# 汉语自动分析中的若干问题与对策*

侯　敏
北京广播学院 播音主持艺术学院 应用语言学系，北京 100024
E-mail: houminxx@263.net

**摘要**：在汉语自动分析中，主要存在以下四个问题：1.词语的自动切分；2.同形词的判别；4.同形词类组的确定；4.多义词的辨识。本文讨论了这些问题并提出了一些解决问题的对策。

**关键词**：自动分词　同形词　同形词类组　多义

## Some Problems and Countermeasures in Automatic Chinese Analysis

Hou　Min
Department of Applied Linguistics , Beijing Broadcasting Institute
E-mail: houminxx@263.net

**Abstract** ： In automatic Chinese analysis , there mainly exist four following problems ： 1. automatic word segmentation of Chinese sentences ; 2. identification of trans-classed words; 3. differentiation of homonymous colligation ; 4. recognition of polysemants . This paper discusses these problems and puts forward some countermeasures to solve them .

**Keywords** ： automatic word-segmentation； trans-classed words； homonymous colligation； polysemy

自然语言处理，实际上包括计算机对自然语言的自动分析和自动生成两部分内容。汉语是一种分析型语言，和印欧语相比，除了书写形式不同外，最明显的特点是无形态。由此也形成了汉语在自然语言处理研究中的特点：语言的生成相对容易，而语言的分析要比西语难得多。因为汉语没有形态变化，所以在生成句子时，既不用考虑词的性、数、格、人称等形式，也不必考虑词与词、句子成分与句子成分之间的一致或对应关系，只要把具有一定意义的词按照一定的线性顺序排列起来就可以，这样的生成程序和西语，尤其是和俄语这样的屈折语相比，显然要少了许多麻烦。但凡事有利必有弊，汉语生成的便利是以分析的困难为代价的。同样是因为汉语没有形态变化，所以在进行自动分析时，缺乏形式上的依据，必须靠句法、语义以及语用常识等多方面知识的综合。下面我们来具体谈谈汉语自动分析中遇到的几个主要问题，并尽可能提出相应的处理策略。

### 1、词语的自动切分

汉语的书写形式不像西文，词与词之间没有间隔，所以就比西文的语言处理多了一道手续：自动分词，即让计算机自动地把汉语语流串变成一个一个词的形式，这样才能进行下一步的句法语义分析及处理。但由于汉语中的构词语素大多是不定位语素，又有相当数量的自由语素，这就造成了切分中的多分字段。如“诊”是不定位语素，它在“会诊”中位置在后，在“诊断”中位置在前，语流中它们又可能交集在一起，如：“很少有医生会诊断这种疾病”，其中的“会诊断”就形成了一个交集型多分字段，也叫后字有定型多分字段；“马

* 本研究得到广电总局社科基金资助，项目编号为 BW9943。

上”可以是一个词，如：“我马上下来”，也可以是两个词，如：“我从马上下来”，“马上”形成了一个组合型多分字段，也叫语段多分型字段。在具体句子中，这些可以两切的多分字段只有一种切分是正确的，那么根据什么、如何去找到这种正确的切分就成了一个颇费斟酌的问题。这是难点之一。还有，汉语中的专有名词，如人名、地名、商标名等，既不大写，也没有任何特殊标记，而且还有一部分与普通名词相同，如人们所熟悉的相声演员“牛群”、中央电视台的节目制片人“时间”、福建的制衣名城“石狮”等等，这不啻给这部分本来就难以处理的问题雪上加霜。这些未登录词的辨别，是难点之二。这两个困难使得自动分词成了计算机理解汉语的第一个“瓶颈”问题。

自动分词属于汉语自动分析中的预处理问题，它必须在系统进入语法、语义分析之前解决。对自动分词中的难题，可采取以下对策。

1）多分字段的切分

对策之一：规则法。既然多分字段在一定的上下文中只有一种切分是正确的，那么，我们就应该可以从该语句中找到这种切分的理据，即条件，把这种条件用规则的形式描述出来，告诉机器，这就是规则法。语段多分型字段的特点是：它的正确切分不取决于字段内的某项，而取决于这一字段的上下文。所以，像“马上”“的确切”这样的语段多分型字段只能在短语的平面上用给出规则的办法来解决。后字有定型多分字段也可以用规则的办法来解决。这种字段的特点是：决定如何切分的是段内的后字，而且这个后字是可以枚举的，所以我们可以在词条中给出具体的后字，形成规则。只要做得合适，规则法可以解决多义切分的大部分问题，只是比较繁琐。[11][12]

