝+Y+叶”看作一个构式，是不合理的。这里的 3 个实例“金枝玉叶、粗枝大叶、开枝散叶”应属三个独立的词项，它们只是“碰巧”在词项的第二、第四字位置共享了“枝、叶”这两个成分。除此之外，并无充分的理由，把这三个词语模式化为一个构式，它们并没有一个统一的超乎内部成分之上的由结构整体提供的意义（构式义）。

例 24-26 中的构式实例对应的变项范畴相对性质较为单一。比如例 24 中的 X 一般为动词范畴；例 25 中的 X、Y 均为名词范畴；例 26 “构式形式”中变项直接用名词（n）标记，比例 24、25 用 X、Y 这类通配符标记更明确。相比例 18-23 而言，例 24-26 看作构式合理性稍高一些。不过，这些例子的整体语义基本跟内部成分的组合语义相当，没有特别明显的语义增殖。从这个角度说，也可以把对应的实例作为普通词语（或短语组合）处理。像例 26 这样的情况，属于介乎构式和词（固定词组）之间的地带，是否作为构式收入构式库，是两可的。

4.2 是短语还是构式

有些语法形式，可能看作构式，也可能看作常规的短语组合，同样存在界限模糊的问题。下表中的例子就存在这方面的问题：

例号	构式形式	构式实例
1.	一+v	一上来 一跺脚 一咬牙 一瞪眼 一笑
2.	差点儿+vp	差点儿考上 差点赶上 差点儿摔倒
3.	差点儿+没+vp	差点儿没考上 差点没赶上 差点儿没摔倒
4.	差点儿+不+vp	差点儿不及格 差点儿不够 差点儿不能和你见面
5.	v+一+下	百度一下 霸气一下 小结一下
6.	搞不好+X	搞不好他说的是真的 搞不好等会就下雨 搞不好买亏了

例 1 中变项不仅是 v，可以是 vp，比如“一进教室、一打开话匣子”。这个组合中“一”对整体语义影响更大，表达了 vp 所代表的动作行为的先发性、短时性，同时必然引发后续的动作行为。这些语义基本上可以说落在“一”这个成分上，或者说由“一”带来。如果把“一”分析为副词，这个格式也可以作为一般的状中式 vp 来处理，不一定要看作构式。

例 2 的情况也类似，其中变项 vp 要求是肯定形式（不含否定成分）。常项“差点儿”起到否定作用，标记后面的 vp 所代表的事件为假：差点儿考上 = 没有考上。这个结构可以分析为一般的状中式 vp。

例 3 的情况跟例 2 不同，多出了一个常项“没”。结构整体的语义解读包含两种情况：如果 vp 代表的事件是说话人期望发生的，如“及格、够”，则“没”为有效否定项。整个结构表达的语义是 vp 所代表的事件为真，即双重否定等于肯定：差点儿没及格 = 及格了；如果 vp 代表的事件是说话人不期望发生的，如“摔倒”，则“没”为无效的冗余否定项。整个结构表达的语义是 vp 所代表的事件为假，即表面的双重否定仍为否定（违反一般逻辑规则）：差点儿没摔倒 = 没摔倒。由于例 3 存在这样的“形-义”错配现象，把例 3 看作构式更为合理一些。

例 4 的情况跟例 3 一样，“差点儿”起到否定作用，“不”也起到否定作用，二者作为两个否定算子加合，使得整个结构的语义为肯定 vp 所代表的事件为真，即双重否定等于肯定：差点儿不及格 = 及格。例 4 适合作为一般短语结构处理，不需要分析为构式。

例 5 “v+一+下”是动词的常规用法。从表中所举“构式实例”来说，这里的特殊性在于，其中变项并不是由典型的动词充当，而是由名词（或至少兼属名词）来充当，其中“百度”更是专有名词。像这样的情况，如果要分析为构式，则应该将构式形式表述为“n+一+下”，以突出这种特殊性。如果在构式库中用“v+一+下”表达这类组合，就需要在词库中将“百度、霸气、小结”等都增加一个 v 词性标记。从构式库的设计理念来说，应该以“n+一+下”来表达这类结构，而不应该去修改词库信息。

