

现代汉语树库标注常见问题举例

詹卫东

北京大学中文系 北京大学中国语言学研究

文档最近更新时间： 2009-07-30

目 录

一 短语的层次划分与功能定性问题举例.....	1
1.1 短语结构的层次划分问题举例.....	1
1.1.1 dp + vp + np	1
1.1.2 d + np + vp	1
1.1.3 np + d + np	2
1.1.4 n + d + n	2
1.1.5 np + np + vp	3
1.1.6 np + vp + vp	3
1.1.7 np + vp + qp	4
1.1.8 vp + np + qp	4
1.1.9 助动词 + v + xp	5
1.1.10 助动词 + 助动词 + vp	5
1.1.11 像.....那样 + ap.....	5
1.1.12 像.....似的.....	8
1.1.13 像.....一样 像.....一般.....	9
1.1.14 跟.....一样.....	9
1.1.15 跟.....一样 + ap.....	9
1.1.16 没有.....可言.....	10
1.1.17 一点到两点之间.....	10
1.1.18 (自)从.....以后/以来.....	11
1.1.19 十一点半.....	11
1.1.20 np + 在 + vp + 时, xp.....	11
1.1.21 xp + 等(等等).....	13
1.1.22 有+个+np+vp.....	14
1.1.23 有 + 个 + np + r + vp	15
1.1.24 好得很、好极了.....	16
1.1.25 x + ude2 + vp	16
1.1.26 有是有.....	17
1.1.27 该轮到我要上课的时候了.....	17
1.1.28 1997年4月8日.....	17
1.1.29 十多个.....	18
1.1.30 有院墙高、有院墙那么高、有三米高.....	18
1.1.31 沙包大的拳头.....	18

1.1.32	我女儿不及你女儿一半聪明.....	19
1.1.33	职务名称 ₁ 兼 职务名称 ₂	19
1.1.34	“X 包括 Y”“X, 如, X1, X2, ...”	19
1.2	短语结构的功能定性问题举例.....	20
1.2.1	np+f.....	20
1.2.2	vp+udel+a.....	20
1.2.3	o+m+q.....	21
1.2.4	是 vp 还是 tp.....	21
1.2.5	三比二.....	22
1.2.6	南偏西.....	22
1.2.7	这下子.....	22
1.2.8	量词重叠式.....	24
1.2.9	形容词重叠式.....	24
1.2.10	“数+量”结构	24
1.2.11	一分多钟、三米多长.....	24
1.2.12	有一天、有一次.....	25
1.2.13	联合结构.....	25
1.2.14	“极度紧张”是 ap 还是 np?	26
1.2.15	很中国.....	27
1.3	省略造成的短语功能变异和层次分析问题举例.....	27
1.3.1	比字句中的成分省略.....	27
1.3.2	定中结构中的中心语省略.....	30
1.3.3	骈合结构中的成分省略.....	31
1.3.4	复句中的成分省略.....	33
1.3.5	跟介词结构有关的省略.....	33
1.3.6	跟述补结构有关的省略.....	35
1.3.7	述宾结构中述语的省略.....	35
1.3.8	主语省略.....	36
1.3.9	谓语省略.....	37
1.3.10	跟“不”有关的省略.....	37
1.4	冗余造成的短语功能变异和层次分析问题举例.....	38
1.4.1	dp+np 组合	38
1.4.2	vp+f 组合	39
1.4.3	pp+usuo+vp 组合.....	39
二	复句判别问题举例.....	42
2.1	有关联词语但没有分句标点的句子.....	42
2.2	“np + vp1, + vp2, +, + vpn”形式的句子	42
2.3	总分、解说类句子.....	44
三	分词问题举例.....	45
3.1	组合型歧义字段的切分.....	45
3.2	“X + 方位词”字段的切分.....	46
3.3	重叠式字段的切分.....	46
3.4	离合词的切分.....	47
3.5	“特别是”“取决于”“倒不如说”等是否要切分开.....	47

3.6	“伸出”“穿上”“激起”“露出”等是否要切分开	47
3.7	“选举权”“群众性”“正确率”等是否要切分开	47
3.8	“外交部长”“课题组长”等是否要切分开	48
3.9	“深蓝色”“淡绿色”“粉红色”等是否要切分开	48
3.10	“纸张”“雨滴”“木条”“活字块”等是否要切分开	48
3.11	“横槛”“人声”“党证”等是否要切分开	48
3.12	“6.7%”“上百”等表数成分是否要切分开	48
3.13	“米每秒”“英里每小时”等如何处理	48
3.14	“有的是”	49
3.15	“大不了”	49
3.16	“一整夜”	49
3.17	“一会儿/一会/这会儿”分别怎么分析?	49
四	词性判别问题举例	50
4.1	介词与连词、动词的纠葛问题	50
4.2	量词与名词、动词的纠葛问题	50
4.3	助词与其他词类的纠葛问题	51
4.4	与代词相关的词性判别问题	52
4.5	兼属副词和区别词的修饰词	52
4.6	“同时”是什么词性	53
4.7	“好多年”“二十多年”中的“多”是什么词性	53
4.8	“一个又一个”中的“又”如何标注	53
4.9	“爬呀爬”中的“呀”是 y, 还是 o?	53
4.10	“约两米高”中的“约”	53
4.11	“一般五寸至二尺长”中的“一般”	54
4.12	“早在 3000 多年前”中的“早”	54
4.13	“三十年来”和“三十来年”中的“来”	54
4.14	拟声词的功能定性	55
4.15	“那就比一比吧”中的“那”	55
4.16	“生育健康”中的“生育”和“健康”分别是什么词性?	55
4.17	“等我醒过来”中的“等”是什么词性?	56
4.18	“这一发现”中的“发现”是什么词性?	56
4.19	“以图”、“报经”、“诸如”等紧缩型词语如何标注词性?	56
4.20	“两尺见方”中的“见方”是什么词性?	56
4.21	“总的来说”“综上所述”是什么词性?	56
五	文言及书面特殊用法的处理	56
5.1	文言	56
5.2	特殊用法	58
5.2.1	“印(任)务”如何处理	58
5.2.2	被注音分开的词如何处理	58
5.2.3	“好——笑”这类被破折号分开的词如何处理	58
5.2.4	文章中的脚注或尾注如何处理	59
5.2.5	正文中的随文注释如何处理	60

本文档对树库加工过程中碰到的短语的层次划分问题和功能类归属问题，以及一些词语切分和词性判定的问题，通过分析具体实例，来说明我们的处理方式。

一方面，本文档能够起到类似于案例法中的案例的作用；另一方面，本文档中一些不容易决断的问题，可以引发大家更多的思考，共同来寻求解决这些问题的方法。

一 短语的层次划分与功能定性问题举例

1.1 短语结构的层次划分问题举例

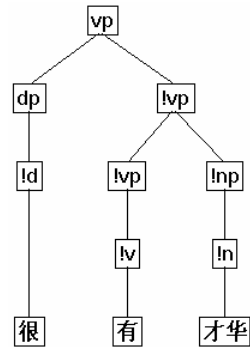
1.1.1 dp + vp + np

从句法关系上说，在上面这个格式中，dp充当状语性成分，np充当宾语性成分，该格式有下面两种切分的可能性：

- a. [dp + [vp + np]] b. [[dp + vp] + np]

这种结构大多数情况下应按a结构模式分析。比如：

- [很 [有 才华]]
- [经常 [看 原版电影]]
- [不 [认识 这个人]]



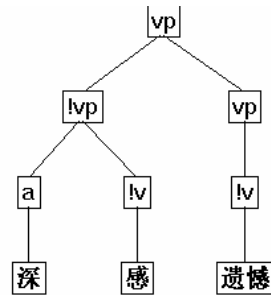
有的时候因为音节原因，dp、ap是单音节成分，只能修饰单音节动词，不能修饰述宾结构，此时应按b结构模式分析¹，比如：

- [[深感] 遗憾]

跟上面格式类似的，还有下面这些格式：

- dp + vp + vp 如：不 能 说
- ap + vp + np 如：认真 学习 历史知识
- ap + vp + vp 如：努力 练习 打篮球

这些格式一般都应该按照a结构模式进行层次分析。



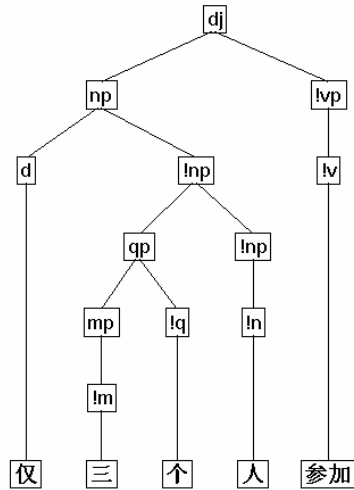
1.1.2 d + np + vp

这个格式中的 d 由“仅、就、只、光”等表示限定量的副词充当。这类格式分析为：

vp [[d np] vp]

比如：

¹ 对于像“飞袭敌营”这样的例子，可以按照 b 结构模式：[[飞 袭] 敌营]，只是这里的“飞”不是 dp，也不是 ap，应该处理为一般动词作状语，跟“袭”组合成为 vp[飞 袭]。为体现这种结构组合的特殊性，我们规定“飞”—“袭”，“深”—“感”应分析为直接由词形成的组合，即 vp→v lv，vp→a lv。



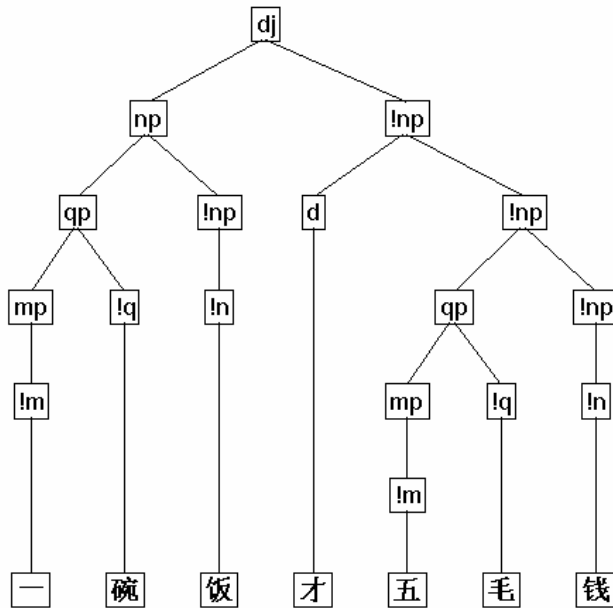
上述格式中vp位置也可能是ap, dj等成分, 层次分析也做类似处理。
 注意: 这个格式中d不需要上升为dp再参与组合, 而是直接参与组合。

1.1.3 np + d + np

这个格式中的 d 由“仅、就、只、光”等表示限定量的副词充当。这类格式分析为:

dj [np [d np]]

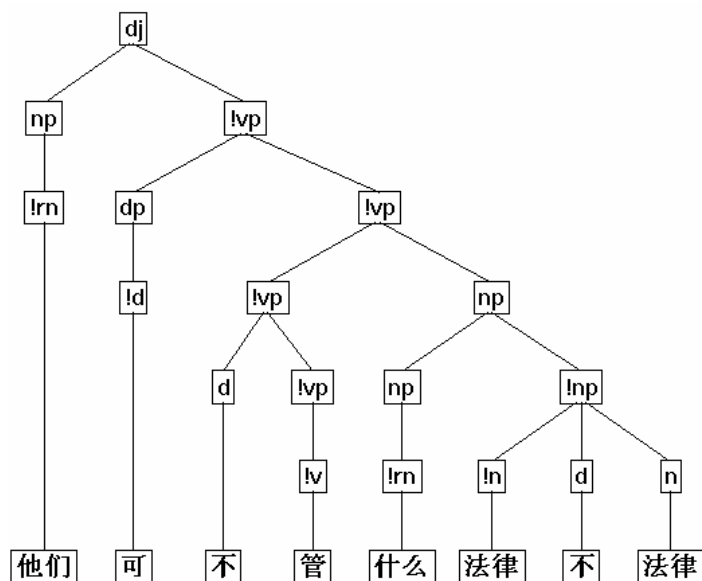
比如:



注意: 这个格式中 d 不需要上升为 dp 再参与组合, 而是直接参与组合。
 这类格式可以看作是省略定语造成的, 可参见下文 1.3 节的说明。

1.1.4 n + d + n

否定副词“不”有的时候可以出现在两个同形名词之间, 形成“n + d + n”格式。比如: “他们可不管什么法律不法律”, 对这种“n + d + n”格式, 按三支结构处理, 整体标为 np。



1.1.5 np + np + vp

这个格式有下面两种切分的可能性：

- a. [np [np vp]] b. [[np np] vp]

这种结构一般按a结构模式分析。比如：

- [这个人 [象棋 下得特别好]]
 [他 [思想 比较保守]]

如果前一个np和后一个np构成领属关系，则按b结构模式分析。比如：

- [[他 父亲] 喜欢下围棋]
 [[我们班 男生] 都报名了]

跟这个格式类似的还有 np + np + ap，处理方式相同。

1.1.6 np + vp + vp

这个格式有下面三种切分的可能性：

- a. [np [vp vp]] b. [[np vp] vp] c. [np vp vp]

这种结构有的可以按 a 结构模式分析。比如：

- [张老师 [说话 很快]]
 [张三 [唱歌 唱得很好]]
 [他 [办事 很有章法]]

有的可以按 b 结构模式分析。比如：

- [[谁 承担这个任务] 还没有决定]

有的可以按 c 结构模式分析，即分析为“三分支结构”。比如：

- [他们 + 一旦斗起来 + 是难解难分的]
 [他 + 说这些话 + 是有来头的]

语料中还有比上述格式更复杂一些的情况，vp 和 vp 之间还可以再插入 np，形成如下格式：

np + vp + np + vp

比如：我们 利用外资 重点 放在 ……

这类格式一般可处理为：dj [dj [我们 利用外资] dj [重点 放在 ……]]
也就是前一个 dj(np + vp)作主语，后一个 dj(np + vp)作谓语，组合为一个更大的 dj。
跟这个格式类似的还有 np + vp + ap，处理方式相同。

1.1.7 np + vp + qp

这个格式有下面三种切分的可能性：

a. [np [vp qp]] b. [[np vp] qp] c. [np vp qp]

这种格式一般应该按 a 结构模式分析。比如：

[我们 [结婚 三十年了]]

其中 vp+qp 构成主谓结构，即处理为主谓谓语句。在 vp 和 qp 之间还可以插入状语性成分，比如“我们结婚已经三十年了”。

值得注意的是，有的例子看上去像是这种结构，但其实不是。比如：

(去年) 我们引进项目一百多个，

应该分析为：

[我们 [引进 项目一百多个]]

其中“项目一百多个”分析为定语后置的 np。可能也有人倾向于把这种结构处理为“引进项目 + 一百多个”，看作主谓结构。但我们不这样处理。因为如果这样分析，“一百多个”和“项目”之间的数量修饰关系就无法直接获得，这在机器翻译等应用系统中，处理起来是不方便的（比如汉英翻译时会造成大量的语序调整问题）。另外，这种格式也可能分析为双宾结构：“引进 + 项目 + 一百多个”。我们也不这样处理。因为如果这样分析，就会带来汉语中双宾动词范围的急剧膨胀。一般来说，语言中真正的“双宾”动词只占动词中很少的一部分，但实际上汉语中许多动词（包括动词性词组）都可以出现在“v + n + qp”格式中²（比如“查获 盗版光盘 一万张”，“打死 武装分子 二十名”，“找出 硬伤 三十处”），如果都处理为双宾结构，弊大于利。在综合权衡后，我们倾向于将这种格式分析为一般的述宾结构，只是宾语比较特别，是数量定语（qp）后置形成的 np。

