从计算机处理的角度看短语结构歧义*

詹 卫 东

北京大学中文系 100871

摘　要

本文分析了计算机对汉语短语进行结构定界和结构关系判定时产生歧义的不同类型，从歧义格式跟歧义实例的关系角度，区分为真歧义、伪歧义和准歧义三种不同情况；从歧义是否造成短语功能差异，亦即结构歧义对外是否发生影响的角度，分为自囿型歧义和他囿型歧义两类。对汉语短语结构歧义做上述类型的区分，可以为进一步解决歧义格式的分析问题提供有效的支持。

Abstract

This paper analyses the ambiguity of determining by computer boundaries and constructional relations of Chinese phrases. The type of ambiguity can be classified from two different perspectives. As viewed from differentiated types of relation between type and token, ambiguous phrases can be classified into three kinds: the true-ambiguity, the pseudo-ambiguity, and the quasi-ambiguity, and as viewed from the influence of ambiguity, ambiguous phrases can be classified into two kinds: the closed-ambiguity and the opened-ambiguity. The author hopes that the classification mentioned above conduces to solve the problem of phrase structure ambiguities in Chinese.

§0 引　言

语言学界对自然语言歧义现象的研究由来已久，并已取得了不少成果[1]。以往的研究主要是从人的角度出发。有没有歧义，是什么性质的歧义，都由人来判断。同样，研究解决歧义的办法也主要是考虑人学习运用语言，认识语言规律的需要。现在，随着计算机自然语言处理理论和技术研究工作的不断发展，自然语言歧义问题就又有了新的研究视角。从计算机的角度来观察歧义现象，跟从人的角度着眼多有不同。突出地表现在两个方面：一是辨识歧义的能力不同。有的语言形式对人而言是有歧义的，计算机却视若不见；有的语言形式对人而言没有歧义，计算机却绞尽脑汁理解出几种意思来。二是消除歧义的能力不同。人具备语言知识、百科常识、联想能力和逻辑推理能力等，对实际篇章中出现的句子大都能消除其潜在歧义，做到准确理解。计算机的排歧能力则大为逊色，不易根据语境条件的制约来确定句子合理的解释[2]。

本文即尝试以计算机的眼光来观察汉语短语结构的歧义现象。之所以选择在短语结构层面讨论歧义问题，是受目前计算机处理自然语言水平的限制。对于自然语言中诸如语义结构关系歧义、语义指向歧义、语用歧义等等复杂的歧义现象，计算机现阶段基本上还处在视若不见的地步，无论在

* 本文研究得到北京大学中文系陆俭明教授、北京大学计算语言学研究所俞士汶教授的指导，特此致谢。

辨识还是消歧方面都是力不从心。而计算机对汉语短语结构的分析，则往往是歧义迭出，常常能把对人而言无歧义的句子分析出多个结果来。如何消除这些歧义结果对提高目前计算机汉语信息处理技术水平就显得十分迫切和必要。因此本文打算尽力对不同性质类型的短语结构歧义做一番剖析和探索。看看计算机在识别和消解短语结构定界歧义和结构关系歧义方面有何特点。

§1 短语结构定界歧义和结构关系判定歧义

计算机对句子进行短语结构分析，包括确定结构体界限和判定结构内部的语法关系，有可能得到多个结果，就是所谓的结构定界歧义和结构关系歧义。大致说来，结构定界歧义也就相当于人对句子进行层次分析时碰到的一个表层线性形式有多种不同层次切分的情形；结构关系歧义也即两个直接成分之间有一种以上的语法关系。前者针对两个以上语言成分的分析而言，后者则针对两个语言成分而言。这一节我们来说明什么样的符号排列可能会造成短语结构定界歧义和结构关系判定歧义。

首先我们以三个符号串的序列为考察对象，研究可能造成短语结构定界歧义的排列格式。假设有 A、B、C 三个功能标记，它们连续排列为 ABC 形式。如果：（1）B 跟 A 先组合后能再跟 C 组合形成更大的组合体；或者（2）B 跟 C 先组合后再跟 A 组合也能形成更大的组合体。二者同时为真。则 ABC 会发生结构定界歧义。如果只存在上述两种情况中的一种，或者 A、B、C 三者根本不能组合成更大的结构体（即不能构成合法的语法形式），也即（1）（2）不同时为真，则排列式 ABC 没有结构定界歧义。举例说明如下：