例 6 中“搞不好”跟上一节例 8-13 的情况类似，也可以把“搞不好”看作一个词收入词库，语义跟“没准”很接近。不过“搞不好”本身又可以像一般述补结构那样分析，比如“班里的纪律总是搞不好”“夫妻关系老是搞不好肯定会影响孩子的成长”。从这个角度看，把“搞不好+X”收入构式库，以跟常规的短语结构用法相区别，就有意义了。如果作为常规短语组合来分析，或者要在组合规则中强调有述补结构（“搞不好”）跟单句组合的情况，或者要在词库中收入“搞不好”这个词，同时描述其特殊的分布。这样描写的代价似乎更高一些。

在构式和传统短语组合之间容易发生界限模糊问题的，还有一些由关联成分连接变项组合成的复句结构。比如下面这些组合形式：

- (1) 不管 X 只要 Y 就 Z (2) 不管 X 总 Y (3) 不管 X 还是 Y
- (4) 不单 X 而且 Y (5) 不单 X 还 Y (6) 不单不 X 反倒 Y
- (7) 不但 X 而且 Y (8) 不但 X 反倒 Y (9) 不但 X 更 Y
- (10) 不但 X 还 Y (11) 不但 X 也 Y (12) 不止+X+就+连+Y+也+Z

就总的原则而言，判断这类组合属于常规的复句（短语结构语法体系中的单位），还是分析为构式，首要考虑的是组合的语义。复句型构式的语义可以分解为三个部分：

- (I) 内部小句（分句）独立的命题义；
- (II) 由关联成分表达的分句之间的逻辑关系义；
- (III) 除上面两部分意义之外的附加意义（如主观态度义、衍推义等等）

如果一个组合的语义只有上面 (I) (II) 两部分，则归属常规复句。第 (III) 部分的语义越突出，则越容易作为复句型构式收入构式库。

因语义的感知有较大的主观性，在具体操作层面，还可以考虑以下三方面的情况，来帮助做出选择：

(1) 组合中的变项在形式上是否有明显的特征，比如复现；

(2) 组合中的常项是否都是专职的关联词语，即考察关联成分是否总是共现（起关联作用）还是可以在其他场合独用（不起关联作用）；

(3) 变项之间语义关联的复杂程度——整体的语义跟组成成分语义之间的联系；

上面(1)-(12)组合形式中(2)-(11)都只包含两个变项，常项成分作为关联词语，都是专职的，变项之间的逻辑关系也比较单纯，可以按照一般复句分析。例(1)和例(12)都包含三个变项，变项之间的语义联系比较复杂，整体有一定的语义增殖，例(1)强调了Y作为Z的充分条件，Y真则Z必为真；例(12)则强调了在一个等级序列中最不可能为真的命题(Y-Z)为真。从这个角度说，例(1)和例(12)可以作为构式。但是，例(1)和例(12)存在两方面的问题，使得“构式形式”的表述存在困难。一方面，“构式形式”中的变项有成分不连续的情况，比如例(1)中的“只要Y就Z”形成的关联，其中Y部分并不一定是个完整的语法单位，而可能是Y中有一部分跟Z发生语法关系，比如下面例(1a)中Y对应的片段里“教她一遍”是一个语法单位，“她”跟Z对应的片段“记住了”是一个语法单位。这就造成了“不管X，只要Y就Z”这样的线性序列化的构式形式表示难以匹配真实语料的实例。另一方面，“构式形式”中的常项并不都是专职的关联词语，还可能有非关联用法，这样就造成字符串匹配成功的实例中，有的可能并不是真的构式。下面例(1b)和例(12b)都不是例(1)和例(12)对应的构式实例。

(1a) 不管啥戏，只要教她一遍，她就记住了。

(1b) 我们先不管他了，只要他不露头，就影响不了大局。

(12a) 老张不止没有给孩子留一分钱存款，就连他本人也没有一件像样的衣服。

(12b) 事情远不止这几件，就连他本人也不清楚犯了多少事。

例(1b)中的“不管”和例(12b)中的“不止”在这里都不是作为关联成分在使用，而是作为一般动词在使用。但纯粹从符号形式序列上说，例(1b)跟例(1a)都能跟例(1)的构式形式匹配上，例(12b)跟例(12a)也都能跟例(12)的构式形式匹配上。