类似的例子如“卖出 苹果 50 斤， 牛肉 300 斤， ……”，都应该处理为一般的述宾结构：[卖出 [[苹果 50 斤] [牛肉 300 斤]]]

1.1.8 vp + np + qp

这个格式就是上面组合格式中讨论的情况，即将上面格式中的 vp 部分展开为“vp np”形成的。为醒目起见，单列出这一类格式。实例：

引进 项目 一百多个

卖出 苹果 50 斤

拥有 资金 500 万

建成 创新基地 15 个

² 我们的认识是：因为这种格式是语用表达的需要造成的，因此对 v 没有句法层面的特别的选择限制，只要 v 能带宾语，理论上就可以进入这种格式。而句法上的双宾语构造对 v 有严格的选择限制，必须一类特定的动词，即双宾动词（ditransitive verb）才能进入。

对这类格式，统一处理为[vp [np qp]]，其中“np qp”组合为定语后置的 np。

1.1.9 助动词 + v + xp

其中xp代表任意成分。这个格式有下面两种切分的可能性：

- a. [助动词 [v xp]] b. [[助动词 v] xp]

这种结构一般按a结构模式分析。比如：

- [会 [说 英语]]
[能 [说 明白]]

1.1.10 助动词 + 助动词 + vp

这个格式有下面两种切分的可能性：

- a. [助动词 [助动词 vp]] b. [[助动词 助动词] vp]

这种结构一般按a结构模式分析。比如：

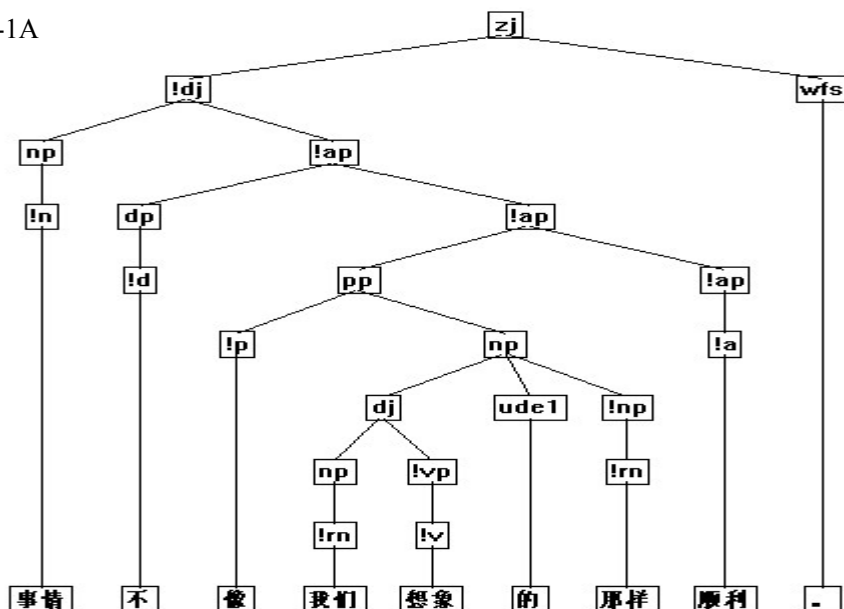
- [敢 [肯定 他说谎]]
[肯定 [值得 试一试]]

1.1.11 像……那样 + ap³

实例：像 我们想像 的 那样 顺利

在标注中可能会出现多种分析方法：

图 1-1A



³ 本文中谈到的问题中有不少是通过 TreeEditor 工具抽取短语结构规则发现的。抽取到的规则中有相当一部分低频的组合模式，这些低频的组合模式有较大可能意味着其中存在分析上的问题。下文图中用红色叉表示有问题的分析方式。

图 1-1B

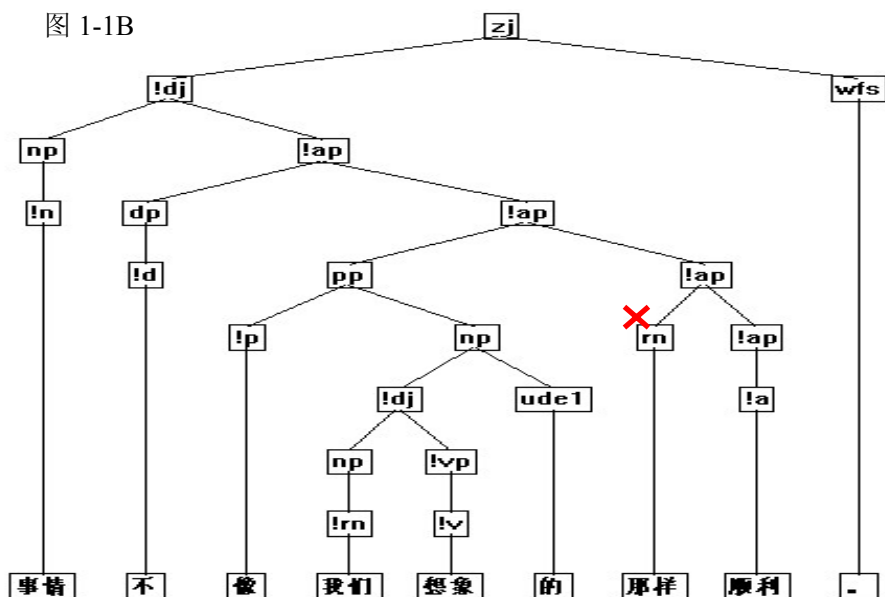
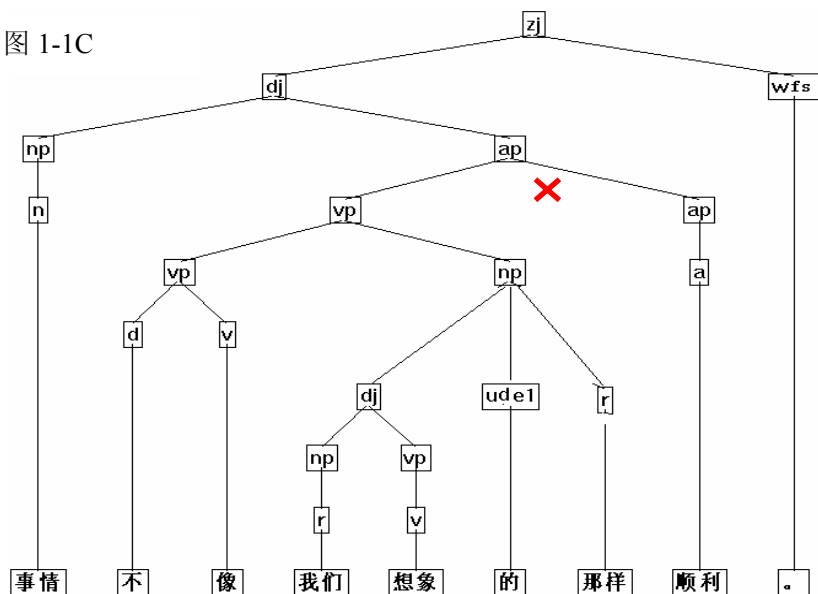


图 1-1C



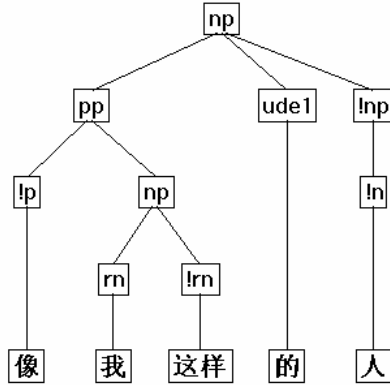
上面例子一方面有层次的问题，同时也牵涉到对“像”、“那样”的词性的认识。

图 1-1C 所示的分析不合理，会造成 $ap \rightarrow vp \rightarrow !ap$ 这样的组合模式，但这种组合模式很难说明 **vp** 跟 **ap** 之间的结构关系（**vp** 作状语？）。

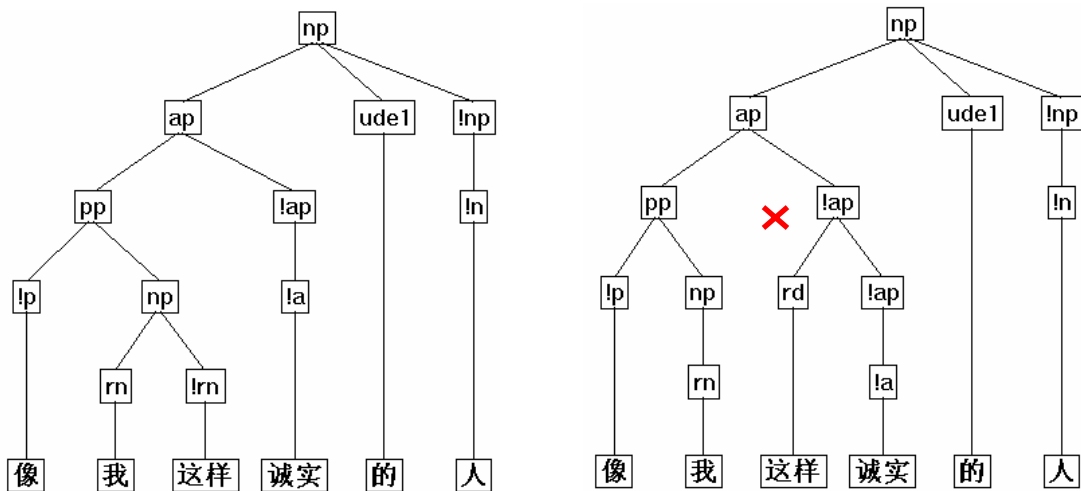
图 1-1A 跟 1-1B 所示的两种分析方式从层次上说都可以接受，主要差别实际上是对“那样”的词性的判别：如果按照图 1-1B 的分析方式，“那样”作状语，不应分析为“**rn**”（担当名词功能的代词）。因此，上面例子应该按照图 1-1A 所示的方式进行分析。“像”在这个结构中分析为介词 **p**，而不分析为动词 **v**。如果句子改为“事情不像我们想像的那样”，这时“像我们想像的那样”作中心谓语，其中“像”分析为动词。这样处理，类似于语法学界通常对“在”的分析。比如在句子“张三在图书馆自习”中，“在”一般分析为介词，而如果句子是“张三在图书馆”，则“在”分析为动词。像“像、在、跟”这一类介词，都保留了很强的动词属性，可以带宾语后作句子的中心谓语。这些词在句中标为介词还是动词，

应根据其所处的句法位置来决定。

有的时候，这个结构中，“那样”（或“这样”）后面的谓词性成分会省略，比如“像我这样的人”，原式应为“像我这样（善良、聪明、诚实、穷困潦倒、……）的人”。其结构分析如下面图示：



如果“这样”后面的谓词性成分不省略，比如“像我这样诚实的人”，其结构分析应如下面左图所示：

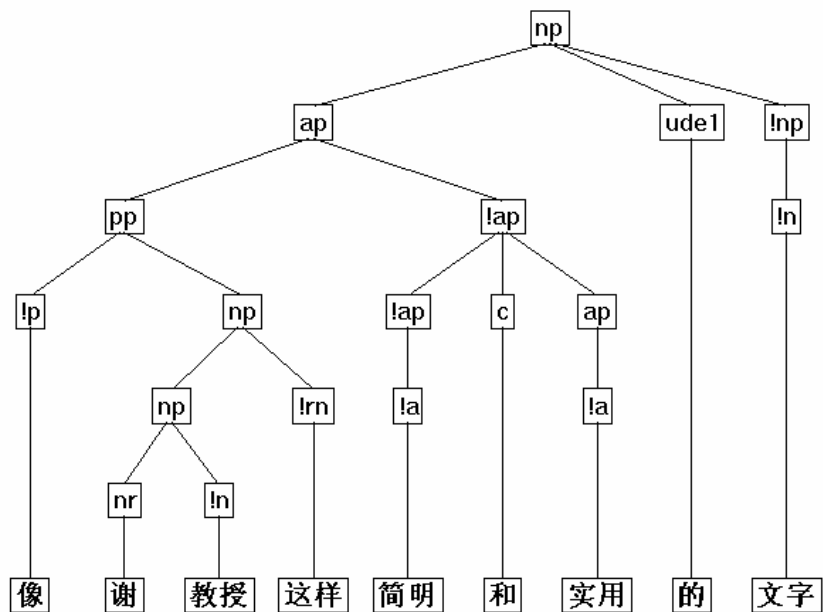


上面两种分析中，对“这样”的处理不同，左图中“这样”为“代名词”*rn*；右图中“这样”为“代副词”*rd*。相比较而言，左图的分析更好一些（跟上文“像我们想像的那样顺利”中的处理是一致的）。因为按照左图的模式，“像 + 我这样”形成结构，介词宾语“我这样”的中心成分为“这样”，如果“我”不出现，形成的结构“像 + 这样”依然可以成立，同时，再与“诚实”组合成“像这样诚实”，也依然可以成立。如果按照右图的模式，“这样”修饰“诚实”，但如果“这样”不出现，形成的组合“像我 + 诚实”不成结构。如此看来，右图的分析从结构模式的类推性上讲，要差一些。因此，我们取左图的分析模式。

有的时候，这个结构中“像”后的宾语成分也会涉及到省略问题，造成结构分析和语义解释之间的一致。比如：

可读者反映，像谢教授这样简明和实用的文字并不多见。

上面这个例句把一个比较复杂的内容压缩成一个简单的结构来表达了。原意应该是“谢教授写的文章简明而实用，像谢教授写的这种风格的文章，并不多见”。这个意思被压缩成“像谢教授这样简明和实用的文字”，其结构分析如下图所示：



其中“谢教授”应该理解为转指“谢教授所写的文章”。(参见下文关于省略对短语结构标注的影响的相关内容)

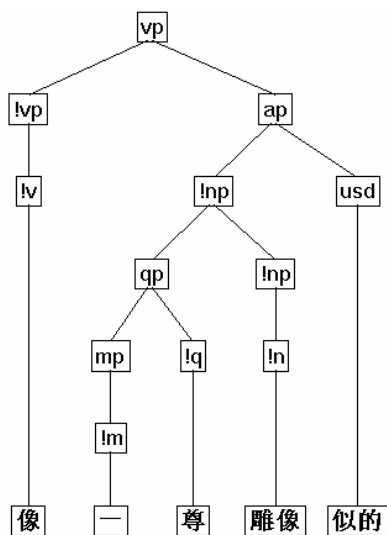
1.1.12 像……似的

“像 X 似的”中 X 可以是 np，也可以是 vp，比如：

np: 像 一尊雕像 似的

vp: 像 着了魔 似的

无论 X 是什么性质的短语，“X 似的”的功能类都可以分析为 ap，“X 似的”一般出现在谓语位置，状语位置、补语位置、以及在“像”后面的宾语位置。



在谓语位置上的“像”都分析为动词。如果“像…似的”处在 vp 前的状语位置上，则“像”分析为介词 p。

1.1.13 像……一样 像……一般

“像……一样”“像……一般”结构，分析方式与“像……似的”相同。“一样”“一般”跟“似的”一样，分析为助词 u。（注意：这里的“一般”跟“很一般”中的“一般”不同，后者是形容词 a）

1.1.14 跟……一样

朱德熙先生在《语法讲义》中分析“跟……一样”结构时，把“跟……一样”分成了两种情况：

- A 脸色跟纸一样
- B 里屋跟外面一样

A 中的“跟……一样”表示比拟，“一样”可以换成“似的”；B 中的“跟……一样”表示比较，“一样”不能换成“似的”。

A 中“跟 X 一样”的结构为：跟 [X 一样]

