

# 有标记联合结构的语言学特征统计分析

王东波, 陈小荷, 年洪东

(南京师范大学文学院, 210097)

**摘要:** 统计和分析了有标记联合结构内部和外部的语言学特征, 内部特征方面主要考察了联合结构的词性序列分布、短语序列分布。外部特征方面主要从联合结构的句法功能分布和联合结构左右边界特征词两个方面进行考察。这些考察一方面为计算机识别有标记联合结构提供了语言学知识, 另一方面为从语言学的角度量化地研究联合结构提供了精确的数据。

**关键词:** 有标记联合结构; 内部语言学特征; 外部语言学特征;

## Analyzing the Linguistic Features of Coordination with Overt

### Conjunctions, Based on Statistical Data

Wang Dongbo, Chen Xiaohe, Nian Hongdong

(School of Chinese Language and Literature, Nanjing Normal University, 210097)

**Abstract :** The article counted and analysed the external and internal linguistic features of COC. The internal linguistic features is the following: The distribution of POS sequence. The distribution of phrases sequence . The external linguistic features is the following : The distributing of Syntactic function and the features of border lexicons. For one thing ,the statistics data will offer the linguistic knowledge for identifying COC, for another thing ,the accurate data will be used to investigate the COC.

**Keywords:** Coordination with Overt Conjunctions; Internal Linguistic Features; External Linguistic Features;

## 1 引言

从整体上说联合结构分为无标记联合结构和有标记联合结构。有显性标记形式的联合结构称之为有标记联合结构,如“建设/vn 方案/n 、/w 项目/n 调整/vn 和/c 概算/n”和“选育/vn 与/c 推广/vn”,其中“、”和“和”就是联合标记。

有标记联合结构语言学特征的详细考察,一方面可以为从语言学的角度考察联合结构提供量化的数据,另一方面可以为自动识别联合结构提供详细的语言学知识。吴云芳的在《面向中文信息处理的现代汉语并列结构研究》博士论文中,以中文信息处理为出发点,全面考察了现代汉语中的并列结构,提出了提出的“有标记联合结构在结构上的平行性和中心语的相似性”这一特征。本文在其研究的前提下,在大规模树库的基础上,更加详尽地考察了有标记联合结构的内部和外部语言学特征。

本文所使用的语料分别为清华973树库和北大人民日报一月份经过分词和词性标注的语料,从前者根据“xx-LH”抽取有标记联合结构,从后者根据已经人工标注的“{ }”标记抽取联合结构。具体的语料样例如下:

清华语料: “[vp-PO 参与/v [np-DZ [np-DZ [np-LH [np-DZ 社会/n 产品/n ] 和/c 国民收入/n ] [np-LH 分配/vN 、/、 再分配/n ]] 活动/n ]] ”其中有标记联合结构在清华树库中被标成了“xx-LH”。

北大语料: “1 2 月/t 3 1 日/t , /w {中共中央/nt 总书记/n 、/w 国家/n 主席/n}

江/nr 泽民/nr 发表/v 1998年/t 新年/t 讲话/n 《/w 迈向/v 充满/v 希望/n 的/u 新/a 世纪/n 》/w 。/w ( /w 新华社/nt 记者/n 兰/nr 红光/nr 摄/Vg ) /w 。”有标记联合结构被用“{ }”出来了。

## 2 内部特征考察

本文考察联合结构的内部特征，主要从下面这三个方面进行：联合结构内部的词性序列分布、联合结构内部的短语序列分布

### 2.1 有标记联合结构内部的词性序列分布

从100万字的清华语料和56万字的人民日报语料中统计出了12362和7193个有标记联合结构的词性序列，表1和表2列出了频次居于前十四位的词性序列。

词性标记序列	频次	具体例子
n+c+n	1776	国家/n 和/c 民族/n
n+w+n	551	老专家/n 、 /w 老朋友/n
v+c+v	533	产生/v 和/c 维持/v
vN+c+vN	358	交流/vN 与/c 合作/vN
n+w+n+w+n	250	技术/n 、 /w 信息/n 、 /w 产权/n
a+c+a	233	炎热/a 与/c 寒冷/a
n+n+c+n+n	219	战略/n 思想/n 和/c 战术/n 问题/n
n+w+n+c+n	157	价格/n 、 /w 年限/n 和/c 用途/n
v+w+v	145	干预/v 、 /w 控制/v
nP+c+nP	130	何其芳/nP 和/c 毕焕午/nP
n+c+n+n	120	国债/n 和/c 金融/n 债券/n
n+w+n+w+n+w+n	109	钙/n 、 /w 磷/n 、 /w 铁/n 、 /w 锌/n
nP+w+nP	95	孙蔚如/nP 、 /w 赵寿山/nP
a+w+a	93	直观/a 、 /w 形象/a