对策之二：查表法。查表法只能用来解决后字有定型多分字段的切分。由于这种字段的特点是“从切分字段本身可以得到切分该字段所需的自足信息”，所以“可以把它们的正确（唯一）切分形式预先记录在一张表中，其歧义消解通过简单的查表予以完成”。[4]

对策之三：排序法。排序法也只能用来解决后字有定型多分字段的切分。排序法实际上是以关键字来结构词典，靠词序的先后安排来解决多分字段的切分的问题。如“会诊”这一词条，可确定“诊”为关键字，把“会诊”放在“诊”这一关键字下“诊断、诊疗、诊治”等词条的后面，这样先查询到的是“诊断、诊治、诊疗”等，匹配不上，才能查到“会诊”，切分结果自然是正确的。这种方法比规则法和查表法更简洁，但词典的结构要改变。这种办法，从本质上来说，不是“分词”，而是“合词”。[13]

2）人名、地名等专有名词的处理

对策之一：直接登录法。将常用的人名、地名收录在词典里。这种方法的好处是准确，操作简便。缺点是：1）不可能将所有的专有名词都收录在册，遗漏的人名、地名等系统就无法识别，这将影响下一步的操作；2）如果大量收录，将会增加系统开销并引起与普通名词混淆等一系列问题。

对策之二：字表查询法。将姓氏用字、人名用字以及地名用字的各种情况统计后列表，用查表的方式解决。[2][3][7][15] 由于人名、地名与普通词语有同形现象，所以字表收取哪些字以及收取的量要选择一个“合算”的“度”。如果量大，召回率高，但准确率下降，一些不是姓名的也会被它“识别”出来；如果量小，准确率高，但召回率会下降，一些真正的人名、地名会被它遗漏。单靠字表法不可能完全解决问题，能在得失之间找到一个比较合算的度，应该是目前最好的选择了。

对策之三：参照确定法。语句中人名、地名出现的前后会有一些语言“征候”，如人名前后常有表职务称谓的“厂长”“经理”“主席”、后面常有“说”“指出”“认为”之类的，

可参照这些可利用资源，确定人名和地名。[2][3][7][15]

## 2、同形词的判别

语言信息处理中的同形词，包括两部分内容，一部分是传统语言学中的同形词（但必须是同形不同类的，像“登报”的“登”和“登在窗台上”的“登”，虽是同形词，但因都是动词，不在其内），其中有同音同形词，如“开个会”的“会”与“我不会写”的“会”，书面信息处理中还有异音同形词，如“地道（dìdào）”与“地道（dì · dao）”；另一部分是传统语言学中的兼类词，如“锁门”的“锁”和“一把锁”的“锁”。实际上，词语同形现象不仅汉语中有，别的语言里也不少。但汉语由于缺乏形态变化，很少有词能依据自身的变化确定词性，所以基于形态的判别方法基本无效，只能根据词的分布环境，运用上下文语境判别法，这就使得同形词的判别更加困难。但任何工程化的汉语句法分析系统都不能回避同形词的判别即兼类消除问题，因为词性是词的最主要的语法属性，如果一个词的词性不能确定，则汉语句法分析就无法进行；如果一个词的词类选择错了，就会导致句法分析的严重错误甚至失败。例如 N|V 同形，V|P 同形等，它们对句法分析的影响是全局性的，即影响到对整个句子的分析。所以，同形词的判别就成了汉语自动分析中的第二个难题。

同形判别是词法平面的问题，一般来说，它应该在系统进入句法分析前完成。判别同形词，可采用以下对策。

对策之一：分布判别法。词类是根据词的语法功能划分出来的，词的不同分布正体现了词的语法功能。所以，根据词的分布来判别同形词，应该是一个主要的、也比较有效的方法。如汉语中副词不能修饰名词，据此可以判定，副词后面的 N|V 同形词应确定为动词；只有名词能受数量短语修饰，所以数量短语后的 N|V 同形词应确定为名词；程度副词只能修饰形容词，所以程度副词后的 A|N 同形词应确定为形容词。当然，具体的语言中问题没有这么简单，但只要仔细分析，条件给出合理，分布判别法应能解决同形判别中大部分共性问题。分布判别主要依据的是词类之间的不同组合功能。被判别的是一类词，作为判别条件的也是一类词。它解决的主要是共性问题。

