

句处理中排歧问题补议*

陆俭明

北京大学中文系 北京 100871

lujm@pku.edu.cn

王黎

武汉大学留学生教育学院 武汉 430072

wangli6502@sina.com

摘要: 排歧问题是句处理中必须面对的一个重要问题, 已有一些学者进行过研究与探索, 提出了多种排歧策略。本文的目的是, 通过对“N+A+的+不+V”这一句法格式所造成的不同句子的句法、语义分析, 补充说明必须充分注意句中某个词语的具体意义在排歧中的重要作用。

关键词: 中文信息处理 句子排歧

Structure And Meaning Disambiguation in Chinese Sentence Processing

Lu Jianming

Wang Li

Dept. of Chinese, Peking, University 100871

Institute of Education for International Students, Wuhan University 430072

E-mail: lujm@pku.edu.cn

E-mail: wangli6502@sina.com

Abstract: It is very important to parse ambiguous sentences in Chinese Information Processing by computer. As a case study, this paper analyses the structure and meaning ambiguity of sentence pattern "N+A+De+Bu+V". In order to help computer to make correct decision on parsing of the sentence pattern, word sense should be paid more attention.

Keywords: Chinese Information Processing, Sentence Structure Disambiguation

排歧问题是句处理中必须面对的一个重要问题, 也是一个难题。以往已有一些学者进行过研究与探索, 并已取得了不小的成绩。本文试通过一组具体例子的分析, 对句处理中排歧所需考虑的问题, 再作些补充, 故谓“补议”。请先看实例:

(1) 香蕉青的不买。

* 本文得到教育部国家人文社会科学重点研究基地、北京大学汉语语言学研究中心2000年重大项目“现代汉语语义知识的形式化模型及语义分类系统研究”和国务院科技部国家“973”重点基础科学研究项目子课题“面向中文信息处理的现代汉语动词论旨结构系统和汉语词语语义分类层级系统研究”(项目编号: G1998030507-1)的经费资助。

(2) 皮儿青的不买。

(3) 皮儿青的不吃。

例(1) — (3), 词类序列相同, 都是“N+A+的+不+V”, 但情况各不相同。我们先来考察一下它们的内部层次构造。例(1)只能作甲切分, 不能作乙切分:

(1) 甲. 香蕉 青的 不买 乙. 香蕉 青的 不买
 1 2 * 1 2

而例(2)则相反, 只能作乙切分, 不能作甲切分:

(2) 甲. 皮儿 青的 不买 乙. 皮儿 青的 不买
 * 1 2 1 2

这为什么?

先分析例(1)。我们知道, 形容词属于谓词, 也有一个配价的问题。“青”是一个一价谓词。如果按乙切分, 意味着“香蕉青的”是一个“的”字结构, 它在句中作主语。可是, 按照朱德熙先生的“VP的”这种“的”字结构的歧义指数公式进行计算, “香蕉青的”的歧义指数该是0, 即: $P = n - m = 1 - 1 = 0$ (P代表“VP的”的歧义指数, n代表谓词的配价数, m代表谓词的配价成分在“VP的”里所出现的数日) 而“VP的”的歧义指数如果为0, 那是不能作主语来指称事物的。由于“香蕉青的”的歧义指数是0, 所以它不能再作主语来指称事物。因此例(1)只能作甲切分, 不能作乙切分。按此切分, 句子意思是香蕉有黄的, 有青的, 青的香蕉不买。类似的例子如:

(5) 梅子青的不摘。| 打印机旧的不要。| 点心太甜的不吃。| 杉树小的不砍。|……

再分析例(2)。为什么例(2)不能作乙切分, 得作甲切分? 按照朱德熙先生的“VP的”这种“的”字结构的歧义指数公式进行计算, “皮儿青的”的歧义指数也应该是0, 即: $P = n - m = 1 - 1 = 0$ 。按理说它也不能再作主语来指称事物。但是, 我们发现, “VP的”这种“的”字结构的歧义指数公式有例外。例外有两种情况(以 $P=0$ 的实例为例): 一是可以指称V的某个已经在“的”字结构里出现的配加成分NP所指事物的领有者。例如:

(6) 孩子游泳的(游泳, 一价, $P = n - m = 1 - 1 = 0$) [可指称孩子的家长]

(7) 他漆了地板的(漆, 二价, $P = n - m = 2 - 2 = 0$) [可指称某个房间或房子]