表1 清华语料有标记联合结构内部词性序列分布

词性标记序列	出现频次	具体例子
n+c+n	701	国家/n 和/c 地区/n
n+w+n	338	方针/n 、 /w 政策/n
v+c+v	311	改革/v 和/c 建设/v
vn+c+vn	238	尊敬/vn 和/c 关注/vn
v+w+v	125	标准化/v 、 /w 规范化/v
n+w+n+w+n	102	政治/n 、 /w 经济/n 、 /w 社会/n
nr+nr+w+nr+nr	92	马/nr 文瑞/nr 、 /w 汪/nr 锋/nr
a+c+a	72	密/a 而/c 得当/a
n+c+n+n	67	专家/n 和/c 人民/n 群众/n
ns+c+ns	67	澳门/ns 和/c 台湾/ns
n+w+n+c+n	64	产业/n 、 /w 产品/n 和/c 技术/n
n+n+c+n+n	61	经济/n 环境/n 和/c 社会/n 环境/n
ns+w+ns	57	北京/ns 、 /w 长春/ns

a+w+a	56	幸福/a 、/w 温暖/a
-------	----	---------------

表2 北大语料有标记联合结构内部词性标记序列分布

根据表1和表2, 频次在前十四位的有标记联合结构分别在两个语料中占据了整个有标记联合结构的37.83%和32.68%。频次为一次的分别为35.88%和30.72%。这两组数据说明了有标记联合结构的词性序列分布呈现两端化, 一方面向频率高的靠拢, 另一方面向频率极端低的接近。

在进行特征分析之前, 首先给出单层单标记联合结构的定义。单层单标记联合结构是有标记联合结构的一部分, 即只有一层句法关系和一个联合结构标记, 例如: “权力/n 和/c 政治/n”就是只有“联合结构”句法关系与“和”一个联合标记的单层单标记联合结构。从频次在前十四位的有标记联合结构的词性序列中可以看出, 单层单标记联合结构分别在两个语料中占居了81.72%和86.43%, 并且这些单层单标记联合结构充分体现出了有标记联合结构的特征: 联合结构中心语的相似性和联合结构的结构平行性。例如: “奢侈/a 或/c 节俭/a, 动物界/n 和/c 植物界/n”和“改革/v 和/c 发展/v, 唐古拉山/ns 、/w 喜马拉雅山/ns”。这些词性序列的分布, 为用规则来识别联合结构提供了充分的条件。

## 2.2 有标记联合结构内部的短语序列分布

根据统计的数据, 在整个有标记联合结构中, 51.84%的联合结构是有一种或一种以上的其他短语组成的, 为了详细的分析联合结构, 需要仔细分析联合结构内部这些短语的分布和组成, 从而为正确的识别联合结构打好基础。因此, 本文根据100万字的清华树库, 用统计的方法详细和具体地分析了联合结构内部的短语。根据统计的结果, 频次出现100次以上的联合结构内部短语分布如表3所示:

短语序列	频次	具体例子
定中结构+定中结构(np-DZ+np-DZ)	2330	[np-DZ 官营/b 工商业/n] 和/c [np-DZ 产品/n 专卖/vN]
定中结构(np-DZ)	1051	[np-DZ 产品/n 分配/vN] 和/c 再分配/n
定中结构+定中结构+定中结构 (np-DZ+np-DZ+np-DZ)	360	[np-DZ 农业/n 发展史/n] 、/w [np-DZ 工业/n 发展史/n] 、/w [np-DZ 银行/n 发展史/n]
述宾结构+述宾结构 (vp-PO+vp-PO)	290	[vp-PO 有/v 影响/n] 、/w [vp-PO 有/v 名望/n]
附加结构+附加结构 (np-AD+np-AD)	159	[np-AD 当前/t 的/u] 和/c [np-AD 长远/a 的/u]
状中结构+状中结构 (vp-ZZ+vp-ZZ)	150	[vp-ZZ 双向/d 选择/v] 、/w [vp-ZZ 合理/aD 流动/v]
状中结构+状中结构 (ap-ZZ+ap-ZZ)	150	[ap-ZZ 极为/dD 重要/a] 、/w [ap-ZZ 极其/dD 复杂/a]
标号结构+标号结构 (np-BH+np-BH)	115	[np-BH “/w 度支/n ”/w] 、/w [np-BH “/w 国用/n ”/w]
述宾结构 (vp-PO)	107	[vp-PO 编写/v 教材/n] 和/c 备课/v