1． VP U NP[3]

这样三个功能类的排列形式，只有一种可能的组合方式即[[VP U] NP]，因为助词 U 是后定位功能成分，只能跟前面的成分组合，不可能跟其后面的成分发生组合关系。这种排列不会有定界问题。例如: [[看 了] 三场电影]

2. PP NP NP

PP 跟 NP 不成结构，NP 跟 NP 发生组合关系后形成的结构体也不可能跟 PP 组合成更大的结构体。这样，这三个类的排列也不存在定界问题。

3. VP AP NP

这三个功能标记的排列式存在两种组合的可能性。a. [VP [AP NP]] 或者 b. [[VP AP] NP]。我们可以在语言中找到相应的实例。如：

（1） 踢新球 组合方式为 a [踢 [新 球]]

（2） 踢碎玻璃瓶 组合方式为 b [[踢 碎] 玻璃瓶]

（3） 踢破球 有歧义， 可以是 a [踢 [破 球]]

也可以是 b [[踢 破] 球]

前两例对人而言都是单义的。第三例对人来说也有歧义。口语中靠重音位置和节奏停顿可以区分开。按 a 切分时，“破”读重音，原调。“踢”跟“破”之间有明显较长的语音间隔（跟“破”和“球”之间的间隔相对而言）；按 b 切分时，“破”可以轻读，“踢”跟“破”之间连得很紧密。在书面上，人也要在更大语言环境中才可能准确判断是哪种意思。

跟结构体定界歧义一样，结构关系判定歧义问题也不是无条件地存在于任意两个符号之间。一般说来，实语类跟实语类组合的短语发生结构关系歧义的可能性大。实语类跟虚语类组合的短语发生结构关系歧义的可能性小。如：

4. P NP

介词 P 是虚词属虚语类范畴，跟属实语类的名词短语 NP 只能构成介宾关系。如：

被公司（解雇） 把茶（喝了） 等。

5. VP NP

VP 跟 NP 都是实语类。二者组合，VP 跟 NP 可以构成述宾关系。如：打击/v 犯罪分子/n；也可以构成定中关系。如：抄袭/v 行为/n

由上面分析可知，跟结构体定界歧义问题有关的排列式只是如例 3 这样的情况。跟结构关系判定歧义问题有关的排列式则是象例 5 这样的组合。而对计算机处理来说，类似上面这些可能造成歧义的排列格式，内部还有不同的类型，各自又有不同的特点。本文下面的主要篇幅就来讨论存在定界歧义问题和结构关系判定歧义问题的短语结构有哪些不同类型和特点。所有的讨论分析，都是我们为解释计算机何以会把对人无歧义的语言形式看成歧义结构所做的努力。我们希望通过这样的探索，能够把计算机分析短语结构碰到的歧义问题的性质进一步澄清，并进而在考虑采用何种消歧策略时能做到有的放矢、事半功倍。

§2　真歧义

象上节例 3 那样的排列式，从类（type）的组合来看，有两种结构定界方式；从例（token）的表现来看，有实际的歧义例子（如：“踢破球”）。计算机在分析这类短语结构时，发生定界歧义，我们称之为真歧义。也就是指功能类组合时存在的结构定界歧义可以投射（mapping）到一个具体的自然语言形式上，类的歧义可以在语言中很容易地找到同形异构异义的实例。再例如：

1. VP　VP　U<了>[4]

这三个功能类的排列也存在两种组合可能：

a [VP　[VP　U<了>]]　　如：　[挤上巴士　[跑　了]]

b [[VP　VP]　U<了>]　　如：　[[引起　争吵]　了]

a 跟 b 对应的两个实例分别都是单义的，只有一种层次结构。上例只能作 a 解；下例只能作 b 解。而下面的实例就是歧义短语了。层次上可以有两种不同的构造，意思上也可以有不同的理解。如：　拄着拐杖走了

可以理解为 a　[拄着拐杖　[走　了]]　　老王没留下吃饭就拄着拐杖走了

也可理解为 b　[[拄着拐杖　走]　了]　　老王的腿好得很快，现在能拄着拐杖走了

做 a 式切分，意为“走了”，是“拄着拐杖走的”；做 b 式切分，可以表示以前不能“走”，现在能“拄着拐杖走”这样的意思，表示一种变化。

还有一种情形是在结构层次上可以做两种切分，但意义理解上没有什么差别。也就是通常所说的多切分的情况。如：　带着一家很快就回到上海了

可以按 a 切分　[带着一家　[很快就回到上海　了]]