因为上面两方面的原因，例(1)和例(12)若作为构式收入构式库，在描写其变项成分，以及变项成分的约束条件，变项成分之间的关联约束条件时，就存在不少困难。

上文1.4节在介绍复句型构式时举的例子“别+说+是+X，就是+Y+也+Z”和“v+也+不是，不+v+也+不是”，相对来说，作为构式描述就容易一些，后一个例子中变项为重复成分；两例中的常项部分都是只能分析为专职关联成分，不能做其他用途。

总体来说，收入构式库的条目，应能准确描述其变项和常项的约束条件，容易归纳整体构式义，这样在匹配实例时就能够避免混入伪构式用例。遵循这样的策略，实质上就强调了构式库中收录构式优先考虑准确性，这样做可能是以丧失一定的覆盖率为代价，但尽可能保证了收入进来的构式都是“精品”。而那些难以处理的问题，暂时仍然留在短语结构的层面，因为后者原本就是开放的。这些问题可以留待以后再考虑寻找合适的处理策略。

4.3 变项的概括度：“a+中+的+a”还是“X+中+的+X”

“a+中+的+a”（特殊中的特殊）和“n+中+的+n”（天才中的天才）是分立两个条目，还是合并为“X+中+的+X”一个构式条目。在构式库建设初期阶段，我们倾向于分立前两个条目，如果有需要，也可以同时收入“X+中+的+X”这一更为概括的条目，并描写前面两个条目跟后面这个条目是上下位关系。之所以这样处理，是考虑到以下因素：

（1）如果只收入“X+中+的+X”条目，而不收入“a+中+的+a”和“n+中+的+n”，就要求在填写“构式形式”后，紧接着就对X进行约束描写，把X的范畴约束条件表达出来。但现阶段的操作方式并不是这样的，仅仅是收录构式，填写构式实例，还没有对构式的常项和变项特征做任何描写。这样的话，构式中的X到底是什么，就不清楚了。而如果用“a+中+的+a”“n+中+的+n”去描写一个构式，就算不对a、n等变项做细节约束，a本身就已经限定在形容词这个类别，起到一定的约束作用了。

在考虑“构式形式”如何表达更为合适的时候，应该多从计算机模式匹配的角度去权衡。假定一个带分词和词性标记信息的文本，用构式库中的“构式形式”去做模式匹配，如何表示构式形式，对于提高匹配的精确率有帮助呢？显然是“a+中+的+a”这样的形式比“X+中+的+X”更严格。

（2）之所以在“构式形式”中同时使用n, v, np, vp等传统语法范畴标记和X, Y等通配性标记，是出于工程上的折中考虑。在具体操作层面，原则是能明确到n, v等传统语法范畴就不用X, Y。同时使用这两个层面的标记，实质上就是在构式描写的准确性和概括性两个方面权衡。只强调一个层面的标记，都不是太合适。对计算机的应用来说，强调准确性更有利于处理；而对人的认知习惯来说，强调概括性也是合理的。因此，允许抽象程度不同的标记来描述变项成分，为构式形式的表达提供了灵活性，同时也使得构式库中的条目可能有一定的冗余。在构式库建设进入相对成熟的阶段，可以再来考虑合并精简的处理。

（3）我们现在做的构式语法知识库工程有别于许多构式语法的理论研究，是希望在“数据（描写）”层面，将构式的研究成果集成到传统的短语句法结构分析中，是把基于不同理念的语法理论本体研究尽量结合起来，而不是强调构式语法观跟传统语法分析方法的区别。要把构式跟以往的短语结构分析衔接起来，在构式形式的表示这个基础环节，就得借用传统语法单位的那些范畴（标记）。否则，就没有结合的基础。也正是因

为如此，在描写构式的变项成分时，我们强调，要尽量使用传统短语结构语法体系中的范畴标记（包括词类范畴、短语类范畴等）。同时，在描写构式的常项成分时，也要注意以词（词库中已有的记录）为基本单位。

4.4 变项的概括度：“n1+的+n2+v+得+a”还是“np+v+得+X”