表3 有标记联合结构内部的短语序列分布

根据表3的统计数据, 频次在100次以上的短语序列占据了整个含短语的有标记联合结构的71.19%, 而频次在10以下的短语序列为12.08%, 因此整体的联合结构内部短语倾向于向高频的短语靠拢, 分布比较集中。

从短语分布上看, 联合结构内部短语主要是下面这些: 定中结构、述宾结构、附加结构、状中结构。定中结构是分布最广的一类结构, 在表3中, 占据了80.00%, 其次是述宾结构, 占据了8.70%。因此, 如果在识别联合结构中引进搭配的知识, 定中结构和述宾结构搭配的知识是最重要的。

根据 100 万字的清华语料统计结果,在 6408 个联合结构短语序列中有 396 个联合结构中嵌套着联合结构,占整个联合结构短语序列的 6.18%。虽然嵌套的联合结构数量有限,但凡是有的嵌套的联合结构,无论在结构上还是长度上都及其复杂,给自动识别该类联合结构造成了很大的困难。

### 3 有标记联合结构外部语言学特征统计考察

联合结构内部特征考察主要是从联合结构的微观角度分析联合结构的各种特征,而外部特征则是根据联合结构所处的位置,从宏观的角度考察联合结构,根据识别联合结构的需要,本文主要从联合结构的句法功能分布和联合结构左右边界特征词两个方面考察。

#### 3.1 联合结构的句法功能分布情况

陈小荷(1999)为实词分类拟定了一个含 8 种句法结构的考察框架,徐艳华(2006)在 8 种结构的基础上增加了一类介宾结构,考察了实词的 14 种句法功能。陈小荷(2006)提出按照“语法功能匹配”的方法来实现句法自动分析,根据自动句法分析的需要,不但考察了每一个实词的句法功能,而且考察了每类短语的句法功能,在考察功能时,由原来的 9 种句法结构,推广到 27 句法结构,句法功能也从原来的 14 种推广到 60 种。

联合结构作为一类短语也有其自身的句法功能分布,这里所说的联合结构的句法功能是联合结构在上一级句法结构中所充当的语法功能。根据清华语料的统计,联合结构所充当的句法功能具体如表 4 所示:

语法功能	频次	例证
宾语	1531	[vp-PO 称为/v [np-LH [np-DZ [np-DZ 信息/n 处理/vN ] 产业/n ] 或/c [np-DZ 信息/n 服务业/n ] ]
定语	1293	np-DZ [vp-LH 量入制出/iV 与/c 量出制入/iV ] 思想/n
带“的”定语	1286	[np-DZ [vp-LH [ap-PO 公平/a 负担/n ] 、/w [ap-PO 轻/a 税/n ] 和/c [vp-PO 稳定/v 税负/n ] ] [np-DZ 三/m 原则/n ] ]
带“的”定语中心语	1261	[np-DZ 计算机/n 的/u [np-LH 运行/vN 和/c 维护/vN ] ]
主语	1020	[dj-ZW [np-LH [np-BH “/w 度支/n ”/w ] 、/w [np-BH “/w 国用/n ”/w ] ] [vp-PO 指/v [np-DZ 国家/n 的/u [np-DZ 费用/n 开支/n ] ] ] ]
定语中心语	967	[np-DZ 有关/b [np-LH 理论/n 和/c 原则/n ] ]
介词宾语	595	pp-JB 在/p [tp-LH 古希腊/nS 、/、 古罗马/nS 以及/c [tp-DZ 西欧/nS 中世纪/t ] ]
状语中心语	575	[ap-ZZ 颇为/dD [ap-LH 失落/a 和/c 愤愤/a ] ]
枚举结构前成分	571	[np-MJ [np-LH [np-DZ 官营/b 工商业/n ] 和/c [np-DZ 产品/n 专卖/vN ] ] 等/u ]
谓语	533	[dj-ZW 人们/n [vp-LH [vp-PO 向往/v 大别山/nS ] 、/w [vp-PO 崇拜/v 大别山/nS ] ] ]
联合结构成分	380	[np-LH 货币学派/n 及其/c [np-LH 政策/n 主张/n ] ]
述宾述语	343	[vp-PO [vp-LH 减少/v 和/c [vp-ZZ 不/dN 加/v ] ] 租税/n ]
方位词前加成分	182	[dl-FZ [np-LH [np-AD 当前/t 的/u ] 和/c [np-AD 长远/a 的/u ] ] 预期/vN ]
状语	131	vp-ZZ [pp-LH [pp-JB 与/p 人民/n ] 、/w [pp-JB 与/p 时代/n ] 、/w [pp-JB 与/p [np-DZ 社会/n 进步/vN ] ] ] [vp-ZZ 相/d 结合/v ] ]