二是有可能指称跟作为V的某个配加成分、居于V后作宾语的第三人称代词“他”(包括“她”和“它”)有同指照应关系的事物。例如:

(8) 你给了他书的[给, 三价, $P = n - m = 3 - 3 = 0$; 该结构可指称某人, 该人与“他”同指]

所以, 例(2)“皮儿青的”虽然歧义指数为0, 但可以指称“皮儿”的领有者, 如西瓜、香蕉、桃子、橘子等瓜果。因此, “皮儿青的”可以作主语来指称事物; 整个句子可以作乙切分, 意思是“皮儿青的西瓜/香蕉…不买”。那么为什么不能进行甲切分呢? 这是因为在我们的日常生活中市场上是不买卖“瓜果皮儿”的, 所以例(2)不能作“不买青的皮儿”这样的意思来理解, 因此在结构上不能作甲切分。

现在来看例(3)。例(3)是有歧义的, 从结构上说, 既可以作甲切分, 也可以作乙切分, 即:

(3) 甲. 皮儿 青的 不吃 乙. 皮儿 青的 不吃
 1 2 1 2

按甲切分，句子意思是“不吃青的皮儿”；按乙切分，意思是“不吃皮儿青的（瓜果）”。例（3）之所以会有歧义，因为“吃”这个行为动作既可以拿“皮儿”作为它的支配对象（即“吃皮儿”），也可以拿有皮儿的瓜果作为它支配的对象（即“吃有皮儿的瓜果”）。类似的例子如：

- （9）馅儿甜的不吃。 [不吃甜的馅儿 | 不吃馅儿甜的包子/团子等]
 叶子圆的不要。 [不要圆的叶子 | 不要叶子圆的青菜/花木等]
 样式陈旧的不喜欢。 [不喜欢陈旧的样式 | 不喜欢样式陈旧的家具/摆设等]
 封面红的不撕。 [不撕红的封面 | 不撕红封面的书刊]

综上所述，“N+A+的+不+V”不是一个很单一的句法格式，而是一个歧义句法格式，它可以有甲、乙两种构造层次。不过这还只是从句法结构这个角度说的，如果从语义结构的角度来分析，情况则更为复杂。

从语义结构的角度看，属于构造层次甲的“N||A+的+不+V”，还可能有三种情况：

（一）受事_[整体]||受事_[部分] 动作_[及物动词] 按（甲）切分。例如：

（10）香蕉||青的不买。| 饭||糊的不吃。

（二）施事_[整体]||施事_[部分] 动作_[及物/不及物动词] 按（甲）切分。例如：

（11）花儿||红的不蔫儿。| 二班同学||近的不住校。

（三）施事 || 受事 动作_[及物动词] 按（甲）切分。例如：

（12）张三||青的不吃。| 妈妈||贵的不买。

属于构造层次乙的“N+A+的 || 不+V”，也还可能有三种情况：

（四）受事 || 动作_[及物动词] 按（乙）切分。例如：

（13）皮儿青的||不买。| 味道甜的||不吃。

（五）施事 || 动作_[不及物动词] 按（乙）切分。例如：

（14）年龄大的||不闹。| 脾气好的||不吭声。

（六）施事 || 动作_[及物动词] 按（乙）切分。例如：

（15）个儿高的||不吃。| 头脑聪明的||不学。

上述六种语义格式中要特别注意第三种例（12）。从内部的构造层次看，例（12）跟例（1）相同，只能作甲切分，不能作乙切分，即：

（12）甲. 张三 青的 不吃 乙. 张三 青的 不吃
 1 2 * 1 2

但内部语义关系跟例（1）“香蕉青的不买”不同。例（1）可以换说成“不买青的香蕉”，例（12）则不能换说成“*不吃青的张三”。类似例（12）的如：

（16）弟弟红的不喜欢。| 弟弟旧的不穿。| 老李便宜的不买。| 王老酸的不吃。

从以上的分析中，我们不难看出：

第一，“N+A+的+不+V”是一个“包含终结符”（格式里的终结符如“的”、“不”）的“内含型”歧义格式，而且是一个“真歧义格式”（关于“包含终结符歧义格式”、“内含型歧义格式”、“真歧义格式”等概念，见詹卫东等 1999，詹卫东 2000），“皮儿青的不吃”歧义句充分说明了这一点。

