

一个语义的结构模型假设

惠新标 陈志强 叶楠 周德雨
上海一知先网络信息技术有限公司, 上海 200120

摘要: 本文尝试从含义表达的层面出发, 对语义及其相关的问题进行数学化的探讨。在对一些与语义中的含义相关的内容作一些简单梳理后, 首先提出一个关于概念的数学模型假设, 在此概念模型基础上提出一个拟全语义空间和语义预范畴的数学模型, 以及语义对象在语义预范畴的表示, 并对它的性质做一些初步的研究。在此语义模型的基础上讨论语义预范畴的形式化的方法工具, 并用类似数学的局部化方法对语义的拟全/形式二重性进行讨论, 初步探讨形式化语义预范畴的性质, 拟全/形式语义子预范畴与表达之间的关系。然后给出这种二重性模型在语义处理, 特别是自然语言处理和人工智能方面的可能应用, 并给出两个简单的例子。最后给出一些基于本文语义结构模型假设的一些基本结论, 以及结合数学、哲学、自然科学、社会科学和人文科学可能的学科综合研究方向和应用设想。

关键词: 语义空间, 语义结构模型, 拟全化, 形式化, 语义预范畴, 可计算模型, 拟无限
分类号: B017, TP391

An Assumption of Semantic Structure Model

Hui Xinbiao, Chen Zhiqiang, Ye Nan, Zhou Deyu
ith-science Information Technology Co. Ltd., Shanghai 200120

Abstract: In this paper a mathematical way of explanation is explored for semantic problems. First a mathematical model for concept is proposed after a brief introduction of semantic meaning. Based on this a pro-complete semantic space is proposed and semantic “pre-category” model is discussed, and how the semantic object is presented in this pre-category, also some of its characteristics are discussed. Then a formalization method is set up for this semantic pre-category, and a localization strategy is discussed to address the semantic duality of pro-complete/formal, also the characteristics of formal semantic pre-category, the relation between pro-complete/formal semantic pre-category and expression. After that, some application using this duality model are discussed for semantic processing, especially Natural Language Processing (NLP) and Artificial Intelligence (AI). Two simple examples are discussed following that. In the end a summary of basic conclusions based on this semantic model is given and its prospect on cross study with mathematics, philosophy, nature science, social science and humanities is raised for future discussion.

Keywords : Semantic Space, Semantic Model, Pro-complete, Formal Method, Semantic Pre-category, Computational Model, Pro-infinite

1 引言

关于语义的研究由来已久^{[1][2][3][4][5][6][7][8][9]}，特别是在人工智能^{[10][11][12]}发展后，对语义在智能化方面的应用研究发展更快。这些研究形成了很多学术分支，它们对跟语义及其应用相关的某些方向起到了很大的推进作用。比如自然语言处理、机器翻译等等。这些语义的研究，往往集中在两个主要的层面上：一个层面是与哲学与认知相关的层面，这个层面的研究往往比较笼统，缺乏一些实际落地的基础，近来主要的进展在认知科学领域^{[13][14][15][16][17]}；另一个层面是直接面向应用，包括计算机语言、编译^[18]，还包括自然语言的处理^{[19][20]}、机器翻译^{[21][22]}等等。采用的数学工具主要以数理逻辑^{[23][24][25]}为主，辅以一些基本的代数^[26]等工具。

分析这两个研究层面，可以发现：这些研究对于语义的表象外延（如认知学中关于语义与神经活动的对应^{[16][27][28]}）和应用（如形式语言^{[29][30][31][32]}、自动翻译^{[21][22]}）取得了不少的进展，但关于语义本身的性质和结构的模型，似乎并没有什么很好的进展，还是处于一种“说不清楚”的状态。

本文重点从语义本身的结构和性质模型出发，给出一些假设，建立相应的数学模型，希望从这个层面给出一些新的研究基础和研究路径，以促进或提示对语义研究和思维的工作机制的研究的进一步深化，拓宽语义研究的应用范围，加快推进语义相关的应用进程。

下面我们先从意识和智能相关的角度对语义进行一些梳理。

本文对一些常用术语和概念的梳理，虽然会尽量采用更有效和便于理解的表述，但不追求严格的定义。其原因将在概念的数学模型相关章节部分解释。

需要在这里说明一下的是，我们对语义的认识并没有一个清晰的“基础”，也没有一个最终的“顶点”，只能以一些不是很清晰的混合的认知为起点。这里也借用 Stalnaker^[33]的说法，我们“从中间开始”。甚至我们在本文中也不追求寻找最根本的“基础”以及到达最终的“顶点”，而是开放越来越接近“基础”或“顶点”的可能性，冀望有兴趣的研究者们可以不断努力逼近。

1.1 意识、语义、思维与智能

意识作为哲学研究的一个基本问题，它的基础和机制一直以来都很有争议^{[34][16][35][36][37][38][39]}，本文基本不对意识的哲学基础这个古老而又常新的问题做太多的探究，也不涉及意识的生成机制问题，以免陷于没有结论的深渊。

但语义、思维、智能这些都是基于意识域展开的研究，因此无法完全回避，必然也会提及一些，只是不作为本文的重要研究对象。这里仅做一些关于意识和语义（不予证明）的基本假设：语义是属于意识域的对象。

本文把语义看作是广义的，即语义不限于语言表达所传递的含义^[49]，而是囊括所有表达形式（比如绘画、音乐、手势、表情等）所传递的含义。即，语义是所有的交流所采用的表达形式中所传递的含义。也就是说，语义是交流的基础。

为了进一步梳理语义与智能^[40]的联系，这里综合进化论^[41]、生物学、生理学^{[27][28]}和心理学^{[14][42][43]}的一些常识，对交流的产生与发展，做一些假设性的推演。由于没有严格的实验或考古的证据，这些假设只作为一种语义和智能之间联系的发生发展思路供参考，对从当前时点出发的纯形而上^[44]的直观思考做些补充。好在它们对后续章节的影响并不是根本性的。

交流的最初发展是基于双向的“刺激—响应”^[42]机制，生物对外界刺激的响应在生理学中有很多研究，是生物基本反应，从生物保持最大程度存活的目的衍生而来。在生物达到一定的复杂程度后，这种“刺激—响应”过程也会变得更复杂。比如太阳直射引起生物趋光或避光的响应，但这样的响应可能并不是生物体一个动作就能完成，比如避光需要找遮阴的地方。这时候生物的响应就会产生一种主动，试探哪里有遮阴的地方，并感知这种试探的反应或结果。这种“主动”，应该是生物（“意识”）发展中的一个重要节点。

如果被试探的对象也是生物，比如在饥饿刺激的响应下发出猎食的试探，实际上这时被试探的对象收到了“刺激”，对应也会产生“响应”，于是就会有猎食的攻防。这种双向的“刺激—响应”也许就是最初的生物之间的“交流”。

随着生物不断进化发展，这种交流也愈发复杂，比如通过碰触、振动、化学物交换等在同类之间交流信息。

同时生物的进化带来了记忆能力的发展，这些交流逐渐形成某种固定的记忆模式。在机体接收到某种感知模式的时候，就能“理解”交流的信息。这应该就是语义的最初形式。

进一步的发展当然就是交流（表达）的方式越来越丰富、能表达的语义的内涵也越来越复杂。交流形成一些“约定俗成”的模式后，就体现出了初步的“智能”。而这些模式的一部分，逐渐进入到基因里，形成“先天”的遗传。

这种记忆和交流能力随着生物的进化进一步发展时，交流中应该能逐渐意识到“对方”这样的“对象”，并慢慢形成“概念”。随后在交流中可以交换对第三方“对象”的概念性信息，比如在同类之间交换“有“食物””这样的信息。

要注意到交流的对方当然也可以作为“对象”被概念化。于是在记忆的帮助下，就会出现当交流的对方并不实际存在时，也可以试图跟概念化的“它”交换信息，并根据记忆和经验预测“它”的响应。这时生物就能发展出一种“虚拟”的试探交流能力。

这种虚拟的交流能力，就是“思维”的雏形了。

同时，在与各种“对象”交流的过程中，慢慢会接收到它们对于“自己”

的信息，并逐渐理解。而意识到“自己”也是“对象”之一后，当把虚拟的对象换作自己时，生物就实现了与自己的交流。这种交流的思维就是“思考”，脱离交流对象的交流和思维能力。

在思考的过程中，如果再把第三方对象也换成自己，根据与别人的关于自己的交流，以及对自己的记忆和经验来研究自己，那么就会逐渐形成“自我”的意识。这方面的研究也还在不断进展^{[34][45]}。

在这个过程中，换位成自己以及自我的存在很隐晦，我们甚至并没有意识到在跟自己交流。但是，如果仔细分析，是不是我们都经常用语言进行思考？这些语言是对谁说的呢？即使有时是使用了图片等其它交流方式，有没有意识到在脑海里的那些图片等内容好象是在向谁展示？是谁呢？是不是自己？这些过程，与我们平时跟他人沟通交流几乎是一样的情景。

智能则也是在语义的基础上，伴随以上的交流和思维发展过程不断增强。

智能是在语义的基础上进行的有思维参与的综合应用过程。从最初的“试探—反作用—感知—反应”的循环，发展到：“感知—处理—判决—目标—规划—指令—行动（一递归）—达成目标”这样一个复杂的完整的智能事件过程。而这其中处理和判决的基础就是基于语义的思维。而规划的结果，是思维给出的语义结论。指令的发出，也是某种对语义结论的表达。如果有多人存在，这个指令，还必须有一定的能让其他人理解的语义基础。除了感知和行动外，其余都涉及到思维的参与，也都是以语义为基础的。智能是个主动与被动结合的过程，不是一个简单的静态概念。而其中生成目标和达成目标的主动性，应该是高级智能的要点。

关于意识、语义、思维与智能的关系，这里不再深入，后面把主要焦点放在语义上。

1.2 语义研究及其进展

语义的本意应该是关于语言所表达的内容的理解。在科尔内的《语义学：从数学基础到语义语用学》^[3]中，关于语义的表述为：从柏拉图起，语义学的核心思想为：认识某物意味着具备一种能解释其组成部分的能力。其中还引用了弗雷格的表述：复杂表达的语义由它的结构和组成部分的语义决定。

这些其实都是一些简单的解释尝试，不能被看作语义的概念定义。而且，这里两种表述都明显采用了一些递归性的方法。这种递归式的表述在对语义的理解中普遍存在，包括涉及到概念和词汇的意义解释等问题^[3]，这种递归表述存在的合理性，在后面概念的数学模型章节中将会有一些解释。

对语义的研究主要有几个方面：

一是关于语义在语言或其他表达方式（比如绘画、音乐）中的体现的研究，即研究表达中的语义究竟是什么的问题。

二是关于语义在思维、理解、智能方面所起的作用及其机制的研究。

三是关于语义在计算机系统中的应用，比如编译、机器翻译、人工智能等。

这几个方面的研究又相互交织、互相促进。虽然在计算机及计算机语言（或形式语言）出现之前，关于语义的研究也有一些进展，主要是关于历时语

义的研究^[5]，但对语义的研究发展得比较快的，主要还是在计算语言^[18]出现之后。

20 世纪 50 年代义素分析和语义场^{[2] [46]}理论产生为标志，语义学才成为一门独立学科。从 20 世纪 60 年代起语义学才逐渐受到重视。目前发展起来的语义学主要有三大分支：语言语义学、哲学语义学、逻辑语义学。而主流的语义学理论有：结构语义学、解释语义学、生成语义学、形式语义学、功能语义学、蒙太古语义学。这些理论几乎都与计算理论^[47]相关。

这些与计算相关的语义学，主要以研究其在计算理论中的应用为主，所以，其发展也在一定程度上受限于计算理论本身的发展。

在乔姆斯基关于语法的研究^{[48] [49] [50]}有了比较大的进展后，语义的研究似乎主要集中在对一个句子的计算表示。即，主要研究用什么方法可以通用地（无歧义地）表达一句程序语言所描述的内容，并可以让计算机理解，应用于实际计算操作中。

这样，关于语义的研究，对语义在表达中的体现理解、作用和性质机制的研究相对就少了一些，更倾向于迁就应用。而且常常受限于哲学、认知科学以及数理逻辑的手段，相关的模型都有一定局限性，经常会引起比较大的争论和分支，出现各种语义学流派^[3]。

这种局限性一方面是因为上述的计算理论和实践的实际需要引起的，另一方面不可忽视的原因则是对语义的本意进行分析太过复杂，相关议题涉及面太宽泛。一般的研究很难取得突破并产生可以应用的成果。

简单列举一下与语义相关的研究领域，就可以知道与语义关联的议题的宽泛程度：哲学、逻辑学、语言学、计算理论、认知科学、人工智能、交流和传播……，在如此宽泛的范围内对一个问题进行研究，也可以看出这个问题的复杂性。

综合看来，语义相关的问题在这些领域的研究都有一些进展，但明显也能感觉到有一些比较大的障碍，进展不是很顺利。仔细分析的话，有些针对应用的研究还是取得了一些成果。但是，感觉几乎所有的研究在碰到一些真正的核心问题时，不知不觉地选择了绕路。

至今，对语义的本质在上述领域中的应用和研究，还是缺乏一些基础和关键的联系。特别是，没有一个关于语义的有效数学模型，可以在这些领域中广泛应用。

这可能跟之前的研究方式和研究应用的目的有关。当然，语义本身的复杂性，也容易让人迷失在纷繁的应用细节之中，很难捕捉到它的真实脉络。

1.3 语义研究的新尝试

前面讲到，关于语义的研究，很多倾向于直指应用的细节化研究，往往有些缺乏整体性的感觉，甚至有些研究有一种被语法结构绑架的意味。试图用一些基本元素和现象，构建整个语义大厦，却说不清语义的结构框架。具体可参

见语义学对语义流派的引用^{[2][3][4]}。

本文尝试在概念和含义的层面，换个角度对语义及其相关的因素展开一些分析，并在意识领域建立起与语义相关的基础数学模型。

本文将首先简单分析和整理这里选取的层面和角度所涉及到的研究对象，以及相关的一些重要问题和困难，然后从属性和概念开始，建立起与语义相关的一些数学模型，给出在数学模型基础上的语义结构和语义分析工具，并进一步在此基础上研究它们的形式化，以及它们与智能、人工智能的关系和可能的应用。

有关概念的研究^{[4][51][52]}虽然也不少，但文献中很少有它和语义的关系的直接而深入的研究。从现在的语义研究多用形式方法的角度看，确实很难找到概念与语义之间的直接联系。

这里将从概念的拟全化（其定义参考后面概念的数学定义）的数学模型着手，尝试寻找语义和概念的数学联系，形成某种语义的数学模型框架。

文献中关于语义的研究，很少从拟全化的角度出发。即使有少数文献一开始会提及完备问题，但很快就会转进到形式域中。比如使用具体词语来代表语义等^{[4][53]}。

这种转进有时候是不知不觉的。因为后面会看到，拟全的语义由于无法直接表达，所以直接对其研究，是有些“无从下手”的。这样，研究者就必然（自然而然）要找到一些形式化的途径进行实际的操作。

这种形式化的方法在后面的相关章节中也会提到。但是本文的处理方式会有所不同：这里尽量区分拟全域和形式域的问题，把这两者看成是不同而又有相互联系的空间，从而希望把一些一直混淆的问题理清楚。

由于本文研究层面和角度的选取，本文的论述和研究方法及其逻辑可能会让很多读者不习惯，（比如更多的可能是提出问题而不是解决问题，比如语言尽量平白少定义专用名词，比如不追求特别的严谨性而倾向易于理解性（原因会有解释），比如更多是框架而不是实用细节理论或技术等等）务请读者谅解和耐心，也欢迎批评和讨论。

1.4 语义与表达的鸿沟

意识与物质之间的鸿沟是哲学研究中一个重要的悬而未决的问题^{[34][54]}，比如亚里士多德的学习理论试图解释物体的区分：物质性和缺乏物质实体的柏拉图型相^[3, p. 57]。这里不打算从那样的高度出发，即本文不试图解决“语义是如何产生的”这样的问题，仅对语义和表达“总是不能讲到位”之间的鸿沟进行一些描述和分析。虽然隐隐觉得这两个鸿沟之间有一定的联系，但因为没有深入找到明确的证据，不在这里赘述，只在后面关于“属性”和“二重性”等问题的探讨时略略提及。

（1）不可名说

我们从几个中国古典的例子开始吧：

“名，可名，非常名。”几千年前，老子的《道德经》^{[55] [56]}中所说的这句话究竟指的是什么？如果用现代的语言，是不是可以这么解释：一个物体对应的概念，可以用某个符号（名称）去表示，但符号表示的概念，却并非是我们日常看到说到的那些具体的物体。

而“白马非马”^[57]，则是用具体的例子，讲了同样的事情：但这其中“马”究竟是什么？为什么感觉好象明明就在那里，却总是说不清楚？

同样的感觉，可以应用于对“愁”字的描述：描写“愁”的文字^{[58] [59]}，放在一起可能都已经比一江春水长了，但“愁”写清楚了吗？或者，你的“愁”，你说得清楚吗？再或者，你说的那个“愁”，是你切身体会的那个“愁”吗？

这种例子还可以举出很多，否则也不会有“只可意会，不可言传”^[60]这样的说法了。

为什么明明在意识中存在而且似乎很明确的某些概念，在表达的时候，却有一种不管用多少字（符号）也说不清楚的感觉？针对一个具体对象或物体的时候，我们也会碰到同样的困难。

我们意识中的概念，与我们对对象的表达之间，存在一条不可逾越的鸿沟？但我们一提到某个概念名称，明明就知道它所指的对象是什么，对表达的语义的理解却没有障碍。为什么会这样？这中间有什么环节被我们遗漏了吗？这个鸿沟究竟是什么？

（2）究竟是什么鸿沟？

为了更详细地分析这个鸿沟的问题，我们先把一些常见的术语：概念、属性、符号、表达、语义的关系，再简单重新整理一下（与之前的类似，这里很多是假设，并没有直接的证据），看能不能从中找到一些线索：

首先，需要解释一下，这里对很多术语的描述，并不是严格的定义，只能算是在有一定公共认识的基础上，从某个角度，对这个对象的一定程度（尽量严密）的描述。这种定义的不确定性，在很多资料里也有提及^{[61] [3, p. 42]}。至于允许这样的描述方法在文中存在的更深层面的理由，在下一节概念的数学定义之后会有进一步的简单解释。

意识域里，一个重要的基础是概念。意识活动，实际上都是围绕概念或概念所指对象，以及它们之间的状态、关系、作用、过程进行的。

概念是对一系列相类物体或对象的属性进行归纳得到的。即，概念的内涵直观上是由属性和属性之间的作用组合构成。（注：这里所使用的“属性”是广义的，具体将在下一节中进一步展开。）

关于概念和属性，这里必须要指出并提醒的是：属性本身也是概念。比如长度、颜色等，它们既是表达某些概念时所用到的属性，而它们自己也是一些概念。这种递归解释的情况在意识领域中经常碰到^{[3] [34]}，我们在这里暂时不对

它进行太过深入的探究，只是先认可这样的表述方式的存在及合法。没有这样的前提，很多理论和研究是推进不下去的。在后面的适当地方，我们会对它的成因、必然性和处理方法做一些讨论。这种递归性，在某种意义上，也是造成表达困难的一个重要原因，即概念的属性需要一层层递归地描述下去，似乎总也说不清楚。

前面提到，意识活动是围绕概念或概念所指对象，以及它们之间的状态、关系、作用、过程进行的。而概念或概念所指对象，以及它们之间的状态、关系、作用、过程组合起来的意义，就是语义。因此，对语义的理解和表达，与概念紧密相关。

表达是利用符号的组合，代表语义中的概念、属性及其状态、关系、作用、过程，最终达到承载语义进行交流沟通的目的。这里的基础是给概念指定人的感觉器官容易接受的符号（比如文字、语音、图片、音符等），用来指称概念（对符号进行赋义），方便交流和沟通。

在上面理出的关系的基础上，我们回过头来再分析一下前面的“鸿沟”问题，特别是“明确”与“说不清”之间的矛盾，可以发现这其实是两个方面的问题。即：语义中概念的指称对象（即概念本身，区别于其它概念的外延）明确，而概念的内涵（属性）说不清楚。

本文接下来就将从这个说不清楚的部分入手，探索概念和语义的数学模型。并给出有关“鸿沟”的一些分析，在此基础上，形成语义的拟全化和形式化双重结构的假设。

2 概念的数学模型

概念的应用由来已久，关于它的研究一直在进行^{[4] [43] [62]}，但是，概念本身是什么，似乎也一直没有合适的解释或模型。这里先用一个简单的表述：概念是可以用符号来表示的某个有一定语义的对象，用于方便表达，提高表达效率。定义一个概念的目标主要有两方面：一方面是规定概念的主要（内涵）性质；另一方面是明确与其它概念的（外延）区别。

如果不这样做（即不用概念）会怎么表达呢，比如我们要提到“电磁波”的时候，就需要把一整套的相关电动力学的物理学^{[63] [64] [65] [66]}放到这三个字的位置，才能讲清楚（也许还未必清楚，因为还需要偏微分方程^[67]那样的数学基础），这样的表达显然无法有效地进行。

概念的来源是我们对很多碰到的类似个体、现象、经验和理论的总结归纳和推导。

概念在自然语言和文字中应用最广，这是因为自然语言文字在这方面有明显的优势：自然语言的文字符号可以通过这种定义表达非常复杂的概念，而音乐（或非语音的其它音频）、绘画、姿势手势、表情等其它表达方式不可能象文字一样可以层层（层次性）地表达或详细地写成一本书那么长的表达去说明一个概念。而且自然语言最先实现了在媒介上进行记录。如果不能记录，全凭人的记忆，是不可能把那么多概念口口相传地全部记在每个人的大脑里。那么，

媒介的发展会不会引起人们对符号和概念应用的根本性变化呢？比如以后有没有可能用视频来表达各种概念，然后把视频发展到比自然语言更常用的语义表达工具？虽然不知道结果，但这个苗头是有一些的，现在的短视频的应用比书本或文字并不弱了，以后会怎么发展，目前还无法作出准确的预期，但视频很可能以后成为语义承载的重要工具。当然，会不会发生根本性的变化还不清楚，目前书写使用的科学符号在知识传递方面还是很有优势，不过也不能否认以后也许会有更有效的视频符号之类？但文字也有其一些特别的地方，容易书写、容易翻阅、快速查找，长期的应用形成的习惯等等。

从这里我们可以看到，概念是把一些本来表达起来非常繁复的内涵经过长期实践归纳在一起。概念一旦形成，它就会有自己的结构。比如新的概念往往会用已有的概念来定义或层层说明。而且随着人类认识的发展，概念定义的层级越来越高，概念的内部结构会越来越复杂。

为了对概念进行数学的抽象，下面我们试着对概念的一般结构做一些探讨。

2.1 属性与映射关系

上一节提到，一个概念的形成，往往被认为是对归于一类的很多表达对象个体内涵的属性的归纳。比如原始的星的概念，就是对天上很多星的各种属性的归纳，归纳出的那些星的共同属性就构成了星的内涵。下面基于此，对概念的内涵属性展开一些详细的分析，希望找出概念的内涵“说不清楚”的原因。

概念的内涵可以用属性来解释，这在其它文献中也有提及^{[3][6][17]}。而属性是花样百变、变化多端、不断翻新的。

这里的属性，是一种泛化意义上的广义的属性，是描述概念或概念所指向的对象的各种（互不相关的）性质。比如结构、组成、类属、大小、颜色、质地、变化等等。

概念的属性可以由属性表达式来表示，比如： $c_1: x_1 = v_1, c_2: x_2 = v_2, \dots$ 。一个对象 O 一般总是有很多的属性，用符号 c 来表示某个属性。通常用对象的属性来描述对象，用来描述对象的多个属性用 c_i 来表示。于是，对象 O 就会有某个属性描述集： $D = \{c_1, c_2, \dots, c_i, \dots, c_n\}$ 。不同的人或者不同的场景下，对对象的描述并不相同，于是对某个对象的属性描述集也有很多个，记为 D_d 。

对对象而言，除了属性之外，它的多个属性之间的状态、关系、作用、过程这些映射关系也是对象的组成部分。比如气态的水和液态的水，我们可能都会用 H_2O 这个分子式来描述，但分子之间的间距以及分子振动的频率等又是各不相同的（当然，也可以把分子间距、振动频率这些解释为另外的属性，但这种解释可能会没完没了，所以，这里就把这些关系和作用看作是与某个或某些属性相关的操作），而且又存在相互关系，比如温度升高，分子平均间距就会增加等等。所以，对某个具体的对象的水，它有属性，还有关于属性的“操作”映射关系。

属性的映射关系可以表示为作用 $m_1: x_3 = f_1(x_2)$ ，以及过程： $m_2: x_4(t) = f_2(x_1, x_2, t)$ 等（这里时间也可以看作是一个属性）。一个对象 O 的属性之间映射关系可以有很多，用符号 m 来表示某个属性上的映射关系。对象属性的映射关系可以用 m_j 来表示。对象 O 上有属性的映射关系描述集： $M =$

$\{m_1, m_2, \dots, m_j, \dots, m_h\}$ 。不同的人或者不同的场景下，对某个对象的属性映射关系描述集也有很多个，记为 M_d 。

我们借用范畴学^[68]的数学表示，把属性和相关属性的操作映射关系整合在一起，就得到了关于一个对象 O 的，建立于它的属性描述集 D_d 和映射关系描述集 M_d 之上的类似范畴的描述 C_d 。（要特别注意数学范畴学的“范畴”与哲学里的“范畴”使用上的区别。哲学里的范畴有一种分类学的意味。本文使用的范畴基本上是指数学的范畴学中的用法。）这里用“类似”范畴的描述的说法，主要是因为属性之间的映射关系，很多情况下并不满足范畴中态射的要求。比如范畴学里的态射，基本上是把范畴里的一个对象转换为另一个对象的。它并不表达对象本身的性质，只表达对象之间的变换关系。这种关系，可以表达为“转换为”，并且可以级联下去，而且可以不同顺序组合、有自转换。它只是表达所使用的工具的很小一部分。即对象之间可以互相转化的那一小部分。通常是抽象对象（比如数学对象）。范畴里的对象常常是集合类的对象。

由于映射关系与范畴学里的态射的差异，这里只能“借用”范畴学中的部分术语，后面我们把这种“类似”范畴的表述称为“预范畴”。

定义 2- 1，放宽范畴定义中的态射要求，只要求其为集合元素之间一般的映射关系，这样类似范畴的对象，这里称之为“预范畴”。

如果仔细思考，可以发现，对一个对象的描述或表达，往往正是利用它的属性，以及其属性的状态、关系、作用、过程这些“操作”映射关系来进行的。即，对象可以用预范畴描述 C_d 来表达。当然，这些预范畴描述往往并不能完全全地要把一个对象表达得非常清晰。但可以一直通过增加属性、映射关系等更详细地表达下去，越来越精确。这将是形成概念的数学定义的一个思路。

(1) 意识属性与物质属性

前面提过，我们采用“从中间开始”这样的策略，这里所说的增加属性、映射关系，更详细地进行表达，有几种：一种是不断丰富，比如我们给一个物体增加电磁性质的描述；一种是不断细化，比如我们对瓷器的釉和瓷的部分分别增加描述；另一种是往“基础”属性分解，比如把物体的颜色解释为吸收光波的频率。

这里对属性的分解再进一步做些探讨。

一方面，物质性对象属性的逐层分解，理论上可以分解到物理的几个基本变量上去。但这样的分解会太过琐碎，难以进行，因此，我们可以采用那些基本变量的某些层面的组合或综合，作为基础属性。这可能就是“从中间开始”的要义。我们不排除科技的发展，解释和分解能力的提升，最终可以达到用几个最基本的物理属性解释整个物质世界，这对语义的结构和处理可能会带来一些深刻的变化。而我们现在的结构只是兼容这样的可能。

另一方面，属性分解的过程很快又会涉及到著名的“意识与物质的鸿沟”。这里我们略加解释，与之前一样，目的并不是解决这个问题，只是解释这里属性表达和分解所遇到的困难，并澄清一些关于属性的混淆。

有些属性，比如颜色，我们可不可以分解为电磁波的波长属性来解释？实际上，这里存在着一些关于对象性质和主体感知之间的混淆。

颜色属性是人的感知，是意识层面的。具体来说，是物体发出或反射的光线，通过对人眼的锥细胞刺激，产生了相应的电化学反应，最终在人的意识中形成了颜色的知觉。这其中视神经的电化学反应，如何形成颜色知觉的过程还未被发现。

而发射或反射电磁波的波长属性，是物体对象的性质，它通过上面分析的生理和心理过程，在人的意识中产生颜色属性。

仔细分析上面的两段，我们可以发现，物体“有”颜色的说法，甚至某个波长的电磁波（光线）“有”颜色的说法，从严格的语义来说都是错误的！物体只是有反射或发射某些特定波长的电磁波（光线）的能力属性，而特定波长的光线只是能刺激人的感官，让人产生颜色属性的知觉，并不是它们自己“有”颜色属性。

这样说的证据在神经生物学中可以得到更进一步的证明：神经生物学可以分辨出一些与颜色的知觉相关的神经电化学反应与刺激。也就是说，我们可以

对某些神经细胞施加电化学的刺激，从而让人产生颜色的知觉^[16, p. 148]。那么，如果说物体或光线“有”颜色这样的说法是正确的，那电化学的刺激也应该“有”颜色了！这显然是荒谬的。

换个角度来说，用电磁波的波长来解释颜色，是从另一方面去解释了对应的物理刺激物的存在和性质，根本上来说，并不是解释了感知。哪怕这样的解释能让色盲的人了解物体因为反射光线的不同会展现出红、绿这样的区别，甚至通过其它辅助手段可以让色盲的人分辨出红、绿来。但并不能帮助他在意识层面获得红、绿颜色的“体验”。

这样，就可以证明，颜色实际上与物体、光线、电化学刺激都不是直接对应的，即颜色属性不可以分解为电磁波的波长属性来代替。

那么意识中的颜色究竟是什么呢？目前的科学水平还回答不了这个问题，能有的都是假设，比如假设它是电化学反应在视神经上产生的一种特别的“场”（也许在弦理论的隐藏空间里）？怎么证明是不是呢？也许需要发现一种意识交流机制，让人的意识可以“直接”传递给另一个人？（比如在没有任何刺激物存在的时候，一个人能把自己的“红”、“绿”、“香”、“疼”的意识直接发给另一个人。这样的话，色盲的人是不是真的可以感知颜色了？）如果真的有这样的通道存在，也许可以在推进意识与物质的统一方面迈出一大步了。也许那个时候，才是填平意识与物质之间鸿沟的真正时点。这里，我们就不多讨论了。

这里我们要强调的是，需要谨慎，在语义层面，有时候表达（比如语言）会把我们不知不觉地带入了歧路！前面分析的颜色属性，是属于意识层面的，而吸收反射光线的频率这些属性才是属于物体层面的，但语言的表达经常会让我们把这些轻易地混淆在一起。所以，在后续的关于属性（我们会把一些属性选做某个形式语义结构的基石）的处理中，一定要小心区分“意识性属性”和“物质性属性”。

实际使用中，很多属性其实是混合的？比如“苗条”，它是某种物理属性和感觉属性结合的混合属性。这些属性的分析或分解更加麻烦，是不是也是很多语义分析所面临的困惑的主要来源之一？

关于属性的基本类别，是否需要进一步分解？比如情感属性要跟意识属性

分开吗？

其实分不分开并不是简单分辨的要求，主要是要看它们的性质和处理工具上有什么不一样才对。即，如果这些属性的性质和对这些属性的处理方法不一样，就应该把它们分开，否则，就没有必要分别。

所以，情感属性和意识属性有什么不一样吗？在处理上我们要用什么不同工具吗？目前看起来似乎并没有什么不一样，因此可以不用对它们做进一步的分解。

这样的问题，同样适用于抽象属性、感知属性等等。

(2) 关系属性

属性之间的映射关系是比较复杂的（后面发展到概念之间的映射关系也是一样）。比如在许多人组成的集体这样的概念中，关于广义的属性“人”之间的关系就很复杂。

由于语义表达中的映射关系情况过于复杂，一般数学的映射关系似乎并不能完全表示这里的映射关系。我们对此作略为深化一些的分析，并把这里的映射关系再做进一步的模型化。

前面提到，属性（和概念）之间的状态、关系、作用、过程都应该属于映射关系，那么，比如自然语言中涉及到形容词、动词的表达，都会涉及到映射关系。从中理出头绪初看起来并不容易。

首先我们注意到，作为状态、关系、作用和过程的映射关系，它们往往本身也是概念和属性。比如美、疼、友谊、共同、打、喜欢、蒸发、思考。这一点，对后续的分析会有很好的启发。

为了讨论的方便直观，我们从一个具体的例子入手。表达“A喜欢B”，显然在A和B两个目标（这两个目标，比如某个集体中的两个不同的人，当然可以定义为两个属性或概念）之间建立了关系，因此它包含了某个映射关系。

但“喜欢”这样的映射关系，直接做数学模型研究似乎是无从下手的。这里提出的一个模型是把这种映射关系进一步分解。这个模型基于之前提到的，表达中的某个具体的映射关系，它也是一个属性或概念。在此基础上，我们把映射关系分解为：“关系+属性+值”的组合。

用这个的模型，我们可以这样表述表达“A喜欢B”中的映射关系 m ：1，映射关系 m 在A和B之间建立了关于关系（其它映射关系可能是状态、作用、过程等）的关系 f ，当这个 $f = null$ 时，表示“A和B没有（这方面的）关系”；2，映射关系 m 的属性 $m_c = \text{“喜欢”}$ ，对于其它的映射关系，这个属性可以是“爱”、“鄙视”等等；3，映射关系 m 的属性 m_c 的取值 m_v 范围为 $(-\infty, +\infty)$ 的实数（也可以用函数变换为有限区域，比如 $(-1, +1)$ ），这个值在这里是作为“喜欢”程度的一个值，包含“A喜欢B”，“A不喜欢B”，“A对B没感觉”这样的一些模糊的描述。

于是，我们把一个映射关系用一个三元组来表示： $m \sim (f, m_c, m_v)$ 。而其中的 (m_c, m_v) 组合，其实指向一个属性 c （这个例子里是“喜欢”）的带有属性值的实例（，后面我们还可以看到，由于属性往往也是概念，所以它指向一个概念，这个概念表示的是一个属性（这个例子是映射关系属性）。这里我们又看到了无所不在的嵌套（递归）。

在映射关系的模型被这样分解后，由于映射关系的属性部分又重新回到了

属性，我们后面在研究映射关系时，就有可能把重点放在 f 的性质上，从而在更高一个抽象层面去研究映射关系。而映射关系的属性部分，依然按照属性的性质去研究。

我们把这种映射关系的属性类型，称为关系属性。

关系属性与前面的意识属性有时并没有非常明确的界限，大部分也属于意识属性或混合属性。它并不是完全与意识属性、物质属性并列的第三种属性。

2.2 概念的数学定义

为了归纳一个概念 C ，需要利用很多对象的个体，记为 O_1 、 O_2 、 $\cdots$ 、 O_n 、 $\cdots$ 。即，要对 O_n 这些个体进行考察。而对它们的考察，根据上面的介绍，又可以用它们的属性预范畴来表达。而一般来说，考察主体的不同，往往对 O_n 的理解和表达也不一样，我们把不同考察主体 S 对物体个体的认识用 O_{sn} 表示。把 O_{sn} 对应的属性预范畴记为 C_{sn} 。而 O_n 对应的属性预范畴记为 C_n 。

为了归纳一个概念，最好能把所有可能的个体都找出来，甚至现在还不存在的那些将来会出现的个体也都找来，但是这做不到！在现实世界里，这里的 n 的取值只能取有限个。同样的，对某个物体 O_n 的属性预范畴 C_{sn} ，在现实中的不同情景和不同主体 S ，只能取有限个属性来描述这个属性预范畴，而这样的描述可以有多个，选取不同的属性集来描述，这里把这些不同的描述记为 C_{snd} 。

这里的某些描述可能选取了非常多的属性，但不管有多少，它也只能是有限的。但如我们前面指出的，有限的属性及其映射关系的描述，是不能完完全全把一个物体描述到“精确无误”的程度的。也就是说，这些描述做不到全面。

但是，概念的应用，一方面要求对所有可能出现的个体，都必须适用，不管它是现在的、将来的。比如说“狗”这个概念，需要能应用于过去的所有狗、将来出现的同类的狗，甚至某些目前所不知道的意义上的“狗”。比如智能狗？

另一方面，要求概念所对应的属性，也必须能适应每个个体。而前面讲到，个体的属性也是描述不完的。这就又出现了很大的要求偏差和冲突。

这些冲突，本质上是概念的拟全化要求与实际现实中有限的样本资源的冲突。

为了达到概念的拟全化的要求，考虑人的意识活动实际的运作，这里提出“拟无限”、“拟存在”和“拟遍历”的概念。拟无限、拟存在和拟遍历是在上节提到的人的意识中基于记忆的虚拟能力的进一步拓展。有了这些工具，再引用范畴学的一些方法^[68]，就可以进行概念 C 的数学上的拟全化定义。下面我们给出这些定义。

定义 2- 2，人的意识活动中，会把只有有限个样本的对象，通过想象对记忆中的样本进行复制，自动扩充成无限个。人的意识活动的这种能力，称之为“拟无限”。

定义 2- 3，人的意识活动中，了解一个对象时，会主要基于对象已知的部分属性。而对未掌握的或不关注的属性，则无冲突地默认为可能的属性取值，并自动扩展到对象全部属性的范围。人的意识活动的这种能力，这里称之为“拟

存在”。

这里要解释一下的是：“拟无限”的应用并不是这里凭空强加给理论体系的，其实在很早以前我们就在理论中应用了，只是并没有深入地关注它。比如我们可以在几何的公理里就能找到它的应用痕迹：在几何公理中，直线的定义是“线段的无限延伸”，这并不是一个公理或定理，是被当作常识和直觉来运用的。所以，这里提出的“拟无限”，只是把这个常识里的人的（意识）能力机制显性地指明了。

利用拟无限工具，我们把物体的个体 O_n 的序列复制扩充到无限个。

利用拟存在工具，我们尝试得出一个主体对某个个体的属性预范畴的拟全表示。这里采用的方法是某个主体 S 对某个个体 O_n 所有属性预范畴的描述取（范畴学中的）协极限，即，找到关于 O_{sn} 的所有属性及其映射关系。

属性本身也是一个概念，按照索绪尔的“共时语言学”在同时轴线上展开，共时语言假设下属性的极限的存在性显然，本文就不进一步展开了。

更详细一些来说，一个主体对 O_n 可以有很多不同的表述，包含的属性和映射关系不一样，细致程度不一样，但相互之间应该没有冲突。这些描述即为前述的 C_{sn_d} ，注意这里的下标 d 与 O_n 的属性没有关系，只是标记不同的描述。而由前面的解释可以知道，在现实中， O_{sn} 的属性描述过程很可能是非常冗长的， C_{sn_d} 甚至无法用有限集写完它的属性。

但既然 O_n 这个个体存在，那么它的属性集及其可能的映射关系就是确定的，不管这些属性能否被全部列出，甚至不管它们是否被知晓。因此，关于这个个体的“完整”属性集应该是存在的，它是不断增长的属性描述集的拟全化结果，这个拟全化过程，正是协极限过程。而这个过程中的有些属性并不是真正被掌握的，所以，实际上这里取极限的过程是借用了意识的“拟存在”的工具，协极限取所有“拟存在”的属性描述。于是，主体关于某个个体的属性和映射关系的拟全描述预范畴可以记为：

$$C_{sn} = \lim_{\vec{d}} C_{sn_d} \quad (1)$$

定义 2- 4，人的意识活动中，对部分属性，在以“拟无限”方式复制时，以“拟存在”的模式自动遍历属性的全部取值范围。人的意识活动的这种能力，称之为“拟遍历”。

接下来利用拟遍历，在前面得到 O_{sn} 的拟全属性预范畴 C_{sn} 的基础上，再根据全部 O_n 的拟全属性集中相关属性取值的“拟遍历”并集，得到每个属性的取值范围。

先要做一下属性集序集的重新整理。前面提到“既然 O_n 这个个体存在，那么它的属性集及其可能的操作就是确定的”，而根据 O_{sn} 的描述和 C_{sn} 的定义，我们可以认为 C_{sn} 中的属性集及其可能映射关系也是确定的，可以用某个序集 I_{sn} 进行某种排序。令 $I = \bigcup_{s,n} I_{sn}$ ，某个 C_{sn} 中属性可以表示成 c_{sn_i} ， $i \in I$ ，而属性的取值表示为 v_{sn_i} 。

注意这里采用的序集与 s 和 n 无关。而前面的每个 C_{sn} 中，其属性可能并不一致，其序集可能包含全部序集 I 的一部分即 I_{sn} ， C_{sn} 中的属性只能用 $c_{sn_{i_{sn}}}$ ， $i_{sn} \in I_{sn}$ 来表示。但在按前述采用了拟存在和拟遍历工具后， I_{sn} 可入射到 I ，对于 $i \in I$ 而 $i \in I_{sn}$ ，可以认为相应的 c_{sn_i} 为空属性，这样 i_{sn} 和 i 就实现了对应，于是对于我们考察的对象而言， c_i 、 c_{s_i} 、 c_{sn_i} 、 $c_{sn_{i_{sn}}}$ 如果不是空属性，都是指向同一个属性，这样我们就不再区分 i_{sn} 和 i 。

于是我们可以用拟无限和拟遍历后的所有 n 对应属性取值 v_{sn_i} （可能是范围值）的协极限来定义 C_{sn} 中序集索引为 i 的属性 c_{s_i} 的取值范围：

$$R_{s_i} = \lim_{\rightarrow n} v_{sn_i} \quad (2)$$

这时候属性 c_{s_i} 及它们之间的映射关系有可能是无限的、有递归和冲突存在的。我们对它们进行进一步的处理，使用范畴学的极限工具对拟无限的个体取极限，去掉一些不同个体相互冲突的部分属性（相应索引号的属性取空属性）和映射关系，而递归则被极限吸收完成拟全化，于是，针对每个主体有：

$$C_s = \lim_{\leftarrow n} C_{sn} \quad (3)$$

这样，以属性 $\hat{c}_{s_i}$ （ c_{s_i} 在经过上述取极限操作之后的象，可以认为是拟全化的结果，有可能为空属性）、属性的取值范围 $\hat{R}_{s_i}$ （经过上述取协极限的 R_{s_i} 进一步取极限操作之后的象，可以认为是拟全化的结果）为基础组成的属性集，加上属性的映射关系，就可以得到一个拟全的预范畴，我们把这个预范畴称作某个主体认识到的物体对象的概念 C_s 的定义。这时需要考虑不同主体 S 认识到的 C_s 是不一样的，因为 $\hat{R}_{s_i}$ 甚至 $\hat{c}_{s_i}$ 在不同的主体 S 的认识中是有差异的。

取 C_s 中所有主体共同认可的部分，就可以形成概念 C 。这个过程正是范畴的极限工具可以做到的，即，我们把主体 S 也用拟无限工具无限化后，就有：

定义 2- 5，概念 C 的定义为：

$$C = \lim_{\leftarrow s} C_s \quad (4)$$

定义 2- 6，这种使用拟无限方法对概念或语义对象的属性进行全面处理的模型，称为“拟全”的。

这里简单讨论一下 C 的性质，它是一个拟全的预范畴，由属性 c_i ，以及属

性 c_i 对应的取值范围 R_i ，组成的集合，以及映射关系 m_j 组成的集合，共同组成的。借用之前属性节中的相应符号，概念预范畴 C 由属性描述集 $D = \{d_1 = (c_1, R_1), d_2 = (c_2, R_2), \dots, d_i = (c_i, R_i), \dots\}$ 和映射关系集 $M = \{m_1, m_2, \dots, m_j, \dots\}$ 组成的，即 $ob(C) = D, m_j \in C(d_k, d_l)$ 。它是不同主体对 O_n 那些（拟）物体个体相关的属性及属性的映射关系的共同认识。这些属性及映射关系一般情况下是不能完全被描述（列举）的，很多情况下是无限的。

属性之间经常是有联系的，特别有时候是组成动态系统的，这时候把它们分列开设定取值范围不能准确反映实际情况！参见后面量子逻辑相关部分讨论。有些属性应该是由多个属性组合而成的，不能分开考察。比如形状。如果一个物体必须保持一直旋转不能停下来，那么形状如何表达？如果运动过程中形状还发生规律的变化，这时候形状与运动又关联在一起不能分开了？这是表达方法的问题，还是基空间本身要做出变化？这种动态应该描述为某种映射关系，如果采用关系属性描述，就是基空间要有相适应的维度。

在一个概念形成过程中，很多时候，有些人可能对这个概念的部分属性的认知是缺失的或错误的（极端一些的，比如在电磁的物理性质被发现前，人们对大部分事物的电磁属性的认识都是缺失的，但这并不妨碍对事物概念的形成）。

“拟存在”的能力，不排斥对不掌握的样本属性的接受以及后续的修正能力，这样保证了即使有不掌握的属性，也能默认接受，并进行概念的渐进成形。即：如果对某个概念的某些属性或取值，某人对它不掌握，则可以先根据他掌握的部分来形成概念，如果有其他属性或取值出现时，与他自己所掌握的部分不冲突，则可以默认接受和借用，并进行概念的属性内涵的补充修正。比如A以前看到的和听说的狗都是黄的，那么在A的关于“狗”的概念中，可能默认（拟无限和拟存在产生的）狗的颜色属性就是（可以是）黄的（注意这里狗的颜色属性是意识属性，即是人看到的狗的颜色是黄的，参见前面“意识属性和物质属性”的提醒，后面类似的情况将不再提醒，请注意分辨），也可能对狗的颜色属性是开放的（未做明确规定）。后来A看到了黑狗，或者B告诉他有黑色的狗，那么在A关于“狗”的新的概念中，就会在新的样本基础上做拟无限和拟存在操作，颜色的属性部分就会用一些映射关系，定义为可以是有一定概率分布的“黄的或黑的”，或者继续保持开放。实际上，这时候相当于部分被“拟存在”的属性实际化了。当实际所了解的样本数比较大或者通过学习得知一些全面的信息时，这样的概念内涵就会比较完善，比如现在我们知道“狗”的颜色属性中没有蓝色或绿色的情况。（以后会不会出现？比如人工培育出绿色的狗？这些在这里仍然可以是开放的。）

例外的情况，可以定义为某种映射关系。比如三条腿的马，是不是马？我们可以看作它本来应该是四条腿的，只是因为某些意外的原因（比如基因缺陷、事故、生病、甚至艺术创作等等），映射成了三条腿。再比如，如果关于“狗”的概念中颜色属性的部分依然保持开放，那么，并不妨碍人们在艺术创作或玩具制造等情况下做出（颜色替换映射关系）蓝色的“狗”。这样处理例外部分，应该还是比较合理的，而且对这里概念的定义没什么影响。

上面的两段其实也说明，概念经常是不断发展、丰富和变化的。某个属性，

对某个概念的某个阶段而言可能是未被接纳的，后来有些情况的发生，可能会被概念以例外接受，再后来情况发生了更大的变化，该属性出现的概率大大增加甚至变成普遍现象，慢慢就会被接纳为概念的属性了。