对策之二：特定词鉴别法。有一些词，它前面或后面只能出现某一类词，这些词就可以作为鉴别的条件。如“的”的后面只能是名词性成分，所以，当“的”后有 N|V 或 A|N 同形时，就可以确定为名词①；“一下”主要用在动词后面表动作的短时少量，它也可以作为一个鉴别词，它前面的 N|V 或 A|V 同形词应确定为动词。特定词鉴别法中被判别的是一类词，但作为判别条件的是特定词，它可以作为分布判别法的补充，解决的也是共性问题。

对策之三：过滤归并法。同形词的类型是不一样的，有些是基本类型，如 N|V、A|N、A|V、A|N|V、V|P 等等，每类中都包含有大量的同形词；还有一些是非基本类型，如“白”除 A|F 同形外，还要做姓氏词 X，而这种 A|F|X 同形的词比较少，这时就可以先把它做 X 的条件找出来，在词条下给出规则，解决了这个带有个性的问题，然后再归并到 A|F 共性规则中去。就是说，把一些词的个性用法先过滤出去，再利用共性规则统一解决。[1]

对策之四：个别确定法。这主要解决的是两种情况。第一种是非基本类型的同形词。如“啊”，是感叹词与语气词同形。这两类同形的词很少，做共性规则得不偿失，应该采用在词条下做个性规则个别确定的方法。第二种情况是，虽属基本类型的同形词，但判定条件

① 以汉英机器翻译做背景，“的”后只能是名词性结构。

与同类型词不一样，这也需要个别确定。显然，个别确定法解决的是个性问题。应注意的是，个性规则的执行要优先于共性规则。

### 3、同形词类组的确定

同形词类组指的是能映射不同结构层次、不同句法关系或不同语义关系语言结构的词类序列。如“A+$NP_1$+和+$NP_2$”，可以映射一个并列结构，如“年迈的爷爷和孙女”，也可以映射一个偏正结构，如“新鲜的蔬菜和水果”，那么“A+$NP_1$+和+$NP_2$”就是一个同形词类组。有形态变化的语言中，词类与句法成分往往是一一对应的，所以一般情况下人们可以根据词类序列以及词的变化形式确定句法语义关系。如英语中，“NP+VP”只能是主谓关系，“V+NP”只能是动宾关系。英语中也有同形词类组，但比较少，如与上述相同的“A+$NP_1$+and+$NP_2$”以及“$NP_1$+V+$NP_2$+P+$NP_3$”（a. He hit the car with a stone.---介词短语做状语； b. He hit the car with a dented fender.----介词短语做定语）等。然而，汉语就不同了，第一，由于词类与句法成分不存在简单的一一对应关系，同一类词在句法结构中可以做不同的句法成分，而在形式上没有任何标记；第二，汉语主要靠“意合”，缺乏“语态”等各种语法范畴的标记和变化；第三，汉语的修饰语只能左递归。因此，汉语中的同形词类组非常多。例如：