“n1+的+n2+v+得+a”对应的构式实例是“他的老师当得好”。之所以看作构式，是这里的“n1+的+n2”跟一般的定中结构（如“他的老师”）同形，但并不能按照一般定中结构的语义解读方式来理解，因此其中“n1+的+n2”也被称作准定中结构。同类型的例子还有：

- (1) 他的相声说得我都睡着了
- (2) 他的篮球打得简直要进 NBA

对于这类例子，在“构式形式”的表示方面，就存在不同选择的可能性：

方式 1: n1+的+n2+v+得+ap
 n1+的+n2+v+得+dj
 n1+的+n2+v+得+vp

方式 2: n1+的+n2+v+得+X

方式 3: np+v+得+X

从方式 1 到方式 3，构式形式表示的概括度在不断增高，即越来越抽象。这样带来的好处是可能对语料中同型例子的覆盖率会提高，但缺点就是容易过度泛化，造成模式匹配时准确率的下降。另外，在构式知识库描述变项间约束关系的时候，方式 1 中，比较容易描写 n1、n2 跟后面的谓词性成分 ap、dj、vp 等之间的语义联系。而方式 3 中 np 的内部情况如何描写，np 跟 X 之间是什么关系，都带来了描述上的复杂性和困难。

从方式 3 到方式 1，之间的关系类似于 C++编程里函数模板跟特化的具体函数的关系。目前一般的句法分析程序依赖的语言结构模型是短语结构语法知识体系，在用组合规则去匹配实例的时候，还是会用方式 1 的那种“构式形式”。方式 2 和方式 3 的好处是抽象，方便描写共性。对人来说，看着概括、舒服，不啰嗦。而方式 1 的好处则是计算机方便使用。

在构式库建设的当前阶段，我们优先考虑使用方式 1 来表达构式形式。这里不妨再举个“被+X”的构式例子。对计算机来说，“被+X” “被+v” “被+自杀”三种表示构式的形式相比，“被+自杀”这种最为具体的形式（抽象度最差）。处理起来最为容易。从理论上讲，如果所有的短语结构实例、构式实例，都像词条一样收入数据库，即让计算机“死记硬背”这些逐项列举的语言形式，那正是计算机的强项，处理起来就容易多了。但显然，因为语法组合形式的开放性，这样“穷举实例”的方法又很难做到。于是，人们还得从具体到抽象，收录“被+v”“被+X”这样的形式，希望能够以简驭繁。

归结起来，对于构式形式表述的概括度，应该认识到，概括是有弹性的。如果构式库中既有概括的形式，又有精确的形式，可以兼顾不同的应用需求，从工程上讲更好用（当然代价是增加工程工作量）。如果一定要二选一的话，优先选择是满足描写的精确度要求，可以损失概括度，而不是相反。

4.5 变项的概括度：m+q 还是 qp

在描写“一锅饭吃十个人”这样的数量比例义构式时，其中的“一锅饭”部分是描写为 m+q+n（即“数词+量词+名词”序列），还是将“m+q”的部分合并为 qp（数量短语）？

前面已经表述过这样的原则：优先考虑精确度（其次才是概括度）。因此，分开描写为“m+q”比合并为 qp 更好。

qp 是短语范畴，主要由 m+q 组成，但也并不限于这一种组合模式，也可能包含两个数量短语并列的结构类型，比如“三天两天”，也是一种 qp。但这种 qp 就不能出现在数量比例以构式中。显然，用短语范畴标记来表达变项，仍然存在“过度泛化”的风险。

从表达精确性的角度考虑，能用词类范畴表达的变项成分，就不应该用短语范畴来表达。

类似的例子还有：“那个兴致高啊”对应的构式形式是表述为“那+个+dj+啊”（dj 对应“兴致高”部分），还是表述为：“那+个+n+a+啊”（用“n+a”对应“兴致高”部分）？

显然，基于上面已经给出的理由，后一种表述形式更好。因为这种构式形式表达得更准确。更有利于表达变项之间，变项和常项之间可能存在的约束条件关系。

4.6 构式形式是否最简

上面 4.5 节提到的例子“那个兴致高啊”还有一个意思相近的表达方式：“兴致那个高啊”。

这就引出一个问题，这也应该看作是一个构式的实例，那么这个构式该如何表示呢？是不是表示为：n+那+个+a+啊？显然，这样表示，可以方便进一步描述这一个构式形式跟“那+个+n+a+啊”之间的联系。