(1) 抽象概念与完备性

关于抽象概念的问题，应该与普通概念区别开来。很多抽象概念是在普通概念基础上推演出来的，不是同一个概念，性质也是完全不一样的，只是借用了普通概念同样的符号或表达方式。

比如抽象的点、线段、直线。抽象的点实际上是把普通的点关于大小的属性取了极限 0（使用点的大小属性可以任意小的性质，达到抽象目的。要注意，严格来说点的大小实际上不是 0，而是无限小），而抽象的线段则是把线段的线宽取了极限 0（同前，实际应该是无限小），抽象的直线是把抽象线段的长度取极限 ∞ 。严格来说，平时使用概念时应该标明抽象与否。只是实际使用中，根据上下文，基本可以知道它是普通概念还是抽象概念，而都标明的话就太过麻烦，所以就不加标注了。这也是语言的使用经常容易误导我们的地方，在实践中遇到或应用时也要象应对意识属性和物质属性那样谨慎，注意分辨。

这些抽象概念往往是在“拟遍历”的过程中，把其中的几个属性取值范围做了不同于取并集的其他映射关系，以上的几个抽象概念就是用了（数学分析 [69] 的）极限映射关系，即

$$AR_i = AR_{s_i} = \{ \lim_n v_{sn_i} \} \quad (5)$$

注意这几个抽象概念的某属性取值范围集只有一个元素，且不同主体认识到的属性取值范围都一样 $AR_i = AR_{s_i}$ 。

对其它抽象概念，这个“拟遍历”过程还可以是其他映射关系，比如关于“0”“1”的离散化等等。

抽象概念与一般概念经常容易混淆，我们应该注意区分抽象概念、一般概念、抽象概念的表达、一般概念的个体，以及它们之间不易察觉的转换。

比如，我们以点为例：在黑板上画一个点，用于解释数学中的点的含义，在这个意义上这个点是个抽象概念；如果我们考察这个点是怎么画出来的，比如是粉笔还是电子笔画的颜块，这时候的这个点就转化为有大小、颜色的一般概念；而我们前面画出的一个点来代表抽象点的概念，实际上就是抽象点概念的一个表达（类似考察文字语义、文字符号与文字字形或材料等细节的关系、差异和转换）。

一般概念很多是不完备的。比如一般的点概念是有大小、形状、颜色等，比如我们定义点的大小是以人眼看到的一定范围立体角来定义，这样点的大小就有一定的范围，但显然不能太大，也不能是 0。具体来说，点的最小界限可以以人眼看不见为准，而最大界限可以设定为人眼看到的立体角大小为某个 Ω 左右，或者点的大小与背景物相比可以忽略。可以看到，这里的最大最小界限，往往并不是非常精确，我们甚至可以用模糊值（参见后续关于属性值的模糊性部分）来表示。由此可见，这里概念的属性值往往不是完备的（数学的完备性定义，即序列的极限不在可取值范围内），很多情况下是开集（比如线段的长度是 $(0, \infty)$ 但不包括 0 和 ∞ ）。

抽象概念很多是由这样的一般概念（或抽象概念）推演过来的，比如抽象点的概念就是把一般点的概念大小无限小化，忽略形状、颜色等属性抽象而来。或者把线段的长度设为 0 等等。这样，往往抽象概念的属性值会取到一般概念的某些属性值非完备开集取值范围的闭包。从而，抽象概念某些属性就成为一般概念的完备化结果。比如上面点的例子。当然也有抽象概念进一步推演出新的抽象概念。比如由线段定义点和直线等。

要注意有些抽象概念在有些属性上可能是完备化的，有些则不是，比如前面提到的线段，它的长度属性的取值是开区间，但是线宽则是完备化的无限小。这些在实际应用中需要小心处理，以免混淆。

有关概念的很多问题，在有些文献中有部分相似的讨论^[4, p. 32]，有一些跟这里有不少异曲同工之处，但是那里的讨论并不深入。

2.3 部分相关问题的解释

这里回过头去解释一下，在上一节中，关于允许对对象的描述可以有不严密性（甚至递归性）存在的理由：由概念的极限化定义可以看出，通常情况下（反例的部分将在形式化部分介绍），我们做不到用有限的属性描述的方法，给出一个拟全概念的完整描述（定义）！

所以，我们只能“从中间开始”：

推论 2- 1，如果从头就要求对所有概念（对象）的严格严密的定义，我们就会无从着手！

比如，对几何的几条基本公理进行分析的话，其实我们还是缺少很多关键的定义：比如点的定义、直的定义、角的定义……

在那里，我们用的定义的方法依然是描述，不是这里的拟全性定义方法。而我们知道，属性是很难（甚至几乎不可能）描述完整的。

那么，为什么长久以来，对概念的形式定义似乎一直是行得通的，虽然有时候有些定义被认为是“不精确”而不断被修正？

仔细分析一下的话，也很有意思：概念的定義的目的实际上主要有两方面：一方面是与其它概念的（外延）区别；另一方面是规定概念的主要（内涵）性质。对概念的定義行得通，最主要是跟第一个方面的目标有关，这又与后面介绍的稀疏性有关，即：由于语义预范畴的稀疏性，我们并不需要很多的属性限制，就能把一个概念跟其它概念明确地区分开来。而概念定义目标的第二个方面，就跟这里讲的问题很关联：概念的定義很难精确，因为在不同情况下，我们关注的概念的內涵性质的重点属性可能是不一样的。于是，当这些关注重点发生漂移时，我们只好对定义进行重新修正。

而概念的內涵属性往往是无穷的，所以，本质上来说，概念是不可能明确而严密的形式定义的。但这不妨碍我们对概念进行形式定义，由于语义预范畴中的子预范畴的稀疏性，这种形式定义的指称往往又是明确的，是我们能理解并与其它概念相区分的。

而形式定义后，人又可以通过“拟无限”、“拟存在”和“拟遍历”的方式自动把那些未定义的属性补充进去，形成一个拟全化概念。

上节中概念的定义可以用来解释为什么写东西总有一种“写不完”或“不可言说”的感觉。比如“愁”，按这里的模型，各人心中各种体现形式的“愁”，它们的属性是写不完的。但“愁”似乎又都能理解。也就是说，在意识中有“愁”这个拟全概念，所以它的指称是明确的；但没有它的拟全的简单（有限）表达，即，“愁”就是说不清楚的，只能用极限来表示。

概念的属性，在作为独立对象时，本身也是某个概念。比如长度，比如器官、比如部件。

而属性之间的关系、作用、过程，这些在广义上也还是概念。比如生长、比如等价、比如烤。

概念由属性及其上的状态和操作（关系、作用、过程）的映射关系组成，而属性又是概念，它进一步由它的属性构成……这似乎又形成了一个循环的递归，应该也是概念的描述通常不能有限的另一个体现。这个递归问题，我们前面就提过，当时没有解释，只是默认这种递归的表述是存在且合理的。

现在，有了拟全概念的极限化的定义，我们可以在这里做个简单的解释，以便在后面语义模型（预范畴）的构建部分进行相应处理。

比如普通概念的点，它可以是各种形状的，比如会有方形，而方形的说明中肯定会有线段的概念出现，然后线段的说明中可能又会出现点的概念……这样循环下去，点的概念就越来越说不清楚了。

而采用这里概念的定义，用拟存在和拟遍历的极限来表达时，点的概念其实对形状没什么要求。而在极限过程中，点成为点的概念的中间环节也没什么问题。这就有点类似广义相对论中，引力会引起时空弯曲，而时空弯曲就会改变引力……，最终总会有个明确的极限结果，由微分几何给出，所以这样的递归并不妨碍引力的定义^[70]，某种意义上来说，这种递归最后被一个拟全化的极限过程“吸收”了。或者再举个比较日常的例子：两面平行的镜子里的影像，A中有B，B中有A，……，最终它们也能建立起一个稳定极限状态在人眼中呈现，而我们其实并不会追究到整个影像的完完整整的细节，就已经能理解了。

概念与实物的对应，是一种映射关系。可以表达成“某实物具有指向该实物的某概念的属性”。比如在“白马非马”的例子中。“白马”是作为某个个体的实例出现的，而“马”是作为一个概念出现的。这两者定义的域就不一样，更别说其它方面了，所以“白马非马”。

在它们二者的关系中，作为实例的“白马”是被认为具有概念“马”的属性的，而且这些属性的取值范围符合概念定义中取值范围，即对马的概念 C_M 中的属性 c_{M_i} 而言，白马的属性值 $v_{白马} \in R_{M_i}$ 。否则，没有这样的一些关系，概念也就没有意义了。

从对拟全概念的定义中，可以注意到，由于人的知识面、理解能力、环境和历史的差异，关于属性集，是有一定的主观选择性的。（即，每个人的“拟无限”“拟存在”中所涉及的属性集是有差异的）。所以，同一个名称的概念（比如“车”，比如“美”），在每个人的意识中，很多情况下并不是一致的，甚至差异很大。即一般情况下， $C_s \neq C$ 。

沟通的过程中却并不妨碍用这些概念进行交流，为什么？

其实一方面，很多沟通的目的往往就是对某些概念的认知的探究和统一，人类知识和思想的发展，也是以此为基础的。另一方面，这些存在于不同主体的概念还是会有一些“共识性”。前面的概念定义其实也是认为，共识的概念是许多主体的一个共同认识。关于这些，我们将在后面有关章节进一步探讨。

关于语义和概念的关系，以及各自的定义、组成分析，两本《语义学》^{[3][4]}都提到了与本文相似的很多观念，只是比较分散、没有全面系统论述，偏重于某一方向。其实明显能感觉到它们之间有一些可以相互联接的部分。后面可以试试把它们统一在这里的同一个数学基础下。比如词库的方法^[3]，实际上应该被划入到可计算语义预范畴部分；动词的有些处理方式^[2]可以划入到概念的映射关系；概念的定义^[4]与这里用属性来定义也类似，只是没有用极限的理念，所以很多问题解释不好。

3 语义预范畴及其生长

“语义空间”（与本文的语义空间不同，因此加引号标注）已经有文献提及过^{[71][3, p. 36][72]}，在已有文献中，对“语义空间”的研究，主要是用语义向量来表示某个语义。这些研究认为语义是个点，这样对“语义空间”内部结构以及点之间的联系的研究就比较表面化，主要是通过点之间的相互关联来研究词汇之间的语义联系，对语义（本身）的内在（结构）研究关注较少。Valiant^{[73][3, p.}

^{34]}将概念定义为样本空间中的子集，再用高斯分布来把点定义扩展到子集，但也没进一步指出它和语义的内在关系和内部结构模型。

另外，这些模型没有把概念的成形过程纳入考虑，实际上并没有在子集上形成有机的“概念”，只是个散装的东西。模型所使用的数学工具也相对比较简单，能进行研究的深度也受到了很多限制。

再者，这些模型几乎都有个共性，就是在提出模型后，很快把问题都转到形式域中去了^{[3][2]}，放弃了在拟全域的进一步研究。

而在形式域中的研究，很快就都进展到了有限状态机的方向，主要用于与语法相关的处理了^[3, p. 85]。

这其中实际有一个隐藏的思路：有限状态机被当成某种“函数”，实际是用于处理“词素”之间的关系^[3, p. 51]。这个思路跟这里把概念模型化成属性和属性之间的映射关系有一定的异曲同工，即这里的“函数”是某种映射关系的体现。但在那些研究中，“函数”也被直接形式化了，失去了很多信息和处理能力。

这里的思路将以拟全化的概念为基础进行语义处理，提出一个在拟全域中的语义（预范畴）的数学（结构）模型。而有关形式化的方法将放在后面的章节，作为拟全化语义（预范畴）模型的某种实用化的逼近操作。

3.1 语义预范畴

由概念的定义，它是由属性和属性的取值范围、以及属性之间的状态、关系、作用、过程的映射关系组成的预范畴经过一定的拟全化过程来表示的。这里在前述概念定义的基础上，给出语义的数学模型——语义预范畴——的定义。

(1) 基空间

这里回顾一下属性的表示，属性可以由属性索引（指称）、属性取值范围和属性映射关系来表示。于是可以用属性来表示某个空间的一个或若干维度（比如长度是一维的，物理世界的物体占用空间是三维的）。再用前面用过的拟存在工具应用到“所有”概念上，得到“所有”概念的“全部”属性，并用它们作为空间维度扩展成一个空间 X 。那么，概念（暂不考虑映射关系）应该就是这个空间里的一个个子空间 X_c 。

定义 3- 1 语义预范畴的基空间 X ，为拟存在的所有概念的属性（作为维度）扩展的空间：

$$X = \prod_n \prod_i c_{ni}, c_{ni} \in d_{ni}, d_{ni} \in C_n \quad (6)$$

这样的基空间有很多维度的属性都是重复的或者是线性相关的，我们可以把这些重复和可以相互表达的维度用线性代数方法消去，最终留下的基空间 X 的属性保持线性独立。

基空间实际的属性维是很复杂的，很多时候不同层次的，甚至嵌套性的维都可能会出现。比如人的概念有意识的“意志”、“理想”、“信仰”，感性的“美”、“慈祥”等属性，这些属性必须在“人”作为一个完整的不可分割的对象时才有意义；但另一方面，人又可以从生理学的角度分解为由各个器官组成，甚至进一步分解为细胞、化学物、物理粒子等等，这些属性的分解又不能以人这个整体对象为基础。而人又是其上一层智慧动物的一个分类属性。因此，基空间的属性维必然是很复杂的，并非简单的线性关系，很多看起来可以直接分解的属性，其实因为整体有另外的属性存在，并不能由子属性线性组合而成，往往它们只能共同组成基空间的维。

这也是语义直接表示无法进行的困难因素。因此，拟全语义往往是无法直接应用的，这将在后面进一步分析，那里还会给出能实际具体应用的实现方式：形式化。

在基空间的基础上，可以把概念的属性子空间嵌进去。这样的带有概念属性子空间结构的组合体，我们定义为语义空间。

定义 3- 2 语义空间由基空间及其上的概念属性子空间结构共同构成。

为了简化标记，在不引起混淆的情况下，我们后面把语义空间也记为 X 。
 这样，一个概念的所有属性的取值范围在语义空间中构成了一个子空间 R ，这个子空间的大小和形状可以用语义空间中的属性表达式（或者借用语言等其它表达式）来表达。借用代数几何的理论^{[74] [75] [76] [77]}，把属性表达式看作类似多项式的方程式，那么子空间 R 是某些关于属性的表达式的联立方程的零解子空间：代数集^{[78] [79]}。

反过来，语义空间中的每个子空间是不是可以对应一个概念呢？这个应该不成立，因为概念的定义有一定的要求，概念未填满整个语义空间。所以，我们只能说语义空间的每个子空间都对应于某个语义，但不一定是某个概念，更不对应于某个现实。当子空间只有一个元素时，我们可以说语义空间的每个点对应于一个特定的语义。相应地，每个（组）属性表达式都表示语义空间中的某个语义子空间，所以它也对应于某个语义。

有关语义空间的其它一些初步性质的数学化分析，我们放在后面专门讨论。

（2）映射关系集

有了语义空间，以及把概念的属性看成是语义空间上的属性子空间后，我们就可以在语义空间中进一步研究概念之间的映射关系。要注意的是，前面提到过，属性也是概念，这里就包括了概念中的属性之间的映射关系。而根据前面关系属性部分的介绍，映射关系也是属性或概念，所以映射关系的属性部分也是组成语义空间的基空间的组成部分。这其中的嵌套性的联系需要认真体会和仔细分辨。

定义 3- 3 语义空间上，概念之间的映射关系集 M 为所有概念之间的映射关系的集合：

$$M_{kl} = \{m_{1_{kl}}, m_{2_{kl}}, \dots, m_{j_{kl}}, \dots\}, m_{j_{kl}}: C_k \leftrightarrow C_l$$

$$M = \{M_{11}, M_{12}, \dots, M_{kl}, \dots\} \quad (7)$$

由于映射关系的种类和数量也是“说不清楚”的，我们仍然借用“拟无限”、“拟存在”和“拟遍历”的方法进行扩展。

类似概念定义时对属性的处理，这里为了避免混乱，也将每一组概念之间的映射关系用统一序集进行编号，即，我们假定 $m_{j_{kl}}$ 和 $m_{j_{mn}}$ 是同一种映射关系。

比如 $m_{j_{kl}}$ 表示“K 喜欢 L”的映射关系，那么 $m_{j_{mn}}$ 就表示“M 喜欢 N”。如果这样的映射关系没有意义（比如说“石子喜欢碳”），我们就把这个映射关系视作空映射关系，按我们在关系属性中所给出的三元模型，即 $m_{j_{kl}} \cdot f = null$ ，注意它

跟“不喜欢”即 $m_{j_{kl}} \cdot m_v = 0$ 的区别。

语义预范畴的映射关系部分是很重要的。在整理好语义预范畴的基空间后，发现：1，映射关系的性质可能对空间的性质有很大的影响。比如映射关系能否

用环来表示，直接决定了能否用代数几何的工具来分析语义预范畴的问题。2，映射关系的多样性及其结构也是语义预范畴中非常重要的组成部分。如果对映射关系的理解很含糊，会把握不好语义预范畴的研究方向。

因此，在有关数学模型和深入分析的部分前，需要把映射关系进行一定的整理，包括在可计算语义预范畴中，需要构建一个关于映射关系的类似泛函^[80]^[81]的“空间”。

做这一步，应该需要把语义中常见的映射关系做一些分类和初步性质的分析。包括：1，一般作用的性质；2，逻辑是不是映射关系的一种？3，类似数学那种复杂的关系，是不是也都属于语义的映射关系范围？4，其他形式化的对象是不是也都是映射关系的范围？

如果2、3、4都是，那么语义的映射关系空间似乎就太复杂了，几乎涵括了全部数学对象，对这样的对象做统一化的处理看起来是非常困难的。那么能不能总结一些通用化的办法，先对一些简单的部分做处理？比如范畴学的函子工具等？

对于映射关系，我们这里采用在“关系属性”部分所介绍的三元模型。三元模型的应用使得映射关系的属性部分又重新回到了语义空间。我们后面在研究映射关系时，可以把重点放在三元模型中的关系 f 的性质上，而映射关系的属性部分，依然按照属性空间或语义空间的方式去研究。

这种把映射关系的属性部分，放在属性空间或语义空间的方式是否可行，我们将在后面作一些初步探讨。

从另一种意义上来说，很多属性是由其它属性转换而来的，而这种转换也可以看作是一种映射关系。因此映射关系表示成属性也是很自然的。如果把这些映射关系都抽象为对某些属性的值的群、环、域上定义的另一个或一组属性的群、环、域的变换结果，则可以把映射关系看作是预范畴上的函数。这样的处理方式其实和三元模型就很类似，有可能改善语义预范畴的性质。

映射关系的关系部分，也是非常复杂，这里也不再展开，也将放在后面的部分章节做一些初步探讨。

(3) 语义预范畴

把概念之间的映射关系（根据前面解释，包括属性之间的映射关系）放到语义空间 X 上纳入到考虑对象，就可以得到一个预范畴 X 。

定义 3- 4 语义预范畴 X ，由语义空间 X 上的概念集以及概念之间的映射关系集 M 构成：

$$\begin{aligned} C_n &\in ob(X) \\ M &= \{M_{11}, M_{12}, \dots, M_{kl}, \dots\} \\ M_{kl} &= \{m_{1kl}, m_{2kl}, \dots, m_{jkl}, \dots\}, m_{jkl} \in X(C_k, C_l) \quad (8) \end{aligned}$$

而语义，正是基于对概念，以及对概念的各种状态、关系、作用、过程的综合的含义体现。于是我们发现，跟语义相关的操作，基本都发生在上述的预范畴的范围内。

有了语义预范畴的模型，语义的一些现象，可以用预范畴中的对象之间的映射关系，以及映射关系的映射关系类似的工具进行研究。一定条件下的两个语义预范畴之间的映射的问题也可以利用范畴工具（比如函子）进行研究。

比如语义表达的过程中，对象以及映射关系的转换肯定是存在的，包括对象性质的变化，比如由爱转恨、由喜欢变成爱、作用改变了状态等等，但这些对象和映射关系之间的关系与变换，是不是有函子性（语义预范畴中的状态变化可以认为是一种与时间有关的动态预范畴？即预范畴本身在不断变化）、或者它们有什么其它属性，这些都需要进一步深入。

很多智能的应用也都是在语义预范畴的领域进行的。比如机器翻译，比如人工智能等。如果在一定的应用条件下，语义预范畴的性质比较好，应该也可以利用类似范畴等数学工具进行研究。

比如在翻译的应用中，不同语言（构成的语义预范畴）中对应于相同概念和映射关系的表达，是不是可以用函子性直接翻译？当然，实际的翻译往往没那么简单，会涉及语法的关系怎么处理、如果找不到完全相同的语义表达，怎么用近似的表达（概念+映射关系）完成翻译、遇到歧义（即概念的映射不唯一）的时候怎么处理等等很多具体问题需要深入研究。

而翻译实际上主要包括两个过程：一个过程是把一种表达方式（对应语义预范畴中的语义）形成的表达进行分析理解，把它所表达的语义映射到一个公共的语义预范畴中的一个语义子预范畴；另一个过程是从公共语义预范畴中把对应的语义子预范畴提取出来，用另一种表达方式生成（相应语义预范畴中的语义）表达。

从这个意义上看，广义来说，所有的交流过程，实际上都包含着翻译过程。

另外，根据后面有关语义预范畴性质的讨论，如果可以把语义预范畴用数学工具来研究，那么它的更多详细性质会得到发掘，相应的应用会更加广泛和深入。

3.2 语义对象在语义预范畴中的体现

这一节简单解释和举例一般的语义对象怎么在语义预范畴中体现？

比如某个概念的实体、两个实体之间的作用、实体状态的变化过程、一个抽象的描述、一个感觉/知觉/意识/情感的描述等等。

按前面关于语义预范畴的模型，这些语义对象应该都体现为某个语义子预范畴。是不是这样呢？这里将举一些例子来澄清这样的一些问题。这些例子也将能为后续形式化后的可计算语义预范畴的构建起指导作用。

要特别注意的是，这里举的例子是拟全语义对象的例子，因此不能用一般的句子来作例子。因为一般的句子是一个形式表达式，实际上表示不了语义对象的完整语义。

是的，不能用一般的句子来作例子！比如，如果我们以“那只杯子是蓝色的”为例子来确定某个具体杯子这样一个对象在语义预范畴中的体现，我们从这句表达里能拿到的关于对象“杯子”的语义的信息是很少的，根本完不成对一个对象的拟全语义建模。

而至于为什么这样一句话可以作为表达使用，而我们还能明白它所指向的是一个“杯子”对象不会产生信息缺失感、疑惑或混乱，我们将在后面特别是形式化部分再去解释（实际上这个句子只表达了对象的一部分局部化的属性）。

例一：我们仍然拿那只蓝色的杯子作为例子（注意这次不再是一个句子所表达的杯子，而是一个实体的杯子，要分清这之间的区别）。那只杯子的语义对象 $O_{\text{杯子}}$ 有很多的属性 c_i ，对于每个属性 c_i ，它有在属性空间中的取值 v_i ，比如它的

颜色 $c_{\text{颜色}}$ 是蓝色的 $v_{\text{蓝色}}$ ，它有高度、有直径、有材质、有盖子、当前装了水

$c_{\text{功能}} = v_{\text{装水}}$ 等等。这些属性取值在语义预范畴中限定了一些属性维的点（如颜

色值）以及另一些属性维的取值范围（如所占空间，它也可以转化为使用高度、直径、位置的值点来表示，或者冗余地使用这些共同表示）。但这些并不能有限地表示出来，比如颜色，进一步的详细描述可能需要对杯子表面的每个点的颜色要分别表示。而材质也可能是不均匀的，需要更详细的表达去精细化，甚至精细到每个原子、基本粒子……而且，这些材质的细节相互之间还有各种力的关联，否则它就不能成为杯子，而是一些碎片。这些相互关联又构成属性之间的映射关系（当然，我们也可以把这些力的关联的映射关系定义为某种属性，归回到属性维，这些变化的可能性和处理上的差异将在后续详细讨论）。杯子的语义对象则是所有这些属性和映射关系的综合体。

但不管需要多么复杂的、无限的表达，它总还是基于语义预范畴的属性维和映射关系，因此，它最终仍然应该表现为语义预范畴中的某个（也许很复杂的无限维的）子预范畴，这个语义子预范畴我们仍然用 $O_{\text{杯子}}$ 表示。

接下来我们考察，这个杯子实体的语义子预范畴 $O_{\text{杯子}}$ 与概念“杯子”的语义子预范畴 $C_{\text{杯子}}$ 是什么关系呢？由前面的定义过程可以知道，“杯子”的概念是由实体杯子归纳出来的一些属性、属性值范围及映射关系所组成的子预范畴。所以，实体杯子对象中关于“杯子”概念中所规定的那些属性，它的属性取值应该在“杯子”概念的属性取值范围内，比如 $v_{\text{装水}} \subset R_{\text{功能}}$ 。其映射关系也是类似的（如果用映射关系的三元组模型转化为属性可能更清晰些）。

简单总结一下的话，类似杯子这样的语义对象在语义预范畴中的体现，可以看作是语义对象在语义预范畴中占据了一个子预范畴，这个子预范畴在有些属性（称为广延性属性？）上会占据连续的一段值（带有一定“体积”），而在另一些属性（称为非广延性属性或参数属性？）上占据一个或多个点。如果这个语义对象属于某个概念的对象，那么它和这个概念的子预范畴有非空的交集，并且应该在概念所涉及的属性维的属性取值范围内都有交点。

上面举的杯子的例子是比较简单的，杯子是一个静态的物质性的对象，所以对它的语义对象在语义预范畴中的体现的分析比较容易。

那么，对一些复杂的语义对象，上述的总结还成立吗？接下来我们再考察两个比较复杂的例子。

例二：这个男人一直喜欢那个女人。

我们先用子预范畴+映射关系这样的模型来考察。这时可以把这个语义对象的实体看作包含两个实体对象 $O_{\text{男人}}$ 和 $O_{\text{女人}}$ ，以及一个映射关系 $m_{\text{喜欢}}$ ，这个映射关系还有个时间属性“一直”。其中的实体对象的语义对象同上一段的分析，比如包括分析它们分别与概念“男人”、“女人”、甚或“人”的关系等。这样，这个例子中的语义对象就可以视作由两个语义对象子预范畴的并和一个映射关系的组合构成的子预范畴。基本上可以使用上一个例子的总结的结论，只需要在以上基础上再组合上相应的映射关系。

如果把映射关系看作是前面所提供的关系属性三元组模型，我们再来看这个语义对象会有什么不一样？这时， $m_{\text{喜欢}}$ 被三元组 $(f, c_{\text{喜欢}}, v_{\text{喜欢}})$ 代替。这时它在语义预范畴或语义空间里，是个什么形态？这时语义空间中有属性维 $c_{\text{喜欢}}$ ，这个男人和那个女人的两个语义对象所呈现的语义子预范畴，通过这个属性维，按 $v_{\text{喜欢}}$ 这样的喜欢程度值联系（这个联系事实上可以把关系 f 消化了？）在一起，成为维度更高的复杂空间。而喜欢的时长属性取值“一直”也是一个属性维，它们共同在语义空间里形成一个新的语义子空间。这个空间的形状似乎很难考察，我们可以通过一些数学变换的技巧来试试。显然“这个男人”和“那个女人”是不完全线性相关的，如果我们通过某种正交化变换，把“男人”和“女人”分别转换成相互正交的属性维，而“ $c_{\text{喜欢}}$ ”则是另一个属性维，忽略所有其他属性维。这时，“这个男人喜欢那个女人”这个语义对象就会变成一个三维子空间的一个点，坐标分别是“这个男人”、“那个女人”、 $v_{\text{喜欢}}$ 。

要注意的是，这里的点是变换过的（类似于降维的投影？），实际的语义预范畴中，它可能是由无限维组成的非常复杂的形状。即，这里某个属性上的“点”，在另外的属性上体现的是“范围”、“体”。但变换前的形状，又不容易描述出来，所以我们在这里采用了变换的方法从另一个角度去表示它。这也是后续在形式化时介绍的局部化的一个重要视角。

当然，这种方式来将映射关系转换为属性维，看起来相当复杂。所以，是保留映射关系用预范畴来做为基础语义结构模型，还是把映射关系进一步属性化，用属性空间做为基础语义结构模型，是需要进一步深入探索和实践检验的。

前者更直观，容易理解；后者空间性质较好，比较容易范畴化，处理可能更方便。具体怎么决定，似乎是个很难的问题。也许二者都可以使用？分别应用于不同的情景？这里把这两种可能性都保留下来，以便进行进一步的研究探索。

可以看出，按以上三元组的模型，例一中总结的结论仍然可以成立。

例三：他享受着那些美好。

这个例子的语义对象中，显然包含实体对象“他”，几个具体的意识对象“美好”，映射关系“享受着”。注意这里的“美好”虽然似乎很抽象，但它并不是以概念的形式出现的。

这个例子似乎与上一个并没有太大的差异，只是对象的性质有些不一样，对象数量也出现了多个的情况。具体可以参照例二进行分析（比如对“享受”

用三元模型时,把跟“美好”相关的属性进行变换,转换为某属性维后,“那些美好”在该维坐标上对应多个坐标值,相应空间就对应多个点),这里不多详析。

例四:人们向往美好。

跟上个例子不同的是,这里的“美好”是以抽象概念的形式出现的 $O_{美好} = C_{美好}$,而“人们”也是统称词,不是具体对象,也是一个抽象概念 $O_{人们} = C_{人们}$ 。因此,这里的语义是表示了两个概念 $C_{人们}$ 、 $C_{美好}$ 之间的一种普遍映射关系 $m_{向往}$ 。除了具体实体对象或意识对象被换成概念对象外,其它也可以用之前的例子(例二)同样的方式进行分析。

由这里的几个例子看,例一中总结的结论:

命题 3-1 语义对象在语义预范畴中的体现,可以看作是语义对象在语义预范畴中占据了一个子预范畴。

基本都可以成立。后面的讨论我们就将以这个结论为基础。

3.3 语义预范畴的共识性和个性化

我们前面讲到,交流需要有一个“共识”的语义预范畴作为基础。如果从一开始参与交流各人的语义预范畴就完全不同,我们就无法交流。比如使用不同语言的人且各人都不会其他人的语言时,他们碰在一起,就无法用语言进行交流(但不排除用非语言的方式进行简单交流,这时候实际上还是有更基础一些的非语言表达的共识语义预范畴)。关于共识方面一直有相关的研究^{[82][83]},虽然研究的方向不太一样。

这个“共识”语义预范畴,可以是一种交流中普遍认同的语义预范畴(所谓的“常识”应该就是指向某些这样的语义预范畴或语义子空间),这在前面介绍概念的定义时就提到过。即一般情况下,对不同个体而言,概念的理解是不一致的^[84],即 $C_s \neq C$,相应的个体的语义预范畴也不同于共识语义预范畴 $X_s \neq X$,甚至个体拥有的语义预范畴也有不止一个。比如熟练运用多种语言的人就可以用不同语言建立起多个语义预范畴 X_{s_i} ;艺术家们还可以用绘画、音乐、诗歌、舞蹈等方式进行表达,也可以认为他们也拥有另外一些表达空间或语义预范畴 X_{s_j} 。当然,个体拥有的这些不同语义空间也会有很大部分重叠、同构或共识的部分,在一个人的意识中相容相通。

而共识语义预范畴 X ,跟它所涉及的人群范围有关。比如世界上所有的人、

一个国家的人、相同宗教信仰的人、一个小村子的人、一种专业的全部人、一个小团体的人等等，都会有不同范畴的共识语义预范畴。而极端的两个团体分别是：全部人类、每个人。我们可以认为这里的共识语义预范畴是这种人群范畴内的所有人的语义预范畴的交集部分。即：

$$X = \bigcap_s X_s$$

我们人类之间显然是可以相互交流沟通的，哪怕是差异再大（争执、对抗、甚至发生战争），这种沟通的共同基础（“虽然我不同意你的观点，但我们仍在用互相能理解的某些东西试图说服对方”）也仍然存在，也就是说，这里的 X 不会是空集。

推论 3- 1 不管多大范围的人类群体，肯定有非空的共识语义预范畴存在。

另外，我们应该接受一个事实：语义预范畴并不只有一个，而是有很多，而且个性化和共识性也是普遍的。如果每个人都有自己的隐私是别人不了解的，那么实际上每个人都有一个个性的语义预范畴；两个或多个人进行交流沟通时，他们语义预范畴中重叠的部分随时就可以组织起一个属于他们之间甚至适用于一次交流沟通的共识语义预范畴；而不同的群体有群体化的共识语义预范畴，比如不同区域、不同宗教、不同种族等等，都会有一些特定的语义子预范畴是其他群体所不掌握或不关注的；如果可能的话，也可以根据通用知识构建一个统一的通用知识库公共语义预范畴（也许也会被按国家或语言构建，还是会有一定差异），后面我们会对构建这样的空间做进一步讨论。

每个人的意识中都有一份关于世界的语义预范畴的复制品（世界图景），并可以有几种不同的解释和表达方法（比如不同语言）。但复制的预范畴中可能会有大量的空白和错误（或者说不一致，未达成共识的部分）。这种思想在其它文献中也出现过^[33, p. 19]。

一般情况下，由于每个人掌握的知识和信息不一样，对事物的理解和看法也不一样，这会造成对同一个概念，甚至同一个属性，每个个体的人的理解会不一样。比如使用不同语言的人，对一个概念使用的名称都不一样，理解差异可能更大些。更不用说，有很多专业领域的概念，不是这个专业的人员可能就是空白了。所以，个性化的语义预范畴里的内容（语义子空间的结构、形状等）通常是相差甚远的。但这些个性化的语义预范畴里，都会有一些相互重叠的部分，这些部分就是形成共识语义预范畴的基础。

我们交流的过程，往往是交流的几方在几方的共识语义预范畴的基础上，寻求对关注的交流目标语义子预范畴 X_{so} 统一化 $X_{so} \approx X_{ro}$ （或形成同构映射

$X_{so} \cong Y_{ro}$ ，如不同语言表达同一概念时），从而形成局部的共识语义子预范畴的过程。交流过程中的争论，实际上就是交流目标对象的语义子预范畴在交流参与者各自的个性化语义预范畴中差异太大导致的冲突。交流的过程，参与者通过改变他人的语义子预范畴或修正自己的语义子预范畴，从而逐渐使得对共同关注的事物的认识所相关的语义子预范畴尽量重叠化。如果交流沟通的对象

目标是概念，那就是各人的同名概念 C_{so} 统一化的过程。

有些沟通过程会形成教材、法规、协议、约定等，就是把那些统一化后的语义子预范畴（或者称为共识）固化下来的方法。我们前面还讲到过思考是人与自我的交流沟通（按这种思路，一个人的个性化语义预范畴有时候也有几份不同的拷贝），这种交流的结果也会对自己的语义子预范畴中有矛盾或缺失的部分进行补充修正，写论文就可以看作是这种交流过程的成果固化的过程。

从以上的分析可以看到，个性化的语义预范畴，实际上是不断发生变化的。人们在学习和交流的过程中，不断补充和修正自己对概念、对事物的看法，甚至产生新的概念，这些都会对自己的语义预范畴的结构和内容产生影响。

同样，这些个性化的语义预范畴产生的变化，也会反映在共识语义预范畴的变化中，虽然在群体比较大时这种变化可能相对要慢得多。因此，语义预范畴本质上都是在不断发展变化的。而我们现在所掌握的（个性化的或共识的）

语义预范畴中，有很大一部分是从进化和历史传承^[85]下来的。

3.4 公共语义预范畴及语义预范畴的生长

上一节讲的共识语义预范畴主要是指每个个体化语义预范畴的交集部分。而这里讲的公共语义预范畴则是另一个方向，它是一种“全部”意义上的“公共”。即把所有共识语义预范畴当作一个“大而全”的语义预范畴，这样，每个共识的语义预范畴都成为其中的一部分。而作为个人隐私，从未与他人交流过的，则不属于公共的部分。即，它是（一定范围内的）每个团体共识语义预范畴的并集：

$$X = \bigcup_p \bigcap_s X_{ps} \quad (9)$$

上一节我们提到，个性化的语义预范畴实际上是不断变化的，公共语义预范畴也由此相应不断变化。长期变化比如二十世纪初的那些科学发现中出现的新概念、新方法大大拓展了人们的公共语义预范畴（的维度），对人们的观念产生强烈的影响和改变；短期的公共语义预范畴的变化比如在某些宣传和舆论引导^[86]方面可以看到很明显的例子。本节我们在语义预范畴模型的基础上进一步探讨这种语义预范畴变化和生长的机制，重点关注生长机制，变化的机制与生长类似，这里就不多分析了。

在我们思维的直观中，语义预范畴中有价值的部分应该是伴随着知识的增长而不断生长和变化的。我们在不断学习和交流的过程中，掌握的概念不断增多，了解的各种映射关系（概念之间的状态、关系、作用、过程等）也不断深化。

这种直观与本文语义预范畴的模型很契合。我们可以认为公共语义预范畴有一个用“拟存在”的属性定义的“全”空间。而我们实际掌握和知晓的部分，只是其中的一小部分，即部分语义子预范畴，包括掌握的属性也是“全”空间所有属性的一小部分。随着人类对新知识的不断掌握，一方面在原有掌握的属性范围内，不断有新的概念语义子预范畴被发掘，或者原有的概念也不断有新的扩展和变化；另一方面，新的研究进展，会带来一些全新的属性维度，大大

延展所掌握的语义子预范畴的范围，对概念产生全新的补充和修正；再一方面，原有的概念之间的映射关系也会不断深化增加。

对个人而言，所掌握的语义预范畴，也会随着个人的知识和交流的积累，出现类似的语义预范畴生长变化过程。

我们实际使用的语义预范畴，由于存在很多的未知造成的不全面，因而不完整和不明确的。其中含有有意义的语义子预范畴的部分，甚至是稀疏的。这样的稀疏，给语义预范畴的增长留下了很大的余地。

一般而言，语义预范畴会伴随知识的丰富而生长。语义预范畴的生长包括两个方向：内涵生长和外延生长。

内涵生长的方式，主要是指概念的属性以及属性取值范围的增加或生长，从而使概念不断地丰富。

比如在量子力学之前，人们认为“原子”是万物的组成基础，于是“原子”的概念在那时候就是一个不具备什么属性的基本元素。或者说可以把“原子”当成是基本属性，或“奇点”。而随着量子力学的进展^{[87][88][89][90][91][92][93]}，我们发现原子其实是有内部结构的，于是它可以用很多新的属性来表达。这个过程有点类似奇点^[94]爆发（blow up）的过程，原来的一个点（0 维的属性）爆发成了很多维的子空间。这个进展，远远不止影响到原子本身，而是把所有物理世界的对象的属性维度都大大扩展了。实际上，原子属性维度的生长是反映在语义预范畴维度的生长上了。

再比如个人对水的概念的认识，一开始只是普遍的水，后来知道水随着温度变化可以有冰的形态和汽的形态，再后来随着对相关知识的学习，知道水分子是由氢原子和氧原子合成的，再后来知道还有重水等等。这些都是在一些已知的语义预范畴的属性范围内，对水的概念的自身的内容（内涵）的不断深入和丰富。

而外延生长的方式主要体现在概念的映射关系方面，即概念之间的关系、作用这些不断地深入和扩展的过程。（外延的另一方面是关于概念与其它概念的区别部分。这是不是主要通过映射关系中的功能部分来区分？属性的不同也可以用于概念的区别，但也可以认为这种不同是某种映射关系。）

比如同样以水的例子来说，开始可能只知道水可以喝，然后知道水可以滋润植物，后来知道水还可以浮起行船，再后来知道高压的水甚至可以做金属切割工具等等。于是对水的概念的功用（映射关系）的认识越来越丰富和深入。这样，语义预范畴中关于水的映射关系的部分就不断增长。

外延生长的另一方面是不断了解概念与其它概念的区别部分。比如还以水为例，一开始我们了解到水是流体，可能并不知道水和油的区别。后来知道同样是流体还油，可以通过颜色进行区别。再后来知道，与水一样是流体而且没有颜色的还有酒精，可以通过气味来区别。于是水的外延边界就越来越清晰。

在对世界的理解深入后，我们发现有了新的属性出现，就需要在语义预范畴 X 的属性维中增加新的属性维度 c_{l+1} 。然后对已有物体 O_n 和概念 C 增加相应的新的属性描述 c_{l+1} 。即概念的语义子空间也需要对新添的维度进行补充描述处理。对概念增加新的属性描述的过程有些可以用“拟存在”的方式自动进行，比如量子力学关于原子的属性的添加，可以自动添加到所有物理对象相关的语义子

空间中。而有些只能逐步展开，比如有些物体和概念与某些属性不相关，而另一些物体与属性的关联可能需要实验确定，比如导电的属性被发现后，有些物体导电，而有些物体不导电，所以只能在实验得到结论后逐步添加。

在以上的过程中，语义预范畴中还会有大量的缺失甚至错误存在。所以，对语义预范畴的维护也将会是一个很大工作量的任务，这些语义预范畴生长、变化和维护的任务通常是潜移默化地完成的。

4 语义预范畴性质初探

上一章我们定义了语义预范畴的数学模型，这样的模型定义的目标显然是希望能在语义处理中使用更多的数学工具，特别是现有的数学工具，以帮助扩展语义处理的应用范围和应用效率。

语义预范畴数学模型的性质究竟如何，是确定可以应用哪些数学工具，特别是现有数学工具的很重要的前提。如果语义预范畴的性质达不到现有数学工具的要求条件，那么就只能考虑：要么在某些情况下做些规范要求达到特定条件，以便应用相关数学工具；要么就需要扩展数学工具的研究以适应语义处理的要求。

基于以上的原因，我们在这里对语义预范畴的性质做一些初步的探讨，为语义的数学处理的以上两种可能情形提供一些粗浅的参考。这里的分析能涉及到的程度都不会也不可能很深入，主要目的是为有兴趣的专家学者提供一些进一步研究的思路 and 参考。

具体来说，这里将从语义预范畴模型中的一些可能应用某些已有数学工具的部分性质着手，寻找使用这些数学工具的条件，比如对语义预范畴模型进行一些适应化的处理和限定。如果有些性质距离使用相关数学工具比较远，则对把数学工具的应用条件放宽到语义预范畴模型的性质的可能性作一些讨论。

比如前面我们在介绍语义预范畴的基空间时，把表达和语义之间关系用代数几何中的代数表达式与几何形状的关系做了个比拟。这里我们将进一步探讨按这个思路继续走下去能走到多远？在什么条件下能在语义处理中使用相关代数几何工具进行研究。或者现有的代数几何理论能不能为语义预范畴的应用拓展一些更宽的条件？或者仅使用代数几何中的部分工具。

语义预范畴的性质与其它数学工具应用可能性之间的关联分析也采用类似的方法。这里将主要针对语义空间性质与代数/几何/拓扑数学工具之间的联系、映射关系的性质与范畴学工具之间的联系、语义预范畴性质与代数几何/代数拓扑之间的联系、语义子预范畴的属性值与模糊数学之间的联系、语义预范畴自有的逻辑体系与各种扩展的逻辑学（比如量子逻辑/甚至模糊逻辑等）之间的联系等做一些初步探讨。

最后根据这些对语义预范畴性质分析的部分结论，尝试对语义相关的长期积累的某些疑难问题做一些解释。

4.1 语义空间

在语义预范畴基空间定义时我们类比过语义空间和代数几何空间，这里再

从另一个角度来考察一下这个语义空间的类似代数几何构造：

前面提到属性的表达方式为： $c_1: x_1 = v_1, c_2: x_2 = v_2, \dots$ ，其实可以把它们看作是语义表达式： $(x_1 - v_1), (x_2 - v_2), \dots$ 的零解。这样可以把 v_1, v_2 看作是属性 c_1, c_2 的谱点。而这些表达式正好又是概念 C 的部分语义的表达，于是把语义表达用多个这样的表达式组合的多项式来表示，就形成多项式环 C ，而多项式环 C 的谱 $Spec C$ 就给出了概念属性的取值范围 R 。在此基础上进一步扩展到所有属性和可能的属性取值范围，就可以得到一个环化空间，从而构成以属性为基空间的环化语义空间，并进一步形成概形 X 。

这样概念 C 就是以语义子空间 R 为基础的子概形，它所对应的环上的层 O_R 即由关于概念的语义表达所构成。这只是一个用类似代数几何来研究语义的简单类比，模型定义和性质都并不严格，这里对这样的数学对象的研究也并不深入，只是提供一个研究探索方向的可能性。

而语义表达中概念所指的实体对象，对应于语义空间中概念子空间的子空间，甚至可以对应于一个点 $x \in R$ 。一段表达所代表的语义，则对应于语义空间的系列点或点集（子空间），以及它们之间的映射关系的总和。

但是表达式很多并不是看起来形如 $c_1: x_1 = v_1$ 这么简单的。比如有些属性的取值范围可能看起来不直观，特别是属性的值本身又是概念或属性的时候，比如生物的种类属性的取值就会有动物，而动物这个值本身又是一个概念。这些属性的值的性质似乎很复杂，其中，按之前的讨论，它们有一部分还可以进一步用概念的属性来表达。于是，最终总可以化为一些比如长度、位置、时间、颜色等等相对比较基本的属性的组合。

如果这样的属性分解过程成立的话，那么，组成基空间的属性的值应该可以用群、环、域或者它们组成的多维向量（实数、复数或多维实/复数的向量）来表示。这样，语义空间也可以用多维（的实/复）几何空间来表达。

复杂的属性的表达式，根据数学分析^{[67][69]}的理论，可以表示成多项式级数，有一些可微函数或其它函数也可以形成环。于是，根据前面讨论，在这样的情况下，语义空间中属性的表达式就可以模型化为环或模。

而在环化的空间上，我们可以定义预层、层，以及谱空间、概形、簇、同调这些数学对象和工具。然后再利用这些数学工具，开展对语义空间的性质和语义处理的分析研究。比如把翻译模型化为在两个代数几何结构之间的同态变换，把对一个事物的逐渐清晰的表达模型化为簇的相交等等。

理论上，所有我们在数学中遇到的数的对象，都是从实践中抽象出来的，所以，应该都对应于语义空间某些属性的取值。因此，象二进制数、半群、群、环、域、矢量、张量、纤维丛，这些代数对象，都应该是属性的取值考虑对象。

从这样的出发点考虑，在某些情况下，我们可能不能把语义空间用属性的取值及其上扩展的表达式全部环化。更不用说，属性的取值很多时候会表达得很模糊，比如说这朵花颜色是红的，味道是香的，其实是一种很宽泛的表达。

（这部分我们放在后面属性值的模糊性专门讨论）

对这些问题，还需要进一步深入分析。这里只能对语义空间能否类比为代数几何模型做一些方向性的探索，这方面更深入的性质需要相关专业的学者们进行了。

另外，很多表达式是关于语义空间中概念和属性之间映射关系的表达，那些表达式更复杂一些，它们能不能达到使用数学工具的某些条件也需要仔细分析。这部分实际上对语义的处理也非常关键，由于它的重要性，我们接下来将专门探讨。