**（1）V+V：**

a. 打算回家（动宾） b. 唱歌跳舞（联合）

c. 演出开始（主谓） d. 回家睡觉（连动）

**（2）N+N**

a. 工人农民（并列） b. 木头桌子（偏正）

c. 鲁迅先生（复指） d. 今天晴天（主谓）

**（3）N+V**

a. 红旗飘扬（主谓） b. 农业生产（偏正）

**（4）V+N**

a. 学习外语（动宾） b. 学习方法（偏正）

**（5）V+A**

a. 行动迅速（主谓） b. 喜欢漂亮（动宾）

c. 说清楚（动补）

**（6）A+$N_1$+$N_2$**

a. 大 / 百货商场（偏正）

b. 大商场 / 气派（偏正）

**（7）数量+$N_1$+的+$N_2$**

a. 一位 / 商店的售货员（偏正）

b. 一座房子的 / 主人（偏正）

**（8）$N_1$+$N_2$+$N_3$**

a. 中国 / 工农红军（偏正）

b. 中国人民 / 生活（偏正）

**（9）V+A+NP**

a. 穿 / 漂亮衣服 （动宾）

b. 说清楚 / 问题 （动宾）

**（10）$NP_1$+VP+的+$NP_2$**

a. 我们学校 / 获奖的学生（偏正）

b. 我们学校选派的 / 学生（偏正）

**（11）V+$NP_1$+的+$NP_2$**

a. 穿着红衣服的 / 姑娘（偏正）

b. 穿着 / 节日的服装（动宾）

**（12）P+$NP_1$+的+$NP_2$**

a. 对这件事的 / 看法（偏正）

b. 对 / 朋友的到来（介宾）

**（13）V+NP+VP**

a. 希望/他来（动宾） b. 派/他/去（兼语）

c. 告诉他/今天开会(双宾) d.去学校/上课（连动）

**（14）V+的+是+NP**

a. 买的是菜（主谓，NP 是 V 的受事）

b. 来的是小王（主谓，NP 是 V 的施事）

**（15）NP+VP**

a. 信写完了（主谓，NP 是 VP 的受事） b. 学生写完了（主谓，NP 是 VP 的施事）

这些只是举例性的，不是全部，而且都是最基本的，如果加上扩展式，情况将更复杂。其中根据所映射词组的结构及关系状况可分为四类：(1)～(5)是结构关系不同的同形词类组，(6)～(10)是结构层次不同的同形词类组，(11)～(13)是结构层次和结构关系都不同的同形词类组，

(14)～(15)是语义关系不同的同形词类组；如果根据所映射词组的功能是否一致可分为两类：同功能的，如(6)、(7)、(8)、(9)等；不同功能的，如(3)、(4)、(11)、(12)等。对于自动句法分析而言，同形词类组，尤其是不同功能的同形词类组的确定是一个极其重要的问题。因为，计算机进行句法分析时，首先看到的就是词类序列，如果是同形词类组，它就是多义的，具有映射不同词组的可能，但在具体的语句中它又只能有一种意义，一种解释。那么根据哪些条件、如何在具体的语句中确定这些同形词类组的结构，就成了汉语自动分析的第三个难题。

同形词类组的确定是句法平面的问题，它应该在系统进行句法、语义分析的过程中解决。由于确定同形词类组的问题十分复杂，所以应将其分散，可根据不同情况、采用不同办法、在各个不同的层面上解决。采用的主要对策如下。

对策之一：次范畴分类制约法。次范畴分类制约法主要解决动词性同形词类组的问题。即通过给动词进行次范畴分类来确定同形词类组。如将“打算”、“喜欢”之类的谓宾动词归成一类，标记为 Va，那么 Va+V|A 就可以确定为动宾关系；将“开始”之类可以带动词性主语的过程动词标记为 Vb，那么 V+Vb 就可以确定为主谓关系；将常常后接表目的的动词的“回家”“去”之类标记为 Vc，那么 Vc+(NP)+V 就可以确定为连动关系；将“希望”之类带小句做宾语的动词标记为 Vd，那么 Vd+NP+V 就是动宾关系；将“派”之类常带兼语的动词标记为 Ve，那么 Ve+NP+V 就是兼语关系；将“告诉”之类可带名动双宾语的标记为 Vf，那么 Vf+NP+V 就是双宾关系。除此之外，两个同范畴的动词相连，就是并列关系，如“唱歌跳舞”、“讨论研究”。

对策之二：语义特征制约法。语义特征制约法是通过给名词、量词等打上语义特征标记来确定同形词类组。如将“木头”之类给出质料的语义标记 Nzl，那么“Nzl+N”就可以确定为偏正关系；将“鲁迅”之类给出专有名词的标记 Nzn，“先生”之类给出称呼的语义标记 Nch，那么“Nzn+Nch”就可以确定为复指关系；将“位”等给出人的语义标记 Lrl，那么它就只能与有相同标记的名词搭配；两个同语义特征标记的名词相连，应是并列关系，如“工人农民”、“工厂学校”。

对策之三：上下文语境识别法。有些同形词类组单靠内部成分的分类和特征难以确定它的结构，如“发现了敌人的哨兵”（V+NP+的+Nrl），好在这类同形词类组所映射的词组外部结构功能不同，就是说，它们只能出现在不同的语言环境中，于是我们就可以根据上下文语境来确定它的结构。如前面有数（指）量词组或者后面有动词，它是名词性偏正词组；如果前面只有一个名词，它就是动宾词组。