B 中“跟 X 一样”的结构为：[跟 X] 一样

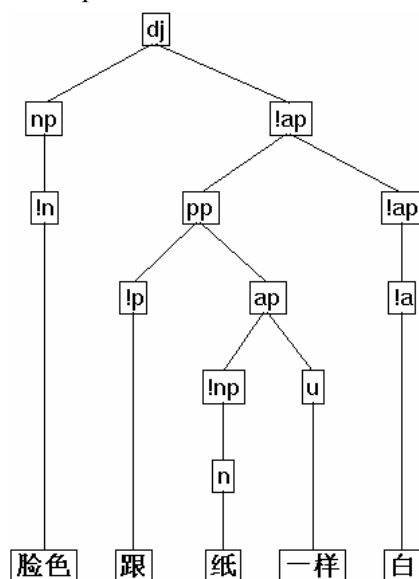
按照朱德熙先生的分析，“跟……一样”表示比拟，则分析为 $vp \rightarrow !vp\ ap$ ，即述宾结构。其中“跟”分析为动词 v（类似上文对“像”的分析），“一样”分析为助词 u（类似于“似的”）；如果表示比较，则分析为 $ap \rightarrow pp\ !ap$ ，即状中结构。其中“跟”分析为介词 p，“一样”分析为形容词 a。

1.1.15 跟……一样 + ap

“跟……一样+ap”也有表比拟和表比较两种情况。比如：

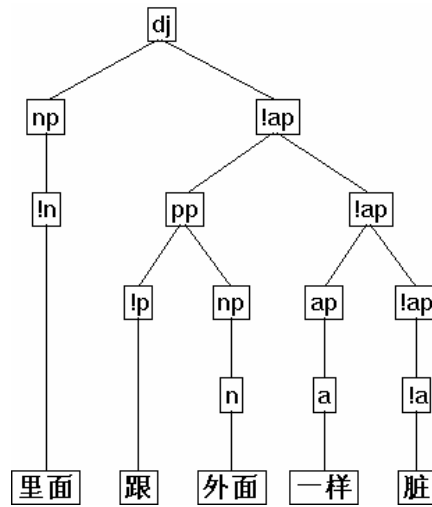
- A 脸色跟纸一样白
- B 里屋跟外面一样脏

如果是情况 A，“跟……一样 ap”分析为（类似上文“像……那样 ap”）：



这种情况下，“跟”分析为介词，“一样”分析为助词，附着在前面的 np 上。

如果是情况 B，“跟…一样 ap”分析为：

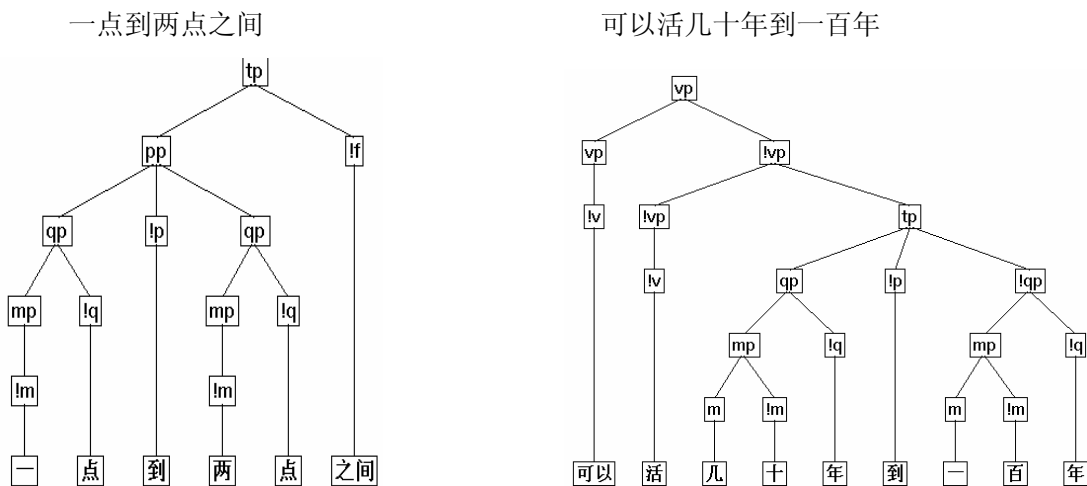


这种情况下，“跟”分析为介词，“一样”分析为形容词，作状语修饰后面的 ap（“脏”）。

1.1.16 没有……可言

比如“没有什么民主可言”。这个格式做三分处理，“没有”和“可言”都分析为动词 v，二者之间的成分一般都是 np 成分。组合模式为： $vp \rightarrow !v \quad np \quad v$

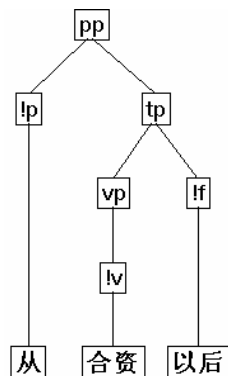
1.1.17 一点到两点之间



“几十年到一百年”后面还可能出现“不等”，形成“几十年到一百年不等”这样的表达形式。组合模式可分析为： $tp \rightarrow !tp(\text{几十年到一百年}) \quad u(\text{不等})$ “不等”分析为助词。

1.1.18 (自)从……以后/以来

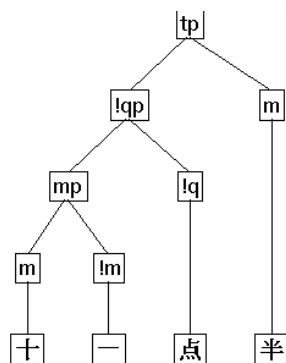
在这个格式中，“从、自从”都分析为介词 p。“以后”“以来”分析为方位词 f。整个结构处理为二分结构。



类似的，“以上”“以下”“以前”等等，均分析为访问词 f。

需要注意的是，从语义上说，“以前”“以后”已经从表达方位转为表达时间，而且可以像“今天”“明天”这样的典型时间词一样独立使用（上面例句中“以后”是后附在 vp “合资”上，不是独立使用的情况），在这种情况下，应该先将 f 上升为 tp，再参与组合。比如“以后他就再也没来了”，其中“以后”的参照时间省略，没有明确表达出来，这时候，应先将“以后”上升为 tp，再跟后面的 dj 组合。

1.1.19 十一点半



“半”分析为数词。

1.1.20 np + 在 + vp + 时, xp

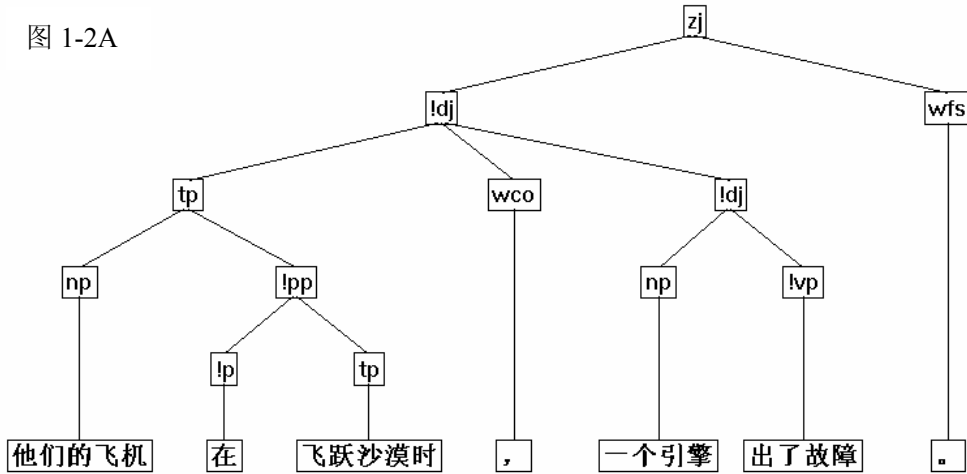
例如：“他们的飞机在飞跃沙漠时，一个引擎出了故障。”

这个格式有可能分析为：

A [np + 在 vp 时], xp

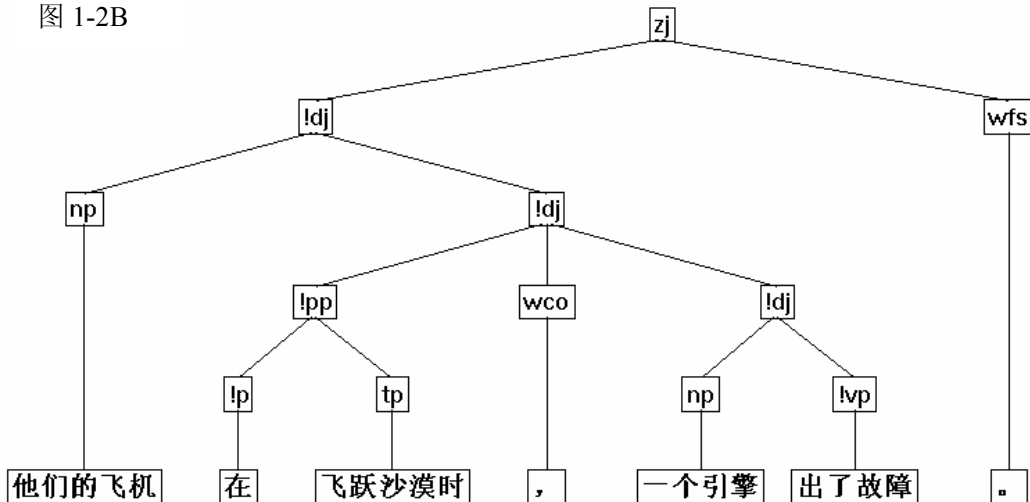
B np + [在 vp 时, xp]

图 1-2A



按图 1-2A 所示分析，tp “他们的飞机在飞跃沙漠时” 作整句的状语，np “一个引擎” 是主语。

图 1-2B



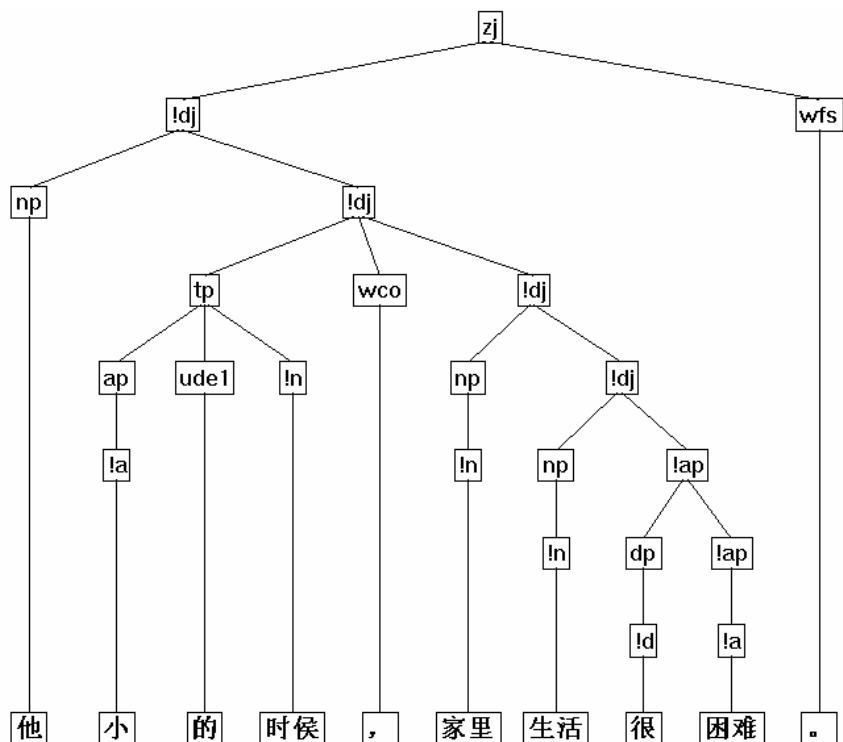
按照图 1-2B 的分析，np (“他们的飞机”) 是统摄全句的话题性成分（逗号后的小句主语是 “一个引擎”，是 “飞机” 的组成部分）。

按照图 1-2B 所示方式分析更合理一些。逗号是整句谓语中的一个成分。

如果碰到这样的例子 “他们的飞机在飞越沙漠时，我们正在海边等待救援”，这时候，“他们的飞机” 不再是全句话题（逗号后的小句主语是 “我们”，跟 “他们的飞机” 不同），就应该按照图 1-2A 所示方式进行分析。

类似地，像下面这样的例子也应做同样的处理。

“他小的时候，家里生活很困难。”



1.1.21 xp + 等 (等等)

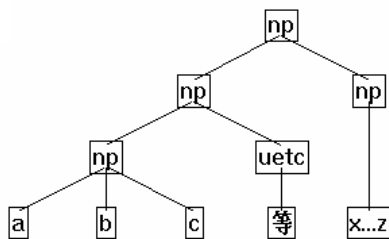
助词“等(等等)”有可能出现在两种格式中：

- ① a、b、c……等(等等) np
- ② a、b、c……等(等等)。

从逻辑上说，格式②是格式①的组成部分，即去掉格式①中最后一个np形成的。

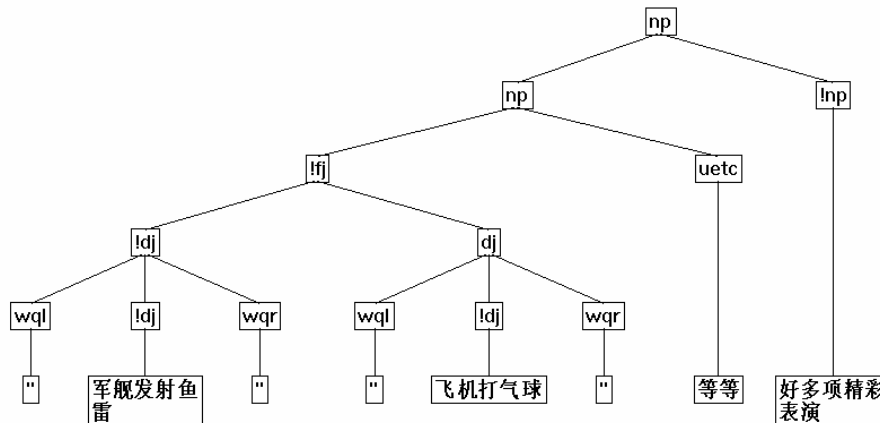
对于格式①，例如：“广西、云南等少数民族较为集中的地方”。需要先把“a、b、c……”捆绑起来，再与“等”组合，最后再与np组合，即按下面图1-3所示的方式进行层次分析。

图 1-3



格式①中“等(等等)”前面的成分也可能是句子(fj, dj)，比如：

我们还观看了“军舰发射鱼雷”“飞机打气球”等等好多项精彩表演
 这种情况下应将“等(等等)”前面多项成分与 uetc 一起组成 np。



格式②又分为两种情况：一种是词语（词组）的并列（即 a, b, c 等是词或词组），另一种是句子的并列（即 a, b, c 等是句子）。

第一种情况下，先把（a、b、c……）捆绑起来，再与“等”组合（跟上面格式①的处理方式相同）。例如：

中国的大城市有北京、上海、深圳、广州、天津、重庆等。

第二种情况下，如果句子过长，会涉及到断句问题。在书面语文本中，可能存在这样的情况：多个小句并列，各小句之间以分号隔开，在最后一个小句结尾处以‘等等’来结束。为便于计算机操作，碰到这样的情况，我们的处理策略是以分号为断句标志（参见下文第五节的说明）。这样，“等等”就只能和最后一个句子组合了。例如：