4.2 映射关系与范畴化

上节关于语义空间的环化讨论，希望能找到语义预范畴展示某些比较好的性质的条件。而语义预范畴中映射关系的性质对语义预范畴的性质影响也非常关键。

映射关系比属性取值要复杂得多。

在由概念语义子空间（子范畴）和它们之间映射关系组成的语义预范畴中，一个比较好的关于语义预范畴的期望是探索它能否在一定条件或一定的模型转换下达到成为范畴的条件，即寻找把语义预范畴范畴化的可能性。

但范畴中除了对象集之外的映射关系要求是态射，而语义预范畴的映射关系非常复杂，似乎很难满足态射的要求。态射是一种变换映射，即从一个对象转换到另一个对象的映射关系，而语义预范畴中的映射关系，则是对语义预范畴中的一个、两个或多个语义子预范畴之间关联关系（状态、作用、关系、过程）的表示，并不一定是变换性质的。

这种关联关系与转换关系的差异应该是我们面临的第一个困难，感觉它可能会直接阻止范畴化的努力……

采取的思路仍然是在“关系属性”中提到的三元化模型 $m \sim (f, m_c, m_v)$ ，把关联关系转化为扩展维度后的子空间的某个维，而只保留某种抽象的“关系”用于表明对象之间有这样的关联性存在。

这样的转化能达到范畴化的目的吗？有两个问题需要确认：1，转换后的关系属性维和抽象“关系”的组合能表示对应的映射关系吗？2，保留的抽象“关系”满足态射的条件吗？

第一个问题在上一章“语义对象的体现”中得到了部分的肯定的解答。“部分的肯定”主要指两个方面：一方面是所选的例子并不保证全面；另一方面是那里的映射关系转换为关系属性维后的表达还是觉得有些生涩。但由于映射关系、关系属性与语义空间的关系的复杂性，它的分析可能需要很多的深入研究，在这里就不再多展开了。

第二个问题在上一章的分析中提到过，三元化后，似乎那个抽象的“关系”被吸收到语义空间里去了。

而且，这里的映射关系本来也并不是A的元素映射到B的元素这样的映射方式，更多是A和B之间的某种联系性质的。因此，在三元化后，主要是转换为增加关系属性维后的空间关系上了，并不体现在对象之间的相互转换或变换的性质。

如果要研究这样的抽象关系本身的性质，可能主要是单向还是双向关系的区别。比如“打”、“喜欢”是单向的，“友谊”是双向的。

简单总结一下，我们试图把映射关系按三元组模型增加关系属性维时，本来想把预范畴往范畴化方向转化的目标，结果转化成了增加属性维的语义空间。映射关系在这种模型下被吸收进了语义空间的基空间维度中去了。也就是说，这种模型可以把语义预范畴简化为语义空间。语义对象则从对象/概念子预范畴

与映射关系的组合的语义子预范畴简化为语义子空间。

虽然这样的简化似乎更简洁，但是，这种转化有一种不直观性，理解起来比较困难。另外，关系属性维的性质似乎与一般属性维还是有些区别，而且处理时采用的把涉及映射关系的对象变换为属性维的模式也更加不直观，这对它们的性质研究也带来了复杂性。再者，三元模型中的抽象“关系”在这种模型下最后被吸收到语义空间中，它的单/双向等性质又该如何表示？是另加一个属性维吗？这些问题都还有待进一步的深入探究。

总之，是不是采用这种映射关系的三元组模型，还是需要进一步探讨和研究的，这里只能把这种可能性先保留着，以期有兴趣的读者进一步深入研究。

这里关于映射关系性质和空间性质的讨论，以及范畴化的努力，对后续的关于翻译的应用部分有很大的影响。

在那里，各个不同的语义预范畴（或映射关系属性化后的语义子空间）当作是一个个集合，而把这些语义预范畴之间相关联的语义子预范畴或语义子空间之间的映射看作类似语义对象之间的变换函子。这样，翻译就类似在不同的语义预范畴（个性化的、普遍的、专业化的、不同语言的、不同表达方式的、不同区域的……）之上的函子变换。

这里的范畴化努力是希望后续的应用可以有更好性质的基础，从而找到一些简洁的应用工具。当然，另一个可能的努力方向是针对预范畴中的映射关系性质直接发展一套类似范畴学的预范畴学，不要求函子关系一定要建立在态射的基础上。

映射关系的另一种研究方向还是利用类似代数几何的方法，即把有些映射关系看成是定义在语义空间上的函数，如果这些函数能组成环，就可以在这些函数的基础上建立环化空间，从而应用代数几何的工具。但映射关系本身也很复杂，是否都能达到环化条件，也需要深入分析。

对映射关系的处理，这里并没有给出唯一的模型，这里提及的几种模型似乎都有一定的合理性，具体怎么处理更优化，需要进一步深入研究，也可能在不同的应用场景下采用不同的模型和方法。

4.3 语义预范畴的代数、拓扑、几何性质

由于语义预范畴往往是包罗万象的，它的属性、映射关系的性质很难一下子研究清楚。比如它的属性值是不是在某些条件下可以环化？映射关系能不能态射化（比如用三元组表达？），在什么条件下可以实现，从而可以把预范畴转化为范畴？这些条件是不是有普遍性？等等。因此对语义预范畴性质的研究比较困难。

这里我们借用代数几何的一些基本结论做一些直观的类比分析。因为前面讲的原因，语义预范畴是否可以用代数几何的模型还不能很明确，这里只是一些不是很严格的类比，比如这里并不要求属性的值集和表达式一定是环结构的，只是借用一些基本的直观。

语义预范畴的模型，其本质是任一个语义对象都可以看作语义预范畴中的某个语义子预范畴。如果我们把无限表达也认可为表达，语义可以通过无限表达式表达清晰，于是每个语义 S 对应的语义子预范畴 X_S 都有对应的表达式集 $E_S = \{f_1, f_2, \dots, f_i, \dots\}$ 。借用代数几何的类比，如果这些表达式是环的元素，那

么它们可以形成一个理想 I ，而表达式集就可以由 I 中的元素生成。但 I 的零解会在语义子空间 X 中形成代数集或簇 V ，这个代数集 V 应该正好就是语义子预范畴 X_s 对应的语义子空间。

根据代数几何的结论，当对一个对象的表达式集包含的内容越多，比如 $E_s \supset E_t$ ，则 $I_s \supset I_t$ ，对应描述出的语义子空间 $X_t \supset X_s$ ：

$$\sqrt{I_t} \subset \sqrt{I_s} \mapsto V(I_t) \supset V(I_s) \mapsto X_t \supset X_s \quad (10)$$

即描述越多，得到的语义子空间或语义子预范畴越精细、指向越精确。这与我们日常使用语言描述的直观是一致的。

多个对象时，相当于子空间的并，它的理想对应于原多个理想中元素的积。即对某个属性，我们需要同时用多个表达式表达对应于不同对象的取值。这也与我们日常的语言表达和描述的直观是一致的。

当我们完整描述一个现实中的具体对象 O 时，由于对象 O 的语义 S 对应的有些属性的取值是某个精确值或片段，而不再是一般的范围，描述集 E_s 中的描述表达式 f_i 就会变成 $c_i = v_i$ 这样的表达式（有些表达式可能并不这么简单，可以视作关于对象的不可约表达式，比如一个动物所占的体积空间是不能分割的），并且 i 遍历这些属性维，于是描述集在这些属性维子空间中的“理想” I_s 就是某个“极大理想” M_s ，语义子预范畴 X_s 的子空间 X_s 相关属性维将指向一个“闭点”

（或“不可约闭子空间”，这里都暂以“闭点”来称谓） P_s ，也就是说，现实中的一个具体对象实际上对应于语义预范畴中部分属性维的“闭点”。这个结论反过来并不成立，即语义预范畴中的“闭点”并不一定有现实中的具体对象存在，这也体现在语义预范畴的稀疏性上。但每个“闭点”显然都是有其表达的。这一点其实也很容易地解决了哲学上讨论的“金山是不存在的”这种描述中，“不存在的东西为什么会成为命题的主词”这样的问题：语义预范畴中大部分的“闭点”有表达，但并不存在。

当然，由于具体对象有大小等属性，具体对象的时空属性这些属性维只能对应于某个有“体积”的闭语义子预范畴，这个闭语义子预范畴的闭点，对应于物理存在。

这里还会经常碰到一个问题：很多情况下，语义对象在某种意义上只能作为整体处理，不能分割。比如我们在把一个人作为语义研究对象时，就会出现这样的情况：比如“他爱上了她”，这样的语义对象中，两个人作为语义子预范畴，显然在考察“爱”这样的关系时是不能做空间分解，也不能进行器官分解，包括一些意识性属性（如性格、信仰等）都不可再分。这与我们把一个人当作生物体进行生理和化学考察时完全不一样。

这种整体性要求是不是可以与代数的素性或几何拓扑集的不可约性对应？即，对表达式而言，在某种应用环境下（对应于某个指定的代数域或代数域扩张），人作为一个整体是不可分解的素表达式环，如果 g 、 h 是素的， f 是素理想，若 $gh \in f$ ，则 $g \in f$ 或 $h \in f$ 。对语义子预范畴或子空间而言，如果 U 、 V 是闭的，且 X 是不可约的，若 $X = U \cup V$ ，则 $X = U$ ，或 $X = V$ 。

更有意思的是，当基础的环或域本身发生改变时，不可约可能就变成可约了。比如 $x^2 + 1$ 在实数域不可约，在复数域则可约。

是不是神奇的：在某些情况下我们只能把某对象 O 当作不可分割的整体来

看，在另一些情况下则可以进一步做属性分解，以进行更详细的分析。

感觉这里应该有一个很深刻的哲学意义的存在：在什么情况下由部分组合成的一个东西成为不可分割的整体？它在什么意义下不可分割？这种组合与一般的乱堆相比有什么本质性的区别？它们在组合的过程中，产生了某种新的拟全化意义了吗？类似于二维投影空间的无限远的“线”改变了空间（原来仿射）的性质？

直觉似乎是有这样的事情发生了。比如通过各种元素、分子组合形成的细胞，开始具备自我修复、自我复制与新陈代谢的功能。人这么复杂的系统，故障率比一般的机器要低得多，基本上是自修复的体系，医生处理的只是很极端的情况。

这基本上都是在某种拟全意义上的完整性才能提供的能力：我是我的基础，我的意识也是我，我也是我的处理对象，我的一切还是由我负责，我的最高统帅还是我……这种嵌套似乎与宇宙或世界的构成有着千丝万缕的相似性，一种拟全性。

人的意识和思维的整体性在这方面更加彰显。感觉好几个哲学家所提及的最终“太极”、“道”、“真理”、“上帝”似乎都与此相关。实际上很多不同的哲学理论似乎是在不同角度讲述同一件事情。

另外，这里实际的表达式可能不能用环来表示，因此，这并不是严格的数学结论，只是个类比。究竟在什么程度上可以应用这种类比，是需要深入研究的。

按这个思路考察一下语义预范畴，会有个很有意思的结论：当我们没有收到任一条表达式时（理想为0），我们拥有整个语义预范畴。换句话说，我们在交流开始前，实际上是宽容任何可能性的。或者，能不能说：人其实是“带着”整个语义预范畴来到这个世界上的，我们学习生活的过程，只不过是把语义预范畴中的子预范畴（比如概念）一点点地“发掘”出来了。所以，学习并不是一个“获得”过程，而是个“激活”过程，那些知识原来早就在你的意识中，只是你不知道它在那里而已。

或者说，我们的意识中本来就（可以通过感知、归纳等）在语义预范畴中自然生成了一个语义对象（比如概念）的存在，只是那些个语义对象怎么表达，一开始并不知道。而生活中，通过“学习”（归纳），把某个符号与这个语义对象建立了联系（对应），于是激活了这个语义对象的应用。这个“激活”过程，是不是就是赋义过程？即把语义对象的意义赋予了某个符号。

上面关于语义预范畴的类代数几何的处理，只是一个很粗浅的尝试，虽然很不严密，但可以看出语义的处理有可能这样与语义预范畴或语义空间的代数

[26] [95] [96] [97] [98] [99]、几何 [100] [101] [102] [103]、拓扑 [104] [105] [94] [106]性质相关联。也许更深入一些的研究能提供给语义处理更多的启发和支撑。有一些研究者其实已经在语义和几何性质方面做了些探索^[72]，但跟本文的思路差异比较大。

比如在拓扑方面，语义子预范畴或语义子空间的形状与概念、对象的关系显然是密切相关的。那么对这些形状的研究，是否能对概念、对象的语义处理带来一些新的思路？

比如同一个概念的不同具体对象之间是否应该满足某种同伦变换的要求？而不同概念之间的差异或“距离”是否可以用某种同调来表示？

而从表达和语义的对应关系来看，有没有可能把表达代数化，然后用代数拓扑^{[107] [108] [109] [110] [111]}的工具来分析研究语义对象之间的关系？

当然，实际的模型和处理方式并没有那么明显和简单，肯定需要投入大量的精力才能把这些细节的性质理清楚，然后才能慢慢（结合形式化，最终）做到实用化。

4.4 属性值的模糊性

在处理语义预范畴中的语义对象时，我们经常碰到语义对象的属性值存在模糊性的现象（我们通常说的语义模糊，实际上就是语义对象的属性值的模糊）。这种模糊性不仅表现在属性取值的方面，还涉及到属性的取值范围、语义对象的属性边界等。下面针对几个例子来略加分析。

例一：（渐变）颜色判定。《语义学》^[3]中举的颜色渐变的例子实际上在很多地方都会出现^[112]。两个渐变颜色的边界往往有三个确定方式：一个是突然给出某个颜色，对它直接识别时采用的边界；另两个是分别从两个颜色中的一个往另一个颜色渐变时所采用的边界。这里的颜色值和边界都会出现模糊性。

例二：警戒水位。设一个警戒位后，实际报警往往是在超过警戒位一定数值才开始；而报警取消往往也是低于警戒位一定数值才取消。

从控制论来说，稳定的负反馈要起作用，总是需要有一点“矫枉过正”的部分，才能让系统保持有相应的反馈控制信号生效。如果系统真能在被扰动的情況下被精确地控制回到希望值，那么这时候负反馈系统的控制信号输入是0，输出也应该是0，这时候系统是处于自由状态的，所以它不可能产生校正，这就和系统被校正回希望值矛盾了。

因此，对动态系统而言，一个精确的分界点是不现实的，反而容易引起波动。

以上两个例子都引起属性值和边界的模糊性。而属性值的范畴的模糊性则更麻烦些。

例三：我们在概念的定义里经常会给出某些属性的取值范围，比如狗的颜色，我们一般会认为有黄的、黑的、白的、黑白相间的。但如果玩具或动画里造出一个蓝色的“狗”，我们是否还承认那是狗？如果承认了，那么狗的颜色定义是不是就太模糊了？

例四：看一个建筑里的床是“小”的，而一个人躺在上面的时候它是“大”的。这样，体积大小这样的属性取值范围是不是也太模糊？

这里我们简单尝试分析一下造成这种模糊性的原因。

我们知道，语义对象本身的属性是非常多的，按之前的概念的定义所讲，如果详细分析的话，它往往是无穷多的。而属性的值往往也是某个连续范围内的任意值，那么，哪怕是实数，也可能是类似 π 的超越数。而我们平时用来描述语义对象（比如某个狗）的时候，是不可能用无限的方法去描述的，只可能采用有限的属性和有限的值（离散值）去描述和分析。

这样一来，我们实际应用时，一方面很多考虑不到的属性会对我们关注的属性产生细节的影响，比如我们考察颜色时，会受周围的光线影响、会受之前

看过的东西影响等等，这样就会对颜色的判断产生扰动，从而产生模糊性；另一方面，我们只能用有限的值去描述属性值也会产生模糊性的问题，比如我们用“红橙黄绿青蓝紫”来描述颜色，而不可能把每种颜色值都起个名称记住，哪怕是8比特的三基色采样，也需要 2^{24} 个颜色名称（当然人眼的分辨率也达不到8比特的精度），这样在红、橙之间的颜色值就必然会模糊。

关于属性取值范围的模糊性其本质也是类似的，比如例三中的蓝色的“狗”，我们可能仍然接受它是狗，那是因为它有太多的其它属性与狗的概念是相符的，而那些属性我们又不能一一地无限地考察，因此造成了我们关注的颜色属性的取值范围发生了模糊。

例四仍然是类似的问题。它是不同语义对象属性的相对性引起的。比如在鲜红的环境里，看到的某个暗红会被认为是黑，而在黑的环境里看到的同样的暗红会被看作红等等。例四的问题是我们在简单情形下，只能用“大”、“小”这样的离散值来描述体积，而不会用一个实数去描述，于是因为上述的相对性，就造成了属性取值范围的模糊。

语义对象属性值、取值范围和边界的模糊性为我们进行语义处理带来了很多的困扰。要解决这些问题，需要多方面的努力。这里简单提及一些可能的努力方向，另一些将在后续的章节介绍，其原因会在这节的末尾做些交待。

一个方面可能还是从代数、拓扑的角度去努力。比如对模糊边界的处理。以二维为例，边界的模糊导致边界不再是一根0宽度的线，是一个有宽度的渐变带。这个带本身的同调就值得研究？它是个隔离带，同调不为0。从外到内的一根线段从0变到1，是个纤维丛。这样的边界会有什么性质？这是拓扑空间的同调，不是上面的环或其它代数结构？模糊值的部分，会形成什么拓扑？这些拓扑会形成同调、同伦等数学结构吧？

另一方面可以用分析的方法。比如对负反馈的“矫枉”模型是不是可以成为某种机制化的模型？给边界定一个明确值，但实际使用时，还要考虑到抵近边界的速度梯度，并在速度梯度方向给出一些与速度梯度大小有关联的阈值余量。这样是不是可以处理一大类的问题？这可以做为模糊数学的核心点之一来拓展模糊数学的应用吧？这种速度梯度可以是物理的，也可以是意识梯度（倾向性？舆论或势的引领？上下文暗示？）。

在鲜红的环境里，看到的某个暗红会被认为是黑，而在黑的环境里看到的同样的暗红会被看作红这样的问题，是不是就可以结合这里的变化的速度梯度来处理？

欧几里得自动机（EA）^[3, p. 203]就是以类似方式处理模糊性。

另外，边界是不是跟控制论也有一定关系？负反馈控制总是要留一点用于稳定控制的差异，一般不可能正好把系统放在最优点上。总是要有一点点用于负反馈控制参考的偏移量的。这点偏移量的范围是不是构成了另一种类型的模糊“最优”界？注意这里的界限跟上面讲的“矫枉”是相反的。但上面讲的是属于“接近”过程，而这里的可以理解为“拉回”过程。

对模糊性进行处理的另一个工具是模糊数学^[113]，但模糊数学目前发展得还远远落后于应用需要。感觉这跟它的发展并未基于一个有明确需求的应用基础有关，不知道与语义研究和处理相结合后，模糊数学是不是能找到一些明确的方向，加快研究步伐。

以上介绍的关于模糊性的处理的方法并不能全面解决模糊性问题，这是因为前面我们分析的造成模糊性的根本原因并没有解决：我们无法把所有涉及到造成模糊性的属性都无限地考察到位，也不可能把所有可能的属性值无穷地列举出来。

也就是说，本质上来讲，模糊性是我们表达语义时不能采用无限属性及无限属性值的问题引起的，也就是我们下一章将要提到的形式化的局部化和离散化引起的，所以，模糊性必然也跟形式化有千丝万缕的联系。因此，我们有必要在后面形式化的部分再进一步对它进行考察，并在形式领域寻找一些处理模糊性的方向。

模糊值处理最理想的状态，是可以把模糊值直接代入到属性值和分类边界中去。包括能直接反映概念和语义子范畴中，但这样的结果似乎需要一个能跟拟全性良好协调的新的数学工具才行（直观上，人在这方面似乎就是可以直接处理？似乎可以根据上下文条件，把模糊性释放到合适的“度”，大部分情况下能在模糊中得到一个清晰的逻辑结论）。

4.5 语义中的逻辑

关于逻辑的研究，应该是语义研究中被最先在哲学的范畴中进行专门研究的部分^{[36][114]}。关于二值逻辑方面的研究，目前应该还是比较成熟的，并出现了一个专门的数学分支《数理逻辑》^[25]。

但虽然关于逻辑的研究已经非常深入和成熟，然而一些困惑仍然时不时地会出现^[115]。很多情况下，似乎总有一些处于“真”与“假”之间的逻辑值存在，不停地干扰人们的思维。

下面我们举几个例子。

例一：无限小数的尾数判决。

如果定义对一个二进制小数尾数中“1”的个数做统计，计数值的奇偶作为某个逻辑值判断，规定奇计数值对应于“真”，偶计数值对应于“假”。然后对这个小数做真假的判断。这对于一般尾数为有限长度的小数是没有问题的。

但对一个无限小数，比如一个超越数，这就会成为一个无法判定的问题了。如果用精确的量化去考察，它也是随着量化精度的不断提高，而不停地变换着自己的真值，但最终值是“真”还是“假”？没有结果，虽然从我们的直觉来说，既然数是确定的，那么尾数也是确定的，这样总得有个奇偶吧，但这里确实没有。

如果进一步的话，还可以假设一个小数B的尾数取值的每一位正好与小数A相反，这象不象量子纠缠？非形式化时它们的“态”相关，在同一个系统形式化（量化精度确定，尾数就确定）后“不确定性”消失，一个数值确定，另一个也必然确定。这个类比似乎不妥。但感觉还是有那么点意思。

以上假设中，更麻烦的是“0.111111……”，这个尾数按上面的计数判断会导致前述的非真非假的结论外，它还有个拟全解“1.0”，按以上规则，应该是偶的，也就是“假”。这样，在传统的二值逻辑下，就会有很荒谬的结论：非真非假=假。

例二：旋转椭圆的长短。

考察一个相关联的属性组成动态系统时，要取其中一个的值或者由关联属性经过某种组合或变换的值，就会有不确定性。

比如我们测量一个在二维平面旋转的椭圆在其中一维坐标上投影的长度，并与某个设定的确定长度（比如长短轴的中值）相比。如果测量长度比设定长度长，就判断系统为“真”，如果测量长度比设定长度短，就判断系统为“假”。于是，这个系统就不能给出一个确定的“真”、“假”的判定。但如果旋转是匀速的或是随机的，那么其不同长度值出现的概率就是确定的，我们可以知道它取“真”值和“假”值的概率。或者说，如果我们知道这个体系的运动规律，我们可以获得有关结果真值的一个“态函数”。

如果有虚（隐）变量与实变量成为一个关联系统，也会是类似的情形。

例三：有理数尾数的假想试验。

我们尝试用某种类似例二的“态”的模式换一种方式来看例一的问题。如果我们假设某个数字是一种弥漫在某个“空间”的态，这个空间或这个态本身在我们的空间外高速运动，我们的空间只有这个空间的一个小小的观察（测量）窗口，每次打开窗口时，我们只能看到这个窗口对应的那个空间的“点”“之前”的那些二进制数字，而其后的部分只能忽略。我们可以假定窗口打开时，我们碰到的“点”在那个“空间”里的分布是均匀的。那么，我们观察到的结果就基本上是个随机的结果了。

如果这个数字是一个有理数，根据有理数的尾数的循环特性，我们可以得到一个例一中二值逻辑的随机结果的基本固定的概率。由前面的均匀性假设，观察次数越多，结果越往这个概率靠拢。

那么，我们可不可以把这个系统看成是一个“波函数”或者“态函数”？每次观察就是一次坍缩？它是真和假的两个“态”的叠加，观察的坍缩结果统计下来分别有一定的概率。

显然，上述的三例都会给我带来逻辑上的困难。三个系统都无法明确给定一个“真”或“假”的确定性判断。为了解决类似这些逻辑上的困难，已经有不少的相关理论出现，比如量子逻辑、投影逻辑、多值逻辑等等。但目前为止还未能形成满意的解决方案，在下一章我们会对这些略加探讨。

这里我们分析一下造成这种逻辑上困惑的原因。实际上，这个原因与上一节中讨论的造成属性值模糊性的原因是类似的。即：语义对象的属性是无限的，属性的值也是无限的，而我们能处理的只是其中极其有限的部分。被关注的有限属性由于受到其它无限属性的影响，表现出一定的不确定性；而对于属性的值来说，当“有限”被限制到只有2个值的数量时，离散化就成为逻辑，于是每个逻辑值实际上包含了无限的可能性。这应该就是造成逻辑困惑的根本原因。

既然属性值有无限多，为什么要离散化为两个值呢？这个看起来似乎很不合理的问题，其实有非常重要的实践基础：这是（人）作为一个整体进行行动选择的必然！即思维指导行动必然要求逻辑化。

比如我们走在路上，虽然方向选择本来是无限的，但处在路上的具体环境和需要完成的事情共同作用的影响下，只有“前进”和“返回”两种选择。再比如当我们被人问及对他的态度时，只能在“支持”和“不支持”之间选择，虽然我们会有很多理由（属性的影响）让我们不要做出这样简单的选择而可以给出很多的中间值。

因此，从某种意义上来说，我们在语义中遇到的逻辑困难，是我们处在—

个拟全化的语义环境中，却需要做一个形式化抉择所引发的困难。这和属性值的模糊性所遇到的困难是类似。困惑实际上是发生在跨越下一章所分析的拟全化与形式化的二重性的时候。

或者，用我们古老的“道”^[55]来阐释的话，这是一个“一生二”过程中的问题，或者是人（只能使用形式工具）认识事物之过程的必然。即人们对事物的认识从最初的混沌（道、太极、拟全）开始试图了解并作出判断，于是有了分析对象（“道生一”），分析的第一步是将对象的属性初步加以区别，从而形成有对立面的区分（“一生二”），这时就可以应用逻辑。但是在分为二类的过程中，慢慢发现二类不足以清楚地认识事物，还需要更多的（属性）区分，于是逼着“二生三”的出现，也就是这里讨论的逻辑困惑问题。上一节的模糊性问题，则是进一步被逼出“三生万物”的问题。而随着对“万物”了解越来越多、研究越来越精细，判断越来越准确，我们不断逼近拟全（完全确定、没有模糊？）。但终究到达不了那里，因为再精细的逼近与整体（拟全）的性质也不一样（这点到下一章的局部化之后会更清晰），一旦到达了最终的拟全，就又回到了混沌的太极，对立统一、“万宗归一”。

既然逻辑的困难与拟全化/形式化的二重性相关，有关这个问题在形式化部分应该也必然会涉及到，因此，有关这方面的部分话题，特别是关于这些逻辑上的困惑的处理方法，我们将留待下一章继续讨论。

这里要补充一点的是：像逻辑这种在哲学领域研究，产生了数学分支，又被哲学作为工具反过来用于研究哲学的路径，会不会继续发展下去？最终哲学和社会科学的研究会不会逐渐引入大量的数学工具？这也是本文所要探讨的重要问题，而且将在文末初步得出一个肯定的结论：包括语义研究等传统社会科学的领域，将会有逐渐利用各种数学工具进行更深入的研究的趋势。

4.6 相关问题的解释

有一个根本性的问题，可能需要在这里再次明确，即：由以上的语义结构模型，语义是存在的，它本身是明确的，人可以通过某些能力实现指称和传递，但是，拟全语义是不可完整表达的。这些结论，我们在讨论基空间属性维时就已经得出。语义处理的很多困难正是这个原因引起的，这里的相关解释，很多也是基于此的。

前面提到，交流是通过语义进行的。那么，上面定义的语义预范畴（由于语义空间、语义预范畴之间密切的联系，在概念表达上都有相互映射，因此为了简便书写，后面我们将以语义预范畴用来统称所涉及的语义空间、语义预范畴，除非有专门针对性讨论时再做区分），就是人们相互理解和进行交流的基础空间。只有当某个“共识语义预范畴”存在的时候，我们的交流才能得以进行。否则，每个人都用自己的一套概念和逻辑，是不可能进行交流沟通的。本文的核心内容也是对该语义预范畴的结构进行数学化的（分析）介绍和研究。

有了上面的语义预范畴定义，我们再反过来探究一下前面提到的一些问题的进一步解释。

1、“拟无限”、“拟存在”、“拟遍历”的形成因素的解释：

这里还是按上节思路借用代数几何的一些直观，先不管语义预范畴是否可以环化或范畴化。在代数几何的空间里，表达式组成的环的理想（解的限制条

件) I 和簇 (零解空间) V 是对应的。根据这个性质, 有: 空的表达 (0 环) $I = 0$ 对应的解空间是整个空间 $V = X$ 。

这时收到的规则 (对象样本的属性表达式), 比如 $O_1: \{c_1: x_1 = v_{11}, c_2: x_2 = v_{21}\}$ 、 $O_2: \{c_1: x_1 = v_{12}, c_3: x_3 = v_{31}\}$, 就是全部规则, 可以应用于整个空间, 即所有其它未限制的属性 c_i 的取值, 都可以认为在不与已知属性冲突的状况下 “拟遍历” 遍布整个空间。反过来, 借用代数几何的角度看, 收到的属性表达式, 组成的理想, 即这里 $\{c_1, c_2\}, \{c_1, c_3\}$ 构成的理想 I , 它对应的簇 V , 对应的所有可能的属性表达对象 O_j 就是 “全部” “拟存在” 对象。也就是说, 我们在归纳开始的时候, 收到的很少的样本就会被当作全部 (无限) 样本的代表, 而可能的 “全部” 对象, 就可以将收到的样本通过 “拟无限” 复制来获得。这应该就是 “拟无限”、“拟存在”、“拟遍历” 过程的起因和发生逻辑。

思考:

所以, 人天生就是数学高手? 人不知不觉中使用的工具, 竟然已经是代数、几何、分析的综合应用工具了! “拟无限” 等的的能力, 是不是就是归纳的基础? 有了这个基础, 就可以实现智能化了? 或者, 这个能力的结果, “概念” 的形成是智能的标志?

或者, 能借用这种拟全域定义的处理能力, 解决形式域的问题, 就是 “智能” 的标志? 而机器不具备把形式域的问题抽象到拟全域来分析的能力, 它只能在纯形式域工作。涉及到拟全域的问题, 只能是人用拟全化方法帮它算好, 再做成形式化, 让它执行。正是因为如此, 所以机器不是智能的, 它只能是个自动机。

我们可以用这个来判断哪些生物有 “智能” 吗?

另一种思考:

是不是应该认为人的意识生来就已经有这个空表达 (0 环) 所对应的全部 (整个) 空间 X 的存在? 我们获取知识和学习的过程, 不过是在这里面不断地 “激活” 那些作为 “概念” 的 “语义子空间” 和其上的 “映射关系”? 人并不是被 “教” 会的, “教” 的过程, 只是让人发现或唤醒了自已意识中本来就有的某些东西?

太阳被我们认识的过程, 是它被入射到我们给它留好的语义子空间中去了。

而机器的玩法却是相反的, 它生来只有空的语义空间, 只能是我们不断在里面 “填空”, 它才能工作? 这是形式空间与拟全空间的本质区别?

2、“说不清楚” 的问题的再解释:

前面 4.3 节中提到, 语义预范畴借用的代数几何结构中, 表达式环和簇 (代数集?) 的对应关系, 也适合于解释描述和语义之间的关系。对一个对象的描述越多, 即属性表达式越多, 通常限制越多 (“或”、“并” 的限制需另行处理), 相应的簇空间 (语义子空间) 的范围越小。即:

$$\sqrt{I} \subset \sqrt{J} \mapsto V(I) \supset V(J) \quad (11)$$

这正好就是: 解释 (表达式) 越多, 语义越明确。最明确的语义, 应该是对应于一个点。

而如果语义预范畴的属性是无穷维时, 语义预范畴的一个点 (语义完全明

确)需要的属性表达式必然是无穷的。或者说,这个语义完全明确的概念,是无法用有限的属性描述表达出来的。

即使语义预范畴的属性是有限维的,当我们对语义预范畴的全部属性及维数还没有完全掌握时,我们就做不到对语义预范畴的某个点实现精确描述。也就是说,这种情况下语义完全明确的概念也是无法完全描述的。

以上,应该就是“说不清楚”的由来。

3、“不可言说”的问题的解释:

由以上“说不清楚”的解释,进一步发展到某些语义的语言表达中,表达者对语义中相关概念(也许是并未成形的概念或属性)的属性(注意属性通常也是概念,也应该有相就在的语义子空间)本身也“说不清楚”时,就会陷入一种“不可言说”的境地。

而这种时候,如果相应的需要表达的语义目标能用其它(非纯语言,比如音乐或舞蹈表达情绪)的方式大致描述或表达,并在交流者之间形成某种语义共识预范畴,那就是“只可意会,不可言传”了。也就是说,“言传”所需要的(用语言表达的)属性或概念还没有被总结或归纳出来(也许以后会有),(所以用语言无法给出表达式,)但语义是存在的(相应的属性也是存在的,也许是“拟存在”),而且在不同人之间还能形成共识语义子空间,借用其它表达方式进行交流。

4、表达能力和“啰嗦”问题的解释:

对一个语义子空间,描述太多之后,更多的描述的边际效应会明显下降。在这里可以解释为:交流对象的语义子空间形状、大小应该是确定的。开始的一些描述“大刀阔斧”地把它的基本形状就“砍”出来了,后面更多的描述实际上就是用“小刀”在修修补补地“削”得更精确一些。因此,从子空间在语义预范畴的占位来看,开始的描述虽然少,但那几“大刀”的效率很高,基本上把大体的位置、形状确定了。而后面的“修修补补”做了半天,实际上对子空间的形状、体积而言变化不大,只是更加精确了。

从表达的实践来说,其实也是这样:一个主要道理,用一本长篇小说写出来和用一首短诗写出来可能效果差不多,但是二者的描述长度却相差很大。中国古代小说中有时插的几首诗基本就把通篇的内容和要义给摘要了。只是诗歌只能给出个大概轮廓,而小说可以给出很多的细节。各种不同的文学体裁的主要差异也许就在这里?要把语义子空间表述得一清二楚很困难,我们只能给出各种不同逼近程度的方案。

把这种思路延伸到绘画里,也许诗歌就可以比作印象派;长篇小说可以比作素描或工笔?最终都落实在表达目的和表达能力上。能做到达到目的的最简表达也是一种表达能力的体现。

比如《江雪》^[58]只需要短短 20 个字,而绘画则需要画山画水画雪画船画人还有垂钓细节,画起来好麻烦。相反,有时候小说里一个章节,往往用一个插图就基本说明白了。

长篇的文字有时候会同时给出很多语义子预范畴,并行进展。这点应该也要注意。表达式其实也是这样。多个表达式,有时可以表达多个不可约的部分,组成相互连通或不连通的语义子预范畴。

即针对同一个子预范畴继续描述可以达到精确化限制,范围应该越来越小(精确)。但长的表达,还可以另辟新的语义子预范畴,并对它进行描述。

所以，内容长，并不一定是“啰嗦”，要看具体是写了什么，是对之前的对象继续描述，还是新立话题或故事。

这一点，在前面相关的地方也应予以解释和注明。

5、“专名”与“通名”的关系解释：

“专名”与“通名”的关系^[9]，在这里可以解释得更加清晰了：概念（“通名”）是语义预范畴中的一个“体积”不为0的语义子预范畴，即它的属性是有一定的取值范围的；而个体（“专名”）是一个（部分组成属性构成的子空间的）“体积”为0的语义子预范畴，即它（的某个子空间）是一个点，它的（某些）属性取值是某个具体值，不再是一个范围。而个体从属于概念的关系，则在语义预范畴中表现为个体的（部分成为点的子空间）“点”在概念的语义子范畴中。概念的从属关系则表现为语义预范畴的包含关系。比如猴子属于动物，而那只黑猴子是（属于）猴子。

6、一些逻辑问题的解释

逻辑上的一些问题，可以化作空间的操作（或映射的操作？）。比如两个表达式对一个对象的描述，在属性空间里产生冲突，那么这两个表达式给出的命题应该就是错的。而如果两个表达式对一个对象的描述，在属性空间中属于两个互不相交的概念，那么这两个表达式给出的命题应该是无意义的。即它对应于语义空间的空集。只有对一个对象的两个描述在某个语义子空间有不冲突的交集存在时，它才可能是真的。

7、多义、歧义、联想、借喻、暗喻的模型

多义、歧义、联想、借喻、暗喻等关联，应该也都反映在引申的语义子预范畴有交集或拓扑连通，或者是某些属性之间有交集或连通。比如部分属性相通，谐音的读音相近，多义字词的关联等。

这样的模型可能对后续相关问题的处理带来一些提示。

5 语义的形式化

一个简单的问题：取极限这个操作，计算机能做吗？

总是能看到关于计算机已经算到 π 的小数点后多少位精确度的报道。会不会有种疑问：计算机算的 π ，不管精确到多少位，仍然是个逼近；而我们拿着符号 π 指称着那个圆周率，虽然具体值是多少并不非常清楚，但是 π 指称着的那个圆周率却是精确的。或者简单说人所掌握的 π 是精确的，而计算机在算的那个 π 的值，虽然很精确，但依然是不精确的！

可以把计算机每多算一位的值，组成一个柯西数列，那么它的极限就是 π 。但计算机到不了那个极限值，也不知道有那个值，人却知道有那个值，就是 π 。虽然具体用10进制来表示是多少我们并不知道，而且与计算机相比，这个具体值一般人比计算机掌握的精确度低很多。

是不是有点矛盾？这其实是拟全和形式之间的矛盾。

形式化要求所有对象有显性且有限的表达，而拟全化则要求对象（性质）的“明确”性。但这个明确又允许量上的模糊性（也许是受现在人的能力限制？）的存在。

由前面的定义，拟全的概念、语义预范畴（和语义子预范畴）的准确描述

都是需要无限的表达式才能实现的。因此，理论上来说，拟全语义预范畴和拟全语义子预范畴是不可精确完整表达的。

人因为有“拟无限”、“拟遍历”、“拟存在”的思维能力而可以捕捉到拟全概念的存在，因此，不受这个问题限制，可以保持思维的进行。因为模型的复杂性，人在利用语义进行思维和意识活动时也会受到很多困扰。我们在上一章节中提出的数学工具就是为了解决人所面临的拟全语义空间中碰到的这些问题。当然，具体能解决哪些问题，能解决到什么程度，还需要进一步的深入研究。

有部分这方面的研究（特别是与拟全化和形式化二重性有关的），将在后面提及。

对于机器而言，由于目前为止的计算模型都是有限状态机^{[47][116][117][118]}。如果我们无法对语义进行有限表达，有限状态机也就无从处理。这样，在计算领域实际应用前面提出的语义预范畴的拟全模型似乎就是个不可为的事情了。

那么，我们就只能退而求其次。能不能有什么办法，给出一个可以使用的、近似的语义预范畴或语义子预范畴的表示方式，能让计算工具（比如计算机）进行处理？

这里的答案是：能。我们可以通过语义的（某种可计算的）形式化，忽略一些“无限”的部分，建立一个有限“基”的可计算语义基空间，以及其上的有限映射关系基空间，以此为基础，不断生长，在一定近似程度上代表语义预范畴。并在此基础上对语义子预范畴做（任意）设定精度的逼近，从而让计算机可以在设定的精度下近似地解决语义问题。

当然，由于这是有限的逼近，这样的解决办法肯定有其一定的限制。比如有些被忽略的部分平时不太起作用，但有些特殊时候会起到很重要的作用，（类似泰勒展开的高阶项）。特别是这个可计算模型对语义的理解（形式化）与人的理解（拟全化）实际上是不一样的。有时候需要这种拟全化和形式化的二重性并行前进，才能处理一些特殊的问题。后面我们将在这个思路下提出语义的二重性模型。

下面，我们先讨论语义的形式化表达问题。

5.1 形式表达与语法

语义预范畴模型能帮助我们理解语义的结构，但不管用代数、拓扑还是几何的数学工具表示，这个模型都是抽象的，或者说它只是在人的意识中存在，不是一种实物性的存在。

语义的主要作用是用于交流，而要形成交流，我们就必须把语义用某种实物性的方式表达出来，以便被对方接收。比如我们平时用的语言、文字、表情、音乐、代码、公式，这些其实都是语义的表达。

而表达的实物性呈现的要求，使得表达只能借助一些实物性的、有限状态（因为人或机器对实物的感知能力是有分辨阈值的，所以必然有限）的工具，比如纸上的符号、振动的声波、机器的二进制电平，因此本质上都是形式化的。

而形式化表达中，一个根本的基础是符号化。即把某些语义，属性、概念或映射关系，赋义给某个可以实物化的符号，由这些符号代表所赋义的属性、概念或映射关系，并通过符号的组合，以表达更丰富的语义。

符号是用来表示难以用基础属性表达的对象的一个符号（概念）展开，

它的表达可能需要一本书甚至不止。人们在应用时不可能每次重复这样的表示，因此用符号来代替，以便于表达。所以，一般来讲，文字的表达效率是比较高的。

这类似于我们用一个函数名称来引用很长的程序代码。或者用一个数学概念表达一长串的定理、命题、推论等性质。

（音乐的表达似乎有些不一样？是用符号了吗？还是直接把某些音频组合与情感直接联接了？但简谱或五线谱都是符号，只是赋义的符号与实际表达的内涵之间似乎还隔了一层？符号的赋义对应于某个声音的振动，并不是喜欢、清寂、悲伤等直接的人的情感语义的属性，似乎从音符的组合到情感语义之间还有什么东西没有被找出来？或者是相反？声音本来与情感语义之间的关联就密切，音符表达的符号化并没有到很高的抽象层面？而文字是从图形进一步抽象出来的，所以能与语义有直接赋义关系？）

形式化表达中，另一个很重要的基础是语法^{[48][119]}。有了符号以后，我们可以用符号来表达语义，但人的记忆有限，能记住的符号不能很多。再加上，很多符号表达的语义之间有一定的关联，可以组合形成更丰富的语义。于是，通过一定规则进行有限符号的组合，就成了表达丰富语义的主要方式。而这里的“规则”，就是语法。

语法是语义（的语言表达）的一个生成规则，是一种限制歧义和无序的工具。对不同的语言系统，语法是可以有相当差异的。在某些情况下，它并不要求非常严格的。这点从诗歌就可以看出来。

另外，除了语言以外的其它表达方式（绘画、音乐、动作等），看起来没有严格的语法规定，但也都有相应的一些规则（也可以称为广义的语法存在）。

语法规则很多是从实际应用中归纳总结而来，而且，随着语言应用的不断丰富化，语法的要求也会有一定的变化。

语法有时候对语义的理解会起到很重要的作用，比如映射关系的属性是单向的时候，对象的顺序就很重要，（例如“他打你”和“你打他”是完全不同的语义。）而这些就必须依赖语法给出其中的分别。

语法工具在相关的文献中^{[7][48]}有非常系统的研究，本文不将它作为研究的重点。

按语义预范畴的模型，要实现语义的表达，表达所针对的对象也应该是跟语义预范畴相关的内容，即构成语义预范畴的属性、属性取值、概念以及映射关系等。

如果部分的属性、概念和映射关系可以由赋义的符号直接表达，那么符号（按语法规则）的组合就是表达复杂的属性、概念、映射关系以及属性取值、未概念化的语义的方法，或者叫表达式。

表达式由赋义符号组合时，不能随意组合，需要有一定的规则，这些规则在语言类的表达中就体现为语法（也可以泛化到其它的表达方式）。它其实也反映在一部分的映射关系中，或者说语法也是映射关系的表达方式的一部分。由于这个原因，本文对语法不单独进行探讨，而是放在映射关系中统一考虑。另外的一些映射关系的表达，则由一些赋义的符号（运算符）、以及一些不在表达式中显性表达的“运算规则”体现。这里我们又碰到了那种嵌套或递归式的问题，符号、运算规则、语法、映射关系相互纠缠……

形式表达系统就是由有限的这样的赋义的符号、语法、规则的结合,所形成的可以表达语义预范畴的属性、属性值、概念、映射关系,以及它们所组合的语义子预范畴的系统。形式表达系统的表达式可以是语言文字的句子,可以是数学公式,可以是程序代码行。表达式可以组合成更复杂的表达式,比如句子组合成段落/文章,代码行组合成完整的应用程序代码等。

各种形式表达系统的表现各异,表面上很难与语义子预范畴构成直接的关联。这里我们简单探讨一下这些形式表达系统的数学模型。

形式表达系统中,最基础的两类表达是属性取值的表达和映射关系的表达。用数学化表达的话,可以写成比如前面举过例的属性取值 $c_1: x_1 = v_1, c_2: x_2 = v_2, \dots$, 映射关系 $m_1: x_3 = f_1(x_2), m_2: x_4(t) = f_2(x_1, x_2, t)$ 等等。而它们实际的表达比如说: x_1 是2块磁铁, x_2 是一条10厘米长的导线, x_3 是把 x_2 卷成的一个方形的环, x_4 是 x_3 在 x_1 的中心旋转。而这段表达指向的语义子预范畴的是一个简单发电机的概念 C_e 的模型。当然,这段表达本身也是被大大简化了的,这里还有很多子概念、子属性、子映射关系可以补充或一层层地分解下去。

这里有几个问题:

1、如果把 x_1 、 x_2 等符号在前例的后半部分用文字直接表达,实际上就已经是两套形式表达系统了。而上面这段例子中后半部分的说明则类似于一个从数学表达系统往文字系统的翻译。比如后半段中,我们也可以认为“磁铁”是被赋义为某个概念的文字符号(由于前面提到的概念描述的困难,这里不再对它的细节追究下去)。

2、如果我们只是看到前半部分的符号和数学表达,实际上根本不明白它们表达了些什么,直到把符号的赋义明确后才能理清语义。(当然,如果是实际使用的表达, f_1 和 f_2 的细节也会有更详细的数学公式,这也是这里前半部分更不容易理解的一个因素。)

3、符号的表达更接近前面讲的语义预范畴的模型。所以语义预范畴模型其实也可以认为倾向于类似一套“形式语言”? (不建议多做这种类比,容易混淆)。用这些数学符号可以让我们把一般的形式表达(比如文字表达)与语义预范畴模型建立对应,从而帮助在语义预范畴层面理解一般的表达。

4、这里的表达已经是形式化的,表达的形式语义子预范畴跟实际发电机概念的完整语义子预范畴是有区别的。这个问题我们将在后面几节更详细地说明,这里只是预先提醒。具体来说,正如前面讲到的,这段表达是被大大简化的,也就是说表达是不全面的,省略了很多细节和一些没有重点关注的属性。比如磁铁的成份、实际工作中需要的电刷、导线的材料及外层绝缘要求……。但我们读到这段时,如果有这方面的基础知识,就会“自动”用我们自己掌握的知识去补齐(“拟遍历”、“拟存在”)这些缺漏。而如果还没有这方面的基础知识,也会默认:“哦,这就是发电机啊”,然后把不掌握的部分先以空缺的模式填进去(“拟存在”)。这是我们后面讲的形式表达的形式语义子预范畴内射至拟全的语义子预范畴的模式。

要补充一点的是:语法规则从某种意义上来说也是某种语义,它的形式化语义也是可以表达的。因此实际应用中,如果需要,也可以化作语义预范畴的属性来处理。

5.2 语义的形式化

关于语义的形式化，文献^{[3][4]}中也都提到过，它们采用的方法和思路与这里有比较大的差异，具体就不多分析了。

这里的形式化，基于前面的语义模型：语义预范畴和语义子空间是由概念（属性）、属性取值（范围）、概念（属性）上的映射关系构成的。由此模型出发，语义的形式化主要涉及两个方面：一方面是表达的有限化，即描述用的属性、以及属性的映射关系取有限集；另一方面是属性的离散化，即属性的取值离散化。