对策之四：固定词组处理法。有些同形词类组所映射的词组用法固定单一，如“参考消息”“参考资料”，虽是“V+N”，但只能是名词性偏正结构。所以就没有必要再去识别判断，做一个固定词组处理更方便。这也可以说是一种优先选择。

### 4、多义词的辨识

多义词的辨识是所有语言在自然语言处理中都面临的一个问题。这里的多义词，不包括由于词性不同造成的多义，因为那种多义可以通过同形判别来区分。但这里的多义除了传统语言学中的多义词外，还要包括同类同形词。如“名字登了报”中的“登”和“脚登着窗台”中的“登”，庄稼收割后运到场上的“登场（cháng）”与剧中人出现在舞台上的“登场（chǎng）”，都是动词，但意义不同，在传统语言学中属同音同形词和异音同形词，然而在语言信息处理中它们与多义词没有什么分别，所以也包括在多义词之列。多义词在静态中

是多义的，但一旦进入动态，在具体的语句中只能有一个意义，例如，“借”是个多义词，有借进、借出、凭借或引用等意义，但在具体的句子中它只能有一个意义，在“去图书馆借书”中只能是“借进”，在“借书给他”中只能是“借出”，在“借古人的话表达自己的心情”中只能是“引用”，它们对应的英语词分别是 borrow，lend ，quote 。那么，如何在具体的上下文中辨识并确定多义词的意义，这也是汉语自动分析的一个难题。

多义辨识属语义平面的问题，在系统中应根据不同情况在词法分析、句法分析或语义分析等不同层面上解决。辨识多义词可采取以下的策略。

对策之一：依存关系确定法。语言中，词语之间往往有一种依存关系，如动词和名词之间、名词和形容词之间，正是这些词语的实体和这些依存关系的层层组合构成了句子。词语间的依存关系也是一些词词义发展的基础，所以根据依存关系来确定词义就成为词义辨识的主要方法。这种方法可以解决大部分动词、形容词以及介词的多义辨识问题。先来看动词。动词有“价”的不同。一价动词只支配一个名词性成分（即所谓“行动元”），那么就根据这个名词性成分来确定动词词义的选择。如“跌”是个多义动词，当它支配的名词是人时取“摔倒”（fall down），是价格时取“下降”（reduce）。二价动词支配两个名词性成分——施事和受事，其中，受事成分与动词关系更近，是 VP 结构内属成分，所以应根据受事来确定动词词义的选择。如“拆”是多义动词，当受事是房屋之类的建筑时应取“拆毁”（pull down），受事是机器之类时取“拆散”（take apart），受事是包裹之类时取“打开”（open）。三价动词具有多义的不多。形容词的意义往往取决于它所修饰的名词性成分。如“肥”，当它所修饰的名词性成分是衣服类时应取“宽大”（wide），是人或动物时应取“肥胖”（fat），是土地之类时应取“肥沃”（fertile）。由于这种多义识别依据的是词语的依存关系，而不是句子线性结构中的某一确定位置，所以，这种辨识条件可以做在词典中，根据不同句式的需要在规则中决定如何选取。

对策之二：语义条件查询法。有些多义词词义的选取条件是根据句子中的某一语义限制，这时，依存关系确定法不起作用，就可以用语义条件查询法来解决。如一些名词的多义选择问题。“黑人”是个多义词：①属于黑色人种的人(Negro)；②姓名没有登记在户籍上的人(unregistered resident)。用依存关系确定法很难区分这两个意义，可以考虑用语义条件查询的办法。如果该词前后有“户籍”“户口”字样的，可以确定为②，如果该词前后有“非洲”“种族”“留学生”等字样的，可以确定为①。语义条件查询法可以解决一般性问题，但不排除有例外。

对策之三：固定词组处理法。固定词组处理法主要是针对个性问题的。有特定意义的词语组合可以用这种方法确定词义，如“拆墙角”（undermine）。运用依存关系确定法不能涵盖的问题也可以采用这种办法。如“看”，当受事是人时往往是“访问”（visit 或 call on）的意思，如“看朋友”、“看老师”、“看同学”。但在“看医生”中，受事同样是人，但“看”的意思却应该是找医生“看病”“咨询”（consult）的意思，这时就可以运用固定词组处理法，做一个固定词组单独处理。固定词组中可以都是具体的词，也可以有变量形式的语法及语义范畴，但要处理的当前词必须是具体的。固定词组处理法方便、快捷、可靠，是解决词义辨识中个性问题的最好办法。关键是做时要把握好层次，以防前后牵扯。