也可按 b 切分　[[带着一家　很快就回到上海]　了]

两种层次都表示一样的意思。

从上面例子可以看出，真歧义的排列式，它投射产生的短语实例，在结构定界和意义理解上，存在三种情况：单定界单义、多定界单义、和多定界多义。其中最后一种情况是真歧义排列式的充分必要条件。再看一个例子：

2. VP　D<不>　AP[5]

有两种定界可能：　a　[VP　[D　AP]]　　如：　[办事　[不　认真]]

b　[VP　D　AP]　　如：　[洗　不　干净]

整个结构体按 a 式组合，是主谓结构，VP 作主语；按 b 式组合，是述补结构。上面这两个实例都是单定界单义的。我们可以再找两个多定界多义的例子。如：

写不好　　a　[写　[不　好]]，不写也不好

b　这个字我 [写　不　好]

解释不清楚　a　论文对语言事实的描写很详细，但[解释　[不　清楚]]

b　小王[解释　不　清楚]事情的原因

这两个短语都是歧义实例。结构定界不同，意义也不同。因此“VP　D<不>　AP”排列式属真歧义类型的组合格式。

短语结构定界有歧义，短语内部结构关系当然也跟着发生歧义。因此，结构定界真歧义格式同

时也就意味着结构关系歧义也是真歧义类型。这里就不另举例了。

真歧义的歧义实例，对人来说也是有歧义的，计算机处理则显然是个难题。

§3 伪歧义

跟真歧义的格式相反，计算机在确定短语界限时，还会碰到我们称之为伪歧义的情况。所谓伪歧义，是指功能类组合时有不止一种可能的定界方式，但映射到实例时，不同的组合方式并不导致意义理解上的不同；或者是实例只有唯一的合理的结构层次构造，亦即是单一结构定界单一意义的。这就是说，对类而言有多种组合可能，对实例而言则或者多切分或者单一切分，意思理解上都没有歧义。例如：

1. DP VP NP

从抽象的类的组合来看，这个排列式可以有两种层次构造：

a. [DP [VP NP]]

b. [[DP VP] NP]

按 a 式组合，是副词性短语修饰一个动词性短语，整个结构体是状中式构造；按 b 式组合，则是副词性成分先修饰一个动词性短语，形成的状中式 VP 再带上一个名词性成分构成述宾结构 VP。下面的实例就可以有这样两种切分。

已经学过语言学 a [已经 [学过 语言学]]

b [[已经 学过] 语言学]

就这个短语而言，a、b 两种切分都可以接受。意义理解上几乎没有差别。是一般所说的多切分结构。而下面这个实例就只有一种合理的组合方式。

很看了几场电影 a [很 [看了 几场电影]]

* b [[很 看了] 几场电影]

必须是“很”修饰整个述宾结构构成状中式 VP。不能是“很”先修饰“看了”，“很看了”不成结构。对这个例子而言，只有唯一的正确的层次组合，不是两可的。再比如：

2. PP VP NP

跟上面的排列式一样，也可以有 a [PP [VP NP]] 和 b [[PP VP] NP]两种组合可能。但对应的实例通常却都是只能按 a 式组合的。如：

[在林子里 [追 野兔]]

[从南方 [跑到 北方]]

[把桌子 [撞断了 一条腿]]

也有似乎按 a 或按 b 切分两可的例子，但两种不同的结构定界，意思上没有什么差别。如：

按原订计划处理小王的问题

可以按 a 式组合： [按原订计划 [处理 小王的问题]

也可按 b 式组合： [[按原订计划 处理] 小王的问题]

上面这两个排列式，都是伪歧义的情形。机器分析这类排列式时，应该比较容易确定 VP 的前后界限。我们可以直接规定上两例都按 a 式组合。这样，就是对类的组合而言也不存在歧义了。这也正是我们称之为伪歧义的原因。显然，结构定界的伪歧义格式在结构关系判定上也不会发生歧义问题。