这种有限化和离散化，主要是围绕人认识到的概念和语义的“核心”进行的，即保留语义预范畴、概念（语义子预范畴）中被认为是重要和关键的部分，达到“够用”的精度。形式化过程是主观的、强制性的。当然，有时候也会受限于技术的发展（比如对电信号的模/数转换的精度要求）。

定义 5- 1，形式化过程是将拟全语义预范畴和拟全语义子预范畴中有限的部分属性提取作为一个形式语义预范畴和形式语义子预范畴的属性集，将有限的部分属性之间的映射关系提取作为形式语义预范畴和形式语义子预范畴的属性之间的映射关系集，并将属性的取值及取值范围（在一定的精度上）离散化。

我们把之前在拟全域的一些定义，对应于在某条件下的形式化对应的定义，再重新整理一遍：

属性：

$$C = \{c_i, i \in I\} \quad (12)$$

其中 I 是有限序集。

属性的取值范围：

$$R_i = \bigcup_{k \in K} \{v_{ik}\} \quad (13)$$

其中 K 有限序集。

形式概念范畴 C 由属性对象集 $D = \{d_i = (c_i, R_i), i \in I\}$ 和映射关系集 $M = \{m_j, j \in J\}$ 组成的，即 $ob(C) = D$ ， $m_j \in C(d_k, d_l)$ 。其中 J 是有限序集。

形式语义基空间：

$$X = \prod_n \prod_i c_{ni}, c_{ni} \in d_{ni}, d_{ni} \in C_n \quad (14)$$

其中 I 、 N 是有限序集。

形式概念之间的关系：

$$M_{kl} = \{m_{j_{kl}}, j \in J\}, m_{j_{kl}}: C_k \rightarrow C_l, k, l \in N \\ M = \{M_{kl}, k, l \in N\} \quad (15)$$

其中 J 、 N 是有限序集。

也就是说，形式化概念 C_f 是拟全概念 C_c 某种截断或近似，形式化语义预范畴 X_f 也是类似的拟全语义预范畴 X_c 的某种截断或近似，或者如果从分析学

的工具的视角看，类似用有限的正交展开项逼近一个函数（这是局部化，不包括离散化过程）。

要注意的是，这里的形式化往往不是理想的，因为前面讲到，属性本身也是概念，它的拟全化定义仍然可以是以无限作为基础的，所以，这里的“有限属性”有可能不是真正的有限。只有当所指的属性本身也是形式的或是有限的，这个形式化才能真正有限。但实际情况下，这很少能得到满足。

比较理想化的形式概念，往往是在一个完整的形式系统中。比如二十世纪初欧洲数学家们对数学的形式化所做的努力，就是期望向这样的理想化的尝试。但是，“完整的形式化系统”是很难做到的。甚至，因为哥德尔所提出的完备性要求^[120]得不到满足，我们事实上是不可能得到理想的形式化系统。它一定会在某些环节上引入某些拟全化概念（属性）作为基石。

另一个要注意的是，形式与否，并不是针对某段表达而言的，而是针对整个语义模型体系。或者说，是针对语义的处理环境的，我们可以称之为“域”，比如拟全域、形式域。典型的比如人的意识是拟全域的，有限状态机是形式域的，而数学是混合域的。

从表达本身而言，所有可以数字化的表达从某种意义上来说都是有限的、形式的。而我们实际接触到的表达，比如语言、文字、绘画、形状、音乐、图象、视频、表情、动作、代义物等，基本上都是可以数字化的，因此是形式的。

比如，我们考察用“ $1/2$ 、 $1/4$ 、 $1/8$ 、 $1/16$ 、……，以此类推”这样的表达，来描述一个极限过程这样的例子。

这个表达既然能在计算机的应用（如这里的文字处理工具）中写在这里，一定是可以数字化的、有限的表达。它并没有真正精确地把极限的全部过程完整地描述出来。

但是显然我们能通过这段文字表达理解那个极限过程。这是因为，人对这段表达的理解是放在自己的那个拟全语义预范畴体系中进行。

再看计算机，这段表达计算机能理解吗？我们显然不能认为这个有限状态机“理解了”这段表达所包含的语义。在它的可计算的语义预范畴里，“以此类推”并不能让它把无限给“推导”出来，因为它永远是有限的。

这里我们再回顾一下有关“鸿沟”的问题。

用上面的形式化过程来解释的话，前面所提到的意识（概念）和对象（应该是表达？不是对象）之间的“鸿沟”，其实就是拟全化概念与形式化解释之间的“鸿沟”。也就是说，拟全域中拟全化的概念，在形式域中用形式化的方法似乎永远也说（表达）不全面和清楚。

而实际上，拟全化的概念，用形式化方法只能不断地逼近，却做不到精确描述。因为在形式化系统范围内，没有“无限”的方法。而拟全化，却往往要求“无限”的介入。

换一种说法的话，我们说的鸿沟，某种意义上是与无限和有限有很大关系的。即：用“有限”的方法，不可能把“无限”的问题精确求解。比如，用数值化的方法，不可能找到三体问题的精确解。但这里不能得出无限和有限是否就是这个鸿沟存在的本质。

从另一个角度看这里的形式化：语义预范畴的形式化，其实是用一些有限符号来可计算地表达语义预范畴中的概念和映射关系。

但这里有个非常重要的问题：事实上，这些符号本身并不能代表什么。只

是在形式化过程中，通过人工“赋义”，把语义预范畴中的某些“意义”赋予这些符号，它们才能开始表达。

即，形式化过程实际上不仅是形成某个形式系统的过程，更是把某个形式系统与语义预范畴相联接的过程。而后者，在某种意义上来说更为根本。

其实，形式化并不仅仅只有这里提到的这种模式，能把一个符号系统与语义对接上的体系，在某种意义上都可以称为是形式化。形式系统可以把一些拟全语义与符号对应上。甚至有些与拟全语义关联太过复杂的形式体系有时不具有可计算性。不过这些不是我们这里考虑的重点。

5.3 形式化语义预范畴的模型

在形式化过程中，我们需要解决的一个横跨拟全域和形式域的数学问题，是语义预范畴的性质转换的问题。即，我们需要研究在拟全域的语义预范畴的性质，能不能形式域中继续保留？可不可以用类似的方法去研究两个域中的关联对象？

换一种说法的话，就是我们在拟全域建立的语义预范畴的模型，在形式域中还成立吗？或者还有多少性质（何种程度）是成立的？特别是重要的性质还可以延用吧？

如果答案是否定的，那么我们是不是要在形式域另外建立语义预范畴的模型？然后拟全域的模型和形式域的模型如何关联？如何相互转换？

这些，对后续的拟全域和形式域相互借鉴和相互配合的二重性处理思路至关重要。这些问题在这里也不能很简短地回答，我们在本章及后续两章逐渐把这些问题展开，并做出一些并非完美的分析和解答。

我们先直观地分析一下：如果所关心的问题的属性正好对应于拟全语义预范畴的基属性，那么，形式化就是产生直接的近似和逼近，有时候不需要很高精度。

如果所关心的问题的属性是由拟全语义预范畴的基属性通过变换得到的，那么，就会出现某些复杂细节的地方（比如奇点附近）在变换后放大的“魔鬼藏在细节里”的问题，可能很高的精度也难以满足要求。

换句话说，形式化逼近是有条件的。或者说，要让形式化模型延用拟全域的模型，成为拟全模型的足够精度的逼近，是有条件的。

这种条件主要是所要研究的属性的对应性，以及所要研究映射关系的类型也必须满足一定的条件。

对不少重要的问题，这个条件是很难满足的。这些将会造成形式化系统的能力限制或系统缺陷。

（1）简单情形

简单情形是指要解决的问题的属性正好是关注的（显性的）语义预范畴的基属性，或者是基属性经过简单变换形成的属性，它们之间的精度保持一种直接的关联性的情况。比如后面我们举的颜色分类的例子。

这种情况下，问题解决的精度与形式化精度保持同步，只要形式化过程中的离散化足够精确，问题的解决就可以逼近到足够的精度。

几何直观上来说，可以把语义子预范畴看作是一个语义预范畴中的几何体，

拟全域中它有一个精确的形状和占空。简单情形对应于由位置、形状和占位基本就能确定这个几何体的性质。在形式域中，这个几何体被用一个一定精度的“刀”削了出来，没有拟全域那么精确了，但是大体性质保持接近，甚至“非常”接近。

在这种简单情形下，可以把语义预范畴比作巴别塔图书馆？目前有效的概念或语义，只是相当于已有的那些书，大量的书位其实是空缺的，很多书位是没有意义的。已有的书非常稀疏。

由于有意义的东西往往有很多限制，语义预范畴中有明确意义的语义子空间其实非常“稀疏”。可以利用稀疏性，部分重叠基本上就能判断一些语义对应的概念，不需要很精确地重合。因为本身概念的分布就不是很密集。当然对于概念分布相对密集的部分，判断起来就更加困难些，需要的条件（属性及映射关系集）也会要多些。

语义的稀疏性带来的一个结果就是：给一个概念做定义，只需要部分属性的描述就够了。所以，很多概念的定义可以很简洁。而且很多情况下，描述一个概念的时候，可以有许多不同的描述方法，感觉每种都可以。其实就是，各种不同描述中属性构成的子空间里，由于稀疏性，已经没有其它概念了，所以不会混淆。

因此，简单情形下，如果上述的几何体能接近拟全域的形状，即，二者大部分是重叠的，基本上就不会产生语义的分歧了。

这种稀疏性对机器理解语义和产生表达也非常重要。

简单情形中，除了属性的部分，关于映射关系的部分也应该遵循类似的要求，即相关映射关系的精度与形式化精度保持同步。这些重要性将在下面部分的讨论中逐渐明确。

特别重要的是：在拟全语义预范畴被（用不同方法和近似，如个性化情形）形式化后，重叠的部分，是否还能有效地指向同一语义（概念）？形式化的缺失部分，会对语义的理解产生多大的影响？

这对交流沟通的实现应该是一个很重要的基础条件。每个人个性化的形式语义预范畴之间，都只有部分交叠，有些甚至缺乏一些关键的属性维度。在这种情况下，能否找到合适的语义重叠子空间（预范畴），保证交流的双方或多方之间，能有一个共同的交流基础和努力目标？下面我们先简单做些分析，并在后续的相关章节给出结论。

要完成机器理解语义和产生表达至少要解决以下数学问题：

- 1、在形式（预范畴）空间中，按属性和映射关系的表述（比如一个遮挡太阳的图片或其它示意、描述），找到的概念及其映射关系（这时应该是形式的），替换为拟全预范畴中的拟全概念，这一过程中的误差可控（类似于分析中，差异只是高阶小项？）；

- 2、对拟全概念，再进行形式化描述，在一定规则下，用形式属性及其映射关系（要注意的是，这里用到的属性及其映射关系，可能已经和上而所提到的一开始的输入完全不一样了。比如关于遮挡太阳的可实现方案描述），来逼近拟全概念及其映射关系。

换句话说，这里对拟全语义预范畴和形式语义预范畴的分析，需要能证明：形式化语义预范畴所指向的概念，有方法可以（在一定误差范围内）对应于拟全概念；而拟全概念（在一定误差范围内）可以有很多的形式化表达，可以根据要求和规则，挑出我们需要的形式化语义预范畴的表达。（前者和后者结合，

其实也是说明：一个概念不需要把所有属性都列出来，就应该能理解；而对一个概念的解释，也可以选取很多不同角度，都能做到基本无混淆的表述，虽然还会有不少不清晰的部分（比如，用手遮、撑个伞、做个棚子、建所房子，都能达到遮挡太阳的目的。当然这里是手段，并不简单是语义的理解）

证明了这点，再有（形式）语义预范畴的工具，就证明了在人工智能领域，我们有可能做到对输入表达的语义理解（由形式化找到拟全化的近似），以及在理解的基础上进行对话输出（由拟全化得到足够逼近的形式表达）。

有关这些方面的论证，我们会在继续做一些准备后，在后续的章节中进行。

（2）魔鬼藏在细节里

在上面的简单情形下，细节的部分占比“小”，所以在逼近的过程中是可以被忽略的。但很多情况并没有那么“简单”。

我们在用类代数几何解释语义子预范畴与表达之间的关系时，就提到过一种“对偶”关系的存在。即：越小的语义子预范畴可能需要越多的表达。有没有隐隐感觉到一丝的不安？

有没有考虑到，语义子预范畴里的“小”，并不是不重要？是不是有对偶的语义子预范畴存在，某些“小”的东西，在另一些空间里就是“大”了？

比如射影空间闭集与仿射空间的交集子空间是射影空间的开集、仿射空间的闭集；频域的一个点，对应于时空域的无限连续波形；等等。

这样，就对语义预范畴的形式化处理提出了很困难的一个要求：语义子预范畴的形式化，并不是“逼近”就能解决所有问题的！有些逼近的误差细节，看起来很细微，并不重要，似乎逼近结果已经与原空间很接近了。

但是，这些误差细节，在某些情况下，正是另一些语义子预范畴中非常重要的东西。

就象是泰勒展式，有些高阶项，在一些情况下可以忽略，但是在另一些情况下，可能就是最主要的作用项。比如量子场论或弦论中的高阶项，实际上是决定物质世界的基础的一部分，并不是可有可无的。

那么，语义预范畴里，有没有这样的对偶存在呢？直觉似乎是有。能找到实例吗？某些小细节影响到全局的事情是常有的。语义预范畴里也有吗？如果有的话，应该怎么解释和模型化呢？那种长篇大论中只有一两个细节成为一个重要事件的导火索的算吗？

更重要的是，如果语义预范畴确实是这样的话，形式化中的“逼近”实际上从本质上来讲是不可行的！只能是针对某些状况（局部）的逼近，不具有全局性。

这种局部化似乎在很多地方也都有类似的形式？比如流形的定义^[100]^[102]、概形的定义^[76]、环的局部化^[95]等等？也许，世界本来就是这样的？不管是物理的还是意识的。

另外，如果这是真的，这倒是更加容易解释那些“不确定”、“模糊性”、“说不清楚”的问题。因为漏掉任一个细节，都可能在其它地方变成一个严重的问题，这样，是不是就更加“说不清楚”了？！

从一定程度上来说，语义预范畴上的映射关系更容易被细节“带”走。在

一个语义子预范畴里的某些属性可能在某些映射关系下几乎不起作用，而在另一些映射关系下则变得非常重要。（举个正切函数把有限角度范围变换为实轴的例子？而正弦函数可以把实轴变换为 $-1\sim 1$ 的循环？无法用同样的量化得到相同的细节。用自然指数函数？或者举书法的例子？同一篇书写，一般是看字的内容，不关注字的形状，但书法却主要关注字的形状，是另一种表达）

这种情况下，语义预范畴不能被当作是一个静态的空间去看，只能是一种适应于应用（上下文）的动态映射关系空间或预范畴。这样的动态映射关系预范畴理解起来似乎更困难一些，好在经历过相对论的洗礼，这种不能绝对地、静止地看问题的能力，似乎还是能接受的。映射关系集本身被看成是动态的空间，这在类似泛函的数学中也是常见的。

从这样的角度看到的语义预范畴和语义子预范畴怎么把握呢？简直就是个万花筒。对它的性质的研究更加困难，但也必须总结出一些明确的性质，否则不就是不可知论了。应该还是要从语义预范畴的模型基础上再做提高模型。

也许，局部化的万花筒模型就是一个好的形式化模型？

针对某类问题，我们在拟全域找到好的处理角度，适应性地调整基属性的选择，然后应用这种局部化的离散化，把这类问题的解决精度与相应的基属性的形式化精度直接对应，从而相对容易地采用“简单情形”控制一类问题的解决。

而针对另外类型的问题，则采用另外的角度，选择另外的基属性，在另外的局部化范围达到相对容易和精确地用“简单情形”解决该类型的问题。

（类似这种方法其实已经在很多实际理论中应用了，比如信号处理的时域和频域的转换；数学问题的代数视角、几何视角、拓扑视角的综合处理等等。）

如果所有的形式化问题都能找到这样的“角度”、“局部化”“万花筒”，做到在该局部化方案下能用“简单情形”无限逼近拟全语义预范畴解，并能覆盖全部拟全范畴，那么，这个形式化模型就是一个（类似流形或概形或局部化环的）可行的语义形式化模型。这正好与语义预范畴想往概形模型靠拢的努力方向有些一致性？（至少有关联性。）

而且，这样的话，可以在局部使用上节简单情形下的数学性质模型，并在“万花筒”中利用该模型完成形式语义模型对拟全语义预范畴模型的覆盖，类似于拓扑的开覆盖模型的思路，或者类似流形或概形中用简单性质的局部化模型覆盖一个复杂的整体。（公式？）

我们把这里提到的局部化做一个定义：

定义 5- 2，用经过挑选的或者经过变换的属性集以及属性的映射关系集为基空间维和基映射维构建形式化语义预范畴基础的过程，称为语义预范畴的局部化。

有了这个局部化的定义，我们回过头来再改进一下前面做的形式化定义，可以把它分解为局部化和离散化两个过程。这里定义一下离散化，然后把之前的形式化定义过程重新描述一下（注意，这里是不是会产生对语义预范畴形式化定义的混乱？本文讲到这里，应该能对这个定义的“不严密”保持平常心了吧）。

定义 5- 3，经过语义预范畴的局部化后，对经过变换的属性值进行离散化以

构建形式化语义预范畴的过程，称为语义预范畴的离散化。

语义预范畴的形式化过程于是被重新描述为：

1、语义预范畴的局部化

局部化属性：

$$C = \{f_{Ci}(c_1, \dots, c_{i_\infty}, \dots), i_\infty \in I_\infty, i \in I\} \quad (16)$$

其中 I 是有限序集， I_∞ 是原拟全语义预范畴的属性序集。

局部化属性的取值范围：

$$R_{\infty i} = \bigcup_{k_\infty \in K_\infty} \{f_{Rik_\infty}(v_1, \dots, v_{i_\infty}, \dots)\} \quad (17)$$

其中 K_∞ 是原拟全语义预范畴及拟全语义子预范畴的属性取值序集。

局部化形式概念范畴 C 由局部化属性对象集 $D = \{d_i = (f_{Ci}, R_{\infty i}), i \in I\}$ 和局

部化映射关系集 $M = \{f_{Mj}(m_1, \dots, m_{j_\infty}, \dots), j_\infty \in J_\infty, j \in J\}$ 组成的，即 $ob(C) = D$ ，

$f_{Mj} \in C(d_k, d_l)$ 。其中 J 是有限序集， J_∞ 是原拟全语义预范畴的映射关系序集。

局部化形式语义基空间：

$$X = \prod_n \prod_i c_{ni}, c_{ni} \in d_{ni}, d_{ni} \in C_n \quad (18)$$

其中 I 、 N 是有限序集。

局部化形式概念之间的关系：

$$\begin{aligned} M_{kl} &= \{m_{jkl}, j \in J\}, m_{jkl}: C_k \rightarrow C_l, k, l \in N \\ M &= \{M_{kl}, k, l \in N\} \end{aligned} \quad (19)$$

其中 J 、 N 是有限序集。

2、语义预范畴的离散化

离散化主要是针对属性的取值范围：

$$R_i = \bigcup_{k \in K} \{S_k(R_{\infty i})\} \quad (20)$$

其中 S_k 是离散化函数， K 为有限序集。

注意这里对变换之前的原属性取值范畴不能做离散化，因为在变换过程中，可能会把属性取值的范围改得面目全非，原来平缓的部分可能变成奇点，原来的奇点可能会成为平滑的部分。因此如果事先进行离散化，就会造成细节的损失，从而无法达到“万花筒”的形式化目的。

用局部化和离散化的方法来表达上一章中的形式化关系，可以表达为：

$X_f = S(L_p(X_c))$ ，或者先完成局部化语义预范畴 $X_l = L_p(X_c)$ ，再在局部化语

义预范畴基础上进行离散化 $X_f = S(X_l)$ ，其中 L_p 表示在 p 条件下对原拟全语义

预范畴 X_c 做局部化，而 S 为做离散化，结果为形式语义预范畴 X_f 。

形式化过程中的离散化，其实可以合并在局部化属性取值范畴的步骤里，

$$R_i = \bigcup_{k \in K} \{f_{Rik}(v_1, \dots, v_{i_\infty}, \dots)\} \quad (21)$$

K 为有限序集。但这样合并后会有一问题不好解释：即这样的形式化很难解释拟全语义预范畴能不能由形式化语义预范畴覆盖的问题，而这个问题又影响到语义能不能作为整体被（我们使用的形式化手段）认知的问题。局部化可以完成的覆盖的可能性，被属性值的离散化打成了筛子。形式化语义预范畴最终是以“网”的模式实现对拟全语义预范畴的覆盖，而局部化语义预范畴才是有可能真正覆盖拟全语义预范畴的（是否实际成立还需要一些数学证明）。所以保留这个中间步骤，可能会对这个问题的解决提供便利。

如果上面的 f_{Ci} 、 f_{Mj} 只是在 I_∞ 、 J_∞ 的序集中挑选出其中的一个 c_i 、 m_j ，则这个局部化的形式化就是上一小节中定义的形式化的简单情形。

这样，我们就可以得出一个结论：

推论 5- 1 形式化是语义对象在局部化所选定的属性经离散化后组成的形式预范畴中的投影。

为了实现真正实用的处理，实际的应用过程中，在完成局部化转换后，还是会用简单情形去覆盖局部化后的语义预范畴，这类似于流形或概形的情形。

需要补充一点的是，“局部化”只是一种说法，并不一定是几何直观上取语义空间里的一小部分的意思，除此外，还更有种“换个角度看问题”的意思。所以，可能用万花筒更确切一些。而一个“好”的局部化选择合适的关注属性，

则是抓住“主要矛盾” [123, pp. 299-340 第一卷]。

比如一个二维的分形的图形，如果用一般的直角坐标系均匀量化，无论量化的精度能达到多少，得出的离散化的结果都不可能再保留分形的结构。而如果我们用分形结构本身作为它的一个属性，再加上分支点、分支长度这些属性，则可以保留分形的结构。但分形结构变换为某个属性，是要在拟全域中进行的转换。这也是一个局部化的例子。

（请别黑着脸看我，不是我想要这么复杂的……不过换个角度想想，有些事情是不是也很有意思的？也许你真的能发现魔法呢？

也许把这些问题看成是形式化产生的副作用更加合理些？就象量子场论，方程就是那个方程，它是拟全的，而你怎么做微扰，怎么把它对应到你观察到的“粒子”，那是逼近和测量这些形式化的问题了。

在拟全域里，一个简单的保角变换，就能把一个点变换成无穷大的面。更不用说还有 Blow up 等等这样的工具。

在形式域里看起来是魔法一样的东西，在拟全域可能就是小菜一碟。是不是太可怕了？

语义预范畴中有关情感、意念的部分，是不是就有些这种性质的？一个无望的眼神，可能就会带来政策的变革，甚至改变整个世界。）

（3） 意识的契约：高级智能的自留地

看起来，语义预范畴没有当初想的那么简单。它的结构可能真的很复杂。一个子空间或子预范畴的形状，不能被静态地绝对化就是一个很大的挑战。

关键还在于，关于语义预范畴的各种不同的性状，都存在于同一个预范畴

内。这该如何表示呢？哪些东西，才是语义预范畴中真正能体现语义核心性质的？

而形式化是个“万花筒”中的世界，这样的性质时刻提醒我们在考虑问题的时候，必须得多留几双眼睛：要素和非要素自己在不断地变换角色，你不关注的那个蚂蚁明天也许就会变成巨人。

能在今天把那个蚂蚁找到吗？

按以上的提法，本来在拟全域中不存在问题的事情，形式化后可能就会有信息损失，而且损失的情况会大相径庭。比如，在拟全域中，我们在选择基属性维度时，只需要看能否相互表达，并且是否有独立性，采用哪组基属性维度的灵活度可以很大。但是，形式化时，逼近一般来说会根据选择的基属性为基础来进行，这样，本来拟全语义预范畴中有些语义子预范畴可以用基属性的变换属性的细节表示时，在形式语义预范畴中，因为近似（比如离散化），这些变换属性的细节由于基属性的近似的精度受限，可能就在形式化语义预范畴中消失了。比如，我们取一个圆周率 π 小数点后 69 位的十进制数字作为某个事物的属性，而圆周率作为基属性带进形式空间时，系统采用了 64 位二进制精度，那么，前述这个事物的属性就失去意义和可操作性。

那么，是不是可以说，在基属性所直接相关的部分，语义空间的几何性质能大体保留，但是细节的部分，几何、拓扑、代数性质有可能发生根本性变化。

但是，这些细节发生的概率相对来说又很小。因此，从全面的角度来看，一般的形式化基本都够用？！再加上稀疏性，绝大多数情况下，这些都够了。这也是我们可以构建可计算语义公共预范畴，并在大多数情况下应用的基础。

有没有什么能应对这种细节变换为要点的机制？比如可变或可兼容的精度模式？（上节中的万花筒模型？）

我们写程序时 DEBUG 和研究工作中对忽略的细节进一步深入的过程，是不是都可以看作是这种细节与要点之间的变换？

如果是这样，形式化似乎天生就不适应于科学研究和创造性工作。

是不是可以反过来说：科研和创造性工作，常常是在很容易被忽视的细节里找到宝藏的？所以，往往很多新发现是在语义预范畴的某个细节的角落里的子空间里找到的？

这些不能一下子被形式化的藏宝（无限）之地，正是人的意识留给人的智能的一些自留地？某些涉及到深层的智能问题，只能逐个问题地寻找相应的“万花筒”。而寻找合适的“万花筒”，正是人类高级智能的一个独特能力？

5.4 拟全域和形式域数学工具的关系

本节主要介绍语义预范畴由拟全域转到可计算的形式域之后，可以使用的数学研究工具。首先介绍一个语义领域常见问题的数学化研究思路；然后再针对一般跨域涉及的数学工具进行一些介绍分析。

(1) 域切换问题

这里尝试从一个语义中常见的问题分析其可能的数学层面的机理，并提出一个针对语义问题的数学化研究方向。

我们所要分析的问题是：在科学技术和社会的发展历史上，有很多次涉及

到人的观念的革命性的颠覆，有些科学家甚至因此而自杀，但为什么对于大多数人，这种观念上的颠覆性改变，似乎并未引起人们太多的问题，生活还在照常地过去。

这个问题与我们在 4.3 节所讨论的问题相关，这个换个角度提问题，并从形式化的视角再做一些进一步的分析。在前面的章节中，我们使用过“类似代数几何”的提法。这里有些问题需要继续借用这种提法。

从语义的模型来说，上述的有些改变，也涉及到有些基础属性的变化，语义的基空间和整个语义预范畴应该也是有很大的变化，比如历史上关于物质构成的元素假说、原子假说、质子中子电子的发现、量子力学的解释等，都对人类的观念产生了很大的冲击。但这些似乎也并没有让人们产生巨大的混乱，引起无法继续交流或思考这样的问题。这其中有什么内在的线索吗？

我们这里借用一下数学的“不可约”的概念。在数论、代数、代数几何中，不可约性是跟素性相关的很重要的性质。本文在之前也已经提到过了，这里用它对上述问题做些更详细的分析。

数学上，不可约性与域的扩张有关。即同一个对象，在不同的扩张域，它的不可约性质会发生变化。

前面我们还借用过代数几何的“素谱”或“谱”的概念，用于模拟语义基空间的结构。这里继续这种借用一下这个概念。

当建立基空间所依赖的域的扩张发生变化（比如增加了新的属性引起）时，原来的某些表达的“素性”或“不可约性”就会发生变化，相应地，谱空间也可能发生巨大的变化。

比如：当原子被认为是不可分的（不可约）的时候，它是“素”的，不同原子组成极大理想。而后来原子被发现是由质子、中子、电子组成时，只有质子、中子、电子是“素”的，它们组成了极大理想。再后来量子力学的发展，发现中子、质子、电子也不是基础时，这一切又改变了。

如前所述，这样的改变，对语义基空间和语义预范畴的构成造成了很大的变化。（如果执于机械性结构，甚至可以认为语义被整体重构了。当然本文的假设并未基于这种机械的思路，所以不存在这样的问题。）

这样的分析，看起来似乎人没有办法在一个基本稳定的语义预范畴里分析和理解问题？这跟我们平时的实际经验似乎不一致！我们并没有在观念有了革命性的变化后马上就思维混乱不能思考。这是为什么？

结合本文之前的内容，可以发现，以上所描述的人的思维可能因为一些基础的变化而崩溃的事情没有发生的原因是：人有一个很特别的能力：局部化。

改用数学的语言来说明，就是在面对一定问题时，人的思维可以局部化为只采用某个限定的扩张域作为基础，而不是要把每个问题都放到“最基础”的域上去分析。

比如，我们考察一个人的性格时，不需要把他拆成器官，而是把这个人当作是一个不可约的整体；但我们用中医的方法考察一个人生病的原因时，就要换一个扩张域，要分解到器官去处理，但不需要把他拆成原子；而我们分析到人体内某个激素的化学性质和药物的作用时，就需要分解到分子、原子甚至量子层面去了。

这种局部化，我们可以暂时称为“域切换”。实际上是对被关注的属性进行处理的“域”也相应做出了限制（也许可以归到“语境”的变化？）。

所以，对一般人而言，我们的日常生活并不需要学懂量子力学，用之前熟

悉的语义预范畴的基础就够了，因此我们并不会感觉到量子力学改变了物质世界的基础以及语义预范畴的一些层面（局部化）的基础对我们平时生活、交流、思考有多大的影响。

当然，关于这个问题，这里基本上也还是假设。要认真分析，还需要对以上的“类似代数几何”的一些概念进一步明晰化，对人的思维具体在什么样的“域”中进行切换，以及如何实现域切换的等等这些问题进行深入的数学化研究。

这方面的研究应该还可以扩展到其它一些领域。比如，域切换和“涌现”现象是否有联系？举个简单的例子：没有引入 i 时，在实数域里实数的很多性质和问题不好处理，而引入了 i 构造了复数域，却把之前实数域不好处理的很多问题解决了。这里的关键点在于：复数的性质并不是简单的实数的数量增加或者实数性质的线性加工与转换，而是因为域的变化出现了全新的一些性质。这应该与现在所说的“涌现”有一定的关联性。引入新的属性维，会引起系统性性质很大的拓展和变化（比如分子组成的细胞、器官组成的人体、两个实数组成的复数）。这些问题可以与 4.3 节提到的问题合并处理，只是一个在拟全域，一个在形式域进行，它们之间应该也有很强的关联性。

（2）跨域的数学工具

这里简单介绍形式化后语义预范畴的一般性性质和特点后，将探讨对应于形式化语义的主要工具。重点研究在拟全域和形式域之间的转换所涉及的工具，以及某些在拟全域中的语义模型，在形式化后变成了什么？还可不可继续使用拟全域数学工具，以及能不能把原拟全域的性质或部分性质带进形式域中继续使用。

语义预范畴系统从拟全域转换为形式域后，一个最直接的问题就是无限逼近问题。形式系统是拟全系统的近似系统，理论上来说，这种近似可以做到任意精度，也就是可以无限逼近。

无限逼近会带来很多问题。逼近的程度有时候决定了系统的能力，即形式系统的精度会影响到系统的基础。比如元宇宙的构建，如果以一个设定的精度去建构，在后来又发现那个精度有问题不够用，那么，可能需要把整个基础掀翻去调整，就象 2000 年的时候的千年虫那样。所以，如果有可能，搭建元宇宙的时候最好就能考虑到这个问题，把精度做成可以无限精确下去的模型。

当然，最好的办法是突破现有的形式系统，能够不受精度的影响。但这似乎在短期不太可能找到答案。

在这里，我们强调的是：可计算语义预范畴是拟全预范畴在一定精度下的近似的离散化系统，因此除了拟全预范畴使用的那些数学工具外，我们还需要一些离散化后的数学工具和计算机工程中常用的数学工具对它进行研究和处理，包括离散数学、数理逻辑、图灵机、信息论、数值分析等。

接下来，我们把重点放到横跨拟全域和形式域的一些问题上。

既然形式域可以从拟全域形式化而来，那么直观上，在两个域中进行分析处理的数学工具，应该有一定的联系。但这些联系面宽理深，不可能在这里全面深入地探讨。这里把相关的数学工具及相应已有的关联的例子简单做一些讨论，希望能激发起这方面深入研究的兴趣。

离散数学 [121] 这种研究模拟/数字表示的专业显然是横跨拟全域和形式域

的学科，这个学科也比较成熟了，这里不多赘述。

数学分析 [69] 与数值分析 [122] [67] 是分别在两个域中处理问题的重要数学工具，它们之间有紧密的联系。后面我们会提到，很多问题可以在拟全域中用分析类工具寻找解决方法，然后再用数值分析方法在形式域中进行计算。其实这种方式已经在广泛的应用。比如前面讲到的最简单的 π 的例子，它在拟全域有个精确的值，但这个值具体是多少，却无法获得，只能在形式域中用数值分析的方法，不断精确化地计算。

除了这些很明显有关联的横跨两个域的数学工具之间的联系外，其它数学工具之间有没有联系？比如模糊数学和离散数学中模糊边界与二值边界的处理之间有没有直接联系？量子逻辑与数理逻辑之间有没有什么直接联系？

更广泛些的，比如拓扑、几何、代数在拟全域和形式域中有没有分别适用的工具？

当然，找出这些工具对应关系的目的是为了解决问题，比如有些问题在形式域容易解决，而另一些问题在拟全域更容易分析。很多时候，需要两个域之间的配合才能解决一些实际问题。比如还是上面关于 π 的例子，如果我们要分析一个对象的性质，可能就需要用 π 在拟全域进行分析，但如果要在工程上应用，最终还是需要在形式域用 π 的一定精度的近似值进行计算。有时候甚至需要在两个域之间来回地切换，才能真正解决问题。

目前，这样的一些来回切换，实际上都是“个案”式地解决的，很多可能是需要在拟全域中有点“碰运气”地逐个解决理论问题的，然后再把方案转到形式域中解决实际问题。如果能找到两个域的数学工具之间直接转换的规律，那就可以分类解决很多问题。

横跨两个域的数学工具之间的值得研究的关系还有不少，比如：1，图灵机的能力范围与语义预范畴之间的关系的探讨，可以了解用形式处理到底能解决语义方面的哪些问题？哪些问题可能是目前计算机能力范围所不可及的。那么，在类似元宇宙或智能机器人的交流方面，我们就能知道如果不改变计算架构，“智能”能走到哪一步。2，形式化对于属性和对映射关系的处理在工具上有什么差异？这些会对系统产生什么样的影响。比如前面提到过，如果一些判定是在映射关系层面做了很特别的变换，让一些平常看起来无关紧要的细节（比如假设有一个判定需要利用 π 的小数点后的第 238 位的值，而与此之前的值无关），在某些情况下变得特别重要，那么系统的基础是不是有可能发生重大的变化？

这里只能就语义的处理所涉及的两个域的数学工具之间的关系做一些粗浅的提问，解决这些问题肯定需要投入大量的研究资源，不是这篇文章所能企及的，这里就只能抛砖引玉了。

5.5 形式化语义预范畴的性质初探

有了形式化语义预范畴的模型后，这里我们对几个与形式化语义预范畴相关的数学问题做一些浅介，并由此对形式化语义预范畴的性质作一些初步的分析。

(1) 逻辑困惑的形式化分析

在前一章中的相关部分，我们已经讨论过语义中的逻辑困惑的问题。这里

用本章的形式化工具再来分析总结一下，理清其本质的问题，并对当前针对这种逻辑困惑所提出的各种处理方法，进行梳理和分析，分析可能的努力方向及发展方向。

我们还是从一个例子开始，从形式化、局部化和离散化的角度分析一个简洁却困难的判断问题。

例一：命题“她是美的”。

我们判断“她是美的”这个命题是否成立，很多时候是很困难的。这个困难实际上是在于属性“美”本身可以分解成很多的子属性，在实际对这个命题进行判断时，我们不可能把所有的属性都考虑到，也无法全面考虑。于是这个命题的真伪实际上是把“美”的很多子属性做了个向“美”这样一个综合属性上的投影，再试图对它进行二值量化。结果就会出现，昨天评判人跟评判对象一起在某个氛围下相处甚欢，他就选择当时关注的部分局部化属性对评判对象进行“她是美的”评判，结果为真；今天评判人跟评判对象一起在另一个氛围下相处很不愉快，于是就选择今天关注的部分局部化属性对评判对象进行“她是美的”评判，结果为假。但对象是同一个对象，对象也没有发生什么特别的变化，评判也都是合理的，然而结论却全然不同。

很明显，这种差异性是对一个拟全对象（属性）进行形式化的局部化，对局部化关注的属性取值，往某个综合属性选择性地加权投影，再进行（二值）离散化的结果。

这种类型的问题一直在发生，会造成很多的困扰。只有当我们所关注的被评判对象的属性很明确、取值范围也落在二值离散化容易判断的合适区域时，逻辑才会显现出数理逻辑^[25]的样子。

通过对以上例子的分析，结合我们在上一章中对几个例子的介绍，可以比较清晰地理出逻辑困惑的本质问题是跨越拟全性与形式化时发生的无限维属性在有限维属性上的离散投影产生的信息丢失的问题。因此，从本质上来讲，只要我们还用形式化手段（比如语言）处理拟全化的语义，这个问题的发生就是不可避免的。

相应的，从这个本质出发，我们也能大致看到改善这个问题的方向：在条件允许的情况下，1、尽可能地从多个角度考察更多的属性，或者选择更能反映核心关切的综合属性；2、采用更精细的量化。

这些也正是量子逻辑、投影逻辑、多值逻辑^{[115] [113] [25] [124]}所依据的深层逻辑方向。

比如量子逻辑。波函数也许有我们不掌握的参数或属性，而我们能观察的属性只是其属性的一部分。测量事实上是把拟全的波函数做了形式化，即把拟全属性在我们关注的属性上做了局部化投影，再进一步加权离散化生成我们可观测的参数。也许这就是“量子坍缩”发生的机制，事实上它是一个降维观察过程。类似我们在“语义中的逻辑”所提到的例二的情况。

观察的时候，只能把系统从拟全域带进形式域，所以，“不确定性”消失。当薛定谔的猫是拟全域的概念时，它可以既死又活。而当它被看作形式实体时，它只能要么是死要么是活。

最新的量子理论中关于复数的虚数维在量子场存在（不可或缺）的结论似乎是对这个假设成立的一个支持证据^{[125] [126]}。

关于非二值逻辑与语义之间的关系的更进一步的研究，这里只能分析一些方向，不再深入，这些更专业的问题需要更多相关专业的学者研究分析。

(2) 属性值的模糊性

顺着上一章中相关部分的分析和前面对逻辑的分析思路，我们可以发现，属性值的模糊性问题的本质和逻辑困惑的本质是类似的。即，它也是形式化时局部化和离散化引起的跨越拟全性与形式化时发生的无限在有限上的投影产生的信息丢失的问题。只不过在讨论逻辑问题时使用的是二值量化，而这里的离散化可以是任意有限量化级的。

同样的，对模糊性问题的改善也应该从局部化的选择和更精细的离散化角度考虑。比如上一章中提出的速度梯度的解决模型实际上就是一种综合的动态局部化方法。

而抽象化则是做一些反方向的努力，尽量把一些容易发生模糊性的属性剥离，留下更容易形式化的部分用于更精确地形式表达。

而我们在交流中能把一些模糊性逐渐变得清晰也是遵循这样的机理。即，交流过程（只能采用形式化表达的工具，比如语言，所以是在形式化域）实际上是对两个或多个主体的语义子预范畴个性的形式化的部分属性及其取值范围进行统一的过程，最终在某个共同的局部化属性的基础上，以共同的离散化标准达成对某个语义对象的统一判断，于是一些模糊性就变得清晰了。

为了能方便语义的处理，我们需要有存在模糊性的条件下，对语义中的属性进行处理的数学工具。比如对两个概念的某个公共属性进行析取或合取的操作。比如对沟通的过程分析中对两个或多个主体的语义子空间个性化的部分属性进行统一的工具。

这些都可以成为比如模糊数学^[113]等专业方向的研究课题。有关这方面进一步的深入只能留待相关专业的学者去实践研究。

(3) 概率的本质

这一节我们尝试用拟全性/形式化的模型来分析一下随机现象和概率^{[67][127]}发生的本质。我们可以从最常见的概率事件为例出发：

例一：抛硬币。

抛硬币的例子是概率论中最常用的例子。但抛硬币真的是个随机问题吗？

如果我们对如下的因素（属性）做一些明确的条件限定：硬币和桌面的材质（比如没有弹性的橡胶泥），硬币的质量、半径、厚度，硬币的密度均匀，击打硬币的力和冲量确定（某个专门机械），击打位置确定，环境温度确定、无风、空气密度确定、起始面确定。在这样的条件下，有没有可能通过物理公式确定地算出硬币在桌面上出现哪一面？我想你一定会在心里嘀咕：这样的条件的话完全可能。

但为什么抛硬币会被当成概率论中最常用的例子呢？因为一般情况下抛硬币时，以上的条件都不能确认，甚至你都不会考虑到有多少属性会影响到抛硬币这个过程。

那么，是不是可以理解为：当一件事情的很多属性我们不知晓、无法掌握

或者因其属性太多而不可能全部掌握时，那些我们不掌握的属性对事件产生的规律性的影响在我们掌握的局部化属性上的表现就是概率。

或者换个说法，概率可以理解为我们未掌握的属性发生的规律性作用在可观测的局部化属性上产生的投影。我们再用一个用过的例子来说明这点。

例二：旋转椭圆。

这个例子我们在“语义中的逻辑”用过，稍微做一点改变：假定这个椭圆在某个二维的面中旋转，而我们无法观测其中的某一维，只能以一个随机的时间窗口观测其在另一维上的投影长度。也就是说，在我们的世界里，我们只能理解这个旋转椭圆是一个有变化长度的一维对象。

这样的情况下，如果椭圆的旋转是有规律的，我们记录下来的长度结果会呈现某种在长轴和短轴之间变化的概率分布。当椭圆的旋转规律很简单（比如匀角速旋转），我们有可能得到它的“态函数”，甚至能猜中它可能是个旋转的椭圆，但由于无法观测到它的全部属性，这猜测并不能确认，只能不断通过可观测的属性维的数据统计去验证。而如果它的旋转规律很复杂，我们甚至连这种猜测都很难给出，也许只能给出它是一个满足某种概率分布的随机事件。

以上这个例子仅仅用了2维的属性，1个可观测（可局部化），1个不可观测。我们实际面对的语义对象，其属性可能会非常多，甚至无限。全面把握几乎不可能，而我们能关注的局部化的属性是极其有限的。于是，我们能掌握的往往就是语义对象在我们可掌握的局部化属性上的投影概率。

由以上分析，可以得出：概率论必然是一门重要的基础学科。它之所以成为一个重要的基础学科，是因为我们无法达到拟全，只能在有限的形式化的局部化属性上对事物进行分析。这样，我们就不可能甩开概率找到一个确定性的世界。事件发生的概率于是成为我们不得不面对的一个根本性的问题，因为很多情况下，我们只能掌握这些。

当然，随着时间的推移，有些原来看似随机的问题慢慢也可以找到其后的部分属性的确定性规律，但总有另外一些属性我们还是掌握不全（比如弦论中的“隐藏维”？），仍然得面对另一些随机概率。

从这个意义上来说，随机过程和概率论也是语义处理所必然需要应用的数学工具。

(4) 形式系统部分性质初探

有了以上的一些介绍，我们在这里试着从形式化的局部化和离散化中能得出关于形式化语义预范畴系统的一些什么样有趣的结论？

我们从一个多维的局部化离散化的形式系统开始。

一个多维的局部化系统，可以看作是多个1维的属性维的积空间系统，即：

$$X^n = X_1 \times X_2 \times \cdots \times X_n \quad (22)$$

一般来说，如果 n 维系统的各维是线性无关的，那么我们就只能在 n 维空间中进行处理。不过，这里的属性维虽然是线性无关的，但它们都是离散化的，即每个属性维的取值都是一定精度下有限的 N_i 。因此，根据积空间的性质，这个局部化系统取值的数量也是有限的：

$$N = \prod_{i=1}^n N_i \quad (23)$$

于是，我们总可以把这些点映射到某个一维的空间上（这里可以有很多映

射转换方案，比如采用“之”字形映射等）。也就是说，我们总有另外的局部化方案，可以把一个局部化的多维的离散化形式系统转换为另一个量化值更多的一维的局部化离散形式系统。（需要注意的是：转换的过程中，中间步骤的量化精度有可能需要大幅提升。）这样，我们就证明了如下的结论：

定理 5- 1 一个可以形式化的问题总是可以局部化为一维问题。

我们在这样的基础上，就会有一个很有意思的结论：不管系统多么复杂，如果选择合适的局部化方案，我们总能在形式化的基础上，实现一维的识别！

用一个实例来说的话：只要人跟人有区别（选定适当的关注属性后，这总能做到），我们总有办法（对人的各种属性进行合适的局部化转换和离散化方案）用形式系统把一个人与另一个人区别开来，即，机器可以把人认出来。推广到一般情况就是：如果不考虑算力、训练样本数量和模型复杂度的限制（，以及获取关注属性信息的难度），模式识别的任务，计算系统总是可以完成的。

有了这个基础，我们更进一步：如果在一个合适的局部化方案下，选择一个合适的离散值把上述的一维的坐标做个分界，我们就可以得到一个二值化的逻辑系统。即：

推论 5- 2 可以形式化的问题总是可以逻辑化的。

把这个推论反过来，就是我们熟悉的结论：‘0’和‘1’可以表达任意的形式系统。

回想一下前面的例子：“她是美的”，这样的问题是不是就迎刃而解了？

所以，哲学的很多早期的思辨特别被逻辑吸引，哲学研究的很大一部分与逻辑相关就不奇怪了。而类似“她是美的”这样的问题产生的基础，我们在前面的章节已经分析过，是人类要思维指导行动的必然需求。

因此，拟全/形式化二重性系统符合人类的思维指导行动的要求。

6 可计算语义预范畴

上一章对形式化的目的和主要功能进行了一些探讨，本章具体探讨如何建立前面提到的有限“基”的可计算语义空间、其上的有限映射关系基空间，如何以此构建可计算公共语义预范畴的框架，以及如何在这个框架上不断生长为可以工程应用的可计算公共语义预范畴。

这里的探讨只能理一个工程的思路框架，具体的工程实践需要大量的精力和资源投入，只能以后有机会慢慢实践了，或者有人有兴趣组织足够资源尽快尝试完成。

这样的构建工程完成后，可以进一步作为工具和基础进行包括后面章节的广泛应用。

工程过程中，也许会碰到很多困难，本来那些困难也应该是本文研究探讨的重要内容，但由于前述的资源 and 精力的限制，这些目前无法进展。也许有些困难会引起本文的一些基础内容的修正甚或重构，没有条件了解这些也是遗憾。

6.1 可计算语义公共预范畴的构建

我们前面讲了语义预范畴的结构，但那只是一个数学的抽象模型，并不能用于计算机实际进行语义分析等具体任务的应用。下面就讨论一个可以具体应用于语义分析和处理等计算任务的形式语义公共预范畴应该要如何构建。