对策之四：主题确定法。主题确定法主要是用来解决专业术语与普通用语之间构成的多义问题。可用建立专业词典，在处理专业领域语料时，专业词典优先查询的办法解决。

对策之五：中性表述法。有时多义词的几个意义很难区分，这时应尽量选择一个概括性大的、具有中性意义的表述形式。这实际是一种不得已而为之的办法。

## 5、歧义现象的消解

这里的歧义现象指的是一个具体的句子可以有两种或两种以上的理解。我们认为，歧义，应该是语言运用中的一种二义性现象，至少应该在句子平面上才能产生。以上四个问题，如果句子中缺乏有效的句法或语义限制，都可能造成歧义现象，现分别举例如下：

（1）白天鹅在湖里游。/ 乒乓球拍卖完了。（不同切分造成的句子歧义）

（2）我要学习文件。/ 这个拍坏了。（同形词造成的句子歧义）

（3）这是一栋新职工宿舍。/ 鸡不吃了。（同形词类组造成的句子歧义）

（4）他走了一个小时了。/ 他背着儿子到老师家学下棋。（多义词造成的句子歧义）

这些歧义现象的消解不可能在句子平面上解决，必须借助于该句子出现的上下文语境乃至情景语境、背景知识等来解决。

另外，还有代词的照应、省略成分的添补、隐含成分的指明以及无标记的名词单复数、动词时态的确定等等问题。这些问题中的大部分与歧义现象的消解一样，需要更广的上下文，而且又都往往属于语用平面上的问题，目前一般以句子为处理单位的系统还难以解决。

总起来看，从第一到第四个问题是当务之急，是每一个涉及汉语分析的语言工程系统必须面对、必须解决的问题，而且，就目前的研究程度和技术水平来讲，其中大部分或至少相当一部分是能够解决的，关键是要一项项脚踏实地具体地去做。而第五以后的问题，因为主要是语用平面的问题，要依赖大语境甚至情景语境的知识，在目前技术情况下解决起来还有一定的困难，其中有些问题在语言中并不是十分普遍的现象，而且对有些系统来说也不是当务之急，所以可以缓一步解决。

## 参 考 文 献


[1]孙茂松、黄昌宁 1989，《汉语中的兼类词、同形词类组及其处理策略》，中文信息学报，第 4 期。

[2]孙茂松、黄昌宁等 1995，《中文姓名的自动辨识》，中文信息学报 第 2 期。

[3]沈达阳、孙茂松等 1995，《中国地名的自动辨识》，计算语言学进展与应用。

[4]孙茂松等 1999，《高频最大交集型歧义切分字段在汉语自动分词中的作用》，中文信息学报第 1 期。

[5]俞士汶 1989，《自然语言的歧义与机器翻译的对策》，中文信息学报，第 2 期。

[6]李维、刘倬 1990，《机器翻译词义辨识对策》，中文信息学报，第 1 期。

[7]宋 柔、朱宏等 1993，《基于语料库和规则库的人名识别法》，计算语言学研究与应用。

[8]冯志伟 1995，《论歧义结构的潜在性》，中文信息学报，第 4 期。

[9]冯志伟 1996，《自然语言处理中的歧义消解方法》，语言文字应用，第 1 期。

[10]冯志伟 1999，《英日机器翻译系统 E-to-J 原语分析中的兼类词消除策略》，中文信息学报，第 5 期。

[11]侯 敏 、孙建军 1996，《汉语自动分词中的歧义问题》，语言文字应用 第 1 期。

[12]侯 敏 1999，《计算语言学与汉语自动分析》，北京广播学院出版社。

[13]孙建军、陈肇雄等 1996，《一种新型汉语电子词典结构》，*The Last Technological Advancement & Applications*, in Singapore.

[14]刘开瑛 1997，《汉语自动分词评测技术研究》，语言文字应用 第 1 期。

[15]刘开瑛 2000，《中文文本自动分词和标注》，商务印书馆，2000 年出版。

[16]詹卫东、常宝宝等 1999，《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 第 3 期。