伪歧义的排列式，无论格式类型还是实例，对人都无歧义，对计算机处理也不造成困难。

§4 准歧义

准歧义是介于真歧义和伪歧义之间的一种歧义类型。对结构定界歧义来说，是指功能类排列式在组合时有多种层次切分的可能性，而对应到某个具体实例，则只能选择其中一种合理的结构定界方式，只有一种意义理解。也就是说，有类的歧义，而具体实例则没有歧义[6]。举例说明如下：

1. PP VP VP

有两种组合方式　　a　　[PP　[VP　VP]]

b　　[[PP　VP]　VP]

a 式 PP 后面的两个 VP 先构成一个结构体，然后被 PP 修饰，整个组合成状中结构 VP。b 式 PP 先修饰相邻的 VP，构成的状中式 VP 再跟后面的 VP 组合。

由这个排列式映射的短语实例，人来判断，要么按 a 式组合，要么按 b 式组合，一般只有一种定界方式，没有同形异构异义的现象。下面对比举例说明。

1A　　[把被子　[拿出去　晒一晒]]

1B　　[[把饭　吃完]　做作业]

2A　　[被黑社会　[敲诈　勒索]]

2B　　[[被老师　批评]　写了检查]

3A　　[在操场上　[跑步　打球]]

3B　　[[在操场上　跑步]　成了我的业余爱好]

4A　　[沿长江　[走了几个省　考察了一些历史遗迹]]

4B　　[[沿长江　旅游]　会有很多新发现]

上面四组八个短语，A 都是按 a 式切分的，B 都是按 b 式切分的。每句都是单义，只是内部结构关系有些不同。按 A 式切分的四句都是状中结构 VP；按 b 式切分的四句中，1B 和 2B 短语内部是连谓结构关系；3B 和 4B 短语内部是主谓结构关系。

如果要计算机来分析这些实例，却未必能跟人一样，该按 a 方式切分的按 a 切，该按 b 方式切分的就按 b 切。由于对一个准歧义格式对应的实例该怎样进行结构分析缺少必要的语言知识，计算机经常是张冠李戴，分析得到错误的结果。

下面再简单看看另一个准歧义格式的例子。

2.　VP　NP　NP

有两种切分可能：　a　　[VP　[NP　NP]]

b　　[[VP　NP]　NP]

两种切分方式构成的组合体的功能类都是 VP。a 式是两个 NP 先构成定中结构再做 VP 的宾语；b 式是 VP 带 NP 宾语后构成的述宾结构又带一个 NP 宾语，述宾结构套叠构成述宾结构 VP。可以看下面的实例：

1A　　[出任　[公司　总经理]]

1B　　[[送　图书馆]　三箱书]

2A　　[买　[猪皮　大衣]]

2B　　[[运　济南]　十吨钢材]

这四个例子中，1A 和 2A 按 a 式切分；1B 和 2B 按 b 式切分。1A“公司总经理”作“出任”的宾语。1B 则“送图书馆”带“三箱书”作宾语。每句都只有一种切分方式，都是单义实例。

短语结构关系判定同样也存在准歧义格式的情况。可参见下文§6 节例 5。这里就不细说了。准歧义的排列式，格式类型有歧义；实例对人无歧义，对计算机则在不同程度上构成歧义。

以上从排列式属类歧义还是实例歧义这个角度，分出了真歧义、伪歧义、准歧义三类。不难看出，虽说是从计算机着眼，但最终还是以人为标准。真歧义的歧义实例,人离开一定语境也无法消歧，是格式固有的歧义；伪歧义的情况，对人和计算机理解都不造成麻烦；准歧义介于二者之间，格式有类的歧义，人具备语法知识，在理解具体实例时不会有问题。计算机如果也掌握了相应的知识，就可以跟人一样做准确的结构定界判断了。

除从类歧义跟实例歧义的不同对应关系考察歧义格式外，下面我们还将从歧义造成的影响是局限于自身还是波及其他短语结构这一角度，来进一步讨论。

§5　自囿型歧义

我们仍然从§1 节例 3 那个真歧义排列式开始。对两种可能的分析结果 a 和 b 来说，最后整个组合体的功能标记都一样，都是动词性短语 VP，对外表现为同样的功能。像这样的结构定界歧义情况，我们称之为自囿型歧义。一般而言，自囿型歧义的影响局限在短语结构内部。再举一个准歧义排列式的例子。如：