诸如：各民族公民不分民族、种族、宗教信仰，都同样地享有选举权和被选举权；
各民族公民的人身自由和人格尊严不受侵犯；……；各民族公民都有保持或改革自己
风俗习惯的自由等等。

经过断句处理后为：

诸如：各民族公民不分民族、种族、宗教信仰，都同样地享有选举权和被选举权；
各民族公民的人身自由和人格尊严不受侵犯；
……；
各民族公民都有保持或改革自己风俗习惯的自由等等。

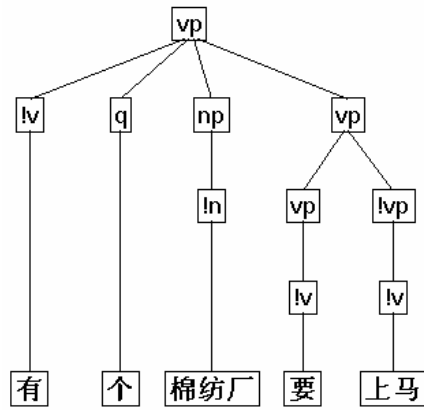
最后一句形成的模式是：dj → !dj uetc

如果句子不长，仍然按照上面格式①中的处理方式，先将几个小句捆绑在一起，然后再将“等等”附着上去。

1.1.22 有+个+np+vp

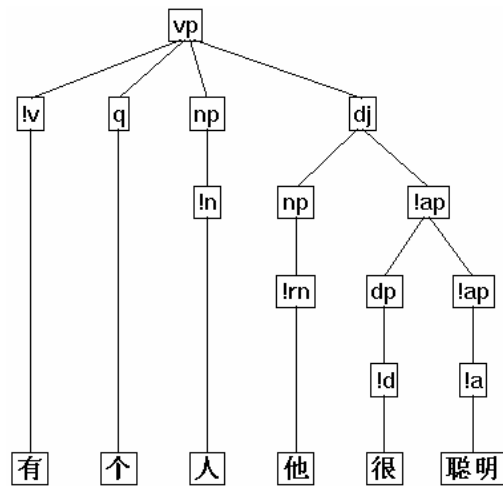
就“有 个 np”部分来说，是省略数词造成的三分结构，就“有 np vp”来说，是“有”形成的兼语结构，按规范要求，应该三分。对于整个“有+个+np+vp”构造来说，应按照四分结构处理：vp → !v(有) q np vp

比如：“有个棉纺厂要上马”

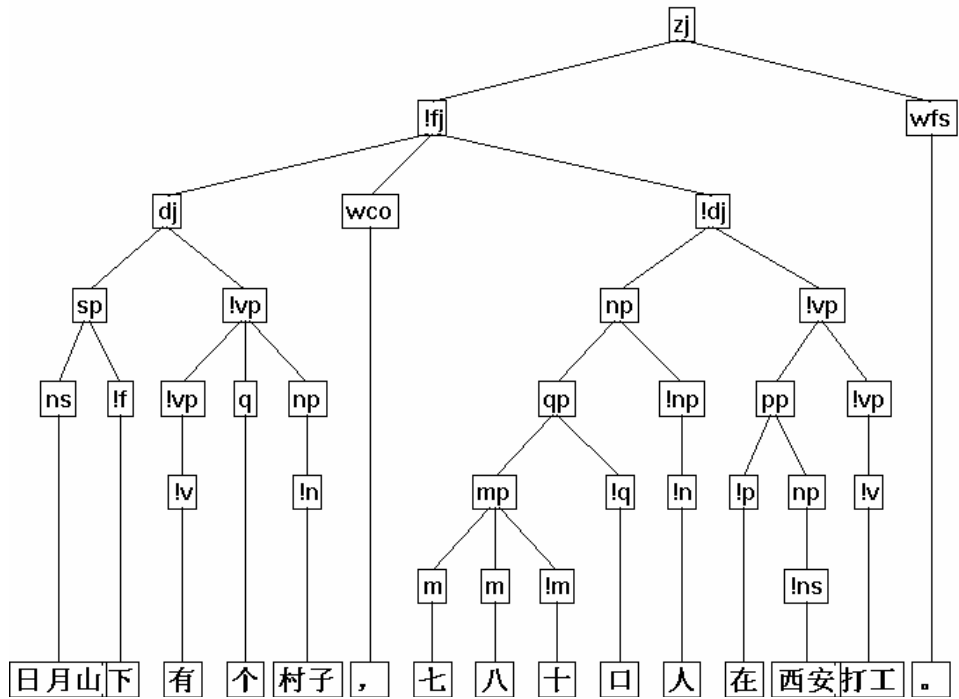


1.1.23 有 + 个 + np + r + vp

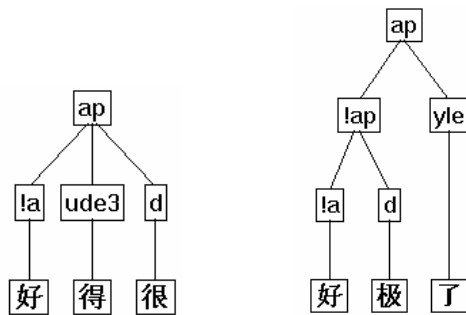
这个格式中 np 和 r 之间有复指关系。比如“有个人 他 很聪明”，也按四分方式处理：



上述格式中 r 的位置上也可以出现 np，通常跟前一个 np 有领属（整体-部分）的语义关系，比如“日月山下有个村子，七八十口人在西安打工。”这个例子中两个 np 之间有逗号分隔开，可按如下方式分析（如果没有逗号隔开，则分析方式同上——四分）：



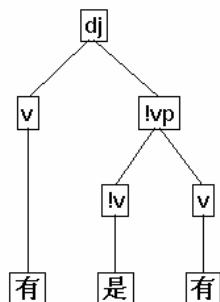
1.1.24 好得很、好极了



1.1.25 x + ude2 + vp

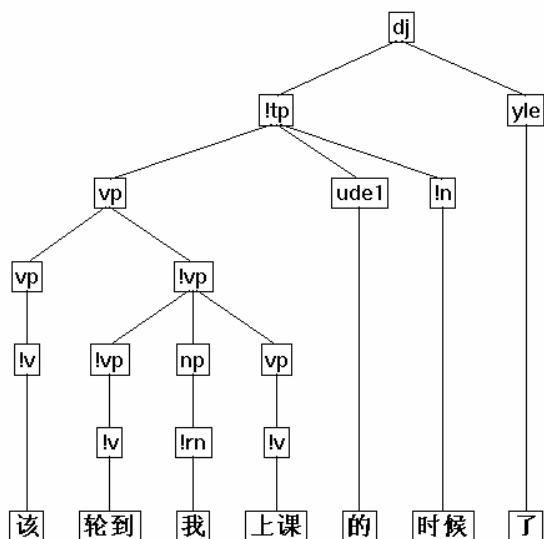
“x + ude2”先组成 dp，然后 dp 作状语修饰 vp。整个格式按照二分结构分析。
 比如“努力地学习” vp [dp [努力地] vp[学习]]

1.1.26 有是有



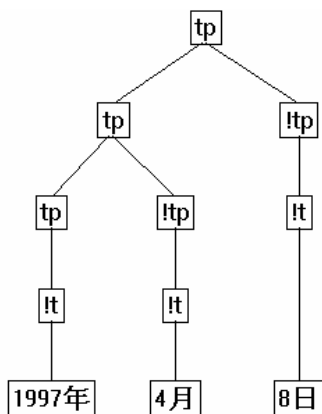
为突出这类格式的特殊性，“是”前后的动词都以 v 的身份直接参与组合，不上升为 vp。

1.1.27 该轮到我上课的时候了

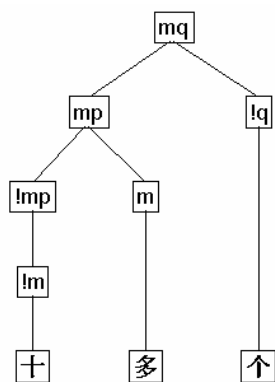


整句大致相当于“十二点了”，即组合规则模式为： $dj \rightarrow !tp \ yle$

1.1.28 1997年4月8日

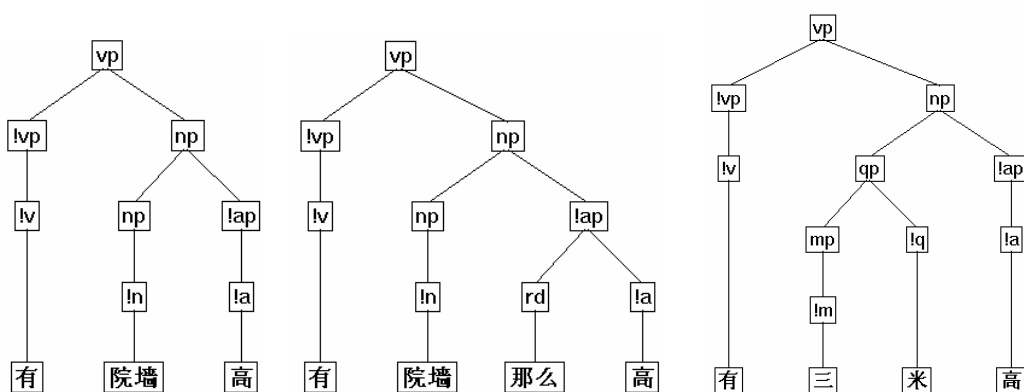


1.1.29 十多个

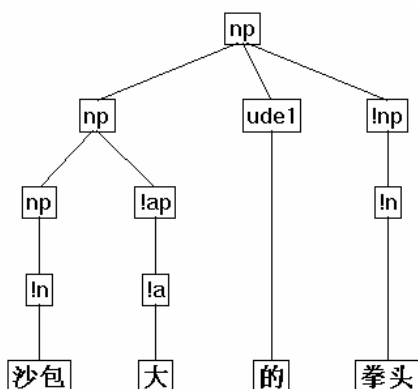


“多”分析为概数词，也标注为 m。

1.1.30 有院墙高、有院墙那么高、有三米高



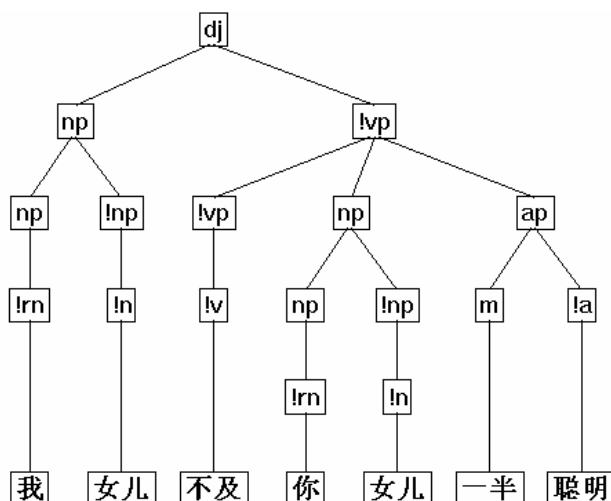
1.1.31 沙包大的拳头



“院墙高”“院墙那么高”“三米高”“沙包大”等分析为 np + ap 和 qp + ap 组合，整体功能都分析为 np。需要说明的是，这些例子中 ap 为中心成分，但整体功能不分析为 ap，而是分析为 np，一方面可以突出这类组合的特殊性，另一方面，从分布上讲，这些例子主

要是出现在宾语和定语位置，而不能出现在状语、补语、谓语位置，因此不宜分析为 ap。同时需要注意的是，这类“np + ap”组合的实例都是有歧义的，这类组合更为常见的解释应该是分析为 dj，比如在“沙包大，沙袋更大”这样的语境下，“沙包大”就应分析为 dj，而不是像在“沙包大的拳头”中那样分析为 np。

1.1.32 我女儿不及你女儿一半聪明



“不及”可以带单宾语，比如“张三不及李四”、“张三不及李四的三分之一”，也可以带双宾语，即把“不及”的具体内容表达出来，“张三不及李四聪明”“张三不及李四能吃苦耐劳”，对于带双宾语的情况，都分析为三支结构： $vp \rightarrow !vp \quad np \quad ap$ 远宾语 ap 也可以是 vp（如“能吃苦耐劳”）。

1.1.33 职务名称₁ 兼 职务名称₂

比如“国务委员兼外交部长”，分析为三支结构：

$np \rightarrow !np(\text{职务名称}_1) \quad c(\text{兼}) \quad np(\text{职务名称}_2)$

其中“兼”分析为连词 c。

“兼”还有动词用法，如“张三兼外交部长”“国务委员能不能兼外交部长”等用例中的“兼”都应分析为动词。

1.1.34 “X 包括 Y” “X，如，X1，X2，…”

分析为多分支结构。比如：

$np \rightarrow !np(\text{两岸旅游交往问题}) \quad wco(,) \quad v(\text{包括}) \quad np(\text{两岸同胞旅行中的安全问题})$

$np \rightarrow !np(\text{家中各类花卉}) \quad wco(,) \quad v(\text{如}) \quad wco(,) \quad np(\text{一串红})$