因为目前的所有计算系统都是基于图灵机的有限状态机，因此要构建可计算的语义公共预范畴，一定是形式化的。

当然，如果以后有其它的非有限状态计算系统，可计算语义公共预范畴也可能不是形式化的。

形式化和可计算的语义公共预范畴构建完成后，也可以反过来给拟全域使用。只是用法的基础（人的能力基础和机器的能力基础）会有不一样。所采用的工具和方法也会有很大差别。

在拟全域中使用形式语义预范畴，实际上是借用人“拟无限”、“拟遍历”、“拟存在”等能力，把形式语义预范畴“嵌入”到或“扩展”为拟全语义预范畴来使用的。

(1) 可计算属性基空间

前面语义的数学模型中，语义预范畴的基空间是属性构成的，因此，可计算语义预范畴基空间的构建元素也只能是属性。

如果我们能把所有描述概念的属性都理清楚，并列出来为语义预范畴的不同的维度坐标，就可以得到一个基础的基空间。

如前面所述，这里能构建的可计算属性基空间，只能是形式化的。因此用来构建基空间的这些属性，首先要能被形式化表达，然后必须要有限。符合这些条件的，可选的比较好的表达是自然语言文字中关于属性概念和定义的字词。

但是这里还有几个问题：

1、可行性。我们首先要解决的问题是：用文字的字词，能否构成有效的语义预范畴？这个问题实际很不好回答，最大的困难在于拟全预范畴和形式预范畴的差异。从拟全域来看，我们理解语义，绝大部分是通过语言和文字进行的，因此语言文字结合人的能力应该能把我们的语义预范畴表达出来，从这个意义上来说，用文字的字词来建立形式语义预范畴的基础维是可行的。从形式域来看，文字的字词是有限的，如果它们在拟全域中代表的语义能“带”进形式域中来，那么，它们应该在很大程度上保证语义预范畴的有效构成。但这其中有个比较困难的问题是，文字在拟全域的语义能不能“带”进形式域来，怎么带进来，这个“带”入会不会产生内涵或外延的损失，甚至整个意义的丢失？

在这里，我们接受上一节所提到的：“这里的形式化不是理想的，它一定会在某些环节上引入某些拟全化概念（属性）作为基石”。也就是说，我们需要认可由人把一些拟全域的属性概念直接“带”入到形式域中来。并以此为基础，构建基础的属性维度。

这样，我们就解决了可行性的问题。但是要记住：我们并不是真的在一个“纯粹”的形式基础上开始形式化大厦的建设的，在这点上，我们并没有比《几何原理》走得更远。这点有可能是一个隐患，或者说，如果有人能在这点上深入推进到能建立一个不需要借助拟全语义的形式系统，有可能会产生一个意义深刻的全新形式语义体系。

2、代表性。根据前面的解释，可计算的模型不是拟全的，因此一定是不完整的、部分缺失的，构成它的坐标的属性维度是挑选出来的，不全的。即，我们需要对每个概念 C_c 选出一个有限序集 I ，及其有限的属性集，形成形式概念

C_f 的属性集 $C_f = \{c_i, i \in I\}$, $c_i \in D_i$, $D_i \in C_c$ ，并由此构成可计算的语义公共基空间。而其有序集选取的过程，同时完成了形式化和公共化的过程。

在选取的过程中，就会涉及到被选取出来的这些属性，能不能覆盖语义空间中所用到的全部属性，也就是它们有没有代表性的问题。从构建的角度看，如果这些挑选出来的属性能把我们碰到的常用的几乎全部语义都能表达到位，就是具有代表性的。但是由于拟全性和形式化的差别，从理论上讲是不能做到语义的全部覆盖，我们本来的目的也是做到拟全语义空间的近似。后面在形式化数学性质分析时我们会提到，形式化其实是一种局部化过程，总有部分细节会在形式化过程中消失（比如前面讲的 π 的离散化，不管取精确到多少位，总还是有后面的位的数字会被我们忽略）。而如果我们正好有某些情形下会用到这些细节，它就会变得重要，而我们会遇到相关属性的语义分析的困难。

所以，在建立语义公共预范畴时，我们只能往尽量覆盖得更多更有效的方向努力。这里我们采用的方法是两个方面：一方面，我们允许这个基础空间生长，这将在下一节讨论；另一方面，我们在构建之初，尽量让它能覆盖得更多一些、更有效一些。

为了让它能覆盖得更多更有效，就需要对尽可能多的情况进行分析判断。但由于概念和概念的属性的基数都是非常庞大的，我们对尽量多的概念 C_f 的属性 c_i 进行相应的分解分析，再从中找到可以用于可计算语义公共预范畴基本维的属性集 $\{c_i, i \in I\}$ 。这将是一个巨大的工作量。实际可行的方法是对一些常用字词和概念先做处理，然后保持一个不断添加和丰富的机制，允许对基空间和整个语义空间不断增补和修订。

3、递归性。由于前面所讲的概念及属性的递归问题，会传导到属性与属性之间的递归问题。所以，这个构建过程，很容易产生递归式的循环，从而无法在开始的时候尽快确定哪些属性可以作为语义预范畴的基础维度。如果任由递归进行下去，就又会变成无限的循环，失去形式化要求的有限性。因此，这个构建过程的开始阶段需要人为参与选择和决定，断开无限递归链。而在后续的空间生长中，也需要有相应的机制保持限制无限循环的出现。

4、基础性。我们开始时并不知道哪些属性是构成可计算语义公共预范畴的基础，这一点，在开始前显然是不确定的，而且看上去很难通过某些理论上的推导得出结果。可行的办法可能是把概念用我们已知的有限属性尽量拆解，同时根据前面的方法尽快切断递归，然后看看需要哪些属性才能基本上把概念拆成由基础属性描述。这个办法的可行，是因为根据字词的有限性，可以知道已有的可文字化的属性一定是有限的，最多也不会超过所有词的数量。

5、语言。既然我们借用文字来构建语义预范畴的基空间，那就会涉及到用哪个语言的文字进行构建的问题。根据之前提出的代表性、基础性的要求，从理论上来说，需要相应语言的文字表达能力和范围要足够大，尽量包含更多的概念和语义内涵和外延。而根据我们在实践中的经验，世界上使用较广泛的语言，都有相当的语义覆盖广度，而且它们相互之间基本都能互译，也就是说，

它们表达的语义空间的能力在很大程度上是相互覆盖的，或者说是近似同构的。这种同构也是语言之间可以相互翻译的基础，如果没有这样的基础，即两个不同语种语义预范畴之间的语义子预范畴根本找不到对应，那么翻译也就无从着手了。因此，从这个意义上讲，用哪种语言来做这个公共语义空间其实关系不大。另一方面，由于一些语言形成的文化在拟全域中有一些无法用简单的语言表达的文化内涵，这些在形式化后只能通过不断生长的方式逐步纳入，因此，为了能更好更快地应用，做多个语言语种的可计算语义预范畴版本在不同地区使用也是可以的。

这几个问题厘清后，就是具体构建的工程问题了。这个问题与语义的映射关系空间可以同步进行，因此我们放在下一节的关于语义的可计算映射关系空间分析之后一并讨论。

(2) 可计算映射关系集

基于与可计算语义基空间类似的原因，可计算语义映射关系基空间的表达，比较好的可选项是文字中关于语义的映射关系的字词。

同样基于类似的原因，可计算映射关系形式空间的基映射关系，也会将一些拟全概念或属性“带”入到形式空间中作为基石。即，这里的可计算语义映射关系空间也不是理想的形式空间。

关于语义映射关系空间的基本问题，与前面语义属性基空间的基本问题也都类似，这里就不再重复展开。

我们前面分析过语义预范畴中的映射关系是由一个三元组来表示： $m \sim (f, m_c, m_v)$ 。而其中的 (m_c, m_v) 组合，其实指向语义空间的一个表示属性的概念 C 。而这里关于属性的可计算模型部分，我们可以并入到上节的可计算语义基空间中处理。这里要处理的重点其实就是映射关系的关系 f 部分。

而在采用了这种三元组模式后，相当于扩展了前面的可计算属性基空间，当空间中有映射关系的对象，在空间中产生交集时，实际上就表明了关系 f 的发生。从这个意义上讲，关系 f 似乎不需要专门的表示。但是这种空间中用交集来表示关系的发生，会有一些限制：比如它很难表示映射关系的方向性，比如A推了B，显然是有主动被动的方向性的，如果只有子空间交集（两个对象与“推”属性组成的空间有非空集）是能表示映射关系的发生，却不好表示它们之间的主动被动的方向性。当然，这也可以用增加属性维的方式来完成，比如给“推”属性增加一个伴随的方向属性，但空间结构就更复杂了。

不用三元组方式的模型也是一种可能的映射关系研究方式，但似乎从构建可计算语义预范畴来看，更加复杂。给那些映射关系建模时，很容易又走回到建立一个基向量空间的方法，与这里的映射关系属性化、空间化又一致了。因此，这里不作为重点去分析，但保留采取其它方案的可能性。

映射关系是状态、关系、作用、过程等概念之间（包括与概念自身）发生的各种关系，因此它是包罗万象的。我们知道的宇宙里发生的所有事情都与它相关，或者说，它都应该能（在一定程度上）表达。从文字来说，所有跟动词、形容词相关的词语（甚至所有非名词性的词语），以及关系符号都是映射关系表达。

因此，它涉及到太多的情形，非常复杂。比如我们在前面语义映射关系集

模型提到过的，数学的表达也是表达，因此数学的工具与方法应该都包含在映射关系集内。反过来，是不是所有的语义映射关系都可以用数学表示呢？

这涉及到几方面的问题：1、数学本身涵盖范围的框定，比如三段论逻辑，算不算数学？2、有些问题，目前还不能用已知的数学知识来数学化，但是数学也是发展的，以后会不会找到数学甚至发展出新的数学分支来表达，也是个未知数。

如果限定于目前认可的数学的范围，以及已有的数学分支，那么语义的映射关系是部分可数学化，部分还无法数学化。这里可以举几类例子分析：

一类例子是数学本身的文字与符号描述，即关于数学的语义映射关系。这类语义映射关系显然是可以数学化的。包括：

逻辑是比较明显的语义中概念之间关系的表达的一种方式，所以，数理逻辑操作应该是属于语义的映射关系空间的组成的。

基本的算术肯定也是语义表达的直接方式，所以也应该是属于语义的映射关系空间的。

其它分析、代数、几何和抽象数学，是不是也要算进语义的映射关系呢？如果从纯数学专业的对话来说，数学概念显然是可以被讨论的，而且，对这些数学概念的数学操作（积分、微分、求边界、同调、变换等等），似乎也都是能通过交流沟通进行表达的。这个意义上来说，所有的数学对象应该都可以看作是概念，而它们之间的操作都可以看成映射关系。

从这样的角度分析，所有的数学工具，应当都被看作语义映射关系的一部分。

再一类例子是关于物质的相互关系的语义映射关系。比如“他打你”，“她写下了很长的文字”，“他用工具把木头做成了桌子”。这些应该可以化为物理过程，然后用物质相互作用、物理公式、数学推导，最终化为数学的映射关系。详细的转化过程这里就不多分析了，有些分析也会非常的复杂，但直观上，这些基本都是物理过程，符合物理定律，在做层层分解后是可以数学化的。

另一类例子是关于思维和情感方面的语义映射关系。比如“他觉得这个理论很深刻”，“他喜欢她”，“这段音乐有种空灵的感觉”。这一类的语义映射关系，分析起来就很复杂。

从某种意义上来说，比如音乐和空灵之间，一个是物质性的，一个意识性的，它们之间存在着著名的“意识与物质的鸿沟”^{[34][54]}，如果我们说有数学模型可以把二者联系起来，把它们的性质都理得很清楚，那么岂不是说已经找到了某种“填平”这个鸿沟的方法了？显然没那么容易。

现阶段要把这类的映射关系直接数学化是很困难的。我们这里只能还是采用之前提到的方法，把一些属性（比如“空灵”）作为独立的拟全映射关系属性直接纳入进形式领域作为基，先不考虑再对它进行进一步的性质分析。否则我们现在就无法前进。等以后各方面的研究进展条件成熟后，再对它们进行深入的研究。这样，我们就可以按照 $m \sim (f, m_c, m_v)$ 三元组的表示来初步把映射关系模型化。

以上的例子并不全面，但由此也可以看出这里的情形太过复杂，我们实际构建可计算语义预范畴的操作中，特别是开始的时候，就要注意处理过程要有足够的适应性和扩展性。有很多问题需要进行针对性的特殊处理。这里举两个例子：

1、能将“喜欢”的关系级联吗？即“A喜欢B，B喜欢C”，能得出“A喜欢C”吗？不能！那么，我们是不是可以得出“喜欢”这个关系，不具有传导性（当然这也是某种性质）。而且在“A喜欢B，B喜欢C”的情况下，我们能得到一个“A喜欢C”的可能性比较大的结论。这应当也是某种规律和性质，所以，尽管不能得出明确的性质，但对一些关系，我们还是能得出一些性质和规律。那么，这样的映射关系，有数学上的联系吗？我们可不可以说：有关系，就有数学？关联本身就是最弱的数学联系？这种联系是映射吗？至少可以形成某种网络？这也是 $m \sim (f, m_c, m_v)$ 这种三元组表示的由来。

2、再来看看在形式域常用的逻辑关系。逻辑关系及三段论其实有两类：一为是推论性的，一类是断言性的。比如“如果A等于B，B等于C，那么A等于C。”这种是由两个条件直接得到一个结果，属于推论性的；而“如果A等于B，那么去完成C。”这种是根据条件去执行另外的任务，属于断言性的。

这里面，只有“如果……”的部分才有逻辑关系和操作。在断言中，“那么”后面的执行部分，其实本身与条件无关，只是由这个断言建立了关系；而在推论中，“那么”后面的结果是与条件有密切关系的，即，如果条件成立，不做这个推论，那个结果也在那里，推论只是把它表达出来，让读者更容易理解了。

证明和推理的部分基本上都是推论式。而行动方案计划一类的，多数是断言。或者说，状态和关系的三段论，推论为主；而作用和过程的三段论，断言为主。

要注意的是， $m \sim (f, m_c, m_v)$ 三元组表示目前还只是一个假设的模型，并未能完成对这个模型通用性的检验。实际的检验需要在构建过程中进一步进行，也有可能在进行的过程中推翻或修改。

关于可计算语义预范畴的模型的基础讨论先进行到此，很多细节需要在具体构建中进一步深入分析。具体的构建方案我们将在下一节进行讨论。

(3) 一个可计算语义预范畴构建方案

前面讲到可计算语义基空间和关系空间的构建都可以基于文字进行。它们的构建可以在对文字表达的概念、属性、关系和定义的处理中同步完成，因此，这里把构建方案合在一起讨论。

由前面的讨论可知，可计算语义预范畴是形式化的，它必然要建立在某种局部化方案的和离散化方案的基础上。因此，可计算语义预范畴的构建需要指出其目标，才能有方向性地进行局部化。

由于没有另外的基础，以及资源的限制，这里准备构建的可计算语义预范畴必须从零开始，规模也不能大，需要有一定的代表性和普适性，能验证本文语义结构模型假设的基本可行性，并在初步验证后，能作为进一步完善或构建其它可计算预范畴的基础。

因此，只能从一个最广泛的公共性基础为出发点，因而可以设定的局部化目标应该是以标准普通话为基础的适用于全中国范围的常识性的日常交流，不涉及专业、地域、民族、宗教等。基于类似的原因，这里设定离散化的目标为满足日常使用要求。

从这个局部化理念出发，一个比较直观的起步构建方案是以词典的条目为基础，每个条目字词都是一个基本概念，而条目下的解释涉及到它的组成概念或属性，以及一些基本的关系（映射关系）。把这些涉及到的概念、属性、关系

一层层地分解下去，应该可以得到一张层次关系网。这个网应该会有递归存在。

关于词义的循环嵌套解释和关于用一些基本词汇生成所有概念的问题^{[53][2]}，

在《语义学：从数学基础到语义语用学》的第4.5节“词素”部分^[3]也有类似基于分类的基础元素的建立思想。并且提到基本词汇表量在850~2700个单词。但文中同样表示，实际的表达并不能完全依赖这些“基元”。这些跟这里应用词典条目基本思路是相容的。

对于递归循环的部分，由人工参与，指定一些基础属性作为语义空间的基维度，指定一些基础关系作为语义映射关系空间的基维度，拆开循环。然后把全部条目中涉及到这些基础属性和基础映射关系的递归都拆开。再对剩下的递归循环做相同的操作，直到解除所有的递归循环，形成一个有限的关系网络。

这时所有被指定为基础属性维的属性就组成语义预范畴的基空间。而所有被指定为基础映射关系维的映射关系就组成了语义预范畴的映射关系空间。

至于具体采用哪本字典，主要涉及到代表性、覆盖性和工作量问题。可选的主要有《新华字典》《现代汉语大字典》等几个主流的基本字词方面的字典，具有较好的代表性。专业方面的词语由于量大且专业性太强，作为基础收录工作量有些过大。覆盖面的问题，留待下一节的可计算语义预范畴生长中，用其他方法解决。

初步准备采用《新华字典》作为构建基本的可计算语义空间的基础。

选定出发点后，接下来就会涉及到最根本的架构问题：基空间和映射空间是分别建立，还是做成一体？这个问题我们在前面介绍语义预范畴时就提到过，由于还没有实践，其中很多具体问题无法详细研究，当时只能把这个问题暂时开放，未给结论。

这里就有两个选择：1，以物质性属性和意识性属性为基空间，而把关系属性为基础构成单独的映射空间。2，不分辨物质属性、意识属性、关系属性的专门空间，而是由这些属性统一组成语义空间的维。

两个选择的差异主要在于：两个独立空间，更适应映射关系与一般属性的性质差异，表达应该能更灵活；一个共同空间，表达使用的处理工具可以得到简化，基本上映射关系被化成关系属性直接在属性维表示，这样，后续的处理结构简单。

独立空间的模式，还可以进一步分解成多个独立空间，比如把基空间分为两个模式的世界：一个以物质性属性为基础，可以分解为基本物理量构成；一个以意识性属性为基础，可以分解为人的基础感觉和知觉构成。这样做看起来可以把不同性质的属性采用合适的工具分别处理，应该效率和效果更好。但这样做也会带来两个问题：1，是不是需要进一步分解呢？如果这样一直分解下去，就会遇到一般分类或分维所面临的很复杂的细节问题，很快就会发现分解本身会错综复杂，没法进展下去。2，物质属性和意识属性本身的界限模糊，有很多属性是混合型的，这些属性怎么处理？空间会因为这些问题越来越复杂，相应的引起后续的处理工具也越来越复杂。

分解空间所带来的第二个问题在前面的两个方案的选择上也同样会碰到：很多关系属性会是与意识属性混合的，分不清界限。这必然导致在分开独立空间处理映射关系时带来很大的复杂性。

由于这样的一些明显的困难的存在，这里倾向于选择采用全部属性维统一

的语义空间方案，这也是前面专门分析映射关系，提出映射关系三元组模型的主要出发点。

在实际进入软件工程的步骤（本文主要对一些原则和基础做分析，这里不对具体的软件工程的部分作介绍和分析）前，另一个需要分析的重要问题是：涉及到具体语义对象（概念或个体对象）时，怎么进行属性分解？是否允许或者要不要进行属性的逐层分解？如果允许，怎么进行分解，能分解到哪一步？实际做的过程中，最终会不会又弄回分类或分维表去了？吴蔚天的介绍 [2] 用语义场的方式，似乎跟我的之前想用的属性分层分类的分维表很类似？但这些都碰到很多困难。

前面提到过，属性有时候也很复杂，整体不等于部分的和，而且是非线性的。有时候某一层面的对象对某些属性是不可约的，对另一些属性又是可约的。不可约的时候，显然是不能再分解下去，可约的时候则可以分解。

因此，合理的方案应该是：属性分解应该根据对象涉及的范围内的一定的局部化方案（注意不是整个可计算语义预范畴的整体局部化方案，是在整体局部化范围内的，针对语义对象及其属性分层的再局部化，这样的局部化（在每层）可能都有很多，类似于语义预范畴中的万花筒）进行，在每个局部化方案下进展到不可约为止。这样的方案原则下，属性分解可能会形成分解层次的参差不齐，这跟我们对语义的理解应该是相符的。

当然，用词典构建的语义空间，显然不能满足语义表达的全部需求。有几个方向完善这个构建的语义空间：一是把一些典型的句式、篇章结构等加进去（灌输式）。这个方式需要一些统计和特别的软件支持，其必要性也需要进一步确认，可以暂时放一放；二是通过学习不断生长（学习式）；三是通过后面关于形式化的局部化的方法，构建多个适应不同情境的局部化形式语义预范畴。这些方向也可以相互结合。下一节将主要讨论第二种方法。

6.2 可计算语义预范畴的生长

构建的语义预范畴，主要是空间，而我们的语义预范畴是由空间与嵌入其中的概念子空间以及其上的映射关系构成。刚构建好的语义预范畴中，其中有效的概念子空间是非常少的，因此，要能实际应用需要进行补充。

这跟拟全语义预范畴的生长是不一样的。拟全语义预范畴的生长是从一个已有预范畴的基础上增补修正式的生长。而这里的可计算语义预范畴的生长是从几乎空白的基础开始生长的，需要大量地补充内容。

生长有两个方向：1，蛮力，大而精；2，巧力，万花筒。简单点说有点类似科普和专业的区别。而蛮力可能更进一步把科普也搞到专业的程度。

前面基于词典构建的基本语义预范畴只是一个可计算语义预范畴的基础，实际的应用必然需要在这个基础上发展。这点很容易理解，本身字和词也在不断增长，因此，语义预范畴的生长是个很平常的事情。之前已经讨论过一般语义预范畴生长的问题。可计算语义预范畴自然也是需要不断补充和丰富，才能适应知识和语义的不断增长和变化。

完成了形式化的可计算的语义公共预范畴构建后，前面关于语义预范畴的生长和语义预范畴的个性化的分析，可以直接应用到这里的可用语义公共空间。

包括对语义预范畴的维护，对可计算语义公共预范畴也都同样适用。对可计算语义公共预范畴的生长和维护，应该也可以使用目前在人工智能中大量应

用的数据推动的模型和方法。

特别是可计算语义预范畴的生长方面。在有了基本的可计算语义公共预范畴和属性维后，可以通过不断输入各种合理的、有用的知识和资料，用适当的算法对这个语义预范畴进行学习训练，从而让它不断地自我生长。

这里的生长将不止是概念和属性的补充和修正，还有它们的关系、作用、过程的丰富。这些内容将对语义的分析起很大的促进作用。

可计算语义预范畴生长的另一个方向是局部化及其组合。形式化本质上都是局部化的。形式化总是在全域的拟全语义预范畴中，挑选了某些被认为是重要的属性作为基础维进行的局部化（这将在后面的形式化数学模型部分更深入讨论）。可计算语义公共预范畴也是局部化的。它是不是可以由更局部化的模块（成分）组合而成？

局部化的公共预范畴肯定有很多的缺失，为了能应对更多的语义处理问题，比如一些专业问题，是不是需要按专业方向实现局部化生长？实际使用时根据情境的不同，装载不同的局部化模块。

如果不包含那些专业的局部化，是不是就需要引用整本整本书去讲一个专业概念？那样的语义处理效率太低。但如果把所有的专业属性都放进公共语义预范畴，会不会又显得基属性维太庞大、太杂乱？太多无效空间，有效语义子范畴更加稀疏？所以专业的局部化可能是比较好的解决方案。

这里的可计算语义预范畴的生长方式肯定不能再基于人工操作了，否则无法跟上社会和科学不断的发展。也不符合本文对语义结构应用于智能系统的要求。

采用人工智能的学习方法，这个可计算语义预范畴的自动生长软件需要完成以下几个主要任务：

- 1、能自动判断需要增加新的属性维度。当一些新的概念（比如新技术）出现时，用已有的属性维度已不能表达，就需要增加相应的基属性维度。这个判断其实是比较困难的，应该需要不少的数学工具参与。

- 2、能自动判断需要增加新的关系维度。这与上面的属性实际上是类似的，不过是关于新的状态、关系、作用与过程，不能由已有的关系来表达的时候。它的判断由于涉及到属性之间的联系，以及跟已有关系之间的联系，所以比新增属性维度更难判断，应该也需要更深入的数学工具的参与。

- 3、在已有空间的维度下，建立概念语义子预范畴。这个任务比1稍微容易一些，是对未接触过的概念进行建模，但可以用已有的属性和映射关系进行足够明确的表达。本质上，这和任务1是同一个任务，只是任务1不能用已有属性表达碰到的新概念。

- 4、对输入的句子、段落、篇章进行处理。对其中表达使用的关系，进行总结，并构建相互联系。把一些歧义、隐喻、暗喻等关联网络慢慢建立起来，对后续输入能自动产生联想，并给出可能的歧义的解释，及其成立的上下文条件及概率。

- 5、要能逐渐分清公共语义预范畴内容与个性化语义预范畴的内容，并针对性地进行局化语义预范畴的动态应用。比如判断上下文宗教语境，对一段表达动态地引用相应的宗教相关的可计算语义预范畴进行语义处理。

- 6、建立可计算的语义理解和分析的基本模型。这个要求有点综合，不是一下子能说清晰的，似乎也不是语义预范畴生长的主要任务，但和语义预范畴生长又关系密切，即语义预范畴的生长机制需要这个模型的支持。应该是更高一

个层级的应用架构。

以上所提的对可计算语义预范畴的自动生长软件的要求是比较困难的，要做好它，往往是需要与本文中其它部分（比如要求中提及的语义理解和分析模型）是密切相关的。甚至生长模型最终就是语义处理模型中的一个必要的部分。

在很长的一段时间里，可计算语义预范畴的生长可能都还需要人工参与。比如一些新的拟全概念的赋义等问题用目前有限状态机的机器应该是无法实现的。而生长模型也需要长时间地维护、跟踪修正。

如果生长的模型做得够好，可计算的语义公共预范畴，也可以几乎从 0 开始通过学习和训练算法，通过数据驱动构建生成？

当然，这部分这里只给了一些原则框架，实际应用与可计算语义预范畴构建一样，需要设计具体算法。并且要利用不少数学工具，判断属性的新维度的必要性（用相关性？）、映射关系新维度的必要性（这里会涉及到网络、拓扑、代数拓扑结构这些问题？）、形成映射关系之间的联系（这部分涉及到对动态网络拓扑问题的处理）、根据上下文环境启用不同的个性化（局部化）语义子预范畴（拓扑的分支化？纤维丛？）、涉及到语义子预范畴的语义分析（代数几何？）等等，其中还要穿插模糊数学和量子逻辑等。另外，有些问题还需要人工参与或人机对话解决，或者更新算法。

6.3 面向社会科学的可计算语义预范畴试验设想

在前面章节，我们讨论了二值逻辑的本质和困惑问题。这里就这方面在社会科学研究中可能的一些相关应用，以及它的简化可计算语义预范畴模型进行一些探讨。

(1) 概念成形过程推测及缺陷分析

在本文前面章节中我们得出的结论是二值逻辑其实是多维（甚至无限维）属性在某个局部化方案下在其中某个（关注的）属性维上的二值化投影。

如果我们把二值化的属性看成是局部化属性投影后的对立面形成的矛盾取值进行分析，比如亮度的“明”“暗”、气味的“香”“臭”、小时候简单判断一个人是“好”“坏”等等，可以发现，正如我们之前所提到的人出于行动选择的要求，人们在对概念的属性（属性本身也是概念）认识的初期，往往都采用二值化处理，然后再逐步多值化。即，人对概念的实际处理操作的认知过程，由于采用的手段本身只能是形式化的，所以正好与拟全概念的形式化过程相反，是由最简单的形式化（关注属性直接局部化、采用二值逻辑的离散化）再逐步细节化（多值逻辑的离散化）和复杂化（多维化）的过程。

于是，正如前几章节讨论过概念内涵的“说不清”，如果回溯人类语言（形式化工具）的发展历程，复杂概念的表达应该是和语言符号几乎同步发展起来。符号的指称在先民的日常生活生产中通过实践逐渐总结而成。同时概念表达又能够将魔鬼藏在细节里，展现出高度的概括性和沟通有效性，其内涵虽然可能蕴含诸多未知细节，但是只要这些细节在生活生产实践中短期影响有限，先民们够用即可，并不追根究底。发展到概念因词害意的时候，再进一步分割辨析概念语义也不迟。

即使发展到现代社会，概念仍然保留了这种由产生而普及的特征：

- 1、为沟通有效性引入，高度的概括性
- 2、可容忍未知细节，并不追根究底
- 3、够用就好，可多版本反复做定义尝试
- 4、定性可以大体理解即可，不追求精确定量

相比语言，数学虽然也在同时间并行发展，但是数学建模及数学完整体系的架构大体是近现代完成的，在此之前中文中的概念因为其高度概括性，绝大部分表现为定性陈述，少量的定量表达如“九死一生”“事不过三”“我们人们”等等，多是泛指，难说是实际概念中数学规律严格表达，大体仍旧属于定性的范畴。如果定量表达，则往往是专门量词“三尺”“五丈”等等来表达，一般概念背后的世界虽然有定量的内容，但是一般不展开。

这种定性表达的方式方法，在远古时代，社会分工不发达，生产力不足的年代，由于大家普遍需要躬身实践，或者即使脱离日常劳作，也能够对实践有近距离粗略的定量了解，概念停留在定性表达层面，沟通双方自己脑补实践常识等定量问题不大。概念的这种特点也就因循下来了。

但是即使是在古代，概念的这种局限性，儒道释中仍旧反复提及，中道而行，无为无不，非常名，不二法门等等，破除人脑中由概念引入的成见，语义枷锁，甚至直接警告有些东西是“只可意会，不可言传”。

纸上得来终觉浅，绝知此事要躬行。

近代随着科技生产力的发展，社会分工细化，各行各业隔行如隔山。知识爆炸的背后也是概念数量指数级上升的过程。人们仍旧承袭了以往的概念定义及认知方法，高度概括的概念中新概念也仍旧停留在定性层面，少有的定量表达概念，大量的数学人名命名定理及指称体系。而遍布语言中的诸多千百年以来反复使用的概念，成为语言游戏，由于沟通双方大概率缺乏一线实践经验，脑补的内容通常往简化偏激方向偏离。好坏掺杂，真假难辨，失败是成功之母等等严重依赖实践认知的常识沟通中常常缺失了。

新冠疫情舆论中“真”“假”概念就呈现出了绝对化偏激化的特点。新冠属于新病毒，进化频繁，专家们也是一边研究一边与大众沟通，这里面对的是人类未知领域知识，错漏疏失难免。但是舆论大众对于未知病毒领域工作流程认知有限，对专家按“真话”“假话”来进行概念标签，忽视未知领域中科学以大量的认知错误通过实践淘汰出一定比例的近似“真理”的机制。“砖家”“反智”等言论大行其道，概念过于概括，冰山下原来依赖大家共同的实践脑补的内容的缺失，是这一现象的主要原因之一。日常使用的时候先入为主形成的概念（属性二值化的）牢笼导致的思维局限。

但由于之前讨论过的人的行动决定需求所驱生的思维逻辑化的习惯是有效的，而且人们是很难改变的。因此，这里考虑有没有可能顺应这种思维模式，提出一个相应的可计算模型，从而在已经经过长期实践可行的思维方法基础上，通过扩张二值逻辑的维数来模拟和拓展人的思考方式，打破思维局限性。

（2）基于属性矛盾取值的可计算语义预范畴建模

基于以上的分析，这里针对性提出一个反向的可计算语义预范畴建模设想，用于适应人类认识过程的社会科学的应用试验。即，我们反过来以属性的二值化矛盾取值作为基础来建立一个可计算语义预范畴，用于某些社会特殊应用，比如辅助日常的决断、人与人的高效沟通、社会问题决策机制等。

前面已经说明，矛盾（逻辑的对立值）通常是概念属性的一种（对关注属性或者说有限维度特征属性的）局部化降维逻辑投影。如果我们把推论 5-2 的结论反过来使用，可以认为在形式域，一般的需要细化为多值的问题可局部化为二值的多维问题，或者说，一个问题除了用多值化细化外，也可以用很多矛盾取值的多属性维细化。即，可近似认为在形式域一般社会范畴的概念对象可以由多维的矛盾属性及映射关系集组成。

矛盾的普遍性原理可以由之前的属性变换投影来解释。每个概念一般都包含矛盾，由有限维矛盾属性来表达。区别于之前的概念尽量明晰化的处理方式，内涵部分的数学模型延迟展开。采用概念的矛盾属性维表达时，为了高效，要求其概念系统数学模型和处理过程尽量做到：

- 1、列举主要矛盾，矛盾可以是内涵中的矛盾，也可以是外延（和作用域都可以表达成概念的某一属性）中的矛盾，或者是作用域中的矛盾，形成有效特征矛盾向量。
- 2、主要矛盾和特征矛盾向量可以随问题对象和上下文语境的变化而随时调整。
- 3、基于逻辑对每对矛盾做数学建模，并形成可计算的形式表达。
- 4、矛盾数学建模困难，未知或者概括性不佳的，采用矛盾对立面的概率模型来概括表达。即按以上对人的认知过程的分析，尽量少增加属性维数，而采用前面章节所讨论的概率模型抑制未知属性的泛滥，在可用的情况下把魔鬼暂时藏起来。
- 5、确有必要增加属性的，可以增加新的矛盾维属性。
- 6、作为日常决断或高效沟通的辅助时，可以把涉及到的相关多维重要矛盾及时展示并提醒尽量全面思考，甄选出最合适的主要矛盾和矛盾的主要方面。
- 7、在实际计算决策处理过程中，未进展到最终步骤前，尽量不要过早做降维投影。保留计算过程中的多样性和信息的丰富性。直至临近决策点时，才做降维投影，提供最终决策的参考。

属性维数过多的矛盾语义预范畴中，如果概念定义中出现矛盾爆炸场景，过多的矛盾会相互抑制，可以用综合化的局部化转换，重新寻找新的“万花筒”，将它退化为低维矛盾属性，甚至于退化到 0 维的无矛盾子预范畴，对应于“道”“太极”等古老的超越性概念。（这里的局部化降维需要非常谨慎，“无为而治”并非万能手段。）

理想的情况下，矛盾对立统一。但是现在不少学科存在有“不可能三角”，类似五行生克。这种矛盾多维度的组合可采用类似上述形式概念成型的方式，不使用可以暂时不展开，不求甚解。由此得到一个可计算的小型化高效的形式矛盾属性基空间。

在这个可计算语义子预范畴中，概念可以用多维矛盾属性维度来表达。如果从概念（外延）区分角度看，则概念必然存在有限维特征矛盾属性维，日常沟通的时候如果建立了这些有限维特征矛盾维度的共识子预范畴，则概念沟通过程中，必要时冰山逐渐浮出水面，沟通也可以逐步达到足够的深度。

在这个模型中，很重要的一个问题是如何保证上面提到的第 2 个原则的实现。人们在交流过程中，经常由于前面所述的概念牢笼而对主要矛盾的选取产生障碍，并容易受到各种影响被带偏方向而产生误解，而且一旦选定了某些属性作为主要矛盾考察时，会产生一种惰性（牢笼），从而作出不客观不合理的判

断。如果在这方面能有机制优化和辅助人们即时调整对主要矛盾属性的判断与切换，将会大大提升人们的沟通能力，并充分发挥共识的作用，为社会关系的提升甚至人类的和平发展提供帮助。

对于这个特别的可计算模型，这里只给出一个可能的研究和应用方向的提示。该可计算模型的具体技术还涉及到很多细节，包括该方案是不是可能失效等问题的分析，这些需要很多专业的深入研究和实践。

7 二重性的世界

本文进展到这里，已经涉及到了几个方面的二重性。主要有：1，意识与物质的二重性。即物质世界有一个在意识中反映的世界，这种对应性就象是物体与镜子里的影子一样，他们有显然的对应关系，但就是跨越不了镜面。2，拟全语义预范畴与形式语义预范畴的二重性。形式化的语义预范畴显然是对应于它所形式化的拟全语义对象，但这之间有关联，也有区别。拟全的语义预范畴是不可有限表达的，形式化语义预范畴可以看作是它的一个逼近。但这种逼近又是有条件的和局部的，有些情况下是会损失关键信息的。3，赋义的形式系统与符号系统的二重性。形式系统的一个基础是符号系统，但仅有符号系统，形式系统是不能工作的。一定要把符号赋义后，形式系统才能开始“有意义地”工作。而根据符号系统的不同、赋义的不同、赋义的范围与能力的不同，形式系统也有不同的工作能力。我们目前处理的形式系统主要是以图灵机为基础处理能力的系统^[128]，因此，这里的形式化也冠以“可计算”的限制。即，这里考虑的形式系统主要是以机器可以处理的系统为基本出发点。

以上三个方面的二重性，都是智能所需应对的不同层面的重要问题。下面我们将对这三个方面的二重性做些探讨，只是希望能对这些问题在智能和人工智能方面特别是语义处理方面的应用提供一些参考思路。

第一个二重性，前面也提过，是哲学领域由来已久的鸿沟问题。我们不指望在这里能把它理清楚。

第二个二重性与第一个二重性有比较深的关联：拟全语义预范畴似乎只存在于意识中，是一种“说不清楚”的意识对象。但它又可以被形式化的体系有限地近似表达，而且在表达过程中有时还能明确所指对象。即，形式系统的部分符号有时可以被赋义为拟全概念，比如我们的语言体系里的 π 和 ∞ ，这两个符号代表的圆周率和无穷大实际上是拟全概念。

但是这种拟全概念的符号指称是特例，一方面我们不可能把所有拟全概念都指定一个符号（那样符号就是不可数无穷多的），另一方面这种符号指称的概念在实际应用中是无法以自身拟全含义直接参与到运算中去的（即，它们即使参与运算，其运算过程是不直接针对它们所代表的语义内涵的，而只是针对符号或所指称的抽象对象，其最终运算结果的语义仍然需要借助形式逼近和近似）。

这就涉及到第三个二重性问题。可以工作的形式系统是赋义的符号系统，但这个系统的工作能力与符号系统被赋义的情况是密切相关的。而赋义过程又是一个非常复杂的过程，它又涉及到第一个二重性问题。这里不再多加探讨，可以参考前面章节中一些关于人的思维能力的基本假设。

我们经常习惯于用同一种方式思考，而不能明确地意识到对象的多面性。

所以，在对象的二重性不是那么清晰而容易分辨时，常常忽略或混淆它的二重性，试图用同一种方法去处理不同的东西，因此，总处理不好。

以下几节对上述的三方面的二重性做些不是很深入的展开，也算是抛砖引玉吧。

7.1 双二重性

意识与物质的关系，也许就是某种二重性的存在。就象前面讲的物体与镜子里的影子那样，他们有显然的对应关系，但就是跨越不了镜面。虽然从光线的物理来解释可以理解物体与影子的关系，但影子仍然跨不过镜面。物质在人的意识中似乎也有这样的“影子”存在，但这个影子和物质本身之间总也有一道“镜面”无法跨越。也许哪一天，我们真的能找到“意识之光”，以及意识的“镜子”，从此不再受“物质与意识的鸿沟”的困扰？

从前面对语义预范畴的分析和模型中，我们可以得出意识是某种与语义预范畴类似的拟全性的对象。意识在某些情况下是不可形式化的。比如点的概念、直线，比如关于情感的一些概念。它们是不可构建的、不可有限表达的。

那么，物质是拟全的还是形式的呢？从表面看，物质作为一种存在，似乎总是可以构建出来的。构建一定是有限的，一定是有基础的。因此它可能是形式的？

甚至，是不是可以说，所有以实物存在的东西，都是可以形式化的？这么说，物质世界最终应该是可以被掌握（和改造）的，也就是说，自动化世界是可以实现的。这应该是乐观的一方面。

但这里有个容易被忽略的点，我们上面提到实物是可以形式化的，实际上是有前提的：即形式化的基础是以某种形式存在开始的，也就是说，我们其实是在物质世界里“认定”了某些物质的“基础”（比如电子、质子、中子？），才做到形式化的。

事实上，如果从物质的量子场或弦论的方程出发的话，物质（波）其实也是拟全的，它应该跟意识是有类似的性质。

我们为了理解，其实把物质这些方程也形式化了。所以并不是物质本身是形式的，而是我们认识物质的方法是形式化的。我们用了某种形式化方法来表达物质，也是一种近似的表达，是某种逼近。当然，我们目前对物质的理解只到这一步，继续深入下去会是什么结果，还未完全明确。

但根据已掌握的情况看，人为了认识和表达意识和物质，其实都是采用了一种把拟全对象形式化的过程。

所在，从物质、意识，拟全、形式的角度来看，这里有双重的二重性问题（如果我们认可意识与物质是有一定二重性的）。

当然，以上双二重性的认识并不一定就是最终结果，随着科学的不断发展，也许有可能在某个拟全的层面，意识和物质就统一了呢？也许世界是由某种拟全的统一形态给出来的，我们能体验到的物质世界，只是它的某个形式化结果？

另一方面，意识与物质的二重性中也还是有一些明显的缺陷：比如对情感和抽象的部分，有时候是找不到对应的物质体现的，它们似乎并不是物质的“影子”。

物质和意识的问题涉及面太广，这里只是为了说明语义的拟全/形式的二重

性问题，把涉及到的一部分问题纳入考虑对象，并不指望能在这里把这些问题理清。只是在这里借用甚或提出一些问题，更多这方面的讨论，可以参考其它哲学文献，就此打住。

7.2 形式与拟全的转换

上节所提到的语义的拟全/形式二重性，可以在语义处理中有大量的应用，这些我们将在后续逐渐展开。这里要解决二重性应用的前提：要实现语义的拟全/形式二重性的应用，首先要解决二者的对应与转换的问题。我们在前面的语义的形式化部分简单做了从拟全语义预范畴到形式语义预范畴的形式化问题。但当时，形式化的细节还没有介绍，因此，对形式化表达的性质介绍得很粗浅。这里，对形式与拟全的转换问题作一些略深入的探讨。

形式与拟全语义预范畴相互转换的问题，其实是语义分析与表达生成的问题。真正的语义存在于拟全域中，而表达则只能在形式域中进行。于是，语义分析的问题，就成了怎么把在形式域中的形式表达转换为拟全域中的语义子预范畴，从而理想表达所指向的语义的问题；而语义的表达生成问题，就成了怎么把拟全域中的语义子预范畴对应到形式域的形式子预范畴，并用一定的方法形成表达式表达出来。

下面将对这两个问题在有人的意识介入的处理过程分别进行一些探讨。也就是探讨人使用语言进行的交流过程中语义分析和表达生成的主要过程，以供后续的应用（包括没有人的意识参与时的应用）参考。

(1) 从形式到拟全

我们从形式表达开始。假设有某个对象 O 的语义 S ，对它进行了形式化的语义表达，其表达式集为 $(E_f)_s = \{f_1, f_2, \dots, f_n\}$ ，注意与之前的拟全语义预范畴的表达不一样的是，这里的表达式的个数也是有限的。

这里继续语义预范畴性质探讨时的方式，借用代数几何的某些术语（注意其严密性有待继续研究确认）来类比描述。表达式集 $(E_f)_s$ 形成了形式域中的

“理想” $(I_f)_s$ ，对应于形式语义预范畴中的“簇” $(V_f)_s$ ，对应于形式语义子空间 $(X_f)_s$ 和形式语义子预范畴 $(X_f)_s$ 。根据语义的形式化部分的介绍，由于局部化和离散化的过程，它应该是语义 S 对应的拟全语义子预范畴 $(X_c)_s$ 的局部化投影后的“网”状子空间和子预范畴。因此，我们可以用一个内射表示二者的联系：

$$(X_f)_s \hookrightarrow (X_c)_s \quad (24)$$

其中由于局部化丢失的属性维，以及离散化损失的除网点外的其他部分，人可以通过“拟无限”、“拟存在”、“拟遍历”能力自动（通过内插、扩展和转换）

补齐，从而实现拟全语义子预范畴 $(X_c)_s$ 指向语义 S 。

这样，人就完成了从形式到拟全的语义分析过程。

$$(E_f)_s \rightarrow (I_f)_s \rightarrow (V_f)_s \rightarrow (X_f)_s \rightarrow (X_{fp})_s \hookrightarrow (X_c)_s \quad (25)$$

这其中，由于形式化语义子预范畴并没有把全部属性的要求都带进来，人的“拟无限”、“拟存在”、“拟遍历”能力所进行的补齐，对缺失的信息部分有一定的随意性，所以人理解的语义与目标语义之间有可能出现一定的误差甚至错误。而由我们前面所讨论过的语义的稀疏性，大部分情况下，这种语义的分析理解是足够满足应用需求的。于是，我们就有如下结论：

命题 7-1 语义的分析理解，是从表达解析出形式化语义子预范畴，内射至拟全语义子预范畴，并借助人的“拟无限”能力补齐完整语义的过程。

(2) 从拟全到形式

从拟全到形式表达的过程基本上是上述过程的逆过程，但是其中有一些部分会有些需要关注的不一样的关注点。

有前面语义的形式化有关讨论的很多结论可以直接使用，这里的探讨就可以省却很多。我们直接从语义 S 和指向它的拟全语义子预范畴 $(X_c)_s$ 开始。按上述讨论的逆过程来说，我们应该是要找到一个从形式到拟全的内射的逆映射，即从拟全语义子预范畴往形式语义预范畴的投射，以获得一个形式语义子预范畴。

但是，我们马上就会碰到一个问题：往哪个形式语义预范畴投射？因为语义的形式化过程中有局部化的过程，而不同条件下的局部化像万花筒一样，可以产生很多不同的形式语义预范畴。

这些不同条件的局部化主要是看语义 S 要往哪些重要属性上去表达，能更加有效。这跟交流过程的上下文，以及表达的目的意向有关，不是由形式化本身决定的。比如在讨论某种专业的问题时，局部化可能主要依据话题在专业领域的属性而定。

除了局部化外，离散化条件其实也是形式化所要考虑的条件，但一般离散化条件相对简单些，能满足所需要的近似精度就可以。由于人对事物的精度的掌握并不是很高，离散化过程一般机器可以做得比人更精确，采样精度更高。由于这方面有专门的学科深入的研究，这里就不多讨论了。

假设我们最终选定了局部化的条件 p ，在 p 条件下我们得到了拟全语义预范畴 X_c 的形式化语义预范畴 X_{fp} 。于是对语义 S 的形式化就是把拟全语义子预范畴 $(X_c)_s$ 投射到形式语义预范畴 X_{fp} 上形成形式语义子预范畴 $(X_{fp})_s$ 。

$$(X_c)_s \Rightarrow (X_{fp})_s \quad (26)$$

有了这个形式语义子预范畴 $(X_{fp})_s$ 之后，就得到它的语义子空间 $(X_{fp})_s$ ，

及其对应的代数集或簇 $(V_{fp})_s$ ，由此可以得到其对应的表达集的“理想”

$(I_{fp})_s = \{g_{p1}, g_{p2}, \dots, g_{pm}\}$ ，而它的表达式集 $(E_{fp})_s = \{f_{p1}, f_{p2}, \dots, f_{pn}\}$ 中的表达式 f_{pi} 就可以由 g_{pj} 根据一定的语法规则组合而成。即：

$$(X_c)_s \Rightarrow (X_{fp})_s \rightarrow (X_{fp})_s \rightarrow (V_{fp})_s \rightarrow (I_{fp})_s \rightarrow (E_{fp})_s \quad (27)$$