第二，“N+A+的+不+V”这一句法格式不只存在着由于不同的结构定界造成的真歧义现象，如“皮儿青的不吃”；而且还存在着虽结构定界相同但内部语义结构关系不同而造成的歧

义现象，如“香蕉青的不吃”和“张三青的不吃”。甚至在实际语言中可以存在终结符（即具体的词语）完全一样的歧义格式，如“男孩儿小的不要”——既可以看作是语义格式（一）“受事_[整体] || 受事_[部分] 动作”，也可以看作是语义格式（三）“施事 || 受事 动作”。若按前者，那么句子可以理解为“（某人/单位）不要小的男孩儿”；按后者，那句子可以理解为“男孩儿不要小的（玩具/水果/糕点/……）”。

第三，朱德熙（1980）认为，造成句法歧义的因素有四个：组成成分的词类，层次构造，显性语法关系，隐性语法关系。詹卫东（2000）进一步认为，这些因素概括而言，“可以转化为两项成分的搭配问题来描写”，即“两项成分能否搭配”，“两项成分搭配时需要满足什么条件”。这些认识和意见完全正确。

但是，除了上面的认识外，我们还需强调这样一点，那就是句中某个词语的具体意义会直接影响一个句式甚至一个具体的句子，是不是歧义句式，是不是歧义句。这也正是本文所要说的中心意思。上面所举的例（1）、（2）、（3）、（12）四个句子充分说明了这一点，请看：

- （1）香蕉青的不买。
- （2）皮儿青的不买。
- （3）皮儿青的不吃。
- （12）张三青的不吃。

先看例（1）、例（2）。二者差别在于句首所用名词不同——例（1）句首名词是“香蕉”，后面的动词“买”可以支配“香蕉”，而“香蕉青的”虽然在某种条件下也可以指称“香蕉（指果实）”的领有者“香蕉（指植物）”，但实际生活中没有买卖香蕉（指植物）的情况，所以例（1）只能作甲切分。例（2）句首名词是“皮儿”，实际生活中一般不存在买卖某种瓜果皮儿的情况，而“皮儿青的”可以指称“皮儿”的领有者，所以只能作乙切分，不能作甲切分。

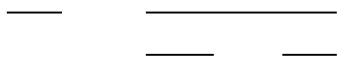
再看例（2）和例（3）。二者的差别在于所用的具体动词不同，前者用“买”，后者用“吃”。例（2），上面已经说了，只能作乙切分，不能作甲切分；而例（3）则可以作甲、乙两种不同的切分，即有两种构造层次。如：

- （3）甲．皮儿 || 青的不吃。
- 乙．皮儿青的 || 不吃。

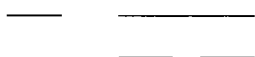
二者差别的原因，正如前面已说过的，“买”，只存在买卖带有皮儿的瓜果这种情况，不存在只买卖皮儿的情况；而“吃”，既存在吃带有皮儿瓜果的情况，也存在吃皮儿的情况。

最后再看例（3）和例（12）。二者的差别在于句首所用的名词不同，前者是指物的“香蕉”，后者是指人的“张三”。可是二者不仅在构造层次上有不同——例（3）有甲、乙两种构造层次（见上），例（12）则只有甲这一种构造层次，即：“张三 || 青的不吃。”而且，即使在构造层次相同的情况下，二者语义格式也还是不同，请看：

- （3甲）受事_[整体] || 受事_[部分] 动作_[及物动词] （皮儿 || 青的不吃。）



- （12）施事 || 受事 动作_[及物动词] （张三 || 青的不吃。）



以上的实例分析充分说明，一个词类序列，是不是歧义句式，一个歧义句式所造出的句子是否有歧义，在很大程度上取决于句中某个词语的具体语义。詹卫东所说的词语搭配的条件，也正是由句中某个词语的具体意义的影响所造成的。其实，在某些词语意义的影响下，按“N+A+的+不+V”这一句法格式还可能造出更为复杂的歧义现象。请看：