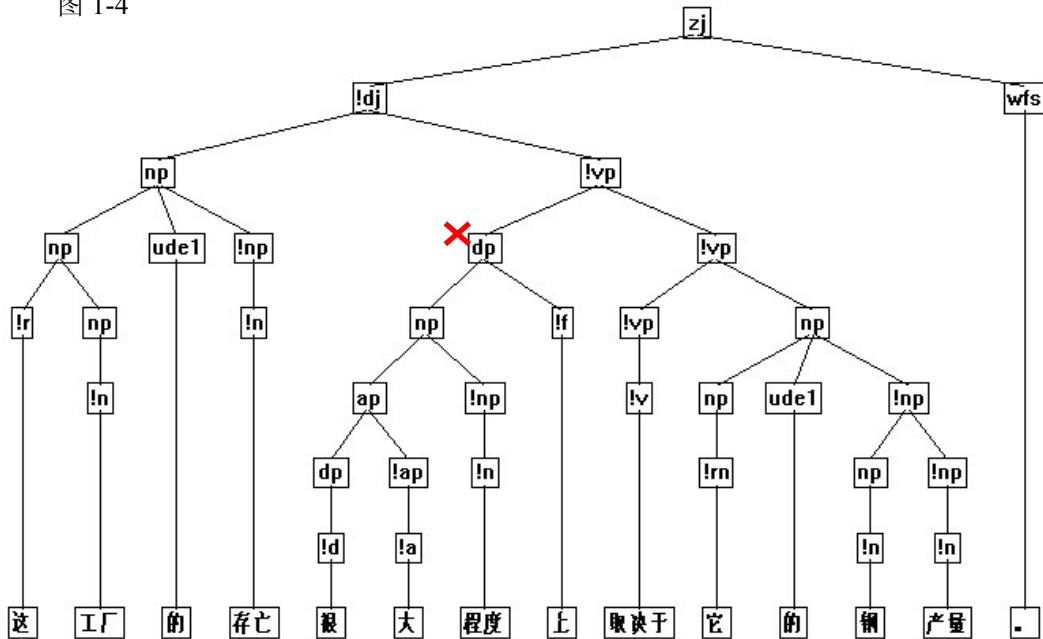
1.2 短语结构的功能定性问题举例

1.2.1 np+f

实例：很大程度上

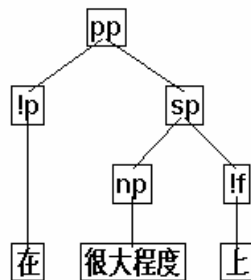
下图将“np f”组合成的短语分析为 dp，是不合适的。

图 1-4



处理意见：np+f 组合为 sp 更好一些，同类实例如“基本上”“理论上”。相应地，如果是“在 + 很大程度 + 上”，应该按照下图所示的方式处理。

图 1-5

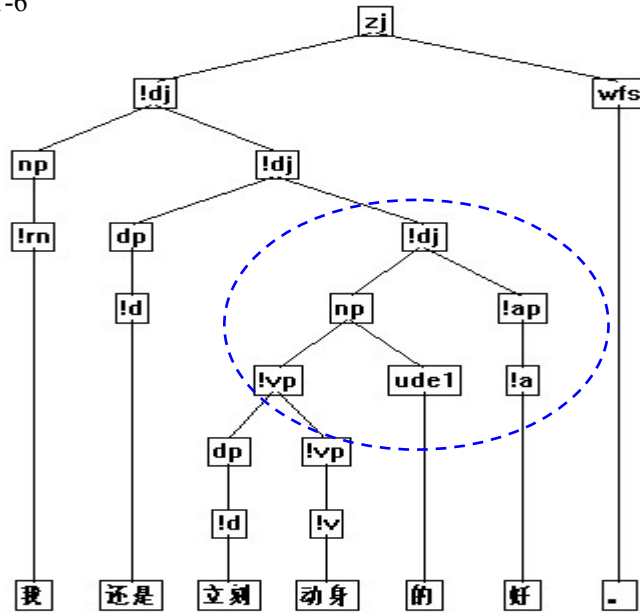


1.2.2 vp+ude1+a

实例：立刻动身的好

把“立刻动身的好”处理为 vp，不是太好。但到底该如何分析，颇费斟酌。我们的意见是，“立刻动身的”可看作 np，跟 ap“好”组合成一个 dj。图示如下：

图 1-6

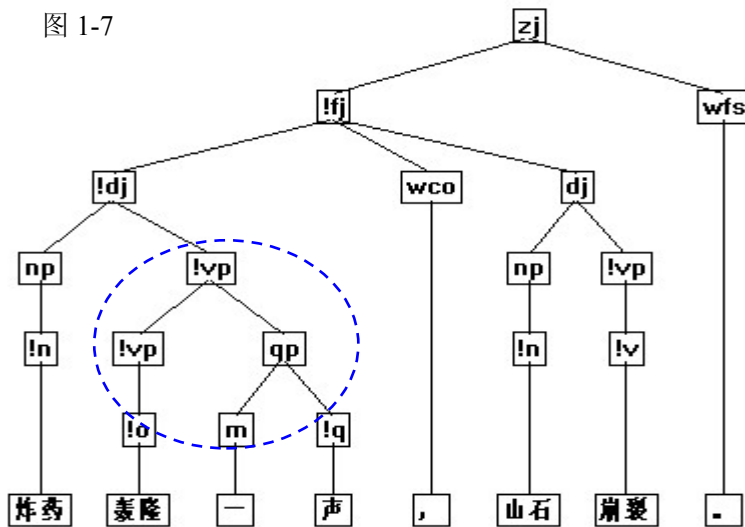


1.2.3 o+m+q

实例：轰隆 一声

“轰隆”是拟声词，后接动量成分“一声”，在这个组合中，“轰隆”的功能类似于动词，因此这个组合宜处理为带动量宾语的 vp。图示如下：

图 1-7



1.2.4 是 vp 还是 tp

实例：等一会儿 我去图书馆

就一般的情况而言，“等一会儿”是 vp，“我去图书馆”是 dj，二者组合在一起构成什么类型的短语结构呢？

vp 和 dj 组合，有两种比较常见的情况⁴：

一是单个动词形成的简单 vp 带 dj 宾语，组成一个述宾式的 vp。比如“同意新生参加演出”，“觉得小王在撒谎”，等等，在这两个例子中，“同意”和“新生参加演出”之间是述宾关系，“觉得”和“小王在撒谎”之间也是述宾关系，vp+dj 构成的是述宾式 vp。显然，这种情况中，要求 vp 一般是单个动词，而且是能带谓词性宾语（dj 是谓词性的）的那一类动词。这时 vp 和 dj 组成的模式为 $vp \rightarrow !vp \ dj$

二是复杂 vp 加 dj，形成 fj。比如“等我一会儿，我也去图书馆”、“看见红灯，你就停下。”、“关完窗户，我又关门。”等等例子。这种情况下形成的模式为 $fj \rightarrow vp \ !dj$

再回过头来看“等一会儿我去图书馆”这个例子。从形式上，这个例子比较接近上面第二种情况，但又不完全一样。仔细推敲不难体会到：“等一会儿我去图书馆”是有歧义的：如果前面的“等一会儿”中的“等”作比较实在的状态动词意思理解，整个句子就有可能分析为 fj。这时候，在口语中，“等一会儿”后面应该有明显停顿；书面上，“等一会儿”后面应该有逗号隔开。而如果这里的“等”已经没有比较实在的动词义，整个“等一会儿”就不再是典型的 vp，而是一个功能上相当于 tp 的成分，这时候，“等”可以拿掉而不影响整句的意思（相当于“一会儿我去图书馆”）；并且“等”后也不能插入时态成分（如“了”），这说明这里的“等”跟一般的动词用法已经不一样了。

因此，对这个例子，我们的处理意见是：“等一会儿”分析为 vp，作后面 dj（“我去图书馆”）的状语性成分，整个结构为 dj，即该例形成的组合模式为 $dj \rightarrow vp \ !dj$ （下文 1.3 小节对“比字句”的分析示例中也有这样的模式）。这样，就既不同于 $vp \rightarrow !vp \ dj$ ，也不同于 $fj \rightarrow vp \ !dj$ 。

跟“等一会儿”类似的，还有“过两天”。这类结构都处理为 $vp \rightarrow !v(\text{等}|\text{过}) \ qp(\text{两天})$ 。从句法功能上说，它们是 vp，但从语义范畴上讲，相当于 tp。

这个例子也可以看作是冗余造成的功能变异，读者可参见下面 1.4 小节关于冗余现象的示例。

1.2.5 三比二

这类表示比分的短语，标为 qp、np 还是其他标记？

这里规定：将这种表示比分的词组统一处理为 np，三支结构。

1.2.6 南偏西

这类表示方位的短语，属于什么功能类？

这里规定：“南偏西”中的“偏”作为中助词，整个词组处理为 sp（处所词性短语）。结构层次分析为三支结构。

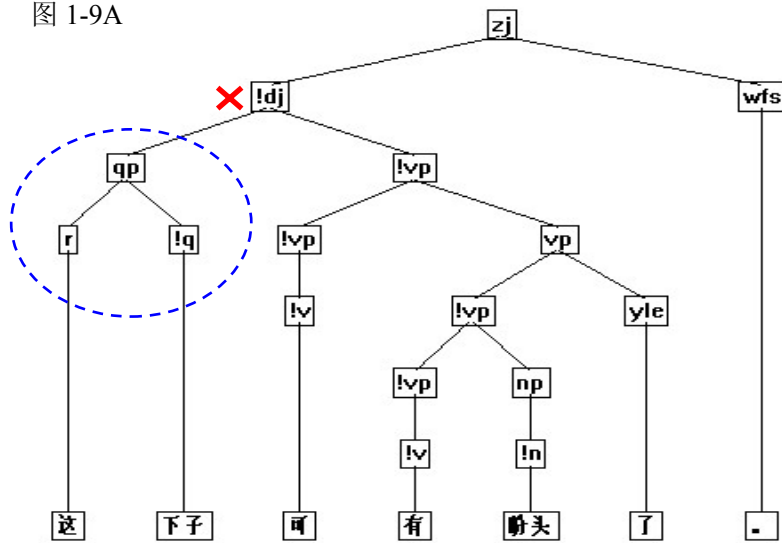
1.2.7 这下子

比如在“这下子可有盼头了”中，该如何分析？

我们的处理意见是：这 + 下子 组成 qp，即 $qp \rightarrow r \ !q$ 。如下图所示（不过，下图中，qp 再跟 vp 组合成 dj 的处理不妥，需注意）：

⁴ 除了这两种比较常见的模式外，也还有 $dj \rightarrow vp \ !dj$ 这样的主谓谓语句模式。参见下文 1.3 小节第一部分对“比字句”的分析。

图 1-9A

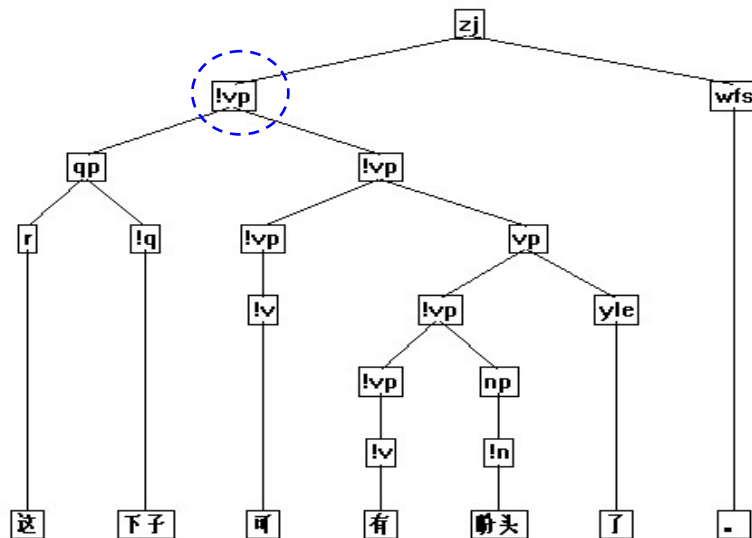


之所以处理为 qp，是考虑到其中“下”是动量词（“子”是子化韵，是附加成分，只有音节作用，没有意义，也可以有“这下儿”的说法），“这下子”可能是“这一下子”的省略形式（随着使用频度的增加，原始形式“这一下子”反而不常见了）。如果是“这”+“一下”，整个组合应该也是 qp，组合模式为：qp → r qp。

如果把“这下子”跟“这回”“这次”“这趟”等 r+q 组合类比，可以发现“这下子”跟这些 r+q 组合之间有很多共性。从动量的“量”上看，“这下”比“这次”“这回”的“量”少。基于这些考虑，我们把“这下子”处理为 qp。

需要注意的是，对上面这个例子来说，qp + vp 组成 dj，不太合适。应该改为 vp。qp 可以作为 vp 的状语出现在 vp 前面，二者组合，形成一个更大的状中式 vp。如下图所示：

图 1-9B



不过，如果在“这下子”和“可有盼头了”之间插入“他们”，情况又如何呢？这时就会出现 qp + dj 组合成什么结构的问题。我们的处理意见是整个结构分析为 dj，即 dj → qp !dj

1.2.8 量词重叠式

比如“条条道路”中的“条条”如何处理？

处理为 qp，即 $qp \rightarrow q!q$ 这样处理，可使得“条条道路”成为 qp + np 组合为 np 的模式。

需要注意的是，如果量词重叠形式前面出现数词“一”，应处理为三分结构。比如“一条条”，应分析为： $qp \rightarrow m q!q$

如果是“一 q 一 q”格式，则分析为 $qp \rightarrow qp!qp$ 。比如：

“这时候，人们才听见船夫一篙一篙下水的声音。”

其中“一篙一篙”分析为 qp，修饰 vp（“下水”）。

1.2.9 形容词重叠式

远远 $dp \rightarrow !a a$

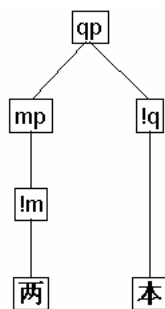
单音节形容词重叠后功能一般是作状语，因而功能标注为 dp。

远远的 $ap \rightarrow !a a ude1$

单音节形容词重叠后加“的”形成的结构一般可以作状语、补语和谓语，因而功能标注为 ap。

1.2.10 “数+量”结构

对于“数+量”结构来说，量词前面的数词应上升为数词短语，然后再跟量词组合。比如，“两本”应该分析为：



1.2.11 一分多钟、三米多长

“一分多钟”处理为：一分多 + 钟 $np \rightarrow qp !np$ (“钟”分析为 n)

“一分多”处理为：一分 + 多 $qp \rightarrow !qp m$ (“多”应分析为概数词)

“一分”处理为：一 + 分 $qp \rightarrow mp !q$

“三米多长”也标为 np，其中“长”标为 a，“三米多”标为 qp。

1.2.12 有一天、有一次

“有一天”是有歧义的。在不同的句法位置，“有一天”应作不同的标注：

(1) 在谓语位置，或者在 vp 前。比如：

i. 你们只有一天

ii. 这次三天的行程，我希望有一天安排参观厂房

在上述结构位置上，“一天”指的是一段时期，“有一天”的意义重心在于“一天”的时间长度。这种情况下，“有一天”分析为 $vp \rightarrow !vp \quad qp$ 即述宾结构的 vp。

iii. 突然有一天，他被人抬回来了

上例中的“有一天”不是指一段时期（“一天”的时间长度），而是指一个时间点（某一天），这里的“有一天”处于“突然”之后，也比较适合分析为 $vp \rightarrow !vp \quad qp$ 述宾结构，“突然”分析为“有一天”的状语成分。

(2) 在句首、或者说主谓结构之前的位置，或者宾语的位置、时间词之前的位置。比如：

iv. 有一天，他从外面回来了。（主谓结构前）

v. 直到有一天，人们再也无法容忍了。（“直到”后面是宾语位置）

vi. 有一天晚上，小王值班，…（时间词“晚上”的前面）

这些位置上的“有一天”都相当于“某一天”或“一天”。“有”在这个结构中的地位比较虚。这种情况下“有一天”分析为 $tp \rightarrow v !qp$ ，即以“一天”为中心成分，整个结构的功能类处理为 tp。

对“有一次”的分析和标注，可仿照“有一天”。在谓语位置、连谓结构的前项位置，标注为 vp；在主谓结构前的句首状语位置，标注为 qp。前一种情况“有”是结构的中心；后一种情况“一次”是结构的中心。比如：

i. 青春只有一次（“有一次”在谓语位置）

ii. 三跳中他有一次跳成功（“有一次”在连谓结构前项位置）

iii. 有一次，我真急了。（“有一次”在句首状语位置）

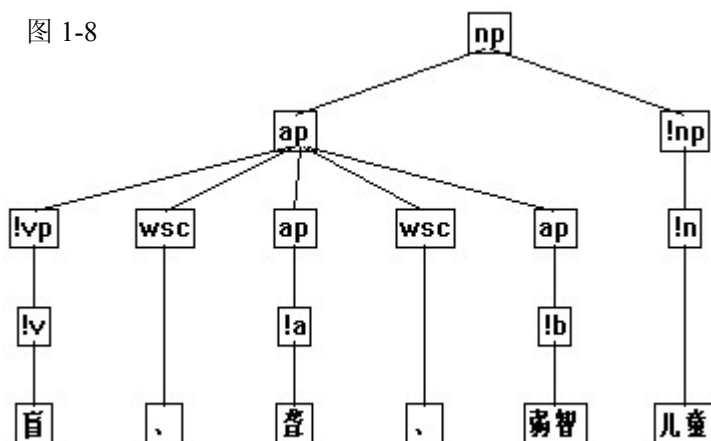
“有一天”“有一次”中的“有”均可换为“没有”，在“没有一天 vp”“没有一次 vp”结构中，“没有一天”“没有一次”均分析为 $vp \rightarrow !vp \quad qp$ 述宾结构，“没有”是结构的中心成分。