对这个问题，有两种可能的考虑。第一种考虑：根据收录构式形式最简原则，这个“n+那+个+a+啊”可以看作是 n 加上“那+个+a+啊”形成的，后者已经作为一个独立的构式收入构式库（例如：那个高兴啊，那个紧张啊）。因此，“n+那+个+a+啊”就可以不再收录。前面作为主语的 n 可以不看作构式的一部分。

第二种考虑：这里的“兴致”跟“高”可以看作类似词的离合用法，即“兴致高”从意念上可以看作一个单位。如果把“n+那+个+a+啊”作为一个独立构式收入构式库，

方便描写其中 n 跟 a 之间的联系，这比把该构式看作是 n 跟构式“那+个+a+啊”的组合更方便一些。

从构式库的精简性角度看，收入“n+那+个+a+啊”条目似乎有些冗余，但从表达构式内部成分的相互约束条件的方便角度，以及构式形式跟实例的模式匹配准确度方面考虑，收入这个构式也有好处。

从原则上讲，收入库中的构式条目应该是形式上最简的。但在实践中，也允许收入少量不是最简的构式形式，允许存在一定的冗余条目。只要有利于提高对构式成分约束条件描写的准确性，方便今后将构式库知识融入到短语结构语法体系，有利于对句子结构的整体分析，就是值得的。

4.7 “构式变体”与“近义构式”

在构式知识库的字段设置中，“构式变体”字段用于表示两个形式和意义上非常相近的构式，但常项成分有差别；“近义构式”字段用于表示两个形式和意义上非常相近的构式，但变项成分有差别²⁴。例如：

(1) “n+中+的+n”跟“n+中+之+n”处理为“构式变体”关系。

(2) “a+什么+a”跟v+什么+v处理为“近义构式”关系。

需要注意的是，“构式变体”这个名称可能有误导作用，在两个构式形式之间，有可能并不能真的确定谁是“正体”，谁是“变体”。这时候只是用“构式变体”这个字段设置来表达两个表达形式 A、B 之间存在一种抽象关系，当 A 构式的“构式变体”字段填值为 B 构式时，表达的意思是：虽然 B 跟 A 不完全同形，但 B 跟 A 只是在常项成分上有区别，B 跟 A 可以共享关于构式 A 的绝大部分信息描述。

在操作层面，一般会选择那个更常见的形式作为“正体”²⁵。比如，构式数据库中应该收入“a+中+的+a”作为一个条目，该条目的“构式变体”字段，可以填写“a+中+之+a”。而不是反过来收入“a+中+之+a”条目，再在“构式变体”字段填“a+中+的+a”。因为前者作为含有文言成分的形式，在现代白话文本中，可能没有后者常见。

4.8 构式实例的形式与意义要求

构式实例除形式上符合“构式形式”的模板匹配要求外，语义上要有共性。不能把一些“伪实例”放进来。（下面举的例子暂且认为“构式形式”没问题）

例 1:

²⁴ 当然，“近义构式”字段描述的情况不限于此，还包括属于“同一个构式家族”的情况。详见 3.1.11 小节对“近义构式”字段的说明。

²⁵ 选择哪一个形式为“正体”，哪一个形式为“变体”，可能有一定的主观性，不同人的处理不一定一致。

构式形式	构式类型	构式特征
------	------	------

上面“构式实例”字段中的第三个例子不是这个构式的实例。这个构式是数量分配义。第三个例子前面可以加“把”，n跟v之间的语义关系跟前两个例子不同。

“构式实例”还应注意跟“构式形式”中的常项部分保持严格一致。

例 2:

构式形式	构式类型	构式特征
------	------	------

例 2 “构式形式”中的常项是“不止”，但实例中的对应常项为“不只”。另一个常项为“而且”，实例中有一个对应的常项为“而”。这些都没有做到严格保持一致。

4.9 半凝固型构式中变项不应超过两项

构式库中收录了一些半凝固型构式，其中变项超过两项，而且用通配标记 X、Y 等表示变项，过于抽象。这样表示构式，不容易从抽象的“形式”想到“实例”，很难概括共同的构式义，也不好描述变项需要满足哪些约束。