注意，这里表达的方式有一定的自由度，只要最终表达式集的“理想”依然是 $(I_{fp})_s$ 就基本能表达出语义 S 。如果在组合的过程运用不好，或者前面选择局部化条件时没有选好，就会出现表达不到位的情况，产生语义表达的偏差、歧义、模糊等各种表达问题。

换一个角度看的话，语义 S 的形式表达也可以看作是用一组形式表达式把形式语义子预范畴 $(X_{fp})_s$ 中的属性取值范围和映射关系描绘到可以与其它语义充分区分开（由于之前提到的稀疏性，这并不十分困难），并且能表明其本身重要的内容（关注的主要属性及其取值范围）的程度。

这里要补充的是，形式表达最终的形态是符号表达式，当然这些符号是被赋过义的（符号的赋义将在后面讨论），并且这些符号的组织方式有一定的规则，即语法。于是，我们有如下的结论：

命题 7- 2 表达是一个赋义的符号系统，按一定的语法，实现对一个形式语义子预范畴的表示。

注意这里有一点很有意思：我们在形式到拟全的内射时只有一种途径，但从拟全往形式投射并生成表达时，由于局部化的不确定性和表达由表达集“理想”元素组合的可变性，可以有很多途径。这也是我们有时候会觉得理解一个表达的语义往往只有唯一的结果（包括歧义、暗喻等部分的总体结果），而对一个语义（想法）进行表达的时候却有很多表达方法的主要原因。

7.3 形式系统能力与界限

进展到这里，我们可以试着回答一下之前提出过的问题：形式化系统能完成拟全语义处理的任务吗？

我们不能直接回答这个问题，但可以按如下的思路尽量把它分析清楚：

首先，一个形式系统能全面代替拟全系统吗？

这个问题，经过本章和上一章的详细讨论，已经比较清晰：理论上形式化语义预范畴能以局部化和离散化的方式，实现对拟全语义预范畴的网状覆盖。于是，有两个问题：

问题一、网状覆盖的空隙会怎么样？在一般情况下，网状覆盖的空隙部分，人是可以通过“拟存在”、“拟遍历”的方式自动补齐，形式化系统可以用内插的方式来模拟补齐。因此，当局部化比较平缓，没有奇点时，只要网格足够细密，这种网状覆盖可以无限逼近这部分（局部化的）拟全语义子预范畴。

因此，在很多局部化条件下，可以认为形式语义子预范畴与拟全语义的局部语义子预范畴性质相当，可以在限定条件下把形式语义子预范畴当作局部的拟全语义子预范畴来处理。即，这种情况下，形式系统可以完成（局部）拟全语义处理的任务。

如果空隙中存在奇点，那么这种“补齐”在当前的局部化条件下，形式化系统内插的模拟就会出现很大偏差。

问题二、局部化覆盖要全面覆盖拟全语义预范畴时需要无限的局部化语义子预范畴。而形式系统实际能提供的不管是多少，也只能是有限个。

由以上两个问题，我们只能得出形式系统（除了在某些情况下能完成局部拟全语义处理任务外）是不能直接代替拟全系统的结论。

其次，形式系统不能直接代替拟全系统的情况下，我们还有办法用形式系统解决拟全域中的问题吗？

我们从局部化的处理方法，可以看到这个问题的希望：前述的网状空隙中的奇点虽然是某个局部化条件下的局部化形式系统无法解决的，但我们可以采用其它的局部化方案。人是可以在拟全域中感知到这样的奇点，参与到局部化选择和属性转换的任务中，挑选合理的局部化方案，再放回形式系统中进行局部化求解。这时，系统又返回到前面所提的问题一中可以解决的一般情况。

从以上的分析，我们可以得出的一个比较确认的结论是：形式系统不能代替拟全系统直接解决拟全语义处理的所有任务，但结合人对拟全域问题的理解和转换能力，在二重性的配合下，经过足够生长（保证网格的密度以及尽量多的局部化方案）的形式系统可以足够精确地解决大部分我们所碰到的问题。

有些拟全域中很困难的问题，我们仍然得耐心等待人类的智慧闪光。

7.4 拟全域与形式域的混淆

拟全与形式的二重性最常见的问题是混淆。首先这里要对“形式”与“拟全”的主要（本质）区别作一些说明，因为太容易混淆了：

“形式”并不是指符号不能有拟全的赋义，而是指不能用形式体系形成拟全表达。用某些符号直接赋义指称拟全概念（比如用 π 表示圆周率），只是一些形式域中的赋义特例，不是形式系统就此获得了拟全表达能力。这里的 π 作为一个运算因子参与形式运算，但是：1，它在参与形式运算时只是一个符号；2，如果要对 π 进行解释或表达，依然是无法在有限表达的范围内做到的；3，还有很多的超越数无法逐个用符号表示。因此，带有 π 的形式系统依然是形式系统，并不因为 π 所代表的圆周率的拟全性而改变形式系统的性质。

拟全与形式的二重性主要体现在二者的相互促进：我们往往拟全地思维，形式地解决。思维要能理解并解决那些拟全的问题，而要付诸行动，必须要有有限步骤的形式方法。只有二者相互结合，才能高效地解决问题。这是我们这里提出拟全与形式二重性的出发点，也将是后续二重性方法的重点。

这里会延伸到一个问题：形式域中还有思维吗？有智能吗？还是仅仅是形式的公式、逻辑、规则，它们是机械可重复的东西，还能算是智能吗？真正的思维只能存在于拟全域吗？这些问题回答起来是有些困难的。也许我们只能说：形式域中能解决一些可以机械化的智能问题。比如在机器处理问题的过程中，最重要的是判断，判断在一定的电磁或机械的条件下，是一个达到条件就执行的过程，这似乎不算智能？但是它确实又帮助人完成了判断，而判断在我们的

认识中，应该算是人的思维和智能范围的。

在拟全和形式方面，经常容易引起混淆的一个问题是：我们经常试图用形式域的直观去解决拟全域的问题。还是举几个例子吧：

例 1：拟全语义下的很多认识和属性与形式语义下的很多认识和属性是完全不一样的。比如对“多少”的认识，在形式语义下，由于可以通过个数统计得到具体的数量，所有多少是有明确的个体计数的；而在拟全语义下，无法做个数计数与统计，所以多少就失去了在形式语义下个数统计的意义，只能用另外的方法去定义或说明。

这对于类似“自然数与其中的偶数的数量哪个多”这样的问题，应该能有所启发。这个问题，从两个角度似乎有截然不同的答案：直观上，按 1、2、3、4……和 2、4……去做计数统计，自然数似乎显然是偶数的 2 倍；而用映射的判断方式，偶数和自然数又能做到一一对应，所以应该是“一样多”。

这里面看似混乱的结果，所出现的问题的核心在于：我们在试图用在形式域的“多少”的概念，去解释一个无限（拟全）域的问题。而我们前面讲到，这两个域的差异，很多情况下不能允许我们用同一个工具在两个域中同时工作。

（这应该也是哥德尔不拟全性定理的一个例证。这个命题在这里不能判定真伪？^[120]）

而计数统计和一一对应，都是形式域中判断多少的工具。

如果要做好上面这样的问题，我们需要在无限域（拟全域），重新定义“多少”的概念和处理方法。（比如用集合论的包含关系等）

例 2：很多语义方面的混乱感，实际上是因为我们用了同样的符号来表达同一个物体在不同域（或不同局部化）的概念或属性。比如“白马非马”中的两个“马”，虽然文字符号相同，实际上映射到两个不同的域中：一个是映射到（局部化）的实际的物体（某一匹具体的马，白色的），另一个映射到的是拟全域中的概念。

由于这种表述太多，而且（为了表达简洁）无法更多地提示域的不同，所以很多时候很容易混淆。（这也有点，在两个域中跳来跳去地对应映射，中间却缺一个“桥”的感觉，应该也是有“鸿沟”的感觉的原因之一）

（当然，对“白马非马”本身所指，也可以有不少不同的语义解释，因为这里规定的前提条件或属性限制并不严格，有很多歧义。这方面就不再展开了）

例 3：很多关于递归的悖论也是这样产生的。比如，我们要是问“概念和全谁包含谁？”这该如何解答？既然有“全”，那么“概念”应该是包含在“全”里才是；但是，“全”也是一个“概念”啊，那么，“全”应该包含在“概念”里，这怎么处理？

实际上，“全”和“概念”都是拟全域里的概念，因此，我们不能用形式域中“如果 A 是 B，那么 A 属于 B”的逻辑。

从以上这些例子来看。拟全和形式之间我们是经常混淆的。怎么处理比较好呢？这可能会涉及到一整套的完整体系的重新整理。也许我们应该把数学和逻辑，拆分为拟全数学、拟全逻辑、形式数学、形式逻辑？并分别重新进行相应的基本概念和方法的定义和研究？

这个课题太大了，这里只能提出个思路。具体要做这样的研究和整理，那会是一个非常庞大的工作，不是这里能马上解决的问题。也许需要一个专门的学科去研究。

但我们至少在应用中要意识到这两个域容易混淆，当碰到问题时，需要仔细分辨一下问题的域究竟是在哪里，然后针对性地在适当的域中使用合适的工具解决，而不至于把自己绕进一个混乱的世界里。

特别是我们后面要强调的，有很多问题，实际上是需要拟全域和形式域同时努力配合，才能真正有效解决问题。

7.5 形式系统与符号系统

形式系统与符号系统的二重性问题，相比上面的问题就没那么复杂了。要点是：形式系统是赋义的符号系统。形式系统需要“赋义”。赋义的符号系统，才是可用的形式系统。这里的“赋义”，指的两个方面：一是符号本身被赋义，二是符号之间有一些运算规则。

这里是不是看到了一个类似语义预范畴的系统？概念与概念之间的映射关系，共同组成了语义预范畴。把符号赋义为概念，运算规则赋义为映射关系，一个有限的符号系统就转化为一个形式语义系统。

这里以“马”为例简单做个符号赋义的解释。赋义前，“马”作为一个图形，被视作图形符号时，本来是可以代表任何概念或对象的，或者不代表作为图形以外的任何其它意义。而马的语义是一个拟全概念，它的语义子预范畴记为 $(C_{\text{马}})_c$ 。在一定的语境（比如汉字系统）下，这个概念只能用形式化方法来表

达，这个形式化的语义子预范畴我们记（表达）为 $(C_{\text{马}})_f$ （注意，这里 $(C_{\text{马}})_f$ 也是个形式符号，这是这篇文章这个形式化表达方式所不可避免的“嵌套”式问题：我们只能用某个形式符号来表达某个语义，只能请读者习惯了）。显然，这个记法太过复杂也不方便记忆，因此，我们借用图形符号“马”在汉字这个形式系统中表示 $(C_{\text{马}})_f$ 这个形式语义子预范畴，即把形式语义子预范畴 $(C_{\text{马}})_f$

所指向的拟全概念 $(C_{\text{马}})_c$ 赋给“马”这个符号。或者说，在汉字系统中，形式表达用的汉字“马”（已经不是前述的图形式符号“马”）对应的形式语义子预范畴为 $(C_{\text{马}})_f$ 有符号属性“马”（图形符号）和语义属性 $(C_{\text{马}})_c$ ，而马这个概念的拟全语义子预范畴 $(C_{\text{马}})_c$ 在汉字系统里的形式化投影 $(C_{\text{马}})_f$ 就此用“马”（汉字形式系统中的汉字）这个汉字来表达。

由此，我们可以简单讨论一下关于典型的由符号组成的语言系统的一些粗浅结论：

如果从符号的有限性和体系性来看，所有的语言系统都是形式系统，不管是自然语言还是形式语言。只是自然语言因为长时间的演化，为了丰富其表达能力，而变得非常复杂，一些规则看起来不是那么规律了。

而我们所说的“形式语言”系统，实际上只是在歧义性处理方面和可计算性方面，比自然语言有更好的应对能力。

语言的本质是用有限的赋义符号尽可能广泛地表达语义，并与一定的主体（人或机器）结合，辅助语义的（逻辑、运算等）处理。

从形式系统与符号系统的关系来看，形式化过程中的“赋义”是很重要的一个步骤，这样就可以把一个符号系统与一个形式系统关联起来。

而我们前面对语义预范畴进行形式化的过程中，还有一个要求，就是可计算性。即形式化后的系统，要能让机器进行相应的计算。

可计算形式系统的形式计算过程，实际上是针对符号的抽象指称，而不是针对符号的语义内涵。我们给出计算表达式 $c = a * b$ 时，并不管它是代表 a 笼 b 只小鸟，还是 a 袋 b 颗石头。这是形式系统中形式规则的基本要求。

从这个角度来看，形式化的语义预范畴，并不是“纯粹”形式的。而是一种混合式的。我们允许很多概念的属性通过赋义直接带进形式语义系统中。

所以，要注意，我们的可计算形式化语义预范畴有两个层面的赋义：一个层面是形式化赋义，即对符号赋以抽象指称和规则的赋义；另一个层面是语义层面的赋义，即对符号赋以属性、概念以及映射关系的赋义。根据上一节的讨论，后面这种赋义甚至可以是拟全概念的赋义（当然，这种赋义的拟全概念语义结果仍然是通过人的“拟无限”、“拟存在”、“拟遍历”的能力对形式对象补充而成的，比如用 π 来表示圆周率），但不改变整个系统的形式性。

由于上述的两个层面的赋义的存在，更容易引发一些二重性的混淆。这点在实际的应用中要特别多加注意。

7.6 图灵停机问题

我们再试着用本文的视角，重新审视一下图灵停机问题，就会发现，问题其实很简单，很清晰：这就是一个典型的拟全和形式系统混淆的问题。

判断一段程序是不是一个死循环，我们可以在它工作的形式系统外，用非形式化（拟全化）的方法给出判据。然后把这个判据形式化，在这段程序工作的形式系统之外，用于对这段程序是否会死循环的判定。

而在这个程序自己工作的形式系统中，它自己是无法判定自己是否是死循环的，因此它根本无法给出让自己停机的命令。所以，在形式化系统内部，是不存在这样的悖论的。而是直接的：做不到。

即，我们只能借用拟全化手段得出的结论再形式化后给出的“监控”程序，来分析一段程序会不会死循环，并在这段程序执行的形式系统外部发出强制命令，让其停机。如果不存在这样的外部强制机制的后门（比如看门狗程序），则只能采取关闭系统的手段。

或者说，有些形式问题，我们只能把它先拟全化，在拟全域用拟全手段得出结论，再应用回形式域中解决问题。（后面的平行双镜是另外一个例子）

这个问题其实与哥德尔的问题是类似的：一个形式系统是不拟全的，它不能得出拟全性问题的结论，并指导自己改变它。

看着是不是很眼熟？我们能穿越时空改变我们自己吗？是不是挺有意思？

“我”是拟全的还是形式的？答案是：“我”这个概念是拟全的，但在这个世界里真实生活的我这个人（在一定程度上）是形式的。所以，“我”能做到的很多事情，我是做不到的。（比如我写了一本书，“我”的思想可以影响未来很多人，但我是做不到这点的？）（能用数学证明吗？我们的行动能力其实与机器差不多，都是有限的、形式的？虽然我们有“拟无限”的思维能力，能知道某些东西的存在，但并不真正掌握/了解它，后者还是需要实际“激活”相应语义子预范畴才能做到？）

8 语言的语义处理

本章虽然主要针对自然语言^{[19][20]}，但并不限于自然语言的领域。这样做的理由是：

从根本上而言，自然语言其实也是一种形式语言，只是由于使用的时间长了，它的语法和表达方式变得相当的复杂和不规范，所以变得“非形式”了。

从自然语言的一个基本点出发，其实可以看出它也是形式的：就字词的数量是有限的这一条，其实已经能基本说清楚了。

从我们之前对“形式”的解释，也可以得到类似的结论：用有限的符号在有限的（时间）范围内，能表达的内容必然是有限的。除非被直接赋义为拟全概念或映射关系，自然语言并不能把拟全概念有限地表达出来，因此，它本质上是形式的。

自然语言（的符号）本身并不承载拟全性，拟全性是（被赋义的）语言被人理解后的语义承载的。

因此，除了某些特殊针对性的问题，这里不再专门区别自然语言和形式语言。（甚至，我们可以认为所有有表达能力的表达方式，都是某种语言。比如绘画、音乐、“肢体语言”、“表情”、“语气”等等）

语言处理的主要目的，是实现交流和沟通。交流和沟通从表面看就是接受语言表达，理解语义，形成新的意见语义，生成语言表达，最终形成共识的过程。当参与交流的主体所使用的语言不一样的时候，就需要有翻译过程参与。广义来说，每个人都有一套或几套语言系统，这几套系统之间，不断地有翻译过程参与。机器和人之间的交流过程也是类似的，只是目前给机器用的语言主要是形式语言。

在这种交流和沟通中，最大的困难还是明知有拟全语义预范畴存在，有时却无法表达的问题。任何形式化的描述，都免不了是“管中窥豹”。在前面我们讨论过，人正是在一种拟全和形式的混合的二重性下掌握和运用语言的。那么，对机器而言，又需要怎样的一些条件和方法去运用语言，实现或辅助“沟通”呢？

本章将重点对语言语义的典型应用、语义运用条件和数学基础、语言语义分析和理解、语义的语言表达生成等方面做一些探讨。

这里并不指望全面解决这些问题，只是在拟全/形式二重性模型框架下，研究语言的语义处理的一般原则，希望能给相关的应用带来一些启发。

8.1 典型应用：翻译与交流

前面已经提到，交流沟通和翻译是语义处理最主要的目的，也是最常见的应用，这里就先从这两个典型的应用入手，对它们的流程做一些分析和模型化。

这里先简单做一个翻译过程的解释模型。需要翻译的任何一段表达，本质上都是在语义预范畴中形成一个语义子预范畴。如果用任何另外的表达方式（不管是拟全的还是形式的）把它足够精度地表达出来，实际上就是完成了翻译。因此，翻译的过程如下：

1、在一个形式预范畴（被翻译表达的形式语言语义预范畴）中，按属性和映射关系的表述（比如一个遮挡太阳的图片或其它示意、描述），找到的概念及其映射关系（这时应该是形式的），内射到拟全范畴中的一个子预范畴，这一过程中的误差可控（类似于数学分析中，差异只是高阶小项？）；

2、对上述的拟全语义预范畴中的子预范畴，再进行往另一个语言的形式语义预范畴（目标语言）的投射，在一定规则下，用形式属性及其映射关系（要注意的是，这里用到的属性及其映射关系，可能已经和上面所提到的一开始的输入完全不一样了。比如关于遮挡太阳的可实现方案描述），来逼近拟全概念及其映射关系。

换句话说，这里对拟全语义预范畴和形式语义预范畴的分析，需要能完成：形式化语义子预范畴所指向的概念，有方法可以（在一定误差范围内）对应于拟全语义子预范畴；而拟全语义子预范畴（在一定误差范围内）可以有很多的形式化表达，可以根据要求和规则，挑出我们需要的形式化语义预范畴的表达。如果用几何的直观来解释翻译过程的话，可以简单描述为：先根据一种语言表达描出一个形式语义子预范畴几何形状内射在拟全语义预范畴中的子空间，再把这个子空间投射到另一个形式语义子预范畴几何形状用另一种语言表达出来。

借用前面形式与拟全转换的数学化过程命题 3- 1、推论 5- 1、命题 7- 1、命题 7- 2，在人工参与的情况下，对某个语义 S 从语言系统 A 的某个局部化条件往语言系统 B 的某个局部化条件翻译的过程可以由下式表示：

$$\begin{aligned} (E_{fA})_s &\rightarrow (I_{fA})_s \rightarrow (V_{fA})_s \rightarrow (X_{fA})_s \rightarrow \\ (X_{fA})_s &\hookrightarrow (X_c)_s \Rightarrow (X_{fB})_s \rightarrow \\ (X_{fB})_s &\rightarrow (V_{fB})_s \rightarrow (I_{fB})_s \rightarrow (E_{fB})_s \end{aligned} \quad (28)$$

弄清了这个过程，并对机器翻译的困难加以分析，再借助一些形式语义预范畴的工具解决，我们就有可能实现在人工智能领域，完成对输入表达的语义理解（由形式化找到拟全化的近似），以及在理解的基础上进行对话输出（由拟全化得到足够精度逼近的形式表达）。即一个完整的翻译过程。

在以上翻译过程模型的基础上，我们进一步给出一个有人参与的交流过程模型。同样借用前面形式与拟全转换的数学化过程，在人工参与的情况下，对某个语义 S 从第一个主体的某个局部化条件下语言系统 A 生成表达，往第二个主体某个局部化条件下语言系统 B 发起交流，系统 B 收到系统 A 的表达后，解析其语义得到语义 S' （交流过程中收到的语义与发起者语义会有些不一样），并形成响应语义 R ，生成表达（这时系统 B 的局部化可能会发生一些变化，比如换了个专业背景来解答等），再发给系统 A（这里系统 A 也可能出现局部化的变化），一个完整的交流过程可以由下式表示：

$$\begin{aligned} (X_c)_s &\Rightarrow (X_{fA})_s \rightarrow (X_{fA})_s \rightarrow (V_{fA})_s \rightarrow (I_{fA})_s \rightarrow (E_{fA})_s \rightarrow \times \\ (I_{fB})_{s'} &\rightarrow (V_{fB})_{s'} \rightarrow (X_{fB})_{s'} \rightarrow \\ (X_{fB})_{s'} &\hookrightarrow (X_c)_{s'} \xrightarrow{\text{交流响应}} (X_c)_R \Rightarrow (X_{fB'})_R \rightarrow \\ (X_{fB'})_R &\rightarrow (V_{fB'})_R \rightarrow (I_{fB'})_R \rightarrow (E_{fB'})_R \rightarrow \times \\ (I_{fA'})_{R'} &\rightarrow (V_{fA'})_{R'} \rightarrow (X_{fA'})_{R'} \rightarrow \end{aligned} \quad (29)$$

两个主体之间实际发生交换的是语言的（形式）表达 E_f 。主体间的交换界

面是上述流程中的×所标记的地方。

主体内部根据输入表达所理解的语义，产生交流响应的语义是交流沟通发生的重要环节，这个环节是在主体内部进行的。

在以上的交流沟通的流程模型基础上，我们可以进一步分析无人参与的形式化系统交流的主要困难点，并找出形式化系统自动交流沟通（或人机系统自动化）的努力方向。

本节主要给出这两个典型应用的一般模型，更进一步的分析将放在后面。

8.2 语境与局部化

语义预范畴的模型在之前已经给出。但拟全语义预范畴是抽象的，它没有一个精确的形式表达。这样就给语义处理带来很多困难。

为了能够交流，以及让机器处理语义系统。我们必须使用形式系统来有条件地表达语义，接受形式系统表达的片面性和不精确性。

对人而言，因为人可以理解拟全语义，虽然还是存在歧义、片面、模糊、误解等困难，这些问题还相对容易些。形式系统或机器无法理解拟全语义，因此会有更多的困难和限制。为了解决这些问题，我们需要在一些具体应用中，使用前面所讲的二重性方法，比如充分利用各种资源，找到好的局部化和离散化方案。

本节主要分析在一般的语义处理应用中，形式化语义预范畴局部化的选择依据。

局部化首先发生在形式化语义预范畴构建过程，这时需要挑选一些合适的属性维作为形式语义预范畴的基空间。可计算的公共语义预范畴的构建之前介绍过。注意它是一个动态的，类似万花筒一样的模型。对于同一个语义子预范畴，根据我们关注的属性重点的不一样，它的形式表达可以有很多版本。

比如，语义要求“遮挡太阳”时，有很多方式都属于对这个语义要求的响应：用手遮、撑个伞、做个棚子、建所房子。实际上，这时候的语义与上下文情境涉及的语义子预范畴和重点属性的关系很大。当然一方面这里语义要求最终是行动，并不简单是语义的理解和表达；另一方面，这里其实是一个局部化的子范畴内的重点属性的问题，不是整个语义预范畴的重点属性的层面。但道理基本是相通的。

一般情况下，在我们关注的重点方向上，表达能够达到一定的逼近精度，就是一个好的形式表达了。

上面这段就提到了我们这里要考察的另一个重点：语义的应用情境。即语境或话题相关的上下文。语义总是在某种语境下进行的，比如涉及某个话题、某个专业等等。在抒情时讲桃花和做植物研究时讲桃花完全是两回事，它们所涉及到的属性维差异很大。

在语义具体应用时，同时关注全部属性不可能，为了提高处理速度，往往是关注的重点属性越少越好。但是由于有歧义、多义、隐喻等困难的存在，为了充分理解语义，又必须保持对足够多的属性的关注，因此怎么进行权衡和局部化就很关键。

注意这里的局部化和建立形式语义预范畴时的属性挑选的局部化是不一样的，相对来说，与语境相关的局部化是一种动态的局部化。这种局部化就是根据语境、话题选择关注的属性，形成当前使用的局部化语义预范畴。

当然，这里的局部化与可计算语义预范畴构建方案也有关。如果语义预范畴的构建方案是“大而全”的方案，那么，在确定语境和话题后，可以从“全”的属性中挑选出一部分来做成局部化；如果构建方案是“专业化”方案，那就在语境和话题确定后装载相应的局部化语义预范畴，与基础公共语义预范畴相结合完成局部化。

一般情况下，在没有事先明确主题时，对话或交流的开始往往使用语义公共预范畴和日常语义预范畴。而随着上下文的进展，会逐渐明确话题，并逐渐建立起语境，于是，系统可以根据对话题和语境的分析，得出当前对话涉及到的专业、领域、方向，找到重点关注的相关属性，以此选择合适的局部化方案，并使用上述方法挑选或动态加载局部化语义预范畴。

由于语境、话题通常也是动态的，因此，要适时地了解所需的合适的局部化方案，除了要有对语境和话题的分析和找到对应的局部化方案的能力外，话题切换的捕捉和识别也非常重要。

对于具体语境和话题的上下文建立和分析，涉及到很多具体问题和方法，比如，语境的建立需要感知哪些条件？它本身有哪些参数？如何根据这些参数选择装载不同的局部化语义预范畴？有隐喻时，是不是应该同时处理多个局部化问题？这些在这里不再针对各种具体情况逐一深入分析，已有的语言处理研究已经有不少的成果可以参考^{[19][20]}，对语义具体应用感兴趣的也可以做更加广泛深入的研究。

8.3 语义处理的理想条件

我们先考察一下一个理想化的形式系统最好能达到什么样的条件，可以最方便地实现语义的处理。

这里当然以所关注的最典型的语言的语义处理系统——翻译系统——来考察这样的理想条件。标准的翻译系统是由两个可以互译的形式系统和互译系统（翻译机？）构成。广义的情况下，我们可以认为任何人的语言体系都是一个独立的形式语言系统，因此，任何交流，不管是使用哪种语种，从广义来说都是一个翻译过程。因此，我们这里举翻译系统为例并不失代表性和一般性。

(1) 理想条件

有人参与的翻译过程模型在前面章节已经介绍过。对那里的翻译过程进行深入分析，就可以很清楚地发现机器翻译产生困难的重点：由于机器翻译的过程没有人工参与，机器不可能产生拟全语义子预范畴 $(X_c)_s$ ，所以上面的数学表达式中中间的那行中的内射和投射就无法实现了。

在人不参与的情况下，如果没有可计算语义公共预范畴，还是要把翻译进行下去，那就只能创造条件直接由局部化的语言系统 A 的可计算形式语义子预范畴 $(X_{fA})_s$ 映射到局部化的语言系统 B 的可计算形式语义子预范畴 $(X_{fB})_s$ 了。

于是，没有人工参与的翻译过程，就变成了：

$$(E_{fA})_s \rightarrow (I_{fA})_s \rightarrow (V_{fA})_s \rightarrow (X_{fA})_s \rightarrow$$

$$(X_{fA})_s \Rightarrow (X_{fB})_s \rightarrow (X_{fB})_s \rightarrow (V_{fB})_s \rightarrow (I_{fB})_s \rightarrow (E_{fB})_s \quad (30)$$

下面我们根据上述的翻译过程，来探讨理想化的翻译条件。

可计算形式语义预范畴，不能指望只有一个标准。因为它本身是个近似，所以理论上可以有很多的近似方法。再加上前面提到的拟全化细节有可能在某些局部化模式下成为重要因素。我们关注的重点属性不同时，其不同的近似方法差异可能非常大。因此可计算语义预范畴可能有多种不同的建构方式。

比如可以用两个不同的自然语言语种搭建两个不同的形式语义预范畴。这两个形式语义预范畴之间，同一个语义对应的形式语义子预范畴形状可能并不相同，因为它们采用的基础的概念和属性可能有一定的差异，这样，它们的“坐标系”就有区别，因此同一个语义在两个形式语义预范畴中的空间形状就会不一样。不同语义之间的映射关系也同样会存在一些差异。但这些并不妨碍在这两个可计算语义预范畴中分别有形式语义子预范畴指向同一个拟全语义子预范畴。或者说，这两个可计算语义预范畴之间应该有某种“同指”的关系。这种同指也是语言之间可以相互翻译的基础，如果没有这样的基础，即两个可计算语义预范畴之间的语义子预范畴根本找不到对应，那么翻译也就无从着手了。

但这种语义的同指之间，有没有人参与，就会产生一个很重要的区别：有人参与时，由于有拟全语义子预范畴 $(X_c)_s$ 的中间转换，同指系统通过内射和投射过程，相对容易地保持语义的同指；而没有人参与时，由于机器无法产生拟全语义子预范畴 $(X_c)_s$ ，需要从 $(X_{fA})_s$ 到 $(X_{fB})_s$ 的直接映射，如果两个语言系统A和B的局部化条件相差很大，两者的属性维之间没有直接的映射关系时，这种同指映射很可能就会效果非常差。因此，如果要机器进行有效的翻译，最好要建立两个局部化语言系统A和B之间尽量同构的条件，而不仅仅是同指条件，从而使从 $(X_{fA})_s$ 到 $(X_{fB})_s$ 的直接映射能够进行。

根据前面对拟全语义预范畴和形式语义预范畴的关系的分析，要研究理想化形式系统，第一个需要理想化的假设条件是：

理想化条件一：我们要基于共同关注的基础属性和映射关系空间来研究两个形式系统，否则就会出现关注点不同而找不到公共预范畴的困难。

从数学的角度看，实际上就是从拟全语义预范畴往不同的可计算的形式语义预范畴投影的目标子空间要相同，否则，在两个形式语义预范畴中的投影形状和信息会失真，部分关注的语义信息将会丢失，从而失去在两个形式系统间的可翻译性。

有了以上的条件，两个形式系统之间就总能找到对应的子空间，用于在一定精度下近似表达相同的拟全语义子预范畴（中所关注的属性空间）。

即我们可以做到两个对应的形式子预范畴内射到同一个拟全子预范畴上，而同一个拟全子预范畴再投射到两个形式系统时，还是对应于这两个对应的形式子预范畴。

接下来我们就讨论可以实现理想的自动翻译的条件：

理想化条件二：如果两个形式语义预范畴系统之间的关系满足范畴和函子

的模型，那么自动翻译就是一个简单的函子映射。

这将是一个非常理想化的可翻译形式系统，基本上是一一对应形式的直译系统。这个要求是很高的。

首先是两个形式化系统分别要是范畴性的，即属性空间是集合空间，属性间的关系要满足态射关系。其次，两个形式化系统之间要满足函子关系，即两个形式化系统之间的属性及其关系的函子变换保持其原有态射关系。

换个说法，就是不同的语言，构成了同构的形式语义范畴。翻译就是不同的同构形式语义范畴之间的函子性操作：概念对应概念，态射对应态射。

这样，翻译系统就是对应的符号、属性、态射之间的一个函子变换。

在本文即将完成时，找到一篇文献报道了采用这种条件下的形式语言建模的尝试^[129]，并且实际应用于软件开发，似乎验证了这种思路的可行性。

(2) 非理想条件

很遗憾的是，除了计算机语言外，实际的语言系统往往都达不到以上两个理想条件。比较好一些的例子比如一些协议或标准^{[130] [131] [132] [133] [134] [135] [136]}，而

一般的语言都离上述的理想条件比较远。特别是自然语言系统，在人类的丰富想象力和拟全语义思维能力下，被“玩坏了”，各种“脑洞”使得被关注的重点属性维的构成随时在变化，有太多的态射规则外的例外，甚至都不能被当作是一个真正意义的形式系统了。

前面讲到，翻译只是语义处理的典型例子，这里对理想条件和非理想条件的讨论，同样适用于一般语义处理的应用。关于一般的语言语义处理条件的非理想化问题，我们下面就最具代表性的自然语言来分析一下，它是一个怎样被“玩坏的”形式系统，有哪些方向可以弥补这样的问题。

自然语言对前述的两个语言（翻译中）语义处理的理想化条件的违背是显而易见的，这里针对自然语言的两类非理想化问题分别做一些探讨。

第一类非理想化条件问题实际上是基础属性空间的不确定性和动态性。往往交流中关注的重点属性有很大的随意性和即时性。这类非理想条件是不可避免的。要解决这类问题，可以采取数学中类似流形处理或概形处理或局部化环的“局部化”的方法，往三个方向努力：一是在某些特定情境下，对重点属性做一些限制；二是根据实际需要，建立一套甚至几套具有普遍性和代表性的基础公共语义预范畴；三是建立一个包罗万象的，高冗余度的通用公共语义预范畴，把各种可能需要的属性都扩展开，但这个语义预范畴将会是异常庞大的，构建和维护将会非常耗时耗力。（参考本文“可计算语义预范畴的生长”部分）

具体的局部化方法可以参考语义的形式化相关的章节。

我们在之前所做的构建可计算形式语义预范畴的探索，实际上就是在实践第二个和第三个努力方向。实际的自然语言的语义分析还是要以一个庞大的可计算语义预范畴作为基础，或者，根据常用的不同情境，可以按不同的重点属性方向构建几个相互补充的可计算基础形式语义公共范畴，比如按不同专业建立相应的公共形式语义预范畴。在实际计算使用过程中，可以根据上下文情境选择不同或同时处理几个可计算语义预范畴，或者在一个比较大而全面的语义预范畴的多个子预范畴中同时工作。

所以，真正可以应用的有意义的自然语言的语义分析系统，需要有可计算的形式语义预范畴成形后才能展开。构建可计算的语义预范畴可能需要的大量人工参与的工作，似乎会是一个阻碍应用的困难问题。

在没有这样的可计算语义预范畴之前，所有其他的努力，包括语义分析的算法设计，都只能是设想，没有办法去实践和验证。

这个困难可能需要组织社会力量或者从国家甚至国际层面立一个大项目去解决。当然，要完成一个小型的可以开始的计算形式语义预范畴，可能工作量也是可以接受的，比如前面所提到的按《新华字典》这样先建立一个最通用化的可计算公共基础形式语义预范畴。然后在此基础上不断生长。

当然，相关的研究也不能静等。哪怕是设想，我们也要看看，如果可计算的语义预范畴已经存在了，我们需要如何设计语义分析的算法。而且，我们也需要这样的一些研究，去指导可计算语义预范畴需要提供哪些关键的要素，才能使得自然语言的语义分析能够得以顺利进行？

本章后面的分析，将以有一个可计算的语义预范畴为基础设定来进行。

第二类非理想化条件问题，是范畴性和函子性条件遭到破坏。翻译所对应的表达要素（属性和映射关系等）非一一对应。在语言的表现上，就是语义上有太多歧义分支，语法上有很多的不规范或不规律。

具体细节可以参见上一节的简单模型的介绍。

这些问题，本来是可以规范化或形式化约束来避免的，但是，这样的约束又会使得运用语言的思维受到极大的限制，会抑制创造性的发挥。

比如诗歌往往就会突破语法及语义子预范畴的边界，采用一种特别的表达方式，或者说，它的语义子预范畴本来就画不出来（不清楚），它追求的往往也是这种模糊性。因此，诗歌从某种意义上来说，是不可翻译的。因为你根本很难找出它的语义子预范畴，或者它就是不想让你轻易地找出其语义子预范畴，至少在现有表达能力的条件下。当然，反过来说，诗歌是在用现有的语言的基础上，尽量突破其表达能力的限制，尝试对那些很难表达的语义预范畴进行表达，所以其必然会采用一些特别的方式，突破语法和语义子预范畴的某些边界，无法用通常的形式化方法去分析。实际上，它是启用了（之前我们介绍的）人类（拟全）思维能力的一些新的赋义，把一些拟全概念以一种模糊的、非形式的方式介绍给人们。由于这些新的赋义并不在已有的公共形式语义预范畴中，这些新方式未普遍接受和理解，所以显得不可捉摸。有很多这种“东方夜谭”式的创造性最终是不被接受、被忽略的。而其中一些幸运的创造性，由于有一定的代表性和共鸣，慢慢被普遍接受，逐渐融入到公共语义预范畴中流传下来。

除了诗歌，还有很多类似创造性也有雷同的经历。所以，从人的智能发展本身的要求来说，这第二类的非理想化条件，也总是会出现的。表达的丰富性，往往和不规范、歧义性纠缠在一起。

要解决这类问题，需要在语义分析、应用模型、语义表达等各环节共同努力。后面我们将在针对具体语义处理各环节的问题时，专门对自然语言这些非理想条件方面的问题分别做些探讨。

8.4 从语言表达到形式语义

根据我们之前的分析，拟全语义的结构模型虽然是明确的，但其中的语义对象的语义子预范畴由于拟全性要求是无法直接表示的，只能抽象地理解。语

义对象的形式语义则可以利用形式化的局部化和离散化方法,用有限的属性维、属性值以及映射关系在形式语义预范畴中以形式语义子预范畴的形式表示出来。

于是,从语言达到形式语义的分析过程,其主要任务就变成从语言表达中提取语义对象被关注的重要属性、属性值及映射关系。即如下过程:

$$(E_f)_s \rightarrow (I_f)_s \rightarrow (V_f)_s \rightarrow (X_f)_s \rightarrow (X_f)_s \quad (31)$$

根据上节的讨论,这种语义提取需要依赖合适的局部化才能有效实现,因此必须在已有的上下文环境中进行,很多语义对象的关注属性已经由上下文语境(的局部化条件默认)给出。特别是一段长的对话或文本。有关语境和局部化的问题,上节已经做过一些分析,这里不再重复。

由于上述类似的原因,形式语义分析所依赖的可计算语义预范畴的生长情况,会对语义的理解起到非常重要的作用。一般语义分析的过程,我们不可能从0开始,语言表达中的字、词、概念如果在每次分析时从头理解,是无法工作的。可计算语义预范畴就是为这种情况准备的,我们可以充分利用可计算语义预范畴(可以是公共的、专业的、个性化的、动态装载的局部化语义预范畴)中积累的属性、概念、映射关系,协助系统进行歧义、隐喻、关联、上下文、语境等方面的语义处理。经过充分生长的可计算语义预范畴,可以做到较全面的局部化覆盖,网状覆盖的网格足够细密,就可以为实现对语义对象的分析提供足够精度、足够有效的局部化逼近,从而获得关于语言表达的语义对象的准确有效的形式语义的形式语义子预范畴表示。

语义分析所处理的语言特别是自然语言,很多不满足上节所提的理想化条件。于是这里就会涉及到上节所提到的非理想化条件的问题。非理想化条件对语义分析的影响是非常明显的,特别是第二类非理想化条件。第一类非理想化条件仍然主要采用合适的局部化(比如前面章节中所讲的基于上下文的局部化)。这里对如何处理第二类非理想化条件方面的问题作一些框架性的分析和探索。

1、尽可能保留歧义分支,把可能的联想和关联尽可能完整地记录下来,这对计算机来讲,比人处理这类问题更加有利。不要过早地截断歧义分支,对每个可能的歧义分支分配一定的可信权重,根据上下文的情境,结合各种暗示,以及历史统计数据,分析歧义可能存在的关联性和可信度。这部分有点类似猜谜,特别适用于类似暗喻、隐喻、双关的分析,其实也可以把猜谜、分析隐喻这些当作是广义的翻译过程。这些问题的具体分析,可以参考前面章节中关于多义、歧义、联想、借喻、暗喻的模型,从部分属性的关联性着手。

2、降低语法要求的强制性。语法是语义的生成规则,是一种限制歧义的辅助性的手段,它有时并不需要严格执行,它跟属性的映射关系常常会有关联,比如某些单向关系对顺序就有要求。类似诗歌和口语这样的语法不规范其实很多,包括一些省略、无意或故意的不规范语法组织等。其实语法在语义的理解中所占的重要性有时候并不是很高(参考^[3,p.6]逻辑结构在句子中传达的信息占比不超过16%,而语法还只是逻辑结构的一部分),一些“创造性”的语言应用,总是首先从语法的突破开始。所以,在语言的分析过程中,尽量先降低语法要求的强制性,可以帮助理解很多不规范的表达。

3、结合关注的重点属性维,有需要的话,可以保留几套不同情境下的局部化语义预范畴体系并行处理,在最终需要给出结论时再做最后决策,甚至最终决策也允许保留几种不同的解释。

4、大量积累数据，做历史数据的统计分析。这里要说明一点的是，不同文献中的数据重要性统计要有一定的策略。有时候不仅要看数据本身在各种文献中出现的概率，更要看文献被大众接受的程度。而由于各个历史时期的人口规模、文化普及程度、对现代文化的影响力、文化交流互通的差异，需要做很多的详细模型。并不是海量地搜索数据做数据出现概率的统计这么简单。比如有些网络上的文章从来没人关注过，那么，它的数据完全可以忽略；而有些文献中出现的数据虽然在其它地方极少出现，但它本身的阅读量非常大，那就应该扩大它的权重。

5、在已有数据的基础上，尽量总结出可以范畴化和函子化的一些基本规律。并把一些逐渐形成公共语义预范畴的态射关系函子化逐步纳入。而把一些“例外”做成“例外库”，用不同的方式进行处理。在人工智能参与的情况下，更需要注意这些性质的不同，不能做成等权处理。

当然，这些方法并不是一个“完全列表”，只是通过以上的问题分析，所能考虑到的可行的方法。对非理想化条件下的语义分析，肯定还会有很多细节的情况需要针对性地处理。

从语言表达进行形式语义分析还有很多细节问题需要解决，比如断词、断句、概念对应、语法分析等等，这些可以借用当前语言处理的非常丰富的已有成果。具体这里就不多分析了，可以参考相关的资料^[21]。

需要提醒的一个重要方面是：在形式语义分析的过程中，一方面需要利用可计算语义预范畴，另一方面还需要根据分析的结果，反向不断补充修正可计算语义公共预范畴及各局部语义预范畴，促进可计算语义预范畴的生长。

这个补充修正过程也需要很多具体工具和方法，比如可以借用现有的语言处理的各种人工智能模型等。这里也无法全面系统地分析总结这些方法，可以提几个可能的启发：比如在做形式语义分析时，当形式语义子预范畴所指符合概念规定时，它击中某概念，反过来，当某个语义子预范畴（以某种简约符号性，如词、词组）经常出现或被重要问题引用时，可以将其归纳为概念。所以，这里需要有某种跟频次与重要性相关的“记忆”机制，而有些“记住”的东西，在随后可能几乎用不上，因此，还需要对有些不重要的东西，有“忘记”的机制，以免记忆的内容太多，占用太多的无效存储。

关于形式语义分析的部分，这里只能粗略地介绍个轮廓，更多细节的东西需要有感兴趣的读者进一步深入分析研究和实践。

从形式语义往拟全语义的入射过程，我们在二重性部分已经分析过，这里不再赘述。

8.5 语言翻译中的语义映射

在前面的翻译过程模型中，我们可以看出，除了语义分析和表达生成外，语言翻译的核心在于：语义在不同的语义预范畴系统中的映射。这里我们对不同条件下的语言翻译中的语义映射环节做一些分析。

(1) 理想条件的语义映射

关于语义处理的理想化条件，我们已经在本章之前的相关部分介绍过，而

且举的例子就是语言翻译。

下面举一个简单模型的例子以进一步阐明前述理想化条件在语言翻译中应用的效果预期。这里的模型跟上面解释翻译过程的可以有些不一样的角度，在那里我们未着重考虑语义子预范畴中的概念、属性、映射关系的细节内容。而在这里我们可以把这些细化一下，但省去部分借用的代数几何术语。并主要考察语义映射环节，即之前理想化无人参与模型中的如下部分：

$$(X_{fA})_s \Rightarrow (X_{fB})_s \rightarrow (32)$$

为了简化模型，假设语义 S 的形式表达涉及的是三个概念 C_j 、 C_k 、 C_l 和它们之间的两个映射关系 g_{jk} 、 g_{kl} 。并假设这三个概念都不再需要进行属性分解（即这个简单模型中我们把概念和属性暂时等同处理）。在语言系统A和B中，我们采用满足第一个理想条件的局部化，得到可计算形式语义预范畴 X_{fA} 和 X_{fB} 。

于是，语义 S 在语言系统A中的形式表达 $(E_{fA})_s$ 涉及形式概念 $(C_{fA})_j$ 、 $(C_{fA})_k$ 、 $(C_{fA})_l$ ，以及态射 $(g_{fA})_{jk}$ 、 $(g_{fA})_{kl}$ 、 $(g_{fA})_{jl}$ 组成语义 S 在语言系统A中的形式语义子预范畴为 $(X_{fA})_s$ 。其中 $(g_{fA})_{jl} = (g_{fA})_{kl} \circ (g_{fA})_{jk}$ 是根据第二个理想化条件映射关系是态射关系，由 $(g_{fA})_{jk}$ 、 $(g_{fA})_{kl}$ 组合而成。

这时，我们进一步利用第二个理想条件，即语言系统A和语言系统B之间可以找到翻译函子 $F: X_{fA} \rightarrow X_{fB}$ 。利用第一个理想条件，系统A与系统B的局部化保证属性、映射关系之间有一一对应，即在两个系统间概念和态射之间也有一一对应。

于是，令

$$\begin{aligned} (C_{fB})_j &= F((C_{fA})_j) \\ (C_{fB})_k &= F((C_{fA})_k) \\ (C_{fB})_l &= F((C_{fA})_l) \\ (g_{fB})_{jk} &= F((g_{fA})_{jk}) \\ (g_{fB})_{kl} &= F((g_{fA})_{kl}) \\ (g_{fB})_{jl} &= F((g_{fA})_{jl}) = F((g_{fA})_{kl} \circ (g_{fA})_{jk}) = F((g_{fA})_{kl}) \circ F((g_{fA})_{jk}) = \\ &= (g_{fB})_{kl} \circ (g_{fB})_{jk} \quad (33) \end{aligned}$$

这样，形式概念 $(C_{fB})_j$ 、 $(C_{fB})_k$ 、 $(C_{fB})_l$ ，以及形式态射 $(g_{fB})_{jk}$ 、 $(g_{fB})_{kl}$ 、

$(g_{fB})_{jl}$ 组成语义 S 在语言系统 B 中的形式语义子预范畴为 $(X_{fB})_s$ ，并与 $(X_{fA})_s$ 同构。进一步根据语言系统 B 的表达规则（这些规则一般已经体现在态射中了）形成形式表达 $(E_{fB})_s$ 。

这样的理想的可翻译条件是很诱人的，可以想象一下：

语义明确后，如果语义表达生成部分可以形式（自动）地完成，那么，所有形式系统都可以自动化。

事实上，我们可以把计算机语言也当作是一种语言。这样，如果能把需求采用某种（合乎）形式化的表达，于是需求的理解就是在可计算语义范畴中的语义分析过程，语义分析生成语义后，语义表达的生成实际上就是程序的设计了。因此，整个过程实际上也可以看作是一种翻译过程。