1.2.13 联合结构

联合结构是一个典型的离心结构，它的功能类可以跟中心语的功能类一致，也可以跟中心语的功能类不一致。例如“东”“西”“南”“北”是方位词 (f)，但“东西南北”并列起来却是一个 np。因此，我们在判断联合结构短语的功能类别的时候，不能简单地依据中心语的功能类别去判断，而主要是应该通过它的外部环境去判断，即看它在句子中占据什么句法位置，充当什么样的句法功能。

例如：为盲、聋、弱智儿童兴办了特殊教育学校。

图 1-8



“盲”是动词、“聋”是形容词，“弱智”是区别词，它们组成联合结构后整体作为“儿童”的定语，因此联合结构的功能是ap。

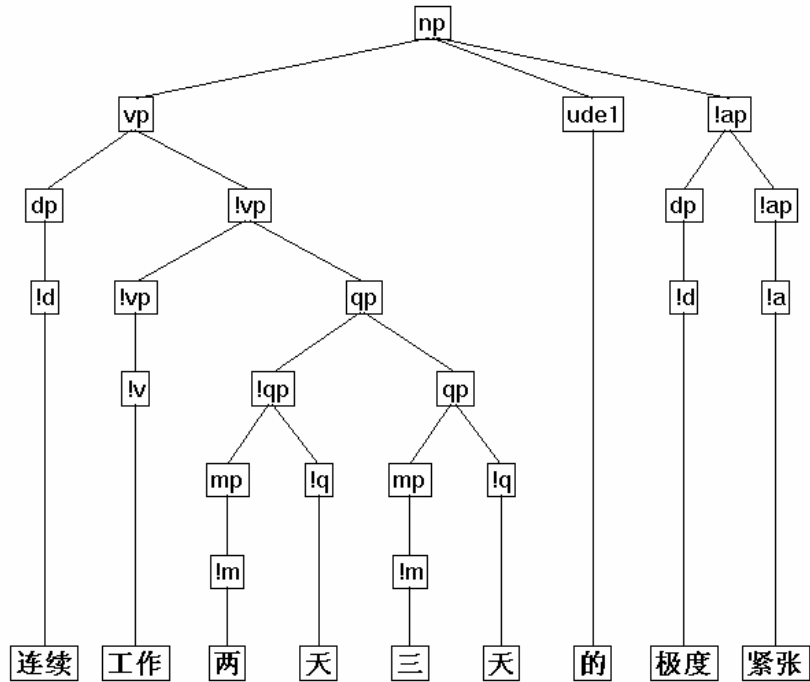
如果通过判断句法位置仍然无法得知结构的功能类别，那么可以根据结构的表述功能进行判断，即是用来指称、陈述还是修饰。如果用来指称，我们将其标注为 np；用来陈述，标注为 vp；用来修饰，标注为 ap。

例如“‘平等、和平、发展’是这次大会的主题。”这个句子。“平等”“和平”“发展”的词性分别为形容词、名词、动词，它们组成联合结构充当句子的主语。动词、名词、形容词都可以充当句子的主语，因此我们无法判断该联合结构的功能类别。但我们可以判断“平等、和平、发展”在句中的表述功能是指称而不是陈述或修饰，因此将整个结构标为 np。

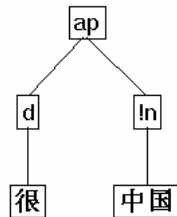
顺便提一下，树库加工规范中规定：一般情况下联合结构的第一个成分是整个联合结构的中心。碰到联合结构内部的词语词性不一致（如上面两例）的情况，也按照上述规定，将中心标在第一个成分上。

1.2.14 “极度紧张”是 ap 还是 np?

在“连续工作两天三天的极度紧张”这个例子中，“极度紧张”是 d+a 的组合模式，同时，“连续工作两天三天的极度紧张”整体应分析为 np，作为中心成分的“极度紧张”应该分析为 ap。



1.2.15 很中国



这个格式整体分析为 ap，其组成成分为 d+a，直接在词性层次上组合，而不在短语层次上组合。

在树库加工过程中，我们注意到，有不少层次分析和功能标注的困难都跟成分省略或冗余有关，因此下面把这两种情况再集中起来，分两个小节举例加以说明。

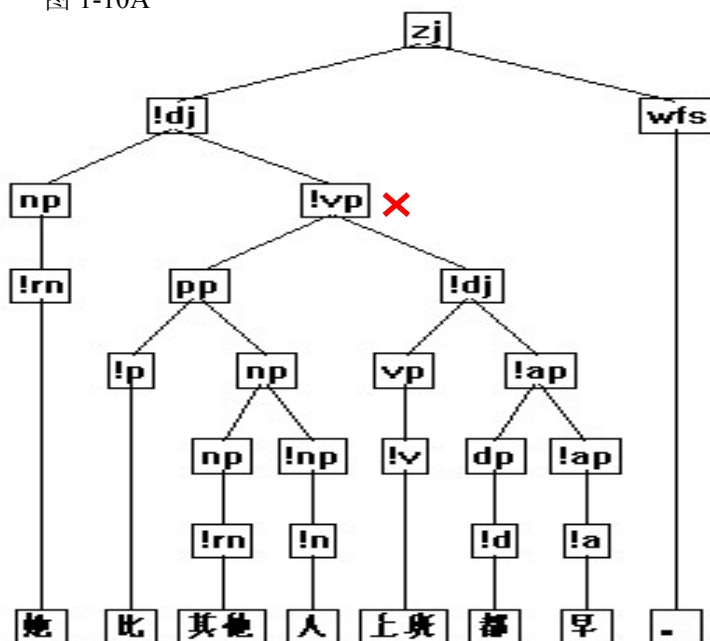
1.3 省略造成的短语功能变异和层次分析问题举例

1.3.1 比字句中的成分省略

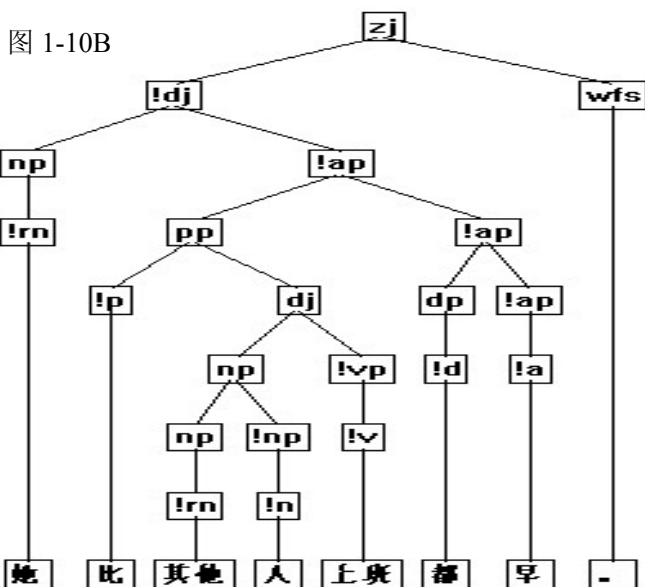
实例：（她） 比其他人 上班都早

如果按照下面图示的分析，就形成 $vp \rightarrow pp \quad dj$ 这样的组合模式。

图 1-10A



这个句子按照下面这种方式分析可能更好一些，形成的组合模式为： $ap \rightarrow pp \quad ap$



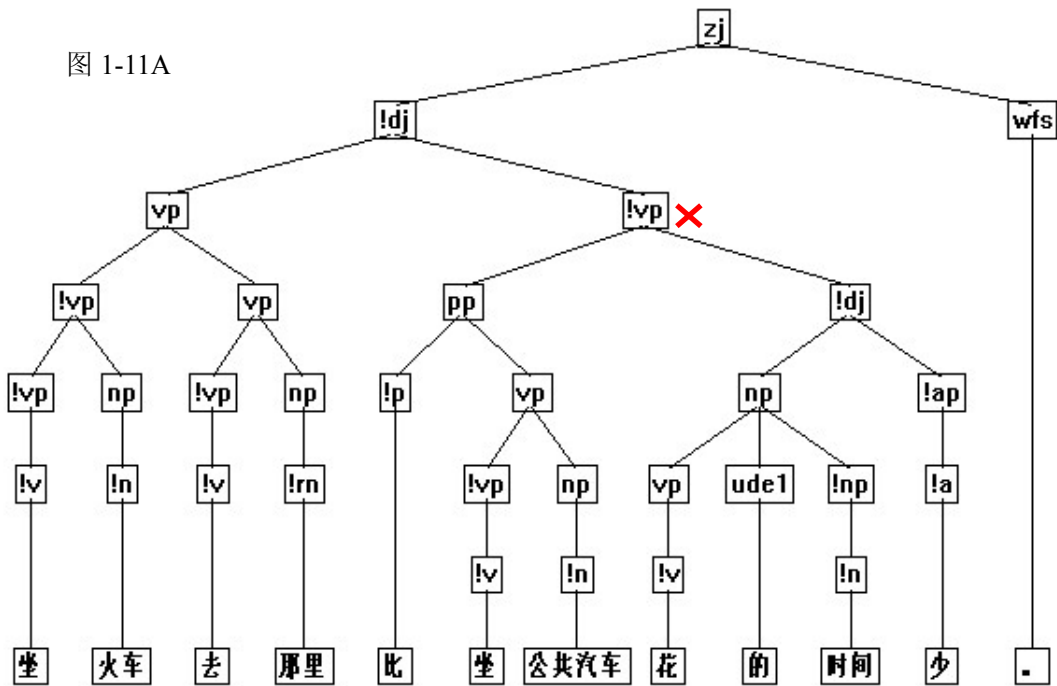
这个例子引起我们对“比”字句构造的关注。就汉语的“比”字句来说，“比 + w + xp”中，一般 xp 是形容词性短语（或一些类型的动词性短语）较多，dj 的情况较少。

按照上面 B 的分析方式，可以体现这种模式组合上的一般性。不过“比”这时候带 dj 作宾语，也有些“奇怪”⁵。一种可能的理解方式是：把这里的 dj “其他人上班”看作是 np “其他人上班的时间”的省略。这样形成的模式就是更常见的 $pp \rightarrow p \quad np$ 。当然，也可以把整句看作是“她上班 比其他人都上班 都早”的省略。

⁵ 其实“比”后带 dj 宾语的例子也是有的，比如“你去 比 我去 合适”，其中“我去”就是“比”后的一个 dj 宾语，只不过这时“比”前的主语成分也是一个 dj。如果主语是一个简单 np，同时“比”后宾语是 dj，我们的感觉是这种组合情形显得有点“怪”。

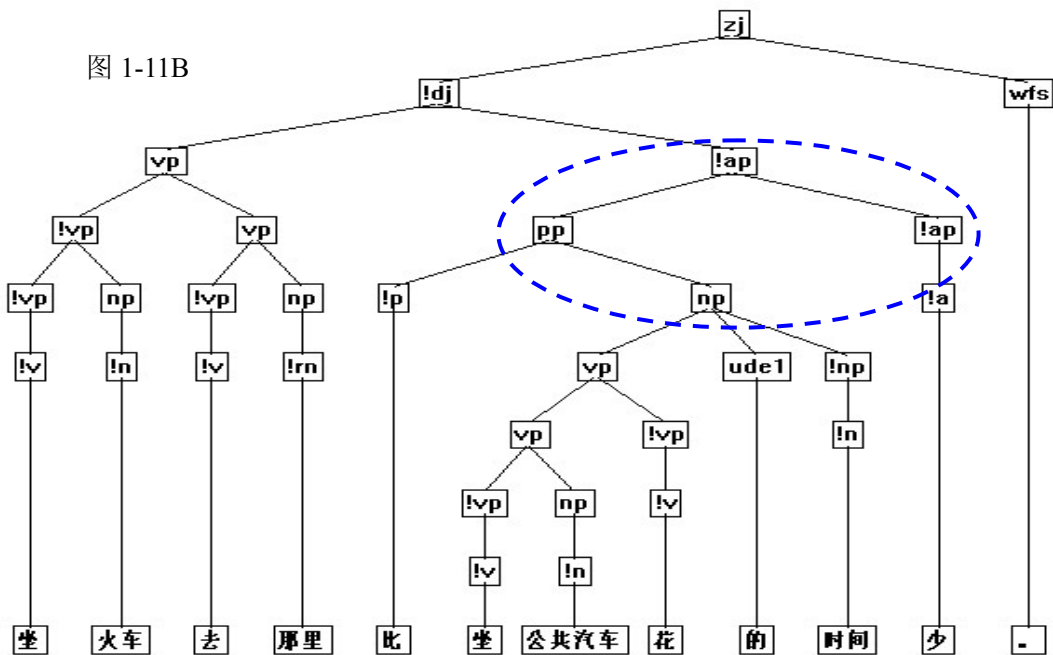
类似的例子：坐火车去那里 比 坐公共汽车 花的时间 少

图 1-11A



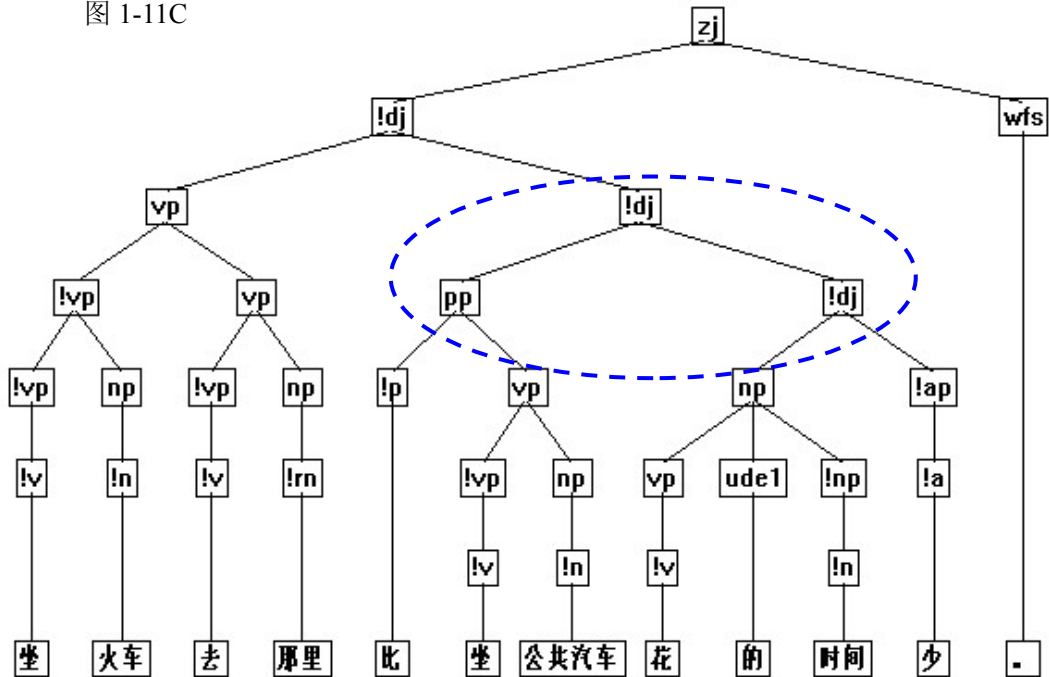
上面的分析方式也造成了 $vp \rightarrow pp\ dj$ 的组合模式，要避免这种模式，可以考虑调整为：

图 1-11B



还可以考虑不改变层次分析方式，而是改变短语功能类的定性，调整为下图所示模式：

图 1-11C



这样形成的模式就是 $dj \rightarrow pp\ dj$ （取代了原来的 $vp \rightarrow pp\ dj$ ）。这种情况下，上例的“比”字句可以看作是“比”字结构（pp）插入到一个主谓谓语句（vp + dj）中形成的，即由“坐火车去那里 + 花的时间少”变成“坐火车去那里 + 比坐公共汽车 + 花的时间少”。