序号	构式形式	构式类型	构式特征
----	------	------	------

显然，上面这两个条目中的实例各自都没有共同的构式义，都不适合看作构式。

参考文献

- Abney, Steven, 1991, Parsing by Chunks, In Robert Berwick, Steven Abney and Carol Tenny (eds.) *Principle-Based Parsing*, Kluwer Academic Publishers.
- Boas, H.C. and Sag, I.A. (eds.) 2012, *Sign-based Construction Grammar*, CLSI Publications, Stanford University.
- Bryant, John, 2004, Scalable Construction-Based Parsing and Semantic Analysis, In *Proceedings of 2nd Workshop On Scalable Natural Language Understanding*, Boston, U. S. A.
- Chang, Nancy, Feldman, Jerome, Porzel, Robert, and Sanders, Keith 2002. Scaling cognitive linguistics: Formalisms for language understanding. In *Proceedings of 1st International Workshop on Scalable Natural Language Understanding*, Heidelberg, Germany.
- Croft, W. & Cruse, D. A. 2004, *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Fillmore, C. J., Kay, P. and M. C. O' Connor (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone. *Language* Vol. 64, No. 3, pp. 501-538.
- Fillmore, C. J., Russell R. Lee-Goldman, and Russell Rhodes, 2012, The FrameNet Constructicon, in Boas, H.C. and Sag, I.A. (eds.) 2012, *Sign-based Construction Grammar*, CLSI Publications, Stanford University.
- Goldberg, A. E. 1995, *A Construction Grammar Approach to Argument Structure*, Chicago and London: The University of Chicago Press. (中译本: 《构式: 论元结构的构式语法研究》, 吴海波译, 冯奇审订, 北京大学出版社, 2007年版)
- Goldberg, A. E. 2006, *Construction at work: The Nature of generalization in language*, Oxford University Press. (中译本: 《运作中的构式: 语言概括的本质》, 吴海波译, 北京大学出版社, 2013年版)
- Goldberg, A. E. 2013, Constructionist approaches, Hoffmann, T. & Trousdale G., eds., 2013, *The Oxford Handbook of Construction Grammar*, Oxford University Press.
- Hoffmann, T. & Trousdale G., (eds.), 2013, *The Oxford Handbook of Construction Grammar*, Oxford University Press.
- Igor A. Mel'čuk, 2012, *Semantics: From meaning to text*, John Benjamins Publishing Company.
- Kay, Paul and Fillmore, Charles J. 1999, Grammatical Constructions and Linguistic Generalizations: the What's X doing Y? Construction, *Language* Vol. 75, No. 1, pp. 1-33.