连谓结构后成分	111	[vp-LW 来/v [vp-LH 完成/v 或/c [vp-ZZ 部分/d 完成/v ]]]
兼语后成份	59	[vp-JY 促使/JY 人们/n [vp-LH [vp-PO 检查/v [np-DZ [np-DZ [np-DZ 国民经济/n 计划/n ] 比例/n ] 的/u [np-DZ 协调/vN 状况/n ]]]] 和/c [vp-PO 剖析/v [np-DZ [sp-FW [np-DZ 计划/n 执行/vN ] 中 /f] 的/u 问题/n ]]]]
带“地”状语	53	[vp-ZZ [vp-LH 世代相袭/iV 、 /w 潜移默化/iV ] 地/u 熏陶/v ]
兼语	51	[vp-JY 依赖/v [np-LH 个人/n 或/c 集团/n ] 承担/v ]
附加结构前成分	41	[np-AD [np-LH 物理/n 、 /w 力学/n ] 的/u ]
连谓结构前成分	38	[vp-LW [vp-LH [vp-PO 建立/v [np-DZ 社会学/n 理论/n ] ]] 与/c [vp-ZZ [dp-AD 历史/d 地/u ] [vp-PO 重构/v [np-DZ [np-DZ 古典/b 社会学/n ] 思想/n ]]]] [vp-ZZ 相/d 结合/v ]]
框式结构中成分	37	[vp-KS 是/vC [vp-LH [vp-ZZ 互为/d 联系/v ] 和/c [vp-ZZ 相互/d 作用/v ]]] 的/u ]
补语	26	[ap-SB 红/a 得/u [ap-LH [ap-ZZ 那么/rD 鲜/a ] , /w [ap-ZZ 那么/rD 艳/a ]]]]

表 4 联合结构充当的语法功能

从表 4 中可以看出,有标记联合结构主要出现在下面这些句法的位置:“宾语、定语、带‘的’定语、带‘的’定语中心语、主语、定语中心语、介词宾语”,占据了整个联合结构句法分布的 71.75%。这些分布数据一方面描述了有标记联合结构本身在整个句法体系中充当的句法功能,另一个方面也为用搭配识别联合结构提供了详细的数据依据。例如,基于介宾搭配实例的有标记联合结构边界识别,根据统计共有 595 个联合结构位于介宾的位置,据观察联合结构的前半部分处于介词和联合标记之间,介词和联合标记可以形成一个框架。”对于/p【金融/n 政策/n 和/c 财政/n 政策/n】”这个例子,根据“‘对于’和‘和’”可以确定联合结构的前一部分,识别出左边界,然后根据联合结构对称性或语义相似性,进而可以确定后半部分的右边界。如果文本中介宾短语识别的正确率和召会率比较高,在识别联合结构的时候完全可以通过介宾短语识别出联合短语。

### 3.2 联合结构的左边界特征词及其分析

统计联合结构左边界特征词的依据是有的词语经常出现在联合结构左边界的外部而不太出现在联合结构的内部,即联合结构的左边界一般不会跨越这些词。根据这一观察到的语言学现象,把联合结构的边界范围限定在以“。!?”结尾的子句范围内,联合结构的左边界绝对不会跨越第一个联合标记,故此考察范围限定在从句子开始到第一个标记这么长的范围内,称为 $\beta$ 。“设 W 是任一个词,  $f(W\_left\_outside)$  表示 W 在  $\beta$  范围内,在联合结构外部出现的频次,  $f(W\_left)$  表示 W 在  $\beta$  范围内出现的频次,词 W 成为联合结构左边界特征词的频率计算公式(吴云芳. 2003)”:

$$P(w) = \frac{f(w\_left\_outside)}{f(w\_left)} \quad \text{公式1}$$

P 的值越高,说明 W 成为左边界特征词的可能性就越大,由于本文统计时用的语料规模是 150 万字(吴云芳统计时所用语料为 56 万字)并且是为计算机自动识别有标记联合结构服务的,所以本文给 P 的经验阈值为 0.8,也就是当  $P \geq 0.8$ , W 就可能成为联合结构的左边界特征词,然后结合人工语言学知识的内省,从中确定联合结构的左边界特征词。表 5 给出了频率值在前十位并且经过人工语言学知识内省后的左边界特征词

左边界特征词	频率值	具体例子
是	0.99417	这/r 将/d 是/v {中国/ns 、/w 南非/ns} 两/m 国/n 人民/n 一个/m 值得/v 纪念/v 的/u 日子/n 。/w
将	0.985401	这/r 将/d {对/p 中华民族/n 的/u 繁荣/an 振兴/vn 、/w 对/p {亚太地区/j 与/c 世界/n} 和平/n 、/w 对/p 跨/v 世纪/n {两岸/n 关系/n 的/u 发展/vn 和/c 祖国/n 和平/a 统一/vn 事业/n 的/u 推进/vn}}
了	0.982394	也/d 带动/v 了/u {水产业/n 、/w 养殖业/n} 等/u 其它/r 产业/n 的/u 兴起/vn 。/w
就	0.981132	就/p {产品/n 开发/vn 和/c 科技/n 投入/vn} 而言/u ，/w
使	0.979167	使/v {学生/n 和/c 教师/n} 之间/f {平等/a 、/w 自由/a} 地/u 沟通/v 。/w
提供	0.96875	并/c 要求/v 其/r 提供/v {证明/vn 材料/n 和/c 与/p 价格/n 违法/vn 行为/n 有关/v 的/u 其他/r 资料/n} ；/w
在	0.959097	/w 英雄/n 的/u 中国/ns 人民/n 在/p {以/p 江/nr 泽民/nr 同志/n 为/v 核心/n 的/u 党中央/nt 坚强/a 领导/vn 和/c 党/n 的/u 十五大/j 精神/n 指引/v} 下/f ，/w
这	0.956522	这/r 是/v 促使/v 价格/n {涨幅/n 趋/Vg 低/a 、/w 涨势/n 平缓/a} 的/u 重要/a 原因/n 。/w
为	0.940711	青少年/n 就/d 可以/v 学/v 到/v 革命/vn 先辈/n 甘愿/v 为/p {祖国/n 和/c 人民/n} 奉献/v 一切/r ，/w
有	0.935294	有/v {重大/a 和/c 深远/a} 的/u 意义/n 。/w

表 5 联合结构左边界特征词

从这十个高频的边界特征词可以看出，边界特征词主要集中在这几类词性上：

动词，例如“是、使、提供、有”，这类动词的特点是它们后面跟的联合结构都十分复杂，作为左边界特征词的动词可以在一定程度上帮助识别联合结构。介词，例如“在、就、为”等。副词，例如“将”。其中还有一类代词，在表 5 中的是“这”，边界特征词有时候不可能很精确，其中在表中关于“这”的例子就是一个很好的说明。

### 3.4 联合结构的右边界特征词及其分布

根据统计左边界特征词的方法，联合结构的右边界范围也是限定在以“。！？”结尾的子句范围内，由于联合结构的右边界特征词绝对不会跨越联合结构的最后一个联合标记，所以考察范围限定在从最后一个联合标记开始到句子结束这样一个范围内，这个范围记做  $\alpha$ 。“设  $W$  是任一个词， $f(W\_right\_outside)$  表示  $W$  在  $\alpha$  范围内，在并列结构外部出现的频次， $f(W\_right)$  表示  $W$  在  $\alpha$  范围内出现的频次。词  $W$  成为并列结构右边界特征词的频率计算公式为（吴云芳. 2003）”：

$$P(w) = \frac{f(w\_right\_outside)}{f(w\_right)} \quad \text{公式 2}$$

根据 150 万字的语料库统计规模，结合自动识别有标记联合结构的需要，本文把右边界特征词  $P$  的阈值也设定为 0.8，根据语言学知识的内省再结合大于或等于 0.8 的  $P$  值，本文确定了右边界特征词的范围。表 6 是频率值居于前十位的右边界特征词。