比较上面三个图所示的处理方式，我们认为图 1-11C 是其中相对最好的一种处理方式。按照这种层次分析，从逻辑语义上说，参与比较的是双方是“坐火车去那里”和“坐公共汽车去那里”，比较的内容（项目）是“花的时间”。其中“去那里”是两个比较项中相同的部分，可以省略其中一个。推而广之，“坐”也可以省略，比如“火车比公共汽车花的时间少”就完全是可以接受的说法。

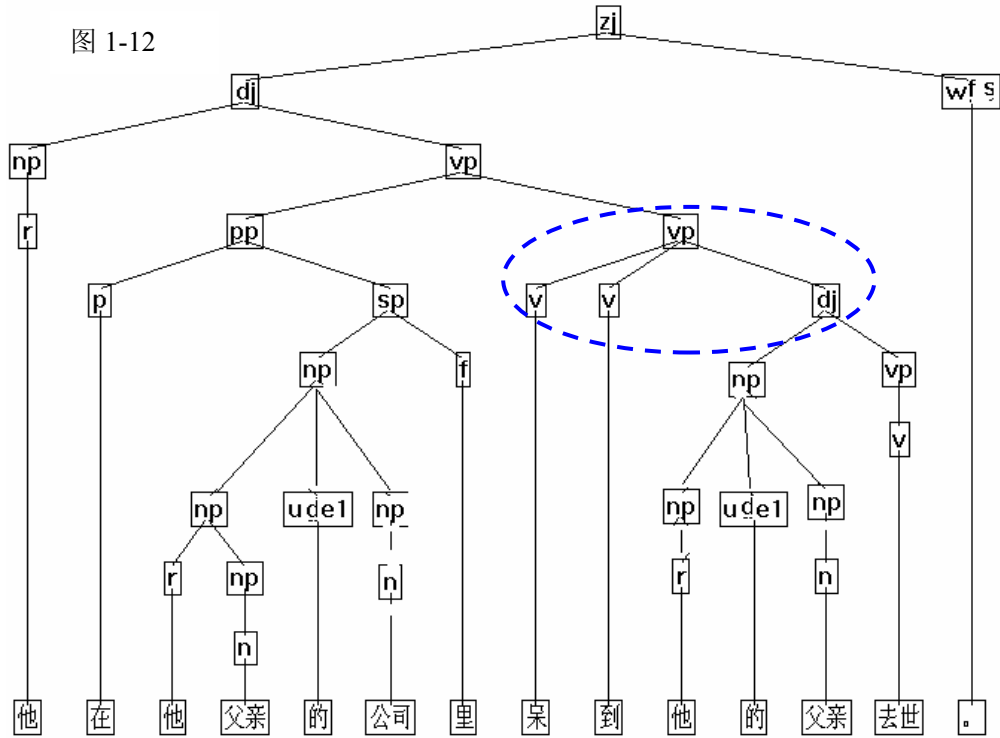
按照图 1-11C 所示的方式进行分析，此例中整个比字句形成的组合模式就是 $dj \rightarrow vp\ !dj$ ，即 vp 作主语，dj 作谓语形成的主谓谓语句。

1.3.2 定中结构中的中心语省略

实例：（他在他父亲的公司里）呆 到 他的父亲去世

孤立地来看，“他的父亲去世”是个主谓结构的 dj，但在上面这个例子中，显然是转指“他的父亲去世的时候”，实际功能相当于是 tp。从省略成分的角度说，“他的父亲去世”是定中结构“他的父亲去世的时候”省略中心语“时候”形成的表达形式。由于省略，导致规则中出现了 $vp \rightarrow v\ v\ dj$ 这样的模式。

图 1-12



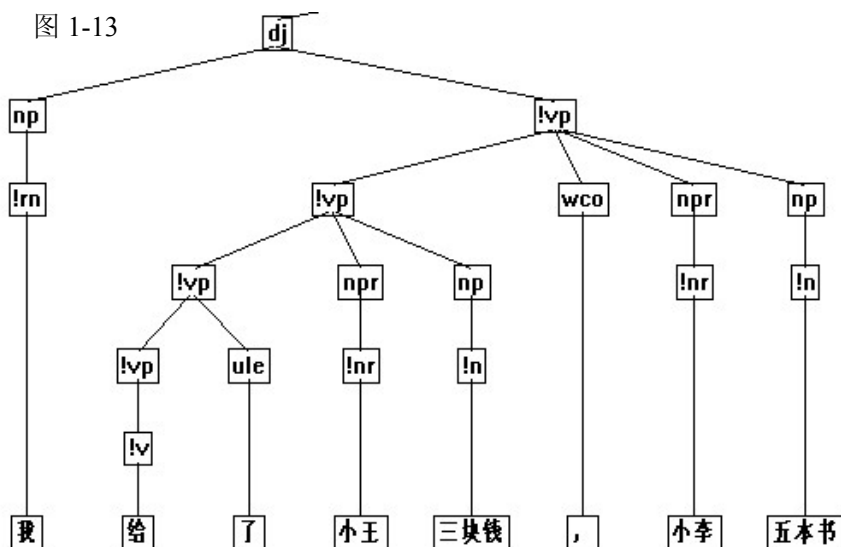
但如果要将上面例子中的“他的父亲去世”处理为 $tp \rightarrow np \quad vp$ 这样的模式，也不合适。目前仍然按照上面图中所示的方式进行分析。

1.3.3 联合结构中的成分省略

实例一：我 给了小王三块钱，小李五本书

“小李五本书”前面省略了“给了”，造成“小李五本书”的层次分析和功能定性发生困难。对这种情况，我们处理为双宾结构再带双宾结构。图示如下：

图 1-13



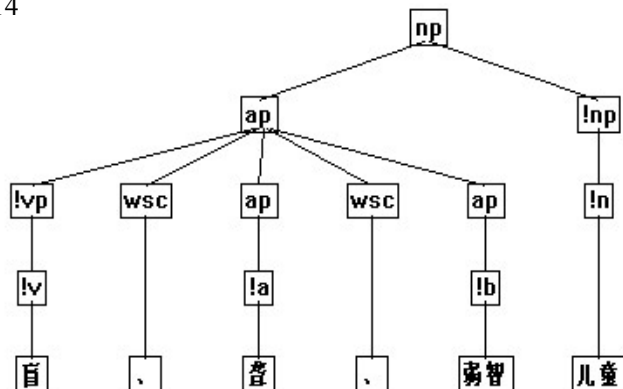
以上处理方式的指导思想是：尽量保证同类型的格式在抽取规则时形成同一个模式。比如上面双宾语结构，根据树图抽取的规则就是 $vp \rightarrow vp \quad npr \quad np$ ，在该树图上这条规则

应用了两次。

实例二：盲、聋、弱智儿童

显然，修饰语在语义上是同类成分，这时应该将修饰语先捆绑在一起，再和中心语成分组合；

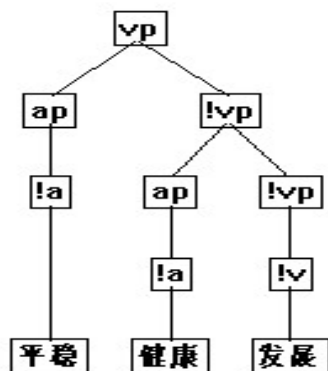
图 1-14



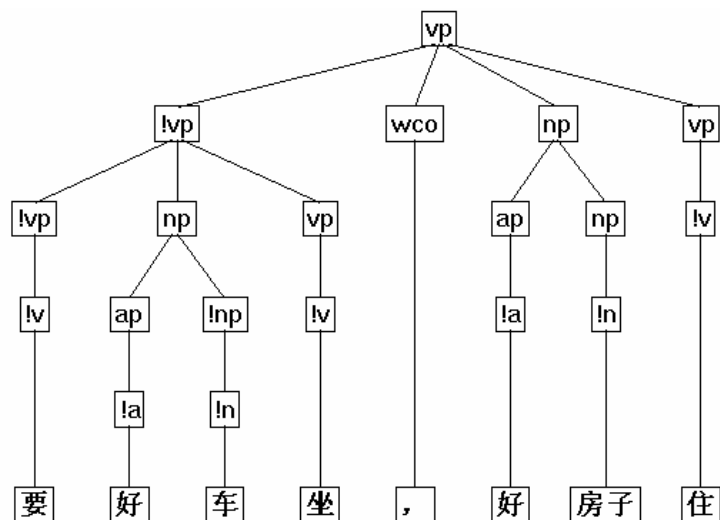
实例三：平稳健康发展

修饰语在语义上不是同类成分，这时可以采用层层修饰的分析方式：

图 1-15



实例四：“会不会要好车坐，好房子住？”

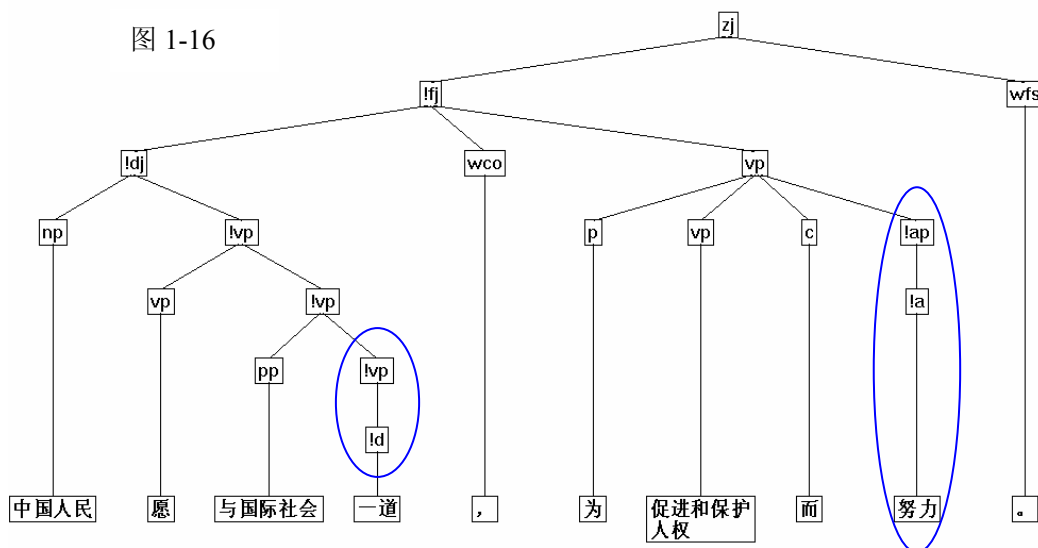


“要好车坐”分析为兼语结构，后一个兼语结构“要好房子住”中的“要”省略。

1.3.4 复句中的成分省略

在树库加工中，有时会遇到这样的情况：一个复句的前一分句中省略了成分，后一分句中可以找到这个省略的成分。这时，我们应采取突出“省略”的分析策略。

例如：中国人民愿与国际社会一道，为促进和保护人权而努力。



这个例子中的“一道”作为副词，如果“一道”后面没有逗号，则可以分析为“一道”跟后一个小句的 vp 发生修饰关系，形成状中结构：

$vp \rightarrow dp(\text{一道}) + !vp(\text{为促进和保护人权而努力})$

另外，也可以理解为“一道”后面省略了其他的 vp 成分，整句的逻辑语义相当于：

(中国人民愿与国际社会) 一道 (做某事)，为促进和保护人权而努力

但在上面的例子中，由于被逗号隔开，“一道”归到前一个小句，因此，在上图所示的分析方式中，我们让“一道”这个副词直接上升为 vp，产生“ $vp \rightarrow !d$ ”这样的奇怪规则，从而突出了其中的省略现象⁶。

另外，值得注意的是，“努力”也可以理解为“努力(做某事)”的省略。但上面的结构分析中没有这样处理，而是认为“努力”以 ap 的身份出现在“为 x 而 y”格式中 y 的位置。

1.3.5 跟介词结构有关的省略

介词结构中有不少省略问题，下面是两个例子。

(1) 介词的宾语省略

“钱被用来买房了”

这个例子中有两处省略，一是“被”后的宾语省略；二是“用”后的宾语没有放在它本该出现的位置，而是移到句首作主语了。整句的语义是“某人用钱来买房了”，经过移位处理后，形成的句子为“钱被用来买房了”。

对于“被”的宾语省略，结构分析为 $vp \rightarrow pBei !vp$ [被 + 用来买房]。也就是

⁶ 当然，“一道”也可以处理为副词和动词的兼类，在状语位置为副词，在其他谓词性位置为动词。

说，一般情况下，介词结构作状语的组合模式为 $vp \rightarrow pp !vp$ ，而对于“被”字结构，组合模式做特别处理。

对于“用”的宾语移位（省略），我们的处理方式是将“用来”合为一个词，标注为动词 v 。

(2) 介词结构中的宾语部分省略

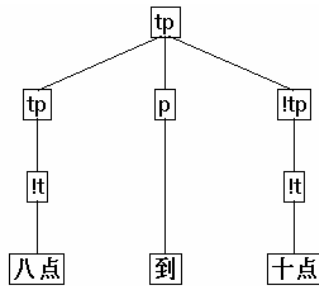
“房租包括水电在内”

这个例子中，“在内”应该理解为“在 … 内”的省略形式。“在”的宾语“…内”部分省略，形成“在内”这一形式。我们的处理方式是将“在内”合为一个词，标注为动词 v 。

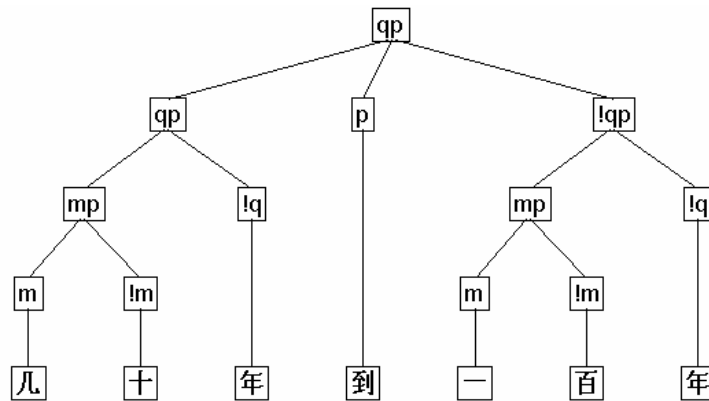
(3) 介词的省略

“八点到十点”

这个例子应该分析为“从八点到十点”省略了介词“从”形成的结构。



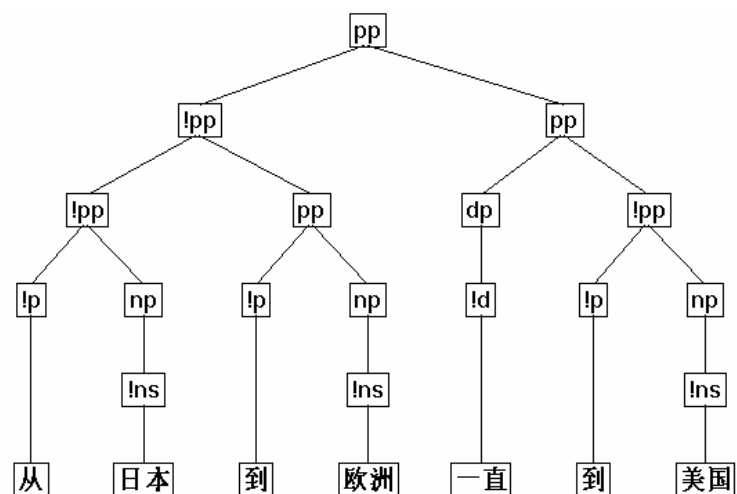
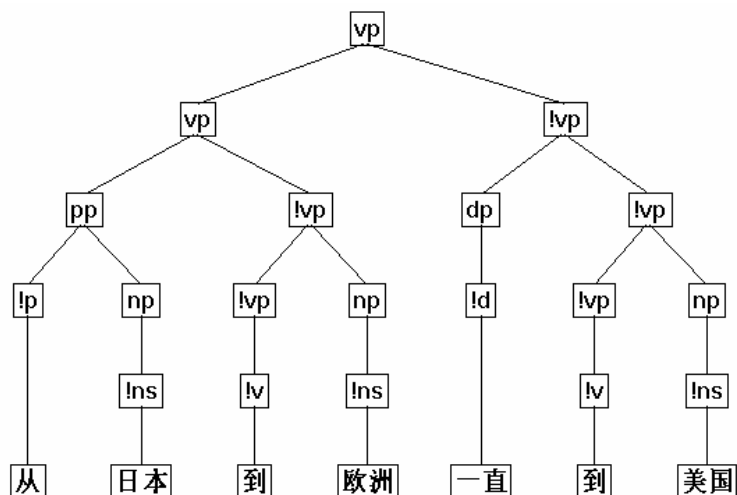
再比如“几十年到一百年”