- Langacker, Ronald W., 1987, *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Vol.1. Stanford University Press.
- Panther, Klaus-Uwe & Thornburg, Linda L., 2014, What do you think you're doing? Exploiting Wh-questions for expressive purposes, Presented in The 1st International Symposium on Figurative Thought and Language, Aristotle University of Thessaloniki, 25-26 April 2014.
- Sag, Ivan A., 2010, English filler-gap construction, *Language*, Volume 86, Number 3, September 2010, pp. 486-545.
- Steels, Luc, ed., 2012, *Computational Issues in Fluid Construction Grammar: A New Formalism for the Representation of Lexicons and Grammars*, ISBN: 978-3-642-34119-9 (Print) 978-3-642-34120-5 (Online), Lecture Notes in Computer Science, Volume 7249. Springer Berlin Heidelberg.
- 陈忠 主编, 2013, 《汉语句式研究与教学专题论文集》北京语言大学出版社 2013 年版。
- 段业辉、刘树晟 等著, 2012, 《现代汉语构式语法研究》世界图书出版公司 2012 年版。
- 甘莅豪, 2012, 《空间动因作用下的对举结构》上海社会科学院出版社 2012 年版。
- 顾鸣镝, 2013, 《认知构式语法的理论演绎与应用研究》学林出版社 2013 年版。
- 郭绍虞, 1978, 汉语词组对汉语语法研究的重要性, 《复旦大学学报》1978 年第 1 期。
- 郭霞, 2013, 《现代汉语动趋构式的句法语义研究》四川大学出版社 2013 年版。
- 李宗江、王慧兰, 2011, 《汉语新虚词》, 上海教育出版社 2011 年版。
- 廖巧云, 2011, 《因果构式的运作机理研究》中国社会科学出版社 2011 年版。
- 廖巧云, 2011, 《因果构式的运作机理研究》中国社会科学出版社 2011 年版。
- 刘大为, 2010, 从语法构式到修辞构式(上、下)《当代修辞学》2010 年第 3、4 期。
- 刘丹青, 2005, 作为典型构式句的非典型“连”字句, 《语言教学与研究》2005 年第 4 期。
- 刘正光 主编, 2011, 《构式语法研究》上海外语教育出版社 2011 年版。
- 陆俭明, 2004, “句式语法”理论与汉语研究, 《中国语文》2004 年第 5 期。
- 陆俭明, 2009a, 构式与意象图式, 《北京大学学报》(哲学社会科学版), 2009 年第 3 期。
- 陆俭明, 2009b, 构式、语块、汉语教学, 载蔡昌卓主编《多位视野下的汉语教学——第七届国际汉语教学学术研讨会论文集》, 广西师范大学出版社 2009 年版。
- 陆俭明, 2011, 再论构式语法分析法, 《语言研究》2011 年 4 月, 第 31 卷第 2 期, pp. 1-7。
- 陆俭明, 2012, 相同词语之间语义结构关系的多重性再议, 《苏州大学学报》2012 年第 4 期, pp. 5-11。
- 牛保义 编著, 2011, 《构式语法理论研究》上海外语教育出版社 2011 年版。
- 施春宏, 2012, 从构式压制看语法和修辞的互动关系, 《当代修辞学》2012 年第 1 期。
- 苏丹洁, 2010, 试析“构式-语块”教学法——以存现句教学实验为例, 《汉语学习》第 2 期。

- 苏丹洁、陆俭明, 2010, “构式 - 语块”句法分析法和教学法, 《世界汉语教学》第 4 期, pp.
- 王寅, 2011a, 《构式语法研究(上卷): 理论思索》上海外语教育出版社 2011 年版。
- 王寅, 2011b, 《构式语法研究(下卷): 分析应用》上海外语教育出版社 2011 年版。
- 薛小芳、施春宏, 2013, 语块的性质及汉语语块系统的层级关系, 《当代修辞学》2013 年第 3 期, pp. 32-46。
- 袁毓林, 2004, 论元结构和句式结构互动的动因、机制和条件——表达精细化对动词配价和句式构造的影响, 《语言研究》2004 年第 4 期。
- 詹卫东, 2000, 《面向中文信息处理的现代汉语短语结构规则研究》清华大学出版社。
- 詹卫东, 2005, 以“计算”的眼光看汉语语法研究的“本位”问题, 《汉语学报》, 商务印书馆 2005 年第 1 期。pp. 64-73。
- 张伯江, 1999, 现代汉语的双及物结构式, 《中国语文》1999 第 3 期。
- 张伯江, 2000, 论“把”字句的句式语义, 《语言研究》2000 年第 1 期。
- 张伯江, 2009, 《从施受关系到句式语义》, 商务印书馆, 2009 年版。
- 张娟, 2013, 国内汉语构式语法研究十年, 《汉语学习》2013 年第 2 期。
- 张立飞、严辰松, 2011, 《现代汉语否定构式的认知研究——一项语料库驱动的研究》高等教育出版社 2011 年版。
- 张晓涛, 2011, 《疑问和否定的相同性及构式整合研究》中国社会科学出版社 2011 年版。
- 周强、詹卫东、任海波, 2001, 构建大规模的汉语语块库, 载黄昌宁、张普主编《自然语言理解与机器翻译》, 清华大学出版社 2001 年版, pp. 102-107。
- 朱德熙, 1982, 《语法讲义》, 商务印书馆。
- 朱德熙, 1985, 《语法答问》, 商务印书馆。
- 朱军, 2010, 《汉语构式语法研究》中国社会科学出版社 2010 年版。

--- 完 ---