1. DP　VP　VP

这三个功能类排列形成的自囿型结构定界歧义，可以有两种组合方式三种内部结构关系，而对应的外部功能标记则都是 VP [7]。例如：

a　[[DP　VP]　VP]

a1　[[大力　培养年青人]　造就了一批人才]

a2　[[努力地　学习]　刻苦地钻研]

b　[DP　[VP　VP]]

[很　[喜欢　看电影]]；　[欢快地　[唱着歌　跳着舞]]

a1 跟 a2 结构定界方式一样，但各自结构体内部句法关系有不同，a1 是连谓，a2 是联合；b 式组合的内部结构关系是状中。三种不同情况，相应的短语外部功能却无差别，都是 VP。这种歧义格式少有歧义实例，属准歧义类型。

跟结构定界歧义一样，短语结构关系歧义也有类似情况。如：

2. VP　VP

VP 跟 VP 的结构关系可以有下面四种情况。

a　述宾关系：　赞成/vp 打排球/vp

b　述补关系：　搬/vp 出去/vp

c　连谓关系：　关灯/vp 睡觉/vp

d　联合关系：　写字/vp 画画/vp

但这四种不同的结构关系对应的外部功能类并没有什么不同，都是动词性短语 VP。

上面四个例子都是单义的，而象 e.“想出来”，既可以理解为述宾关系也可以是述补关系，就是歧义实例了。比较:

e1. 他在里面呆久了，想出来。　（述宾关系）

e2. 这个问题我终于想出来了。　（述补关系）

再比如 f.“骑马打球”这样的例子，可以理解为

f1. “骑着马打球”，“骑马”是“打球”的伴随方式；

也可理解为：　f2. “骑马”和“打球”两种运动并举。

前者是连谓结构关系，后者是联合结构关系，也是歧义实例。

需要说明一下，象“打下去”这样的例子，也可以有两种理解：

(Ⅰ)　我们终于把上山的敌人打下去了

(Ⅱ)　我们两家不能再这样打下去了

“打下去”内部语义关系不同，“下去”在(Ⅰ)中表示位移趋向；在(Ⅱ)中表示动作行为的延续。但从结构关系上说，两种情况下都属述补关系。因此这样的例子不是我们讨论的结构关系歧义的情况。

3. VP　AP

VP 跟 AP 组合，内部结构关系可以有两种情况。

a　述宾关系：　喜欢/vp 安静/ap

b　述补关系：　洗/vp 干净/ap

相应的外部功能类则只是 VP 一种。

上两例都是单义的例子。而象“说清楚”，就是可以有两种理解的歧义实例了。既可以理解为“说”的内容是“清楚”（例如：我问他讲得清楚不清楚，他说清楚），也可以理解成“说”的结果是“清楚”（例如：这个问题你得说清楚）。前者是述宾关系；后者是述补关系。功能上都是动

词性短语。

自囿型歧义的情况，由于内部歧义不影响向外组合，缺乏相应的外部功能差异表现，因而这种类型的歧义不大容易根据外部语境排歧。与之相对的就是内部歧义连带会有不同的外部功能表现的情况。我们称之为他囿型歧义。下面讨论这类歧义情形。

§6 他囿型歧义

他囿型结构体定界歧义是指结构体的组合方式不同，对应的外部功能标记也不同的情况。这样的歧义排列形式是比较常见的。简单举两例如下：

1. VP VP NP

有两种结构定界方式，三种可能的功能标记：

a [VP [VP NP]]

a1 外部功能标记为 VP [喜欢 [看 京剧]] [提高 [修理 技术]]

a2 外部功能标记为 DJ [游泳 [治好了 他的关节炎]]

b [[VP VP] NP]

b1 外部功能标记为 VP [[贯彻 执行] 党的方针政策]

b2 外部功能标记为 NP [[骑马 射箭] 技术]

可以看到，这个歧义格式可以有 VP（a1、b1）、DJ（a2）和 NP（b2）三种不同的外部功能表现。

2. NP VP NP

可以有两种结构定界方式，相应的则有两种不同的功能标记：

a [[NP VP] NP]

外部功能标记为 NP。 例如： [[方言 调查] 课] [[语法 分析] 方法]

b [NP [VP NP]]

外部功能标记为 DJ。 例如： [小王 [看了 十本书]]