右边界特征词	频率值	具体例子

都	0.981735	{民主/a 法制/n 建设/vn 、 /w 精神文明/n 建设/vn 和/c 其他/r 各项/r 事业/n} 都/d 有/v 新/a 的/u 进展/vn 。 /w
等	0.981402	中国/ns 愿意/v 加强/v 同/p {联合国/nt 和/c 其他/r 国际/n 组织/n} 的/u 协调/vn , /w 促进/v 在/p {扩大/v 经贸/j 科技/n 交流/vn 、 /w 保护/v 环境/n 、 /w 消除/v 贫困/an 、 /w 打击/v 国际/n 犯罪/vn} 等/u 方面/n 的/u 国际/n 合作/vn 。 /w
了	0.971319	中国/ns 伟大/a 的/u 领导人/n {孙/nr 中山/nr 、 /w 毛/nr 泽东/nr 、 /w 周/nr 恩来/nr 和/c 邓/nr 小平/nr} 的/u 榜样/n 极大/a 地/u {启发/v 和/c 鼓舞/v} 了/u 非洲/ns 人民/n 。 /w
结合	0.965517	把/p 经济社会/n 的/u 革新/vn 与/p {国家/n 行政/n 机构/n 的/u 改革/vn 以及/c 党/n 的/u {革新/vn 与/c 整顿/vn}} 结合/v 起来/v 。 /
时	0.948276	当/p 他们/r 接到/v 政府/n 送/v 来/v 的/u {食物/n 、 /w 衣服/n 和/c 燃料/n} 时/Ng , /w
提供	0.96875	地铁/n 给/p 纽约/ns 数百万/m {“/w 上班族/n ” /w 和/c 游客/n} 提供/v 了/u 方便/an , /w
也	0.93007	加强/v 自身/r 建设/vn , /w 认真/ad 履行/v 职责/n , /w 继续/v 加强/v 同/p {各/r 地区/n 、 /w 各/r 部门/n} 的/u {沟通/vn 与/c 协作/vn} , /w {各级/r 政府/n 、 /w 各个/r 部门/n} 也/d 要/v 继续/v {重视/v 和/c 加强/v} 语言/n 文字/n 工作/vn 。 /w
是	0.91844	“/w 明年/t 中国/ns 外交/n 工作/vn 的/u {重点/n 和/c 努力/a 方向/n} 是/v 什么/r ? /w 将/d 会/v 有/v 哪些/r 举措/n ? /w ” /w
在	0.892857	要/v 继续/v 发挥/v {各/r 民主党派/n 、 /w 人民团体/l 和/c 各界/r 爱国人士/n} 在/p 政协/j 中 /f 的/u 作用/n , /w
的	0.801342	向/p {伟大/a 而/c 光荣/a} 的/u [中国/ns 共产党/n]nt 致以/v 崇高/a 的/u 敬意/n 。 /w

表 6 联合结构右边界特征词

像左边界特征词一样，右边界特征词也分布在“动词、副词、介词”上，如“是、提供，都，在”等，所不同的是，右边界特征词主要集中在助词上，表 6 中的助词有“等、了、的”等。其中助词中的“等、等等、的”作为右边界特征词的时候，与联合结构结合非常紧密，基本上是联合结构的右边界特征词了。

## 4 结语

本文在大规模语料的基础上，用统计的方法详细考察了有标记联合结构内部和外部的语言学特征，这些考察一方面为计算机识别有标记联合结构提供了语言学知识，另一方面为从语言学的角度量化地研究联合结构提供了精确的数据。

### 参 考 文 献

- [1] 陈小荷 2002 “把…HVP”的计算分析，世界汉语教学第 1 期，47 页。
- [2] 陈小荷 1999 从自动句法分析角度看汉语词类问题，语言教学与研究第 3 期，70 页。
- [3] 冯志伟 2004 《现代语言学流派》，陕西：陕西人民出版社。
- [4] 吴云芳 2003 《面向中文信息处理的现代汉语并列结构研究》，北京大学博士论文。
- [5] 徐艳华 2006 《现代汉语实词语法功能考察及词类体系重构》，南京师范大学博士论文。
- [6] 俞士汶 2003 《计算语言学概论》，北京：商务印书馆。
- [7] 詹卫东 2000 《面向中文信息处理的现代汉语短语结构规则研究》，北京：清华大学出版社，广西科学技术出版社。