也就是说，如果自动翻译能完成，那么自动程序设计也能完成。它们是一回事！

理想化的条件一般只有人工设计的形式语言系统和机器语言能够达到，程序的编译就基本上是这样一个理想化条件的翻译系统。

（2）非理想条件的语义映射

前面提到过，语言系统特别是自然语言系统，往往是非理想化的，甚至是被“玩坏的”。我们实际的翻译系统，就是要总结和适应这些被“玩坏的”规则，尽量把它们都找出来，并针对性地找到对应的解决办法，完成一般条件下的语言的翻译。

当翻译的理想条件不满足时，上节中两个形式语义子预范畴直接互译的效果就会非常差。这时一般需要用其它无人参与的翻译流程模型。

在没有人参与的情况下，一个比较好的解决方案应该是有一个可计算语义公共预范畴，能足够精度地逼近（我们所关注的局部）拟全语义子预范畴，于是，可以用这个语义子预范畴 $(X_a)_s$ 来替代拟全语义子预范畴 $(X_c)_s$ 。

$$\begin{aligned} (E_{fA})_s &\rightarrow (I_{fA})_s \rightarrow (V_{fA})_s \rightarrow (X_{fA})_s \rightarrow \\ (X_{fA})_s &\hookrightarrow (X_a)_s \Rightarrow (X_{fB})_s \rightarrow \\ (X_{fB})_s &\rightarrow (V_{fB})_s \rightarrow (I_{fB})_s \rightarrow (E_{fB})_s \end{aligned} \quad (34)$$

基于这样的模型，我们就可以找出非理想化条件下翻译的一些方案。从模型看到，翻译的语义映射环节的重点就在如下的部分：

$$(X_{fA})_s \hookrightarrow (X_a)_s \Rightarrow (X_{fB})_s \rightarrow \quad (35)$$

这样的情况下， $(X_a)_s$ 对 $(X_c)_s$ 的逼近程度就会影响到系统翻译的准确性和覆盖面等性能。 $(X_a)_s$ 应该是由前面章节的可计算语义预范畴的构建方案构建，

并不断保持生长和维护，其生长情况对翻译结果影响很大，因为，按上面的翻译过程进行的翻译，语义子预范畴对拟全语义子预范畴逼近程度越来越好的情况下会越来越接近人的翻译能力（极限）。甚至，由于机器的记忆能力和联想处

理深度和速度的优势，在有些方面可以做得比人工翻译更好。当然，这也需要系统其它方面的能力（比如语义分析和表达生成能力）足够强。

对第一个理想化条件的违背，导致局部化系统 A 和局部化系统 B 的基本属性维很难保持一致。在上述的语义映射的入射和投射过程中必然会产生一定误差和不精确性。一个解决方案是尽量扩大两个语言系统形式化时的属性维数，使得两个相互翻译的系统间保持有直接的基本属性映射，从而不至于某个局部化系统中的形式表达所表示的属性在另一个局部化形式系统中没有对应的属性维。另一个解决方案是动态局部化，比如不同专业学科的研究对象就会产生非常大的关注重点的差异，在一个学科中简单带过的一个小问题，可能是另一个学科所研究的全部。这种关注的差异度，往往只能采用动态的局部化的方式去处理，即我们有时候会同时需要好几套可以在需要时动态装载的不同的基础属性维组成的局部化形式语义预范畴去解决一个涉及范围广泛的翻译。总之，在翻译的语义映射环节，我们要尽可能选择最优化的局部化方案，使得系统 B 与系统 A 的局部化语义子预范畴结构尽量接近，从而使翻译尽可能保持原意。

对不满足第二个理想化条件的处理（注意第一个理想化条件不满足时，第二个理想化条件就已经缺乏基础了），很多相关的一些解决方案已经在前面的章节提到了（比如总结出一些局部或部分满足相关条件的情形和类型，以便在尽量大的范围内使用函子性处理大部分问题，留下的一些情况再具体问题具体分析），并在语义分析的过程中就已经应用上了，比如对多义、歧义、联想、借喻、暗喻等关联的处理，对语法的处理等等，这里不再重复。

这里要提醒的是，翻译中针对非理想化的问题所采取的应对措施，需要综合应用，各个环节相互配合才能达到最佳的效果。比如在翻译中，语义分析中可能保留了多个歧义分支，涉及到这里语义映射时，几个不同分支可能需要多个不同的局部化形式语义预范畴或子范畴来支持在目标语言中的歧义、联想、暗喻等。比如在翻译的语义映射时，需要尽量选择能比较好地与被翻译的语义子预范畴相匹配的局部化语义子预范畴，而这些局部化的选择，又会对后续的语义的语言表达生成产生影响。

8.6 语义交流响应

有人参与的过程中，语义交流响应基本上取决于人的知识面、智力与思维模式等，这里主要对无人参与的语义交流响应过程做一些初步的分析。

首先，我们在有人参与的交流过程模型的基础上，给出如下的无人参与的交流过程语义模型：

$$\begin{aligned}
 (X_c)_S &\Rightarrow (X_{fA})_S \rightarrow (X_{fA})_S \rightarrow (V_{fA})_S \rightarrow (I_{fA})_S \rightarrow (E_{fA})_S \rightarrow \times \\
 &\quad (I_{fB})_{S'} \rightarrow (V_{fB})_{S'} \rightarrow (X_{fB})_{S'} \rightarrow \\
 (X_{fB})_{S'} &\hookrightarrow (X_a)_{S'} \xrightarrow{\text{交流响应}} (X_a)_R \Rightarrow (X_{fB'})_R \rightarrow \\
 &\quad (X_{fB'})_R \rightarrow (V_{fB'})_R \rightarrow (I_{fB'})_R \rightarrow (E_{fB'})_R \rightarrow \times \\
 &\quad (I_{fA'})_{R'} \rightarrow (V_{fA'})_{R'} \rightarrow (X_{fA'})_{R'} \rightarrow \quad (36)
 \end{aligned}$$

这个无人参与的交流过程模型，与有人参与的交流过程模型最大的区别在于：有人参与的交流模型中，由于人能够理解拟全语义，因此，交流响应是人

在拟全域的语义预范畴的基础上进行的；而无人参与的交流模型中，由于形式系统不能理解拟全语义，只能在某个形式域的可计算语义预范畴的基础上进行。即：上面流程模型中的第三行所表示的：

$$(X_{fB})_{S'} \hookrightarrow (X_a)_{S'} \xrightarrow{\text{交流响应}} (X_a)_R \Rightarrow (X_{fB'})_R \rightarrow (37)$$

对输入表达分析出的在某个局部化语义预范畴上的语义 $(X_{fB})_{S'}$ 首先被内射到某个更广泛或更通用的可计算语义预范畴中的语义子预范畴 $(X_a)_{S'}$ ，然后在系统B中根据一定的交流响应机制形成可计算语义预范畴中的响应语义子预范畴 $(X_a)_R$ ，它再被投射到系统B的某个局部化（注意可能跟之前局部化不一样，比如话题切换或转换专业背景等）语义预范畴上得到局部化语义子预范畴 $(X_{fB'})_R$ 。

流程模型中，除上述被分析的第三行的“交流响应”环节外，其余部分与本章其它节的内容是一致的。因此，这一节我们试着重点对这个环节做些分析。

简单讲，交流响应就是根据输入的语言表达形成的形式语义的刺激，作出反应。而实践中的刺激和反应是变化多端的，是需要根据具体应用找到对应的交流响应机制。比如要具体分析某个语义（比如命令、提问、指令）下具体系统（如工业系统、娱乐系统）不同的应对方法，这些已经到实际的工程应用领域。

因此，从这个意义上讲，形式系统的交流响应不是本文所涉及的主要内容。它的内容与应用相关，非常多，应该有专门的研究。这里仅从拟全/形式的模型出发，试着提供一些一般性的原则和框架。

显然，这个交流响应的一般性过程，应该是分析输入的语义子预范畴的属性和映射关系所指出的要求，根据当前数据与上下文参数，用适当响应（逻辑、函数）模型，算出响应语义子预范畴。响应模型则需要具体问题具体分析，比如生产过程、日常生活、专业讨论等差异很大。也许参考局部化的方法，建立一些专门的（局部化）响应库，可以慢慢积累起一些通用的处理能力，用于类似通用机器人之类的应用。

从以上的分析中，可以看出，可计算语义预范畴的生长程度对这里的交流响应很重要，前面提到的语境和话题也很重要。它们都是对语义响应所依赖的公共的和局部化的基础。

交流响应机制通常是被动机制，但交流的发生是否也需要考虑有主动的交流？比如规划或定时任务，当然，它们实际上仍然可以看作是对规划或定时语义的响应。对形式系统主动性的研究实际上主要是对目的性的研究，即：如何让形式系统在某个规定的目标下主动工作，包括发起交流？

形式交流中另一个重要的问题是，是否需要避免死循环的机制？这里的死循环可能会表现得比一般简单有限状态机更复杂些：它可能会在一个很长的交流或规划序列上陷入死循环，而不是那种容易发现的简短的循环或嵌套。对于这种问题，应该用什么样的研究方法，找到合适的发现机制和“拯救”机制？

另一个问题是，形式交流体系会不会形成（隐）八股？由于形式化的系统

实际上是在人给出的有限指导机制下运作，它会不会慢慢对每个输入的要求形成背景、需求、方案、依据、计划、行动等规则性千篇一律的响应输出？如果会出现，它会是局部化的（某种情况下需要，比如写规划、建议书之类）还是全面的？需要避免这种情况吗？还是我们可以愉快地接受形式系统的这种“可预测性”？我们需要它有“创造性”和“想象力”吗？如果需要，又应该怎么去实现？

也许我们根本不需要担心？比如人类所写的任意一本书，都已早早列在“巴比塔图书馆”的书架上了，但我们在发现一本绝妙思想的优美的书时，仍然欢欣鼓舞……而“巴比塔图书馆”，正是一个与我们这里讨论的形式语义预范畴类似的典型的不断生长的形式系统。

8.7 语义的语言表达生成

语义表达的生成，是翻译和交流中除了语义分析外的另一个非常重要的问题。这个问题对于翻译和交流的不同情形有相通性，也有一些区别。显然的区别在于：翻译只是把理解的语义子预范畴用另一种语言重新表达出来；而交流则是在理解对方语义子预范畴的基础上，形成相应的交流响应语义子预范畴，并将响应的语义子预范畴用语言表达出来。这个区别会引导出另一个重要区别在于：翻译的表达生成过程中采用的局部化方案应该希望尽量与被翻译的表达相近或同构；而交流中则因为可能产生的话题或语境的转换，要求改变所采用的局部化方案。相通性在于：最终的目标都是把相关的语义子预范畴用某种表达方式（翻译中是另一种语言，交流中一般是自然语言为主）表达出来。

用我们前面举的翻译过程例子来说明的话，交流和翻译的不同在于交流会在对方输入语义子预范畴的基础上生成一个新的响应语义子预范畴，最终表达的语义子预范畴对象发生了变化。如果两个交流的人用的不是同一个表达方式（如不同语言，或者一个人用哑语等等），那么翻译的过程会随同交流的过程同时发生。

所以，语义表达的生成，重点还是把语义子预范畴用适当的表达方式形成表达的过程。即在某个局部化条件下完成如下过程：

$$(X_f)_s \rightarrow (X_f)_s \rightarrow (V_f)_s \rightarrow (I_f)_s \rightarrow (E_f)_s \quad (38)$$

而我们之前所讨论的理想化条件及非理想的问题，在语义表达中都将会涉及到。下面将对这些问题做一些相关的探讨。

与语义的分析类似，语义的表达生成也不可能直接在拟全语义预范畴中完成，必须要有一个形式化的输出过程。而形式化语义预范畴与拟全语义预范畴的矛盾问题在这里同样会遇到。其解决方法也类似于上节所提到的，采用数学上的“局部化”方法。根据实际交流或翻译的上下文情境，选择相应的重要属性维或子空间构建一个可计算形式公共语义预范畴（或子预范畴），并在此基础上形成一个关于目标语义的形式化语义子预范畴。（几何直观上就是完成了拟全语义往某个可计算的公共形式语义预范畴的投影）。

经过上述的过程，结合前面所提到的构建可计算形式语义预范畴的介绍，我们实际上是获得了一个关于目标语义的重要属性维和映射关系为基础的形式语义预范畴中的一个子预范畴，从数学上，它可以用属性维为坐标，结合映射关系，形成一个确定的描述。（几何直观上就是对一个形式语义预范畴中的子预

范畴几何形状的描述)。把这个描述用指定的方式表达出来，就完成了语义表达生成的过程。

这里的“指定方式”，在理想条件下应该是一系列的形式规则。类似于计算用的形式语言或者编译过程中的表达式。

但过于形式化的表达不便理解，再加上如果是翻译或交流的话，使用的是自然语言，它们已经不是那么“形式化”了，我们必须应对这些问题。我们这里主要是研究语言语义的机器化处理，不管是翻译还是交流，是基于机器翻译和人机交流，所以，对人的思维中的某些创造性和艺术性发挥，以及一些“巧妙”的语言运用的关注将不是重点，重点在于如何更清晰准确地理解语义和生成表达。

因此，我们的出发点也会是先从尽量简化的着眼点入手，那就是把表达尽量标准化。以这一原则作为基础，可以形成一系列的方法，帮助实现语义表达的生成，比如：

1、尽量采用标准的语法。从语义子预范畴出发，用尽量标准的语法，组织所涉及到的属性、映射关系，从而生成符合语法的表达。这种生成方法，如果从翻译来看的话，会与“直译”有比较大的区别。它首先是保障“义”的准确性，对原有句式和语法的情况会予以忽略，按标准语法重新组织属性和映射关系，尽量保持生成的表达的语义子预范畴的形状与目标一致。

对机器（和人机对话）而言，最好是采用某种规定的形式语言。如果能形成一种广泛应用的通用的统一形式描述体系，将会大大减少这方面的工作量。

2、为了简单化表达，可能会尽量采用分解工具，把语义尽量用多个简短的句子级联表达，而不追求用一个复杂的句法把语义用尽量少的句子来表达。这会降低表达的效率，但这可能是有效性和准确性之间必要的权衡。

比如，要翻译“他在吃饭”，生成表达应该很简单；但如果要翻译“他在吃你做的饭”，其实就成了要组合“他在吃饭”、“饭是你做的”这两层语义。要尽量有效地翻译（生成表达）的话，就需要把这两个语义组合在一句里，这时候用什么样的语法结构就变得重要了。而简单化表达，就可以直接表达成两个句子。

3、尽量少用缩写等一般自然语言中常用的增效方法，除非缩写的歧义性很小，或者已经是非常广泛采用的缩略语。

4、非语言的重要条件的增补表达。比如涉及到语义的表情、语气、音乐氛围等等这些非语言的表达，尽量换成相应语义的语言表达来进行补充。

5、把隐喻和双关等语义明确化表达。减少歧义的可能性。

以上方式会抑制创造性和艺术性，不会产生一些“巧妙”的表达，更不太会产生类似于诗歌一样的艺术性表达。

有些特殊情况，比如想要通过一些隐喻、创造性、艺术性来表明机器对语言处理的“智能性”不在这里做为重点，有需要的可以做针对性的例外处理。其实，根据我们之前对人的思维能力的分析，这种所谓机器的“智能性”都是假象，基本上是都是人的思维的某些总结和机器的“依葫芦画瓢”，并不是（机器）具备真正的“智能性”。

当然，我们这里只是探讨一些基本的原则，并不能解决所有细节问题，甚至大部分问题目前没有解决方案，都还需要深入分析才能找到比较好的解决方案。比如：

1、一语双关的时候，需要表达的语义子预范畴对象可能就有两个。而在一

个表达方式系（一种语言）里可能有交集，在另一个表达方式系（另一种语言）里可能没交集。这样翻译的准确性就要比较考验了。采用前述的方法把双关直接（用两个表达）明确表达出来，是可以实现语义的比较准确的表达，但在某些语境下就会失去它的表达效果。

2、如果语义中有隐喻，用对其表面的语义最有效的属性集表达时，对其隐喻的语义表达就未必合适了。比如一种表达方式下的俏皮，在另一种表达方式下可能就很难实现等等。这对诗歌的翻译之类经常会出现问题，不太好解决。

如果极端一些，某个作者写作水平极高，在一整篇长篇小说中隐藏了几个类似于隐喻的线索，必须有相当的文化背景积累才能理解，你无法用截短的方式把其中的隐喻说清楚。而更麻烦的是，翻译的目标语种中，关于这几个隐喻的表达，完全没有相应的文化背景的通用性，按照明面的语义翻译，这些隐喻将消失。那么，这篇长篇小说该如何翻译呢？（以前的方式往往是在译本以外，再介绍相应的文化背景，或增加注释之类，但有些“味道”肯定是消失了）

3、这涉及到另一个比较难处理的问题：从之前的分析来说，表达不管长短，都会对应于一个语义子预范畴。那么表达或翻译的时候，是（对整篇表达的整个语义子预范畴）整体进行表达或翻译呢，还是按什么样的规则截断？（即我们按段、按句表达或翻译是不是会走样？怎样从全篇的角度把握表达或翻译。如果是，那么机器如何把握这些？）

当然，如果针对翻译，这涉及到整个翻译过程，不光是语义的语言表达环节。实际上是选择我们把一整段输入表达先形成了个完整的语义子预范畴，再行把语义映射到目标语言进行下一环节的完整的表达生成；还是我们把输入逐句构成语义子预范畴，映射到目标语言进行表达，并最终拼成整篇？这个部分放在前面的章节介绍会感觉有缺失的环节，所以安排在这里。

按词的翻译最简单，但由于语法和词义的歧义的存在，按词翻译显然是很难准确的，实际基本不可行。

合理的方式应该是按句，最多按段翻译。但是这可能会有问题：在没有完全掌握全篇文章的时候，翻译一句或一段有可能会有偏差。即按翻译当时的语境做的翻译，可能会在后面发现有些背景条件错解了。

当然，一般情况下看，这种上下文影响很大的情况不是很多。大部分翻译，按句进行基本上就可以了，按段进行一些上下文的修正基本上可以覆盖绝大部分的需求。需要在更广范围的上下文语义下对段落进行调整的是很少的。

但事实上，这种表面上的一般情况下，可能存在某种不被注意的潜在处理逻辑：即翻译者可能已经掌握整篇的语境，并已经在翻译过程中不自觉地用上它随时决定自己的局部化方案了。这是因为译者在翻译前往往实际上已经多次通读全文，把握了整篇文章的上下文。这在某些重要且相对不易懂的资料（比如哲学文献）翻译的过程中经常出现，有时甚至会有几个相差较大的不同翻译版本，可以看出译者对原文的理解会形成一种语境，影响到词、句、段的翻译。

这基本上可以与语义子预范畴的模型相对应。

只是，对于特别长篇幅的文字等，语义子预范畴与表达生成的对应就显得不那么直接和直观了。

比如长篇小说。按前面的分析，我们可以把长篇小说所讲的内容对应为一个语义子预范畴。由于篇幅很长，这个语义子预范畴就会非常复杂。长篇小说往往以一条或几条线索对许多的人和事进行描述，人物和事件的性质可能在整篇中都有可能出现多次的变化。这些都应该以什么样的方式体现到语义子预范

畴呢？最终的语义子预范畴会长成什么样？

然后，怎么证明按以上方法分析的语义子预范畴就是这篇长篇小说的语义子预范畴？按理，如果把这个语义子预范畴表达出来，它应该就是这篇长篇小说才对。但是，由于词句歧义性、表达方式存在一定的自由度、形式化后表达本身的不清晰性，由一个语义子预范畴表达出来的文字，可能跟原来的长篇小说差别非常大。

这一点，可能会带来关于语义子预范畴验证的很大困难。事实上，我们从小学开始做的提炼文章的中心思想的做法，就是把较长篇幅的文字的语义子预范畴用较短的文字表达出来，并且要求新的表达尽量接近原来的语义子预范畴，特别是重点部分。

由于语义本身的特点，以及我们使用概念拟全性的方法相类似的机制（即对概念的形式化的描述往往只讲其中部分属性，能跟其它概念分开就可以），对一篇文章的提炼总结，可以相对于原文来说非常地短。有意思的是，对一篇长篇的很短的描述，基本上可以抓到它的核心，你一听就能知道是哪篇文字。

反过来，对一篇文字的解释，也可以写得非常地长。而且，可能不管多长，你都会觉得它还是那篇文章要表达的东西，没觉得是把它摊得太大了。

这里又带来另一个问题：如果以上这种把文字压缩和扩张，对它的语义子预范畴的影响不是特别大的话，那么语义子预范畴的严密性就会比较差。可能语义子预范畴的核心部分只需要不多的表达就能基本给出，而大部分文字，对它的限定只是在核心部分之外做了一些小的修修补补。

更麻烦的是，这种随意性，会导致语义子预范畴本身很难表示。

实际上，某些抽象的语义子预范畴似乎并没有什么好的表达方法？它似乎总是在心里的某处飘渺地存在着，不可言说。

如果要用什么方法表示，与用语言表达效果其实都差不多。

4、某些艺术类作品的翻译，需要照顾到艺术表现力，这里也会出现一些翻译的困难。比如情绪和动作也是表达，也可以用语言翻译（比如小说中有很多用文字来描述情绪和动作，诗歌对情绪和情感的描述就更多了，几乎是想尽文字的各种办法，但写再多也还总是觉得不到位）。但也翻译不到位，象音乐那样。文字对音乐的描述也有这种感觉，音乐的某种情绪感染能力，文字总写不好。也许音乐就是更合适描述某些情绪（所指向的语义子预范畴的属性）？

另外一些关于语义生成的问题，比如语法与语义的关系等，有两种认识^[3]：

蒙太古语义学和生成语义学，前者认为二者关系不大，后者认为语法结构应该高度相似。这里应该更认同前者，即语义预范畴的表达可以有很多不同方法，比如不同语言、音乐、绘画等。具体细节这里也无法面面俱到地探讨，只能留待以后或他人做深入研究了。

8.8 形式系统自动化的本质

我们在介绍理想化自动翻译的条件时提到过，如果达到那样的条件，在语义和需求明确的条件下，就可以实现自动程序设计，形式系统都可以实现自动化。

任何一段表达，本质上都是在语义预范畴中形成一个语义子预范畴。如果用任何另外的表达方式（不管是拟全的还是形式的）把它足够精度地表达出来，

实际上就是完成了翻译，如果表达的结果是某种执行机构的执行，那就是自动化。

工业自动化系统的本质，实际上就是把生产过程全面地形式化。然后用形式系统完成整个体系的流程和沟通。

这其中沟通的部分，比如人与机器间、机器与机器间的沟通，实际上也是（利用形式语言）进行形式语义的分析和生成过程。

对于工业系统而言，自动化的两个理想条件是有可能满足的：

1、因为某个工业系统的运行环境和目标可以很明确，可以把需求做到足够精度，对各种要素的分析也可以做到基本完整，于是，对具体工业系统而言，它所关注的语义预范畴的重点属性就明确了，其要求的精度也明确了。于是，可以针对该系统形成一个局部化的可计算的形式语义预范畴。这就解决了两个理想化条件的第一条。

当然，这里给出的可计算语义预范畴可能不是通用的，即在数学上它不是“全局”的，而是“局部”的。比如一条自动化生产线生产运行所使用的语义预范畴和维护这条生产线所使用的语义预范畴可能就有不同。

2、关于语言中规范的例外情况，对于自动化系统来说，解决起来相对简单一些，只要对它们设计一套可以满足功能要求的形式语言就可以。当然，为了人机接口的方便化，以及机器与系统之间的沟通方便，最好能设计出一套通用的，一般人也容易掌握的形式语言系统。甚至语言可以类似自然语言，人可以直接和机器对话，只是语法规则的要求要相对严格，不能有例外和自由发挥。

如果系统的需求描述比较完整，那么自动化系统主要是实现已知需求的应对执行，基本上不需要太多考虑例外的灵活性，因此这里的应用可以尽量形式化，不需要太多考虑创造性和艺术性。这样，理想化条件的第二条也能实现。

当然，这里也还会遇到一些问题。

比如第一个理想条件和第二个理想化条件就有冲突：第二个理想化条件中最好要求能有一个通用的形式语言体系，而第一个理想化条件要求针对不同语境有不同的可计算语义预范畴，这就会产生矛盾。解决的办法是在一个形式框架下，允许切换语境，并在切换语境时自动装载“局部化”的可计算语义预范畴。

再比如人机对话的要求，要形式语言尽量接近自然语言，这个要求也是有矛盾的：形式化一定会涉及到一些抽象和严格的规则，而人习惯的自然语言很难达到那样的标准。如果放开系统规则的严格性适应性人的需求，那么系统很快就会变得“杂乱”和庞大。比如不同的人可能用的语言不是同一个语种的，那么就需要把所有可能的语种都研发一遍，等等。这会使自动化系统因为这些要求而不堪重负。解决的办法是对人也要有相当的约束，至少需要做一些专门的培训，才能与机器对话。

实际的自动化系统还会有很多具体问题，这里只能给出一些基本原则和可能，实际的实践还是有大量的工作要做。比如针对各种不同用途的传感器与执行机构的研制^{[137] [128]}。比如前面提到的一个可以适应于描述需求和人机对话的

尽量通用化的形式语言系统，就是涉及面非常广的一个艰巨的任务。

更进一步，如果把语言的范围扩大一下，包括到立体影像、语音、音乐等，语义分析和表达也使用这些方向的局部化方案，那么，语义处理系统就可以成为一个可行的虚拟世界——元宇宙——的基础。

9 语义与人工智能

人的思维与人工智能的一个重要的区别在于：人能够处理无限，而人工智能只能处理有限^[54]。这是由于目前的计算模型是按图灵的有限状态机实现的，所以它不能处理涉及到“无限”的问题。让机器去取一个序列的极限它是算不出来的。有限状态机不能表达无限，所以做不了取极限的操作。

当我们用符号指代某个有限属性定义的概念（为符号赋义）时，我们就开始应用形式化系统了。应该说，数学和所有自然科学，都在一定程度上被形式化了。比如在定义数学的概念时，往往是用确定的几个属性来限定它，而忽略掉所有其它不被关注的属性。这种定义模式，正是数学化的抽象的核心要义所在，而它碰巧又正是形式化的基础。

在这个过程中，有些无限性的（拟全）内容可能被“默认”由赋义环节带进了系统（比如数学中点的定义等）。这些我们在之前对形式化的介绍时都有涉及，这里不再多讨论。这里我们将主要讨论语义的形式化与人工智能的关系，从这个角度探讨人工智能的本质，它的能力限制，以及它可以发挥和发展的领域和空间。

9.1 语义的形式化与人工智能

人工智能^[10]在很多领域取得了十分显著的成效，特别是在一些专业的方向，如图象识别、语音识别、对弈等领域。但在这些领域取得的成果，有被不断放大的趋势，似乎人工智能真的无所不能地智能了，出现了它可以代替人甚至超过人的智能等论调。真的是这样吗？我们先看看人工智能是怎么做到“智能”的。

人工智能的一切的基础，是基于计算模型即图灵的有限状态机模型。而有限状态机只能处理有限的形式化问题。“0”和“1”被作为所有计算的基础，它就决定了所有计算所表示的对象只能有有限属性（由有限个“0”和“1”组成），不管它的数量有多大。

在这样的基础上的人工智能，无论如何发展，也只能是形式化的。它可以无限地逼近拟全化的那些概念，但永远也达不到“拟全”的那个点。

如果我们把人的智能作为参考，由前面假设的人的意识具有“拟无限”能力，从而可以得到拟全概念。而人工智能由于机器的形式化限制，做不到这一点，不能处理“无限”问题，所以，它“智能”的能力一定是受到极大限制的。

再详细一些来看，以识别类的人工智能为例，目前的人工智能不管怎么“智能”，它的算法基础，基本还是以各种改进形式的神经网络为基础。所谓“智能”的提法，很多是基于神经网络有“隐参数”，所以它能完成一些人所想不到的方案去解决问题，甚至其所采用的“解决方法”人都无法理解，所以认为它的“智能”甚至比人都高明。

但是神经网络的各种模型依然是人提出来的，那些隐参数虽然人不能全部掌握，但它还是与系统模型一起，经过大量数据的校正，实现了某些参数与处

理函数的组合，最终实现某种分类能力。

人工智能中人不可掌握的“隐参数”这些，并不是真的由机器“想”出来的。不可预测性的来源是数据，而不是机器的“思维”。所以，所谓的机器智能实际上是数据的不确定性（不可预测性）产生的，并不是机器真的“智能”了。

机器也没有“看懂”数据，只是根据人的算法的要求，对数据进行处理，实现了数据分类的参数提取，以及分类功能本身等规定的目标。

如果我们把人工智能的系统中的参数（不管是显性的还是隐性的）当成是属性集或属性的取值，而把它们形成判断构成的函数集当成是一些映射关系集，那么人工智能的系统，是不是就可以认为是经过训练所获得的一些语义表达。

这样，我们就可以把人工智能的系统转化为一个语义分析和理解的系统。而这也符合我们对人工智能能完成的任务的直观：识别语音、图像所代表的不同的“意思”，即分析语音和图像所表达的广义的语义，并加以分别。

而这正好与我们前面讲过的相吻合：概念的定義的目的实际上主要有两方面：一方面是与其它概念的（外延）区别；另一方面是规定概念的主要（内涵）性质。

其中的第一个方面中，如果把概念改为一般的语义子预范畴，是不是正好就完成了模式识别的任务？

结合我们前面对语义的模型和分析，以及拟全与形式化二重性的分析，就可以得出：从语义处理的角度来看，人工智能不过是某些限定条件下的形式化的语义逼近。

为了更进一步弄清人工智能和语义预范畴的关系，我们在下一节对两者的联系做一些进一步的探讨。

9.2 语义子预范畴与隐藏参数

前面讲语义模型的时候，我们就分析过，语义表达实际上是在语义预范畴里形成一个语义子预范畴的过程。也就是说，所有的语义表达，可以表示为由一些属性和映射关系组成的语义子预范畴。如果用几何直观来简单刻画的话，那么语义就是由属性和映射关系组成的某个语义预范畴中的语义子预范畴的形状。

人工智能的方法是用隐参数，让神经网络在不断训练的过程中，慢慢习得和校正分类的隐参数。然后，用这样的参数，对某些特定对象的分类就会非常有效。但这些参数究竟是什么，我们并不知道。（比如后面例子中的参数，它可能隐含在某些参数的表达中，而我们意识不到那就是某一个明确的形状。如果是一些不规则曲面之类，那就更难以识别了）

对某些特定对象，这样的分类所需要的参数可能并不多，所以经过一定数量的学习后，有限的参数就可以进行有效的分类。

到这里，其实就已经可以发现：如果人工智能训练出来的参数集及相应的判断函数集，正好可以描述一些语义子预范畴之间的不规则边界，那么，语义表达和人工智能的参数训练是不是就是在做类似的事情了？两者都是通过一定的表达来区别不同的语义子预范畴。在语义预范畴里观察的话，人工智能应该就是根据输入的数据，一点点地描绘和逼近语义子预范畴的边界形状，最终根据描绘出的形状和空间的距离，实现分类。

下一章中有个简单的颜色分类的例子可以更详细地说明这种联系。后面我

们将把人工智能系统等价于某个特定条件的语义处理系统。

下面我们从语义预范畴出发，对人工智能完成的任务换个角度理解一下。

语义预范畴中的概念，既是“稀疏”的，又是“拥挤”的，而且是“万花筒”性质的。如果对属性做详细分析，就会发现，有效的概念（在语义预范畴中）占用的空间并不多。但很多概念的属性又有交叉重叠的部分。究其原因，能为人所认识的对象的概念，并不是可以随意选取，往往都是有一些条件的。比如从空间来说，地球上的东西才更可能被认识清楚，这样，空间属性的部分，地球就相对拥挤了，而从宇宙空间来看，存在于地球以外的其它地方的概念就很少了，所以总体而言，有效概念其实很“稀疏”。

由稀疏性，如果要对特定的某些语义子预范畴做有效的聚类或分类，所需要的参数有时可能会是很少的。

但是，由“拥挤”性，这些参数又可能很不直观，人们并不能直接由熟悉的属性维度找到这些有效的分类参数，由于语义预范畴的“万花筒”性质，它可能是某些属性的一些复杂组合。比如目标可以用一个椭圆分类，而我们接触到的属性维度是直角坐标系的两个轴，那么直接用 x, y 轴来划分聚类就显得很困难；而如果我们有一个参数可以直接表示椭圆的内外，即如果我们能把属性维变换到椭圆，那么分类就会非常简单。

也就是说：如果我们能找到稀疏语义子预范畴足够多（不需要尽可能多，但要在尽可能多的属性中选取或尝试组合）的属性，并用这些属性的组合和变换，可以有效地描出针对某些对象的分类边界，并把这些边界用另外的属性或参数表示出来，那么，这些参数就可以用于有效的分类。

或者用形式化中局部化的语言来说：我们找到某些特定条件 p 的局部化方法 L_p ，可以把分类对象 O_i 在语义预范畴 X_c 中的语义子预范畴 $(X_c)_{O_i}$ 形式化为形式语义预范畴 $X_f = S(L_p(X_c))$ 中的稀疏语义子预范畴 $(X_f)_{O_i}$ ，就可以帮助应用形式工具的人工智能轻松实现分类目标。

而人工智能学习训练的结果，实际上就是用隐参数实现了属性的变换组合（即局部化），从而生成了一些用于分类的语义子预范畴的有效边界。在下一章的例子中，可以更明确地看到这样的结果。

所以，从语义预范畴看人工智能的分类，可以看到两个方面的有趣的机会：一方面，我们可以从语义预范畴的模型出发，寻找子空间的边界，使用拓扑、几何、代数等等手段，根据语义预范畴的“万花筒”性质，找到显性的边界。然后把这些边界的形状函数化、参数化，提供给人工智能系统用于更精确的分类。

另一方面，我们可以从人工智能所获得的函数模型和隐参数出发，描出语义子预范畴的一个大致边界形状。然后采用语义子预范畴的拓扑工具进行修正补充，得到更精确且有明确意义的分类边界。再把得到的更精确的边界的属性和映射关系参数化，作为（新的模型、）参数和判断函数送回到人工智能系统中，得到更精确的结果。

换一种说法的话，人工智能的“智能”，就是在语义预范畴中找出更合适的属性集和映射关系表达式，用隐参数反推出用于有效分类的“万花筒”。

9.3 人工智能突破的语义方法

通过以上两节关于语义处理与人工智能关系的讨论，我们这里可以探讨一下，语义处理的模型有没有可能帮助人工智能实现一些突破。

我们先简单探讨一下当下比较热的“数据驱动”模型，看看它能不能是人工智能的救星？从以上的分析来看，这种可能性看起来是要让人失望的。数据驱动的只能是在某个人工智能模型下的参数的精确化。从语义预范畴来看，我们实际上是试图在确定的模型或判断函数的基础上，用训练隐参数的方法，来尽量逼近语义子预范畴的分类边界。

而从形式语义预范畴的“万花筒”的性质来看，这是受到很多限制的，有些“万花筒”属性可能在（数据）离散化时，因为未被重点对待，其信息就已经缺失了，再怎么从（数据驱动得出的）参数里努力也“榨”不出有效的分类边界。

而如果我们认可人工智能是在寻找逼近的（有效的）“万花筒”属性，那么，我们就可以相应地根据一些中间表现，结合语义预范畴处理的其它方法，不断调整包括获取属性的方式，把某些原来（数据中被）忽略的属性用另外的传感器或变换方法让它更加有效，从而方便人工智能找到那个“万花筒”，这样就有可能帮助人工智能更高效地实现目标。

从形式语义预范畴角度来说，人工智能找万花筒的过程，实际上是一种寻找局部化解决方案的过程。如果通过找万花筒的积累，人工智能慢慢拥有足够多的局部化语义预范畴解决方案，可以最终实现针对大部分专业领域，随时根据专业语境找到对应的局部化模块，就可以完成上一章中所提到的形式系统的自动化任务。

这种动态判断和调用不同局部化模块的方式，实际上相当于机器也（模拟人）完成了部分局部化的任务，虽然是预先做好局部化模板的方式。或者可以理解为：针对局部化这个任务能力本身，人工智能也可以做局部化的形式化逼近。

进一步，如果我们可以找到某种拟全语义预范畴的数学处理方法，能通用化地轻松找到语义子预范畴的形状和参数，那么我们就找到了一种通用化（普适性）的人工智能（参数提取）算法。这样，人工智能就可以从碰运气式的找万花筒的任务中解放出来，由语义预范畴的数学模型直接给人工智能提供局部化的判断函数和参数。这是一种比较理想化的模式了。

更进一步，如果能把人的“拟无限”的能力，赋予机器，从而突破计算模型的图灵有限自动机的限制，人工智能应该就有处理“无限”的能力。那时候，机器是不是真的可以象人一样地智能了？这似乎有点“天方夜谭”了？

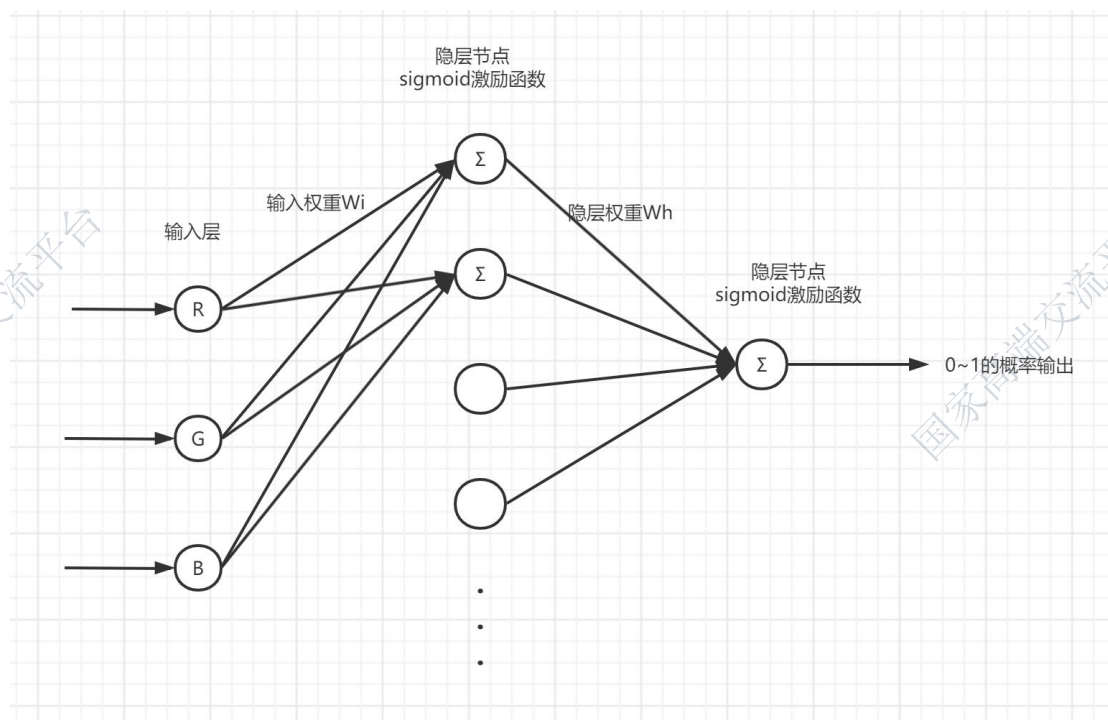
10 两个简单的例子

10.1 颜色子空间的分类

这个例子是一个极其简化的语义子预范畴的例子。我们将尝试进行红色的识别。用语义的语言来说，这里所关注的对象是颜色子预范畴，目标是在颜色

子预范畴里分辨出其中的红色子预范畴。为了简化，我们在这里忽略与颜色无关的其它语义属性，以及属性的映射关系，只保留三基色的属性维所组成的语义子空间。即，这里只处理跟颜色有关的局部化语义空间。这样要素属性明确，便于形式化。

从人工智能的角度看，这也是个非常简单的模式识别问题，用神经网络比较容易计算。例子采用神经网络感知器模型（结构见图表 10- 1），用 2000 个随机生成的数据，通过人工标引对模型进行训练，得出的结果可以实现对红色的识别。这个模型会有由一些隐参数决定的判别函数，在复杂系统里，这些函数是什么我们无法直观地了解。



图表 10- 1 本例采用的人工智能神经网络模型

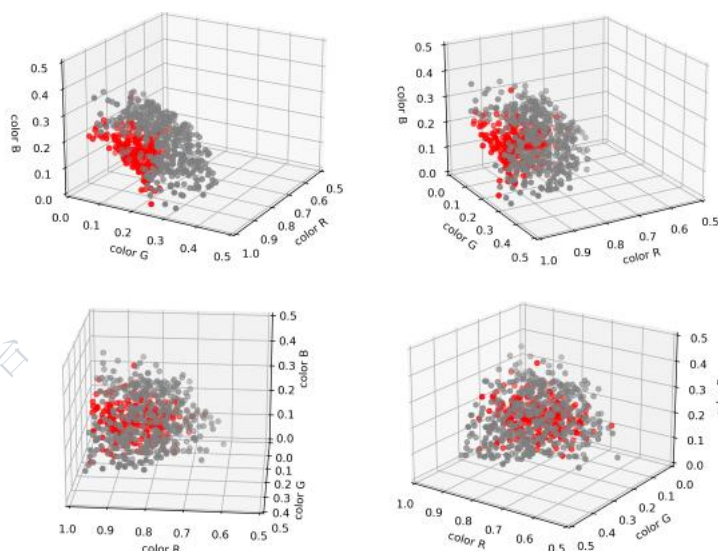
我们再从语义预范畴的角度来看这个问题。如果把颜色用红、绿、蓝三个属性维来表示，颜色语义的取值范围就是一个三维的语义子空间，我们把它称为（局部化的）颜色语义空间。颜色语义空间中的每个点表示一种不同的颜色。如果对每个属性维做 8 位的采样进行离散化，颜色语义就可以用一个三维的每维取值为 0~255 的形式化语义空间来近似地表示。

在这个颜色语义空间中，如果我们能找出红的边界，即在这个形式的颜色语义空间 X_f 中，把红的语义子空间 $(X_f)_{\text{红}}$ 描出来，也可以实现红色的辨别。要注意的是，红色的判断并不是按红色属性维的取值来直接决定的，所以它的判别函数也不是只关注红色坐标轴这样非常简单。

但根据我们对颜色的直观，颜色语义空间中，红色应该是集中在“纯红”的点附近。也许就是红绿蓝三维中以纯红的顶点为中心的某个曲面。

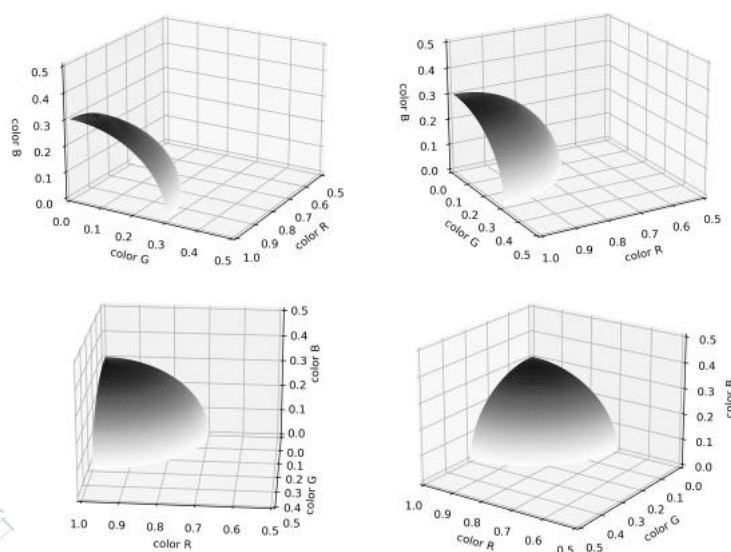
在这个简单的例子里，我们可以按这个思路，在形式语义空间 X_f 中来验证。其结果，基本可以确认这个直观的假设能成立。人辨别出来的红色语义子空间

$(X_f)_{\text{红}}$ ，基本是上述的一个曲面内。但是，由于前面所提到的“模糊性”问题，这个曲面在不同人、不同识别情形下（比如先看一个“很红”的颜色，再看一个不那么“红”的颜色，与先看一个“不红”的颜色，再看同一个不那么“红”的颜色，其结果可能会不一样）会有差异。因此，它基本上是一个带有“厚度”的曲面（模糊分界面）。（参见图表 10- 2）。



图表 10- 2 人工辨别的红色分界面附近采样点

为了更清晰地说明属性的“万花筒”问题，这里我们做个更简单化的假设：假设这个曲面是个半径为 R 的圆弧面。那么如果我们把属性做一些变换，把这个圆弧面离“纯红”的中心点的距离变换成颜色语义空间的某个属性（比如用极坐标系表示颜色语义空间的属性），那么，用这个属性来做“红”的判断，就非常简单了：用以“纯红”为中心的极坐标的极轴和一个参数 R 就可以实现“红”的识别。（如图表 10- 3 所示）。当然，实际情况往往没那么简单，如这里的例子，其分界面实际是上图的不规则形状。



图表 10- 3 简单化假设的圆弧面

现在回过头来考虑一下人工智能的隐参数决定的判别函数，我们是不是可以猜想这些隐参数决定的判别函数构成的识别面，会不会与上述的曲面很接近？也许就是它的近似？如果是这样，我们是不是可以认为：

在语义预范畴中用解析（甚至直观、猜测等方法）给出一个光滑的语义子空间曲面形状（比如这个例子里的曲面）；而人工智能就象是用一系列（特定形状）的刀一步步地在语义预范畴中削出一个形状出来。后面这个形状应该逐步逼近前面的解析曲面。

解析曲面可以由一个或多个由属性为变量的公式来表达（每个公式可以看成是由属性组合而成的新属性或特征参数），而人工智能“削”出来的曲面则由（训练生成的）隐参数来表达。理想状态下，它们二者应该是尽可能地接近。

这个想法是可以通过对隐参数的计算来验证的。验证的结果是：确实如此！隐参数构成的前向过程函数组如下：

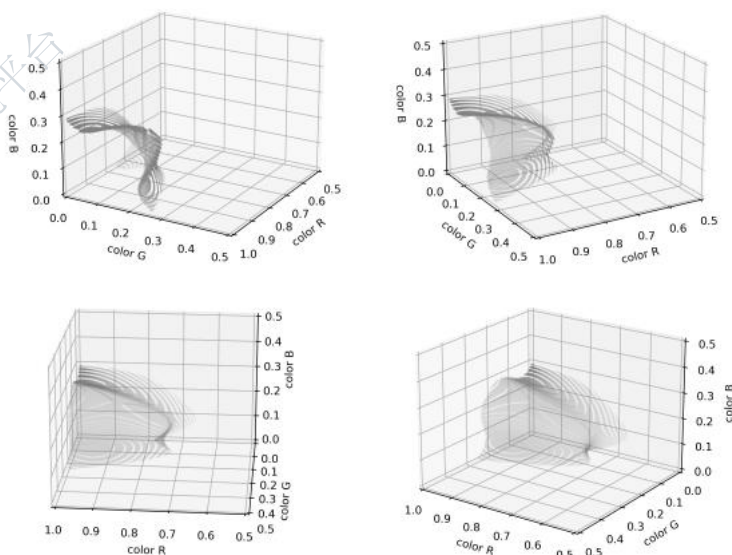
1、从输入到隐层节点的函数组

$$\begin{aligned}h_1(r, g, b) &= 4.38r - 11.31g - 17.63b + 2.97 \\h_2(r, g, b) &= -5.38r + 11.31g + 14.65b - 1.33 \\h_3(r, g, b) &= -2.74r + 5.41g + 8.61b + 2.60 \\h_4(r, g, b) &= 3.12r - 4.75g - 9.55b - 3.23 \\h_5(r, g, b) &= 6.36r + 21.14g - 24.20b - 5.60 \\h_6(r, g, b) &= 2.79r - 35.39g - 2.2b + 1.61 \\h_7(r, g, b) &= 38.42r - 3.3g - 12.27b - 25.32 \\h_8(r, g, b) &= -1.25r + 9.66g + 6.02b - 0.21 \\h_9(r, g, b) &= 5.25r - 15.97g - 14.26b + 1.97 \\h_{10}(r, g, b) &= -3.54r + 32.50g - 29.55b + 0.86 \\a_n(h_n) &= \frac{1}{1 + e^{-h_n}}\end{aligned}$$

2、从隐层节点到最终概率输出的函数

$$\begin{aligned}o(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}) &= 6.06a_1 - 8.03a_2 - 7.51a_3 + 3.78a_4 + 9.17a_5 - 6.61a_6 \\&\quad + 10.82a_7 - 7.33a_8 + 6.07a_9 - 16.81a_{10} - 4.10 \\p(o) &= \frac{1}{1 + e^{-o}} \quad (39)\end{aligned}$$

由以上函数描出的判别边界的图形如下图表 10-4：



图表 10- 4 根据人工智能结果描出的红色分界面

从以上的结果看，我们之前的语义预范畴与人工智能关联的设想，特别是关于隐藏参数与某些局部化转换属性的关联，基本是成立的。

而且，如果我们在上述的分界面示意的几何空间里，以分界面为准，把分界面以下和分界面以上的点（按要求的足够精度离散化）按一定顺序（如“之”字形顺序）分别列成一系列串接起来，并以跨越分界面的串接点为分界判决点，这样，这个形式系统就可以形成关于“红”的一维的投影，并完成“红”与“非红”的逻辑判断。也就是说，这个例子同时也验证了前面的定理 5- 1 和推论 5- 2。

以上语义预范畴与人工智能对应的对比关联分析，给我们提供一个思路：如果可以找到从拓扑的显参数到人工智能的隐参数之间的对应方式（及其反向过程），那么就可以在语义预范畴模型和神经网络的形式化方法之间建立起一个相互促进的方式。这样就可以用分析、几何、代数拓扑或者代数几何等方法来做人工智能的研究，或者反过来用人工智能的成果提示语义研究甚至数学的研究了。类似解析解与数值解之间的关系那样。

如果做不到这种程度，可以做到由拟全模型研究得到一些指导神经网络进一步发展的新方法也很好了。

再次些的话，如果从拟全模型得到一些问题用形式方法不可能解的结论，也可以指导少做一些无用功，也是一种进步。比如类似克莱因瓶有没有可能是“削”不出来的？如果一个语义子预范畴的边界是克莱因瓶，（在 5 维以下？要查一下代数几何的簇的平滑内嵌定理）用“内外”进行分类的判别是无法实现的，即这种情形下判断问题是无解的。

例子里还有关于模糊性的问题需要进一步用模糊数学或者其它方法去完善。模糊性的问题在实际工作中经常碰到，有些是需要具体分析的（比如从不同的方向接近某一个目标的过程中，判别的标准会产生差异），有些应该可以通过模糊数学等一些数学的研究，找到一些比较通用化的方法。

10.2 元宇宙里的平行双镜

(1) 平行双镜

现实生活中很少能对无限有直接的体验。两面平行的镜子是个比较直观的例外。镜子里的镜子和对面的镜子形成了一个相互影射的递归循环，一直无限地反射下去.....