下面再看看判定结构关系时，他囿型歧义的情况。

他囿型结构关系歧义指结构关系不同组合体功能类也不同的歧义情况。例如：

3. VP NP

这是现代汉语中比较常见的一种歧义组合。VP 跟 NP 组合，内部可以有述宾和偏正两种结构关系，分别对应 VP 和 NP 两个功能类。例:

a 述宾关系－－VP 炸/vp 碉堡/np

b 偏正关系－－NP 研究/vp 机构/np

上面两个例子各自都是没有歧义的。而象“出租汽车”、“学习文件”、“炒饭”这样的例子，就都是有两种理解的歧义短语实例了。

4. NP VP

NP 跟 VP 组合可以有下面两种结构关系，分属两个功能类。

a 主谓关系－－DJ 大会/np 正式开始/vp

b 定中关系－－NP 城市/np 管理/vp

上面两个例子各自都只有一种意思。而象“企业赞助”这样的例子，内部就有主谓和定中两种可能的结构关系，是歧义实例。

5. AP VP

AP 跟 VP 组合也可以有两种结构关系，分属两个功能类。

a 状中关系－－VP 认真/ap 工作/vp

b 主谓关系－－DJ 干旱/ap 带来了饥荒/vp

这种格式组合，结构关系歧义一般少有歧义实例。从格式歧义和实例歧义的对应角度讲，也就是所谓准歧义的情况，属于结构关系判定准歧义组合格式。

下面我们以一个他囿型歧义的短语实例在特定语境条件下，歧义可以自然消除的现象来说明，他囿型歧义格式跟语境存在相互制约的关系，比自囿型歧义相对容易找消歧条件。

本节例 1.VP　VP　NP 排列格式，其实例＂学习训练方法＂是一个歧义短语。可以有两种结构定界方式：a.　[学习　[训练　方法]]　整个短语是述宾构造，功能是 VP。

b.　[[学习　训练]　方法]　整个短语是偏正构造，功能是 NP。

我们构造语境 A：教练们正在学习训练方法。－－>按 a 式定界

B：这种学习训练方法很好。　－－>按 b 式定界

显然，在语境 A 中，＂学习训练方法＂受副词＂正在＂的修饰，功能类一定是 VP，NP 短语一般不出现在这个位置；而在语境 B 中，＂学习训练方法＂受指量短语＂这种＂修饰，功能类应是 NP，VP 短语一般不允许出现在这个位置。歧义自然得以消解。

当然也存在有的语境对两种功能类短语都允许出现的情况，比如在＂学习训练方法比较好＂中，主语位置就是 NP、VP 两可的，既可以指＂方法比较好＂，也可以指＂相对于学别的东西，学训练方法要好一些＂。

但不管怎么说，他囿型歧义的格式，毕竟是不同的意思连带有不同的外部功能表现，相对于自囿型歧义格式的短语，消歧条件要多一些。

§7　结　论

本文对短语结构定界歧义和结构关系判定歧义的不同类型作了分类描写。现列表比较如下：（见下页）

通过上面的分析归纳，我们对计算机辨识短语结构歧义的特点就有了更清楚的认识。计算机之所以会把对人无歧义的短语结构分析出多个结果，就是因为大量的表面没有歧义的短语实例,其深层结构格式却是真歧义或准歧义。在从模式向实例映射的过程中，人掌握的语言知识和其他知识都可以帮助人进行准确判断。计算机必须具备同样的知识，才可能也对一个自然语言的表层线性序列，做出正确的结构分析。针对不同的歧义类型，相应的消歧策略应该有不同的考虑。自囿型歧义跟他囿型歧义比较，他囿型歧义相对容易在更大的语境中自动消歧，而自囿型歧义的消解对外部语境的依赖却难以把握，难度也就大一些。比如：“踢破球”是自囿型歧义的实例。它在语境(1) “踢破球不如踢新球好玩”中是[踢　[破　球]]结构；在语境(2) “踢破球会被罚款” 中是[[踢　破]　球]结构。人区分这两种不同的意思，更多的恐怕是根据常识或者说是意义关联上的因素。而这些知识目前还难以形式化表达使计算机也能够掌握。真歧义、准歧义、伪歧义三类情况比较，其中伪歧义可以统一规定处理方式，适用于所有的实例就行了；准歧义因为少有歧义实例，这本身就说明这类组合式的排歧条件相对好找一些；真歧义最复杂，排歧一般都要跨出结构体自身范围，必须在更大语境中找制约歧义的条件。综合起来看，目前比较适宜选择自囿型准歧义的格式，解决其结构定界和结构关系判定问题。解决问题的关键则在于充分准确地概括归纳，歧义格式映射到单义实例时应满足什么语法语义条件。同时，针对计算机处理的需要，还要将这些条件形式化，供计算机自然语言处理系统使用。