(17) 师傅差的不喜欢。

例(17)可以有甲、乙两种构造层次，每一种构造层次都可以表示多种不同的意思：

(17) 甲. 师傅差的 || 不喜欢。 乙. 师傅 || 差的 不喜欢。



例(17甲)“师傅差的”这个“的”字结构可以指称“某个徒弟”(意思相当于“不高明的师傅所培养出来的徒弟”)，而“某个徒弟不喜欢”可以表示两种意思——(a)“某个徒弟不喜欢什么东西或什么人”；(b)“某人不喜欢某个徒弟”。这也就是说，“师傅差的不喜欢”可以表示两种意思。这是因为“师傅差的”与“不喜欢”之间在语义上既可以分析为“施事——动作”的关系，也可以分析为“受事——动作”的关系。而例(17乙)不仅“师傅”与“不喜欢”之间既可以分析为“施事——动作”的关系，也可以分析为“受事——动作”的关系，而且“差的”既可以指人，也可以指物。因此，例(17乙)表示的意思更多：

- (18) a. 不喜欢差的师傅
 b. 师傅不喜欢差的人
 c. 师傅不喜欢差的东西

本文关于句处理排歧问题所要作的补充、所要强调的意见，就是上述“一个词类序列，是不是歧义句式，一个歧义句式所造出的句子是否有歧义，在很大程度上取决于句中某个词语的具体语义”这一点。我们希望能引起中文信息界的关注。其实这类例子是很多的。举例来说，谁都承认，“对+NP1+的+NP2”是个歧义格式，这种格式的特点是 NP2 是一个二价名词。“对校长的意见”就是按这个格式造出的歧义句，然而按此格式造出的短语并不一定有歧义。请看：

- (19) a. 对校长的意见——对||校长的意见 对校长的||意见
 b. 对住房的面积——对||住房的面积 *对住房的||面积
 c. 对校长的生死——对||校长的生死 *对校长的||生死
 d. 对住房的意见——*对||住房的意见 对住房的||意见
 e. 对土地的感情——*对||土地的感情 对土地的||感情

a 句有歧义，因为 NP2“意见”是一个二价名词，而“校长”既可能是意见的提出者，也可能是意见的针对者。b、c 句都不可能存歧义，因为 NP2“面积”或“生死”都不是二价名词。d、e 句虽然 NP2 分别是二价名词“意见”、“感情”，但“住房”、“土地”只能成为“意见”、“感情”的针对者，不可能成为“意见”、“感情”的持有者。那么我们能不能认为，当 NP2 为二价名词、NP1 为指人名词时，“对+NP1+的+N²P2”一定有歧义呢？袁毓林（1992）和李小荣（2000）对此作了肯定的回答，他们都认为，当 NP2 为二价名词、NP1 为指人名词时，“对+NP1+的+N²P2”有歧义，如“对校长的意见”。可是实际上并非完全如此。请看：

- (20) a. 对弱者的同情心——*对||弱者的同情心 对弱者的||同情心

对皇上的敬意——*对||皇上的敬意 对皇上的||敬意

b. 对校长的观念——对||校长的观念 *对校长的||观念

对乞丐的信仰——对||乞丐的信仰 *对乞丐的||信仰

例(20)各例之所以没有或者说不让人感到有歧义,就是句中某个词语的具体语义所起的制约作用的结果——a例中的“弱者”,只能是“同情心”的针对者,不大可能是“同情心”的持有者;相反,b例中的“乞丐”,只能是“信仰”的持有者,不大可能是“信仰”的针对者。这里,我们还可以进一步体会到,在具体词语语义制约的背后,还有人的认知和民族心理的因素在起作用。

其实,句中某个词语的具体意义在排歧中的作用,早已有学者注意到了。众所周知,朱德熙先生《汉语句法中的歧义现象》(1980)对汉语里的句法歧义现象作了非常精辟的分析和论说。这里需要指出的是,朱德熙先生在这篇论文中已经注意到这样一点,那就是按歧义句式造出来的句子都可能有歧义,但“不一定都有歧义”,“这是因为组成这些句子的词义上互相制约,消除了句式本身具有的产生歧义的可能性”。可惜朱先生一直未就这一问题展开进一步的论述。

目前一般认为,今后句处理总的发展趋势可能是“大词库,小规则”。词库里的每个词,既要给以详尽的语法知识,又要给以详尽的语义知识。今后词汇的语法信息应比现有的词汇语法信息词典进一步细化。拿体宾动词来说,最好还需进一步补充(a)是否粘宾动词,(b)要求带什么样的体词性宾语,(c)宾语在简单、复杂上的要求如何,(d)是否要求带某种特殊的状语或补语成分等句法信息。而词汇的语义信息,也不应只是标明该词在语义上的上下位关系、同义关系(实质是近义关系)、反义关系和整体部分的关系等,还应标明该词语在语义上的特点,该词语与其他词语在语义上的搭配条件和限制。譬如说,“吐”,只有“吐出来”的说法,没有“*吐进去”的说法;“咽”则只有“咽下去”的说法,没有“*咽上来”的说法;这都是由“吐”、“咽”的词义特点和词语之间的制约作用所决定的。