这是不是有点类似引力引起时空的变形，变形的时空又导致引力的变化，变化的引力又导致时空新的变形.....

当然，由于光线强度、平行的绝对性保证、人的观看角度等多方面的因素的影响，镜子的序列最终不会是真正的无限，但这应该是我们直观感觉到的很有说服力的无限的例子了。

由于不是最终的无限，我们看到的“无限”只是在我们脑海里完成的，所以，它也应该有“拟无限”的性质。

在这个例子里，至少我们接触到了三个常见的疑惑：1，无限；2，相互包含的递归究竟哪个“大”；3，相互影响的体系里，最终会是什么结果？

第一点上面讲过了。

第二点我们在之前分析“全”、“意识”、“概念”、“语义”等相互包含的问题时就碰到过。在这里也一样，每个反射的影子都包含在对面的镜子或影子里，而且是真包含，那么，最终，这两个镜子的影子谁比谁“大”呢？在这个例子里就可以看出，这其实是个没有意义的问题。概念的映射关系其实应该也是类似的。

第三点我们之前分析过，应该象广义相对论^[70]那样，找到它的极限表达，把二者（甚至更多的影响因素）综合起来考虑，这些因素的影响如果能收敛，会有一个最终的极限表达式。它往往就是我们的拟全化结果。

(2) 元宇宙里的平行双镜

如果把平行双镜的问题带到元宇宙里，就会产生问题。因为元宇宙从打造的基础，就是个形式系统。而形式系统中没有无限，因此如果不借用其它系统，在形式系统的处理方法中，按照反射的模型，它只能逐（次）个地（递归地）计算每次反射的结果。

如果用形式计算方法一遍遍地算反射，首先出现的问题就是：每次反射的结果需要等待计算完成，这就必然产生反射的延时问题。这里产生的计算延时显然要比光线的反射要大得多，于是在元宇宙里，我们就体会不到真实世界中平行双镜的反射情景。

更严重的是，由于这个计算理论上是无限次，所以它就会很快耗尽计算资源。如果要尽量把它算得更象真实世界，甚至这一个小小的问题的处理，就能把整个元宇宙搞宕机了。

这应该如何处理呢？

而在类似元宇宙的形式化空间中，这样的形式计算的递归问题应该会经常碰到。

如果就形式方法本身，这个问题可能会非常难以处理。能想到的办法要么

就用差的体验凑合（比如只做有限步的递归，强制中止；延时长就长了，人能接受就行）；要么就只能放弃（改变规则，不允许双镜的出现）。

这些似乎都不是什么理想的方法。怎么办？

（3）拟全解的形式化：二重性的应用

元宇宙中的平行双镜问题，在形式域中，似乎无法找到可行的解决方案。那我们试试在拟全域中能不能解决？

我们回到现实或理想条件下的平行双镜问题，看看能不能为这个问题在拟全域找到解？

很显然，这个双反射问题的解在拟全域下其实是相对比较简单，它可以模型化为一个线性（好象不是线性的？随着反射次数增加，距离的增加会使得反射物的比率不断增加？到最后接近于平行？）的无限级数的数学问题。由于它是收敛的，我们用数学分析方法比较容易得到它的无限级数的解的解析公式。

（如果需要实际简化模型算一下，可以把问题简化为：边框宽度占比 $1/8$ ，镜子的距离保证首次反射缩小 $1/2$ ，反射率正好保证边框和镜面的光线亮度保持不变，人是透明的，镜子中只有边框和镜面的反射像，没有其它东西，人在正方形的镜子中间，正面面向一个镜子观察。计算边框和镜面的反射像在看到的镜子中的分布情况。实际只需要计算沿镜子中心点到某边中点的这条线上的分布。应该不难算（考虑到非线性问题，应该也不容易，汗……可以把边框理想化为 0 ，背景是某个固定单色，或者把边框当作背景）。严格一些的话，需要对双眼分别计算，可以假设镜子边长是双眼距离的 10 倍。）

然后，我们试着把它再形式化，看能不能应用回元宇宙中？

由于元宇宙的形式性，它的空间和时间分辨率一定是有限的。我们就用这样的条件对双镜问题的拟全解进行形式化。

即，我们利用拟全域的解析公式，直接对元宇宙中每个相关的时空点（这里就是在每个时点上元宇宙的“人”看到的镜子二维面上的点）计算它的显示值，而不再把整个无限反射的模型带进元宇宙中。

迎刃而解！

这是一个很简单的线性的例子，但可以说明怎么在拟全性的概念（语义）空间做好数学模型，再应用到形式化的计算中。

本来在形式的计算中直接按反射模型运用形式逻辑，需要一遍遍直至无穷地计算反射的影像，而且会有比较大的延时。看似一个小问题小细节，却会耗尽系统的资源。

放在拟全性的语义预范畴里，我们可以得出这个（无穷反射）问题的数学解，不再需要把递归过程展开。再把这个数学解形式化到形式计算中，从而节省了计算资源，也避免了无穷递归的问题。

这里实际要说明的是：有些形式化空间中的问题，我们可以先借用拟全化的语义预范畴求得拟全解，然后再把它形式化到形式语义预范畴中，就能找到问题的解。

这其实就是我们前面讲到的二重性问题。利用二重性中两个域的对应，采用拟全域和形式域里的不同方法的组合，共同解决问题。

这种方法其实我们在很多情形下已经在使用，只是人们并没有关注到拟全空间和形式空间方法之间的转换。

上一节的颜色空间的例子，其实也是这样的二重性方法，在两个域中同时处理，共同促进问题的解决。

从另一个角度来看，这种问题解决方法，实际上也是利用了类似数学中的全局/局部性质（环的局部化）的结合来解决问题的办法。是前面所提的寻找“万花筒”的解决模式。

理论上，所有“无限”的问题，在形式空间中都无法解决。但实际工作中，有很多涉及到无限的问题，可以通过一些其他方法辅助后在形式空间中处理，而其中的有限/无限空间和方法之间的转换几乎不可察觉。

这些问题在数学中最为常见。而数学并没有在拟全和形式空间之间做过很好的区别处理和研究（也许应该做一些这方面的研究？）。也就是说，二重性的应用，很多情况下是不知不觉的。

11 小结与展望

本文涉及的面比较广，而很多问题都并没有在文中深入研究下去，因此不可避免会有很多疑问和有待进一步深入的问题点。

下面将对本文所介绍的内容做一些简单的总结。主要包括两个方面：一方面是文中探讨得出的一些基本结论和断言；另一方面是文中所涉及到的但没有深入的问题，以及可以继续探讨的方向。

这些断言和方向的得出是有一定的理论基础的，但因为涉及面比较广，相应的研究也不是特别深入，会有偏颇，也可能会有一些偏差。用本文的理论来解释的话：由于关注的角度不一样，我们每个人看到的都是“万花筒”的一部分可能性，即局部化性质，对于那个整体的体系，还远未掌握。特别是本文对数学的应用，最多只能算是建立了数学学科与语义学的联系，远未深入。所以本文并不指望能对所有涉及的方面都能给出精确解答，只是希望本文能启发一些对语义（进一步扩展到意识、认知、智能、自动化）“万花筒”局部和整体性质研究的新思路。

11.1 一些简单的断言

- 1, 语义是人们之间进行交流的基础。它也是人进行思维和思考的基础。思维和思考可以看作是人与自己的交流。当思考所交流的对象是人自己的意识时，就产生了自我意识。
- 2, 人的思维有“拟无限”的能力。这个能力帮助人形成一个拟全化的处理系统。归纳法是以“拟无限”为基础的，演绎也是需要利用“拟无限”的。
- 3, 思维必须有“无限”的参与，有限的基元组成的形式系统，无法进行拟全化思维。拟全概念一般是不可直接描述的。不要试图对非形式的概念做简短的精确定义。比如“杯子”，比如“马”。但人通过“拟无限”，又可以获得对语义子预范畴的拟全认识。所以，这也是个很有意思的结果：事物是可知的，但又不是俱细的。
- 4, 概念是人使用“拟无限”、“拟存在”和“拟遍历”的能力，对认识的对象的属性及映射关系进行的一种（范畴学意义上的）极限和协极限的组

拟全化的结果。

- 5, 语义的基空间是由属性构成的维组成, 语义是由语义空间和内嵌的概念和概念之间的映射关系所构成的拟全预范畴。人所掌握的拟全语义预范畴是不断变化生长的。人们之间有共识语义预范畴 (它是交流得以进行的基础), 每个人也有个性化的语义预范畴。
- 6, 语义系统中求解 (语义分析) 的过程, 实际上是一个根据表达寻找语义预范畴中的语义子预范畴的过程。
- 7, 多义、歧义、联想、借喻、暗喻等, 也都反映在其引申的语义子预范畴有交集或者连通。
- 8, 人们进行交流的目标, 本质上是修正各自的语义子预范畴从而形成共识的语义子预范畴, 即达成共识。知识的传承就是一个典型的共识传递的交流过程。
- 9, 这样的语义预范畴和语义子预范畴有一定的代数、拓扑、几何性质, 可以使用相关的数学工具进行研究。
- 10, 语义预范畴中的属性和概念有模糊性, 可以利用模糊数学类似工具进行相关研究。语义预范畴的有些映射关系逻辑可能是某种复杂逻辑 (类似量子逻辑?) 的投影。概率是在有未掌握的属性时, 事件在可观测的局部化属性上投影的表现。它们所产生的问题往往都是在跨越拟全/形式域时发生的。
- 11, 拟全的语义, 是存在的, 但无法用有限的表达式完整表达的。从概念看, 就是概念的指称明确 (外延可区分), 而概念的表达总不完整 (内涵说不清楚)。
- 12, 在形式化语义预范畴中, 可以使用另外的一些形式数学工具, 比如离散数学、数值分析、数理逻辑等。
- 13, 形式系统的解 (形式表达), 是在某些限定条件下的局部化和离散化, 实现对拟全解的逼近, 这个逼近过程在一定条件下可以做到任意精度。一个“好”的局部化方案, 就是抓住“主要矛盾”。
- 14, 对一个多维的局部化的形式系统, 总有另外的局部化方案, 可以将它转化为一维形式系统。选择合适的局部化转换方案, 设定合理的边界, 总能把一个局部化离散化的形式系统转化为一个逻辑系统。
- 15, 可计算的语义预范畴, 是挑选了一些属性和映射关系作为要素, 从拟全预范畴局部化的形式语义预范畴。它是不完整的, 在形式化或离散化过程中, 会丢失部分属性或映射关系的信息。
- 16, 语义的拟全/形式二重性是并行存在的。它们之间有一种类似内射/投射的过程可以互相转换。但是它们的性质相差非常大。
- 17, 形式系统是赋义的有限符号系统。对符号赋义的过程是目前无法解释的过程, 类似意识的形成或意识与物质间的鸿沟的关系。
- 18, 所有实际使用的表达系统都是形式系统, 包括自然语言、文字、绘画、音乐、舞蹈、视频、表情、语气、机器语言等, 我们可以把它们称为广义的语言。
- 19, 形式系统表达所覆盖不到的拟全的细节部分, 人会利用自己的“拟无限”、“拟存在”能力补齐。所以形式系统对拟全系统的表达, 人可以理解。
- 20, 如果世界是拟全的, 而我们突破不了表达的形式性的话, 那么我们将做不到对世界所有细节的了解 (到能清晰表达的程度)。即我们能了解世界

“在”那里，但说不清楚它“是”什么。

- 21, 人工翻译与机器翻译有一个重要的区别：人工翻译过程中，可以利用人理解的拟全语义子预范畴实现两个不同语言的形式系统之间的转换；而机器翻译过程中，两个不同语言的形式系统之间需要进行直接的映射，因此难度要大得多。
- 22, 理想化的语言处理条件下，可计算语义预范畴是明确的，而语义预范畴是范畴化、函子化的。这种条件下翻译只是一个函子操作（程序编译就是这样的？）。实际的条件远非理想化，翻译的问题实际上就是为了应对这种非理想化条件。
- 23, 自动化系统，是需求、过程和工具的形式化过程。当需求和语义明确后，自动化系统与自动翻译系统在语义处理上是等价的，只是自动化系统还有传感器作为表达来源、有执行机构执行表达生成的任务。
- 24, 很多问题的解决，可以利用拟全/形式的二重性。比如在拟全域中找到拟全解，再采用局部化手段，形式化到形式域中解决具体问题。也可以反过来，在形式域中找到近似解，再反推到拟全系统的拟全解。
- 25, 人工智能系统，是形式化的语义处理系统。上述的二重性在人工智能系统中可以有广泛的应用。
- 26, 如果要想实现拟全语义的直接处理，或者说要在人工智能中应用拟全语义模型，那么就需要对计算模型进行升级，跳出“有限状态机”的模型。

11.2 后续研究方向和应用设想

这里对本文的语义模型所涉及到的需要进一步研究探讨的问题方向和可能应用的方向做一些简单挑选整理，以供参考。

(1) 语义模型的性质研究

本文所提出的语义模型，在各方面都还远非完善。特别是关于这个语义模型的性质，还有很多的问题需要研究。

语义模型本身的数学基础就并非严谨，借用了不少并非严格的数学工具，仍然需要更多的数学研究进一步严谨化。

这个语义模型与代数几何存在一些关联：（对一个语义对象，）表达越多，对应的解的语义子预范畴越清晰。这与代数几何解的簇的概念很接近，但是不是可以直接应用？特别是，语义的表达有没有代数群的性质？能不能把语义模型环化，或者在某些条件下环化，从而应用代数几何的概形等数学工具对语义的性质进行进一步的研究？

类似的，目前我们把语义模型定义为一种“预范畴”，它是不是有条件范畴化？如果能在一定条件下范畴化和函子化，我们在文中的其他部分的内容中所提到的理想化条件是不是就可以实现了？那样，某些应用可以大大简化。如果可以，需要什么样的一些条件呢？（是机器语言的那样的条件？）

文中我们探讨过数学中的局部化在语义模型中有很明显的应用。语义模型的局部化性质是不是可以更深入地探讨下去？这种局部化与流形、概形、环的局部化有什么关系？

语义模型有什么样的拓扑性质？拓扑中常用的一些方法，比如粘接、对应、

商、边界、同调等，是不是也应该反应在语义的映射关系中？这样，有些关于奇点、无向面的一些特殊几何与拓扑也会进入语义领域，会不会给语义处理带来很大的突破空间？

同样，代数方法，群、环、域这些，是不是也可以应用到属性和映射关系的关系上，从而结合几何与拓扑，在代数几何层面对语义处理带来新思路？

关于概念的映射关系，是不是要专门研究？关于语义表达映射关系的性质，有很多问题需要研究，比如语义表达中的映射关系是不是都可以数学化？怎么数学化和怎么分类？简单的直观告诉我们，所有的数学都是可以表达的（比如老师可以教会学生），因此语义表达的映射关系应该是包括所有数学运算、变换、算子、函数、映射、映射关系等等。那么，它们能不能用这里的模型都表示出来？有没有更多的语义表达映射关系呢？有没有当前的数学处理不了的语义表达映射关系？这些问题，在这里不可能全部解答，也许需要很长时间的 research 才可能取得一些进展。

关于某些映射关系中的逻辑，是不是可以借用量子逻辑的概念，甚至进一步推广到更一般的逻辑系统，甚至某种复杂拓扑的逻辑系统，而把我们当前使用的逻辑系统看成是它的某个投影？这样的逻辑系统会给语义模型的性质研究带来什么样的便利？是不是能发掘出一些关于语义的新的性质？

属性和取值的模糊性是语义处理中普遍遇到的难题。它们为什么可以长期存在，并且虽然对语义的理解造成不便，却并未造成致命的问题导致语义系统不能工作？这些模糊性有什么的规律？是不是可以借用类似模糊数学的工具对语义的性质做更加深入的研究？

(2) 相关数学工具的研究

上一节中关于语义模型性质的研究，几乎都离不开相应的数学工具。这些研究过程中，很可能现有的数学工具^{[138][139]}会感觉“不够用”，那么是不是也需要进一步对相关的数学工具进行更深入的研究？

特别是这里的模型与数学中很多工具要求的理想条件还有比较大的差异。对此有两个方向可以进一步探索：一个方向是尽量在模型中提炼可以符合工具条件的部分，尽量充分使用现有工具；另一个方向是把数学工具中的条件放宽，研究弱条件下的新的数学工具的应用。这里主要讨论后一个方向。

比如，这里提出的“预范畴”，它的映射关系没有范畴那么好的态射性质，但感觉仍然有某些范畴的性质可以应用，是不是可以进一步研究在这种条件下映射的数学性质？

再比如上节提出的，代数几何的概形是不是合适成为语义模型的工具？如果不合适，有没有类似的方法，比如放宽环化空间的要求等条件，从而把相应的更有效的数学工具应用到研究语义模型性质？

关于模糊数学和量子逻辑，可能有更多需要深入研究的内容。这两个学科本身还不是很体系化，研究的深度似乎还没到可以大量解决实际问题的程度，包括这里关于语义的模糊性、语言逻辑的不确定性等很多困难问题。这些学科的进展，也许能为语义处理，进而为人工智能、自动化系统提供更多的解决思路 and 方案。比如这里提出的，模糊性是不是可以从接近的方向和速度梯度去建模对模糊值的明确化判别？逻辑是不是某种更复杂的“逻辑”的投影？

关于拟全化和形式化的数学基础，也是一个值得深入研究的方向。目前数学上在这方面也存在一些混淆，典型的比如经常试图用一些形式域的方法解决拟全域的问题（比如文中提到的自然数与偶数哪个多的问题）。把拟全域问题和形式域问题做个梳理，然后把拟全域问题的方法和形式域问题的方法做些分别的研究，再探索它们的差异和结合的主要研究方法。这样说不定能总结出一些被忽略的规律，找到一些新的途径。甚至，数学上如果能进一步理清拟全域和形式域工具的区别，就可能对数学本身产生很大的推动。

从以上这些列举的可能方向例子可以看出，很多数学工具和专业方向的研究，都可能推动语义模型的研究和应用，这里肯定不可能全面地列出来，很多具体数学研究方向会与实际问题结合在一起。

（3）相关学科的数学化研究

本文旨在通过用数学方法研究语义，来解决有关语义、自然/形式语言、人工智能、自动化方面的问题。但同时，由于语义与意识、认知的紧密关系，实际上也把数学更多地带进传统人文和哲学学科的领域，也希望给出一些进一步打开用数学方法研究人文和社会科学之门的提示和尝试。

比如，通过语义和意识及认知的关系，是不是可以把语义的数学化研究成果，进一步拓展到推动对认知和意识的研究？是否有可能通过对语义的数学研究，建立心理学研究的某些数学模型？

当然，由于本文在语义的数学研究方面的尝试也只是个开头，是粗浅的，并不能很深入地指出相关学科与数学的具体结合点和方向，只是为有兴趣的学者提供一个思路。

（4）跨学科的综合研究

本文涉及到的很多问题，可能是需要跨学科的综合研究来解决的，有些甚至涉及到某些古老学科的传统难题。

文中提到的人的“拟无限”能力，是本文语义模型的一个基础。但对“拟无限”的能力本身我们并没有深入研究，比如它具体是什么，它有什么性质，它是怎么来的？这里只是用一些简单的描述假设了它可能的缘起，但显然并不是严格的证明。对它的研究可能涉及到生理、心理、认知、意识等问题的深层的综合研究。比如它会不会是人有某种自动由一个环的理想生成完整的子环的能力？但这依然是在另一些基础上的问题，不是根本性的。

另一个在文中提到的问题是“赋义”的问题。赋义是将符号系统转换为可工作的形式系统的一个重要步骤。但“赋义”具体是个什么过程？它是不是一个横跨了意识与物质的鸿沟这个古老的哲学悬疑的问题？

文中提出的拟全/形式二重性问题，又是个什么性质的问题？它会不会带来一个认识论的灾难？人对某个语义的理解，是借用语言等工具对属性进行描述的，这些工具描述属性时，实际上是在形式域进行的。而我们通过前面的研究，形式域中，对一个语义的认识，是不可能做到完全的。那么，也就是说，我们要把某一个语义子预范畴“说清楚”是不可能的。这样的不可能，会带来不可知论吗？

为了解决机器的智能化问题，我们有没有可能制造有拟全化能力的机器？这也许是真正智能机器的方案？那么，这样的机器应该有什么样的特性呢？目

前拟全语义本身都只能是抽象的，不可形式化表达。那么，拟全化的要求如何通过机器实现？

更进一步的话，文中提到语义有意识属性和物质属性，那么语义的领域里，物质世界与意识世界究竟是什么样的关系呢？物质世界是语义的某种表达吗？如果在拟全层面能找到某些机制把语义中的意识与物质统一起来，是不是也是某种程度上超越物理的统一？它能填平物质与意识的鸿沟吗？它与物理正在努力的“力的统一”有什么关系吗？

（5）哲学方法论的研究

拟全化和形式化的二重性会不会对方法论产生（较大的）影响？如果数学上严格理清拟全域和形式域的区别，科学研究方法和体系会不会因此也需要重新分类？或者拟全化和形式化在思维中就是形影不离的，根本无法将二者分开？

拟全和形式对智能和认知有没有什么实质性的影响？

从符号系统到形式系统转换的赋义究竟是什么性质的过程？为什么被赋了拟全语义的符号，在形式系统中并不能进行拟全性的操作？形式规则是“智能”的基础，还是仅仅是一些机械的工具？

很多事物，在对它们进行研究的时候，受条件限制，我们并不能找到一个坚固的基础。比如语义预范畴，按照这里的模型，它可以由一些属性构建。但具体哪些属性可以选为“基”呢？这就让人犯愁了。我们可以模仿软件设计，从底向上，或从顶向下，两个思路去构建。但问题是，鉴于现在的理论和科技发展水平，顶在哪里，我们不知道，底在哪里，我们同样不知道。所以，我们能采用的方法，只能是混合式的，“从中间开始”。

有了这样的一个思路作为前提，我们在构建的过程中，也要考虑到模型向下和向上延伸的可能性。也许科技的发展真的能帮我们找到“顶”或“底”了呢？

所以，我们现在的语义预范畴的基本属性，是需要包含物理的和意识的一些属性，尽量往上下延伸，并允许这种混合性在很长时期的存在（甚至一直存在？）。

同样，对整个语义结构的理论，我们也并不刻意追究以什么样一个最基本的可以证明完全坚实的东西作为基础，也不刻意追究我们要到达一个什么样的最终顶点。甚至这里给出的答案是：我们根本到不了那里。但是我们可以不断接近。

我们这里选择的起点是我们的日常认知，“从中间开始”。然后同时往基础和顶点两个方向努力，不断探索。但并不宣称我们找到了明确的基础或到达了顶点。（也许本来就没有什么“顶”或“底”，或者“顶”就是“底”，“底”就是“顶”，就是“太极”？）

由此，是否可以扩展到关于一些哲学理念的相容性研究：很多哲学家激烈辩论的问题，有没有可能是同一个理念的不同局部化体现？我们分别从某个万花筒里看到了它。比如本文提出的拟全语义预范畴模型，是不是与黑格尔的“绝对精神”，与布拉德雷的“真理”最终是相容的？我们可以真正掌握的对它的认识永远只能是形式化的、局部的，我们会知道它的存在，对它越来越清晰，但不会完整地知道它是什么。进一步推广出去，可能还涉及到“道”、“太极”

等等，也许这些哲学理念都是相容的，可以类似物理学中力的统一那样逐渐纳入一个统一的框架之中。

另一个可以研究的哲学意义的方向是：从本文语义结构模型中共识语义预范畴（推论 3-1）衍生出的共识，能否成为一种出发点，助力解决后现代主义^{[140][141]}，特别是虚无主义所面临的困惑^{[142][143]}？

即：我们并非毫无稳定的基础，至少我们在互相交流时使用到的共识语义预范畴（没有它，我们显然就无法交流。而我们用来说服有分歧的对方所采用的某些理由，其实应该都潜藏着一些不易察觉的共识基础）可以作为基础。虽然它只是某种中间点，也并非一成不变，但至少在人类存在的前提下可以保持长期稳定。

从这个中间共识出发，世界并非虚无，我们也不能怀疑一切、否定一切。世界的终极不应该是毁灭，和平是有其实现的基础的。

（6）自然科学研究

本文研究的语义模型，有没有与自然科学的研究起到一定的相互启发的作用？比如拟全/形式的二重性和波/粒二象性有没有关系？有没有可能波函数是某种拟全方程，而所有实验遇到的实体粒子都表现为某种形式性？测量中的波函数的坍缩是不是拟全公式的在物质形态上的形式化投影？意识领域的一些现象和问题会不会与量子场论、弦论或者更一般化的物理理论有关？

生理学中神经、细胞，以及电化学反应，它们最终会影响到概念的形成过程吧？这些与语义模型有没有什么直接或间接的联系？人的传感器和感觉系统能力是不是组成语义预范畴中（意识性）属性维的基础？

再比如，大脑皮层的功能分区，是不是人类应对拟全语义形式化时的局部化要求所进化出来的处理结构？

这些问题应该都可以成为自然科学的一些可探索的研究方向。

（7）工程应用研究

语义的工程应用在文中已经举了一些例子：自然语言处理、人机接口、自动化系统、人工智能的应用、元宇宙的应用等。这里不再一一讨论。

文中对这些工程应用的探讨都还是很浅的，并没有深入。在这几个应用方向上，语义模型应该可以有很大的应用发挥空间。

这些应用有很多是跟拟全/形式二重性相关的，如果能找到一些拟全/形式之间转换的一般规律和方法，可能大大促进语义模型的工程应用。

本文提到的可计算公共语义预范畴的构建、生长与维护，以及各种专业的局部化形式子预范畴的相应的工作，以及有关语义分析和语义表达生成的具体工程，也都需要投入大量资源去实现和不断完善。

（8）人文应用研究

由于语义与认知、思维的关系密切，语义模型在人文、哲学、社会方面应该也会有广阔的应用前景。

比如对语义的研究，不可避免会涉及到对语义表达能力、表达极限和表达

方法的研究, 这些可能对文学和写作中的表达产生影响。

交流沟通中达成共识的本质是在语义预范畴中, 找到共同的语义子预范畴的过程。因此, 把语义模型应用于交流、谈判、协议, 应该可以提高交流的效率和谈判的效果。

而如果把语义模型中形成共识语义子预范畴的部分进一步深入研究和推广, 并应用到形成社会共识方面, 它应该能有效提高传播与舆论引导 [86]的能力和效果。除了以上几个人文社会方面的应用外, 还可以应用于很多其它领域, 比如语义模型在心理学、教育等领域的应用等, 这里不再一一列举了。

参考文献

- [1] 海德格尔, 在通向语言的途中, 商务印书馆, 2004.
- [2] 吴蔚天, 汉语计算语义学: 关系、关系语义场和形式分析, 电子工业出版社, 1999.
- [3] 安德拉斯·科尔内, 语义学: 从数学基础到语义语用学, 机械工业出版社, 2021.
- [4] Jonh I. Saeed, Semantics (4th Ed.), 外语教学与研究出版社, 2021.
- [5] 费尔迪南·德·索绪尔, 普通语言学教程, 商务印书馆, 1980.
- [6] Katz, J. and Jerry A. Fodor, "The structure of a semantic theory," *Language*, vol. 39, pp. 170-210, 1963.
- [7] Stalnaker R, Assertion, Syntax and semantics, New York: Academic Press, 1978.
- [8] Guillermo Del Pinal, "The Logicality of Language: Contextualism versus Semantic Minimalism," *Mind*, vol. 131, no. 522, pp. 381-427, 2022.
- [9] 陈嘉映, 语言哲学, 北京大学出版社, 2006.
- [10] 廉师友, 人工智能导论, 清华大学出版社, 2020.
- [11] Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, 人民邮电出版社, 2017.
- [12] 周志华, 机器学习, 清华大学出版社, 2015.
- [13] Paul Thagard, Mind: introduction to cognitive science, Massachusetts Institute of Technology Press, 1996.
- [14] 梁宁建, 当代认知心理学, 上海教育出版社, 2003.
- [15] 熊哲宏, 认知科学导论, 华中师范大学出版社, 2002.
- [16] 克里斯托夫·科赫, 意识探秘, 上海世纪出版集团, 2005.
- [17] 王少楠、丁鼎、林楠、张家俊、宗成庆, "语言认知与语言计算——人与机器的语言理解," *中国科学: 信息科学*, 卷 52, pp. 1748-1774, 2022.
- [18] 易绵竹、南振兴, 计算机语言学, 上海外语教育出版社, 2005.
- [19] James Allen, 自然语言理解, 电子工业出版社, 2005.
- [20] 双锴, 自然语言处理, 北京邮电大学出版社, 2021.
- [21] 许罗迈, 基于人工神经网络的机器翻译, 科学出版社, 2007.
- [22] T. Poibeau, 机器翻译, 机械工业出版社, 2019.
- [23] 张锦文, 公理集合论导引, 科学出版社, 1991.
- [24] 陆汝钤, 数学·计算·逻辑, 湖南教育出版社, 1993.

- [25] H. B. Enderton, 数理逻辑, 人民邮电出版社, 2007.
- [26] Michael Artin, Algebra, 机械工业出版社, 2008.
- [27] Michael S. Gazzaniga, Richard B. Ivry and George R. Mangun, Cognitive Neuroscience, 中国轻工业出版社, 2017.
- [28] Peter Dayan and L. F. Abbott, Theoretical Neuroscience, The MIT Press, 2001.
- [29] T.B. STEEL, Formal Language Description Languages for Computer Programming, Amsterdam: North-Holland Publishing Company, Amsterdam, 1966.
- [30] B. S. Manjunath, Philippe Salembier and Thomas Sikoka, Introduction to MPEG-7 Multimedia Content Description Interface, John Wiley & Sons, 2002.
- [31] Z. Duan, X. Yang and M. Koutny, "Framed temporal logic programming," *Science of Computer Programming*, vol. 70, no. 1, pp. 31-61, 2008.
- [32] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi and Peter F. PatelSchneider, The Description Logic Handbook, Cambridge University Press, 2007.
- [33] Stalnaker R, "Our knowledge of the internal world," *Oxford: OUP*, 2008.
- [34] E. T. T. Susan Blackmore, 人的意识, 中国轻工业出版社, 2021.
- [35] 斯坦尼斯拉斯·迪昂, 脑与意识, 浙江教育出版社, 2018.
- [36] 维特根斯坦, 逻辑哲学论, 商务印书馆, 2013.
- [37] 弗洛伊德, 梦的解析, 上海三联书店, 2008.
- [38] Nini Praetorius, Principles of Cognition, Language and Action, Kluwer Academic Publishers, 2000.
- [39] 尼采, 偶像的黄昏, 商务印书馆, 2013.
- [40] 李大烈, 智能简史, 生活读书新知三联书店, 2020.
- [41] 达尔文, 物种起源, 商务印书馆, 1995.
- [42] 罗杰·霍克, 改变心理学的 40 项研究, 人民邮电出版社, 2014.
- [43] B. N. T. Jose Luis Bermudez, Philosophy of psychology—Critical Concepts in Philosophy, Routledge, 2013.
- [44] 康德, 未来形而上学导论, 中国人民大学出版社, 2013.
- [45] Daniela Dover, "The Conversational Self," *Mind*, vol. 131, no. 521, pp. 193-230, 2022.
- [46] 唐代兴, 语义场导论, 四川大学出版社, 1998.
- [47] 迈克尔·西普塞, 计算理论导引, 机械工业出版社, 2015.
- [48] Noam Chomsky, Syntactic Structures, Mouton & Co.'s-Gravenhage, 1957.
- [49] Noam Chomsky, Language and Mind, Harcourt Brace & World Inc., 1968.
- [50] Noam Chomsky, On Nature and Language, Cambridge University Press, 2002.
- [51] Mark Pinder, "Conceptual Engineering, Metasemantic Externalism and Speaker-Meaning," *Mind*, vol. 130, no. 517, pp. 141-163, Jan 2021.
- [52] Julien Murzi and Brett Topey, "Categoricity by convention," *Philosophical Studies*, vol. 178, pp. 3391-3420, Oct 2021.
- [53] 赵汀阳, "动词存在论与创造者视域," *中国社会科学*, 卷 8, pp. 24-45, 2022.
- [54] R. Penrose, The Emperor's New Mind, Oxford University Press, 1989.

- [55] 沙少海、徐子宏, 老子全译, 贵州人民出版社, 1989.
- [56] 冯国超, ““道可道, 非常道”新解,” *中国社会科学*, 卷 6, pp. 36-51, 2022.
- [57] 王献唐, 公孙龙子研究, 北京理工大学出版社, 2020.
- [58] 赵昌平, 唐诗三百首全解, 复旦大学出版社, 2013.
- [59] 王晨, 大宋之愁-苏轼与辛弃疾的诗词人生, 上海科学技术文献出版社, 2020.
- [60] 庄周, 庄子, 团结出版社, 2017.
- [61] McCarthy and John, "Ascribing mental qualities to machines, Formalizing common sense," *V. Lifschitz Ablex*, pp. 93-118, 1979.
- [62] 郭纯洁, 语义概念学研究, 湖北人民出版社, 2020.
- [63] 朗道、栗弗席兹, 力学, 高等教育出版社, 2007.
- [64] 王正行, 近代物理学, 北京大学出版社, 2010.
- [65] David J. Griffiths, *Introduction to Electrodynamics*, 世界图书出版公司, 2005.
- [66] J. D. Jackson, *Classical Electrodynamics*, 高等教育出版社, 2004.
- [67] K. F. Riley, M. P. Hobson and S. J. Bence, *Mathematical Methods for Physics and Engineering*, Cambridge University Press, 2006.
- [68] Tom Leinster, *Basic Category Theory*, Cambridge University Press, 2014.
- [69] W. Rudin, *Principles of Mathematical Analysis*, 机械工业出版社, 2014.
- [70] Sean M. Carroll, *Spacetime and Geometry*, 世界图书出版公司, 2007.
- [71] Osgood, Charles E., William S. May and Murray S. Miron, *Cross Cultural Universals of Affective Meaning*, University of Illinois Press, 1979.
- [72] M. M. Yuri I. Manin, *Semantic Spaces*, arXiv:1605.04238v1 [cs.CL] 13 May 2016, 2016.
- [73] L. G. Valiant, "A theory of the learnable," *Communications of the ACM*, vol. 27, no. 11, pp. 1134-1142, 1984.
- [74] Qing Liu, *Algebraic Geometry and Arithmetic Curves*, Oxford University Press, 2002.
- [75] J. P. May, *A Concise Course in Algebraic Topology*, 世界图书出版公司, 2019.
- [76] Robin Hartshorne, *Algebraic Geometry*, 世界图书出版公司, 2011.
- [77] Ravi Vakil, *Foundations of Algebraic Geometry*, math216.wordpress.com, 2017.
- [78] Joe Harris, *Algebraic Geometry --- A First Course*, Springer, 1992.
- [79] I. R. Shafarevich, *Basic Algebraic Geometry*, 世界图书出版公司, 2017.
- [80] 夏道行、吴卓人、严绍宗、舒五昌, 实变函数论与泛函分析, 高等教育出版社, 2010.
- [81] 薛小平, 应用泛函分析, 哈尔滨工业大学出版社, 2002.
- [82] R. Stalnaker, "Common ground," *Linguistics and Philosophy*, vol. 25, no. 5-6, pp. 701-721, 2002.
- [83] Jens Kipper, "When Lingens meets Frege: communication without common ground," *Philosophical Studies*, vol. 178, pp. 1441-1461, May 2021.
- [84] Williams B, *Descartes: The Project of Pure Enquiry*, New York: Routledge, 1978.
- [85] Joseph Bjelde, "Knowledge is Teachable.," *Mind*, vol. 130, no. 518, pp.

- 475-502, April 2021.
- [86] 古斯塔夫·勒庞, 乌合之众, 新世界出版社, 2010.
 - [87] David J. Griffiths, 量子力学概论, 机械工业出版社, 2009.
 - [88] P. A. M. Dirac, The Principles of Quantum Mechanics, 科学出版社, 2008.
 - [89] J. J. Sakurai and Jim Napolitano, Modern Quantum Mechanics, 世界图书出版公司, 2011.
 - [90] Mark Srednicki, Quantum Field Theory, 世界图书出版公司, 2010.
 - [91] A. Zee, Quantum Field Theory in a Nutshell, 世界图书出版公司, 2013.
 - [92] Barton Zwiebach, A First Course in String Theory, Cambridge University Press, 2009.
 - [93] Elias Kiritsis, String Theory in a Nutshell, 世界图书出版公司, 2013.
 - [94] V. I. Arnold, V. V. Goryunov, O. V. Lyashko and V. A. Vasil'ev, Singularity Theory I, Springer, 1998.
 - [95] M. F. Atiyah and I. G. MacDonald, Introduction to Commutative Algebra, Addison-Wesley Publishing Company, 1969.
 - [96] John Stillwell, Naïve Lie Theory, 世界图书出版公司, 2011.
 - [97] 李克正, 交换代数与同调代数, 科学出版社, 2017.
 - [98] Brian C. Hall, Lie Groups, Lie Algebras, and Representations, Springer International Publishing, 2015.
 - [99] Sheldon Axler, Linear Algebra Done Right, Springer, 1997.
 - [100] Manfredo Perdigao Do Carmo, Riemannian Geometry, 世界图书出版公司, 2008.
 - [101] 唐梓洲, 黎曼几何基础, 北京师范大学出版集团, 2011.
 - [102] Richard L. Bishop and Samuel I. Goldberg, Tensor Analysis on Manifolds, Dover Publications, Inc., 1968.
 - [103] Karen E. Smith, Lauri Kahanpää, Pekka Kekalainen and William Traves, An Invitation to Algebraic Geometry, 世界图书出版公司, 2010.
 - [104] 尤承业, 基础拓扑学讲义, 北京大学出版社, 1997.
 - [105] Colin C. Adams, The Knot Book, American Mathematical Society, 2004.
 - [106] J. R. Munkres, Topology, 机械工业出版社, 2018.
 - [107] J. R. 曼克勒斯, 代数拓扑基础, 科学出版社, 2006.
 - [108] Allen Hatcher, Algebraic Topology, Cambridge University Press, 2002.
 - [109] Klaus Janich, Topology, 世界图书出版公司, 2011.
 - [110] D. Siersma, C. T. C. Wall and V. Zakalyukin, New Developments in Singularity Theory, Springer-Science+Business Media, B. V., 2000.
 - [111] Ib Madsen and Jorgen Tornehave, From calculus to cohomology, Cambridge University Press, 1997.
 - [112] Andres Y. Lee, "Modeling Mental Qualities," *Philosophical Review*, vol. 130, no. 2, April 2021.
 - [113] 张博侃, 模糊数学, 北京大学出版社, 2021.
 - [114] C. C. Irving M. Copi, 逻辑学导论, 中国人民大学出版社, 2007.
 - [115] 王伟长, 量子力学的逻辑与逻辑多元主义, 中国社会科学出版社, 2021.
 - [116] 张江, 图灵机与计算问题.
 - [117] Kornai and Andras, "Euclidean Automata", Implementing Selves with

- Safe Motivational Systems and Self-Improvement," *Mark Waser Proc. AAAI Spring*, pp. 25-30, 2014.
- [118] Kornai and Andras, "Finite automata with continuous input," in *Short papers from the sixth workshop on Non-classical Models of Automata and applications*, 2014.
- [119] Beau Madison Mount and Daniel Waxman, "Stable and Unstable Theories of Truth and Syntax," *Mind*, vol. 130, no. 518, pp. 439-474, April 2021.
- [120] 欧内斯特·内格尔、詹姆士·R·纽曼, 哥德尔证明, 中国人民大学出版社, 2008.
- [121] 李晓培、陈小亘, 离散数学, 复旦大学出版社, 2016.
- [122] 李庆扬、王能超、易大义, 数值分析, 清华大学出版社, 2008.
- [123] 毛泽东, 毛泽东选集, 人民出版社, 1991.
- [124] 罗铸楷, 多值逻辑理论及应用研究, 国防科技大学出版社, 2003.
- [125] D. T. M. W. T. P. L. A. T. N. G. A. A. & M. N. Marc-Olivier Renou, "Quantum theory based on real numbers can be experimentally falsified," *Nature*, vol. 600, pp. 625-629, 2021.
- [126] C. W. F.-M. L. J.-W. W. C. Y. Z.-X. S. Y. W. M. G. H. D. F.-T. L. Q. Z. C.-Z. P. X. Z. A. C. C.-Y. L. a. J.-W. P. Ming-Cheng Chen, "Ruling Out Real-Valued Standard Formalism of Quantum Theory," *Physical Review Letters*, no. 1, 2022.
- [127] K. L. Chung, A Course in Probability Theory, 机械工业出版社, 2017.
- [128] 钟义信, 机器知行学原理, 科学出版社, 2007.
- [129] 苗德成、奚建清、苏锦铤, "基于范畴论的形式语言模型建模方法研究," *计算机应用与软件*, 卷 32, 编号 4, pp. 19-23, 2015.
- [130] "ATSC: "Digital Television Standard", " 9 1995.
- [131] "ISO/IEC 13818-1: "Coding of moving pictures and associated audio - Part1: Systems", " 1996.
- [132] "ISO/IEC 13818-2: "Coding of moving pictures and associated audio - Part2: Video", " 1996.
- [133] "ISO/IEC 13818-3: "Coding of moving pictures and associated audio - Part3: Audio", " 1996.
- [134] "ETSI ETR 290: "Digital Video Broadcasting (DVB); Measurement guidelines for DVB systems", " 1997.
- [135] "ETSI ETS 300 468: "Digital Video Broadcasting (DVB); Specification for Service Information (SI) in DVB systems", " 1998.
- [136] "PCISIG: "PCI Local Bus Specification", Revision 2.2, " 1998.
- [137] 拉. P. N. 拉奥, 脑机接口导论, 机械工业出版社, 2016.
- [138] 郑鍊, 当代数学的若干理论与方法, 华东理工大学出版社, 2002.
- [139] Thomas A. Garrity, All the Mathematics You Missed, 清华大学出版社, 2002.
- [140] 伊利亚·普利果金, 确定性的终结 : 时间、混沌与新自然法则, 上海科技教育出版社, 1998.
- [141] 凯文·凯利, 失控, 新星出版社, 2012.
- [142] 劳思光, 当代西方思想的困局, 华东师范大学出版社, 2016.
- [143] 洪汉鼎, "论哲学诠释学的阐释概念," *中国社会科学*, 卷 7, pp. 114-139, 2021.
- [144] Milo Phillips-Brown, "What Does Decision Theory Have to Do with

Wanting," *Mind*, vol. 130, no. 518, pp. 413-438, 2021.

[145] Francesco Berto and Peter Hawke, "Knowablity Relative to Information," *Mind*, vol. 130, no. 157, pp. 1-35, Jan 2021.

作者简介:

惠新标, 男, 1971 年生, 博士, 教授级高工, 研究方向: 视音频通信, 语义结构。Email: hxb@figure-it.com, hxb@ithscience.com。

陈志强, 男, 1984 年生, 学士, 研究方向: 视音频算法, 人工智能。

叶楠, 女, 1973 年生, 博士, 研究方向: 视音频通信, 确定性网络。

周德雨, 男, 1969 年生, 硕士, 研究方向: 视音频通信, 金融算法。

作者贡献声明

惠新标: 主要研究者, 整体思路提供者。

陈志强: 人工智能部分研究, 实验及数据图表提供者。

叶楠: 可计算语义部分试验思路提供者。

周德雨: 论文讨论、协调及修改。