歧义类型	自囿型歧义	他囿型歧义
真歧义	1. 跟结构定界和结构关系判定都有关 2. 格式有歧义，也有歧义实例 3. 歧义对外影响小	1. 跟结构定界和结构关系判定都有关 2. 格式有歧义，也有歧义实例 3. 歧义对外影响大
伪歧义	1. 只跟结构定界有关 2. 格式无歧义，也无歧义实例	1. 只跟结构定界有关 2. 格式无歧义，也无歧义实例
准歧义	1. 跟结构定界和结构关系判定都有关 2. 格式有歧义，无歧义实例 3. 歧义对外影响小	1. 跟结构定界和结构关系判定都有关 2. 格式有歧义，无歧义实例 3. 歧义对外影响大

附注：

[1] 参见赵元任(1959)《汉语中的歧义现象》
载《中国现代语言学的开拓和发展》,清华大学出版社 1992 年版;
朱德熙(1980)《汉语句法中的歧义现象》,载《中国语文》1980 年第 2 期;
吕叔湘(1984)《歧义类例》,载《中国语文》1984 年第 5 期;
黄国营(1985)《现代汉语歧义短语》,载《语言研究》1985 年第 1 期;
邵敬敏(1994)《歧义分化方法探讨》,载《九十年代的语法思考》,
北京语言学院出版社 1994 年版;
冯志伟(1995)《论歧义结构的潜在性》,载《中文信息学报》1995 年第 4 期。

[2] 参见袁毓林(1993)《自然语言理解的语言学假设》,载《中国社会科学》1993 年第 1 期;
宁春岩(1985)《自然语言理解中的几个根本问题》,载《语言研究》1985 年第 2 期。

[3] 本文用到了一些英文字母来标记汉语词和短语的功能类。现将我们的功能分类和标记体系简要说明如下：

词类标记： 名词 n 代词 r 处所词 s 时间词 t 方位词 f
数词 m 量词 q 区别词 b 状态词 z 形容词 a
动词 v 副词 d 介词 p 连词 c 助词 u
语气词 y

短语标记： 名词短语 np 时间词短语 tp 处所词短语 sp
数量短语 mp 数词短语 mcp 动词短语 vp
副词短语 dp 介词短语 pp 形容词短语 ap
主谓短语 dj

有关功能分类的详情可参看：
俞士汶(1993)《关于计算语言学的若干研究》，载《语言文字应用》1993.2；
俞士汶(1992)《关于汉语短语结构体系及其描述方法的说明》内部资料；
周强 俞士汶《汉语短语标注标记集的确定》，载《中文信息学报》1996 年第 4 期。

[4] 尖括弧中的“了”表示格式中的助词 U 是特定的。不包括“了”之外的其他助词。下同。一般认为“了”有助词“了 1”和语气词“了 2”两个。跟在动词性成分后面的兼具助词和语气词性质。这里我们不细分，笼统地处理为助词。

[5] 这个格式中“不”的性质有不同看法。有人认为是副词，有人认为是中缀。我们暂且处理为副词。

[6] 这里我们因强调准歧义格式跟真歧义和伪歧义格式的不同，因而称准歧义格式在语言表现上没有歧义实例。也许更保守地说法应该是很少有歧义实例。所谓“言有易，言无难”。从理论上讲，准歧义格式跟真歧义格式应该只是相对的歧义实例在数量的差异，而不应有质的区别。

[7] 严格说来，DP VP VP 组合还有外部功能标记为 DJ 的情况。如：[[常 跑步] 能长高] 就是一个主谓结构短语。这也就是说，这种格式并非完全是自囿型歧义。事实上，区分自囿型歧义和他囿型歧义，着眼在歧义是否对外部结构发生影响，并不局限于刻板地把某个歧义格式归类。