要突破句处理的难题,就日来说,我们认为,下面几项研究工作似乎应尽快进行:

1.汉语动词的题元结构系统的研究,包括建立面向汉语动词的题元结构理论框架、不同动词的题元结构的形式化模型,不同的词语参与组合后可能发生的变化研究,以及对各个具体动词(按义项出条)进行动名句法形式选择和动名题元角色标注;

2.汉语词的语义分类层级系统研究,包括建立适用于汉语的语义特征分析的理论框架以及建立能与动词题元结构系统配套的现代汉语词汇语义层级系统;

3.词语之间的组合类型与组合规则的研究,这包括同类的词与词之间彼此组合可能形成的种种关系的类型及其具体规则的研究、非同类的词与词之间彼此组合可能形成的种种关系类型及其具体规则的研究,以及同类的或不同类的短语与短语之间彼此组合可能形成的种种关系类型及其具体规则的研究;

4.着力研究句法、语义的多功能性,从功能范畴的角度设想并研究尽可能多的“语义框架”(Semantic frames)、“句式框架”和认知域,并建立相应的各种模框(frames)和图式(schemas);

5.各种排歧研究,包括各种排歧类型和针对不同歧义类型所采取的不同的排歧策略;

6.在上述研究的基础上,建设汉语知识库,特别是明确标有词的语义特征和词的句法特征的词库(可分别建库),同时开展跟句处理相关的实验和应用系统的研制开发。

参考文献

- [1] 冯志伟：“论歧义结构的潜在性”，《中文信息学报》1995年第4期。
- [2] 郭锐：“语法的动态性和动态语法观”，在“商务印书馆语言学出版基金发布会暨青年语言学者论坛——21世纪的中国语言学”（2002.1.17-18.,北京）会上发表。
- [3] 李小荣：“从配价角度考察介词结构‘对于……’作定语的情况”，《配价理论与汉语语法研究》，陆俭明、沈阳主编，语文出版社，2000。
- [4] 陆俭明：“‘的’字结构和‘所’字结构”，《语法研究和探索》（1），中国语文杂志社编，北京大学出版社，1983。
- [5] 陆俭明：“‘VA了’述补结构语义分析”，《汉语学习》1990第1期；又见《陆俭明白选集》，河南教育出版社（现改为人象出版社），1993年。
- [6] 沈家煊：“句式和配价”，《中国语文》2000年第4期。
- [7] 袁毓林：“现代汉语名词的配价研究”，《中国社会科学》1992第3期。
- [8] 詹卫东：“词的语义分类在英汉机器翻译中所起的作用以及难以处理的问题”，《语言工程——全国第四届计算语言学联合学术会议论文集》，陈力为、袁琦主编，清华大学出版社，1997。
- [9] 《面向中文信息处理的现代汉语短语结构规则研究》，清华大学出版社、广西科学技术出版社，2000年。
- [10] 詹卫东、常宝宝、俞士汶“汉语短语结构定界歧义类型分析及分布统计”，《中文信息学报》1999年第3期。
- [11] 张伯江：“现代汉语的双及物结构式”，《中国语文》1999第3期。
- [12] 张伯江：论“‘把’字句的句式语义”，《语言研究》2000第1期。
- [13] 张普：“信息处理用现代汉语语义分析的理论与方法”，《中文信息处理应用平台工程》，陈力为、袁琦主编，电子工业出版社，1995年。
- [14] 朱德熙：“的”字结构和判断句，《中国语文》1978年第1、第2期。
- [15] 朱德熙：“汉语句法里的歧义现象”，《中国语文》1980年第2期。
- [16] Goldberg, Adele E.: *Construction: A Construction Grammar Approach to Argument Structure*, The University Chicago Press, 1995.
- [17] Kay, Paul.: *Construction Grammar*. Jef Verschueren et al. *Handbook of Pragmatics: Manual* Amsterdam/Philadelphia: John Benjamins Publishing Co. .1995.