(4) 整个介词结构的省略 — 联合结构

从日本到欧洲一直到美国

对这个例子，应根据语境做不同的处理，如果在中心谓语位置，分析为 vp ，如果在连谓结构前项或状语位置，分析为 pp 。



1.3.6 跟述补结构有关的省略

(一) 表状态的述补式结构中，补语标志助词“得”有时候可以省略。例如“看不很清楚”应该理解为“看得不很清楚”的省略形式。因此，这类结构的组合模式应分析为：

$vp \rightarrow !v \text{ ap}$ (看 + 不很清楚)

作述语的动词不上升为 vp ，而是以 v 的身份直接参与组合。

(二) 补语一般由谓词性成分充任，但有时候补语中谓词性成分省略，造成体词性成分充任补语的格式。比如：

出去一看，只见豆子撒得满地

其中“撒得满地”是述补结构，但补语是“满地都是”的省略形式。这时就形成了名词性成分“满地”作补语的组合模式：

$vp \rightarrow !v \text{ ude3 np}$

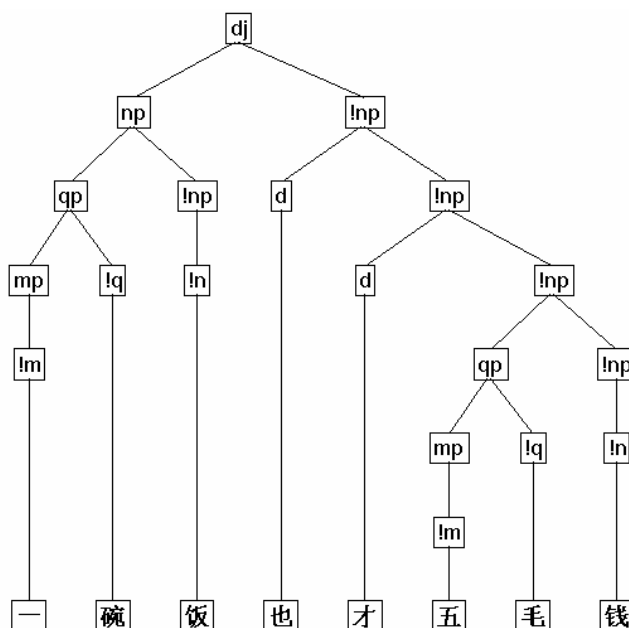
1.3.7 述宾结构中述语的省略

述宾结构中的述语省略会造成下面这样的句子形式：

i. 一碗饭才两块钱 (原式：一碗饭才卖/花两块钱)

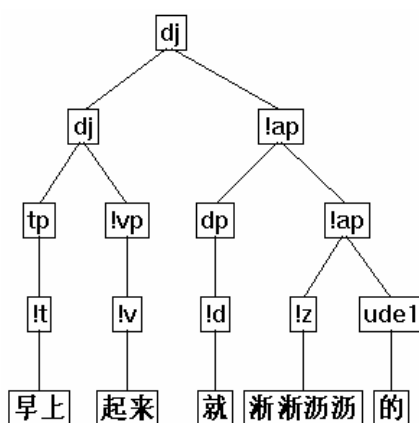
- ii. 一碗饭也才两块钱
- iii. 一碗饭不也才两块钱

上面例 i 的分析方式已经在上文 1.1.3 中作了说明。例 ii, iii 跟例 i 相比, 在于主、宾之间的副词性成分增多了。下面给出例 ii 的分析方式 (例 iii 的分析方式与例 ii 类似):



1.3.8 主语省略

汉语中“主”“谓”之间的句法约束相当薄弱, 因而也造成大量主语省略的组合格式。说汉语的人习以为常, 可能在语感上并不认为 (或察觉到) 这里面有省略存在。但是, 从句法结构分析的角度来看, 现有的结构分析框架很难对这些存在主语省略的表达形式进行句法结构标注。比如“早上起来就淅淅沥沥的”, 分析起来就比较困难。下面给出一种分析方式:



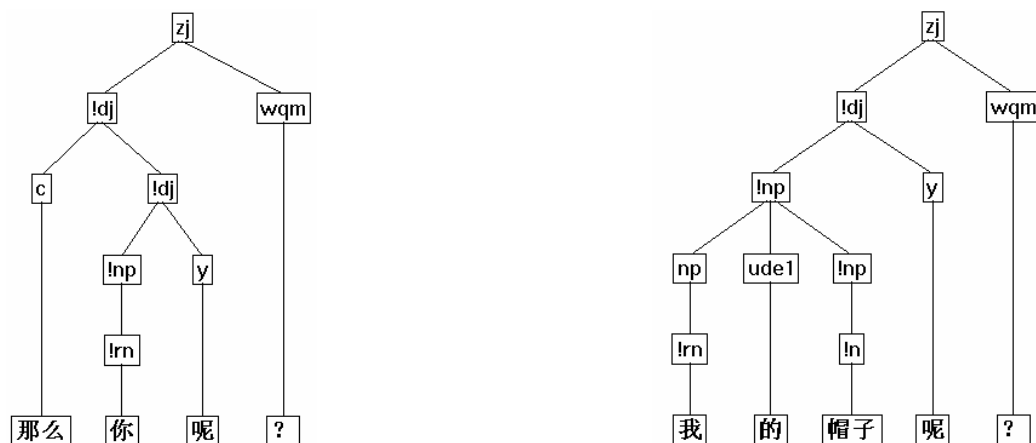
“早上起来”在这里转指“早上起来的时候”, 但句法结构分析的时候很难体现出这一点, 上面的处理中, 只是简单地按照一般分析“tp+vp”结构的处理方式将“早上起来”分析为 dj。这样, 就造成了一个很奇怪的 dj → dj !ap 组合模式。像这样的组合模式特例, 目前暂时这样标注, 留待以后再做专题研究。

类似的例子比较典型的还有跟“是”有关的主语部分省略结构。比如“打架不是好

孩子”、“我是羊肉”，等等。前者应理解为“打架的孩子 不是 好孩子”，后者应该理解为“我点的菜 是 羊肉”⁷。但是，在汉语的实际话语交际中，像这样的省略情况比比皆是。

1.3.9 谓语省略

谓语省略中典型的例子是汉语中的一类疑问句。比如“那么你呢？”“我的帽子呢？”。对这类结构，分析方式如下：

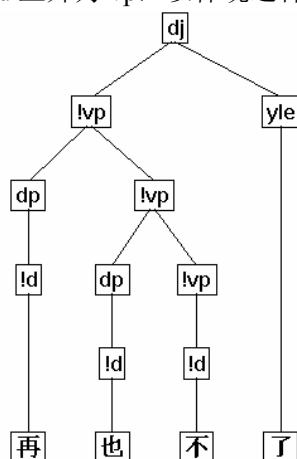


1.3.10 跟“不”有关的省略

“不”是否定副词，一般用在状中结构的状语位置。有时候状中结构的中心语省略，使得整个状中结构就剩下一个“不”，比如：

- i. 再也不了
- ii. 不！我要……

例 i “不”后可能省略一个动词，比如“再也不迟到了”“再也不说谎了”。这种情况下，可以让这个“不”从副词 **d** 上升为 **vp**，以体现这种结构的特殊性。分析为：



⁷ 这里只是提供一种可能的还原方式，并不是说“原式”就一定如此。省略式还原为原式涉及到很多理论问题。这里不去深究。无论是否承认这里存在省略，以及假设何种原式。这些例子在句法结构分析时跟一般‘常态’的结构形式不一样，存在分析上的困难，是不争的事实。

按这种方式分析，就会得到规则模式： $vp \rightarrow !d$ 。

例 ii 中的“不”比较复杂，既可能跟例 i 的“不”一样，也可能是表达说话人的“态度”，是命题否定，而不是单纯的“vp 否定”，这个“不”的独立性比较强，实际上从语法角度讲虽然是来源于 d（副词），但在这里主要是篇章功能，而不是在语法结构中起作用了。

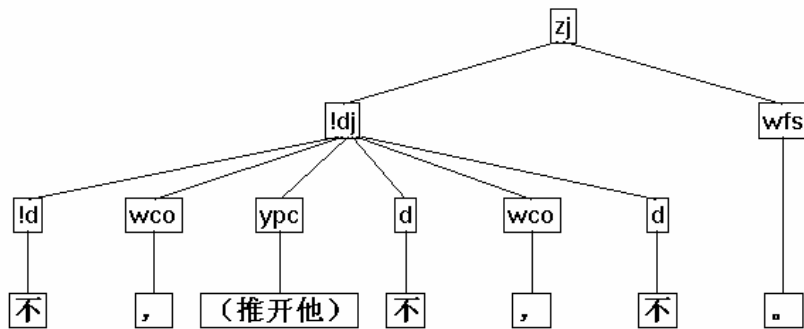
“不！ 我要……” 相当于 “我不同意你说的！ 我要……”

可以把“不！”分析为： $dj \rightarrow !d \text{ wem}$ 。由“不”和叹号一起组合为一个特殊的 dj。

在口语中有时还会有几个“不”连用的情况，比如：

例 iii. 不，（推开他）不，不。

这是剧本语料中一个角色的一句台词，其中“（推开他）”是说台词的同时做的动作。像这种多个“不”连用的情形，按如下方式处理：

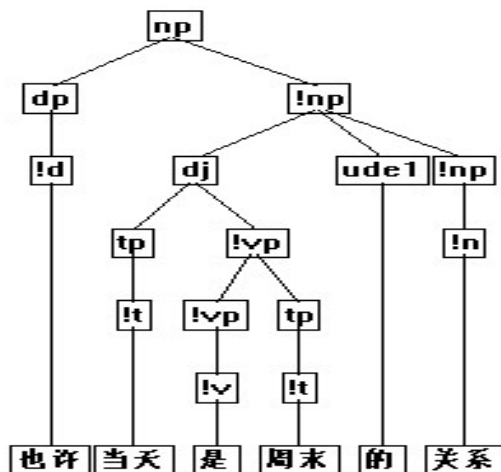


1.4 冗余造成的短语功能变异和层次分析问题举例

1.4.1 dp+np 组合

实例：也许当天是周末的关系，半圆形的观众席上，早已坐满了慕名而来的游客。

图 1-17



“也许”作为副词，后面一般跟谓词性成分（vp,dj 等）。但在上面的例子中后跟以“关系”为中心词的 np，形成的组合模式 $np \rightarrow dp \ !np$ ，这个组合显得很“怪异”。从语感上说，例中的“关系”似乎是冗余的表达形式，这句话如果说成“也许（因为）当天是周末，……”，

从语感上说，会更“规范”一些。

对于上例，也可以有另一种解释方式：这句话也可以说成“也许因为当天是周末的关系，……”，这时候组合模式就变成了“d（也许）+ c（因为）+ np（当天是周末的关系）”，这种组合模式也显得相对“规范”一些（类似的例子：“也许因为你，……”）。

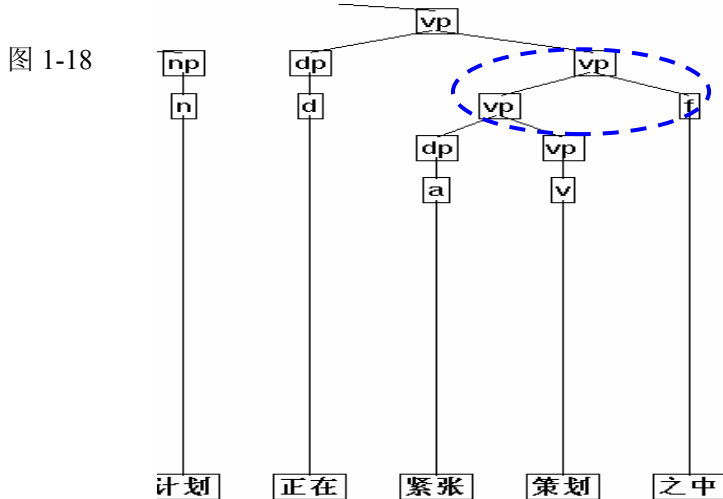
按照后一种解释方式，上面这个例子也可以说是属于“省略造成了功能变异”。

对上面这个例子，目前的处理方式是分析为 np → dp !np 维持这种“怪异”的现状，以便今后从树库中提取类似的“怪异”组合模式，再对此做专门的研讨。

1.4.2 vp+f 组合

实例：……计划正在紧张策划之中

在这个例子中，“之中”是冗余成分，直接说“正在紧张策划”就可以表达相同的意思了。“vp（紧张策划）+ f（之中）”形成的短语，一般作为 tp 或 sp 的较多，比如“工作中”“学习中”“选举中”等等，这里的“vp+f”语义中心不在“中”，而在“vp”的动作过程。一般来说，这个格式中的 vp 都是过程性的动词成分，而不能是非过程性的动词成分（状态动词），比如一般不说“*（正在）知道中”“*（正在）懂得中”。

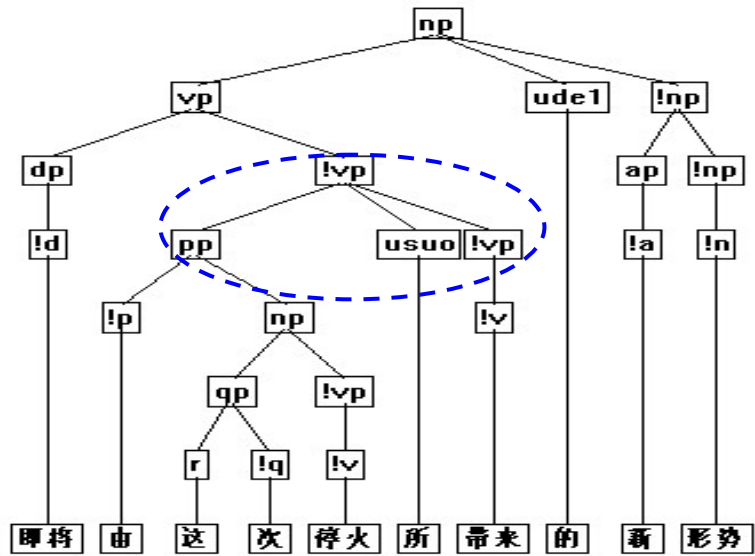


对于上面的例子，目前的处理方式是分析为 vp → !vp f，保持这种不常见的组合模式，目的也是留待今后再进一步考察。

1.4.3 pp+usuo+vp 组合

实例：即将由这次停火所带来的新形势

图 1-19



从表达上看，上例中的“所”省去也不影响意思和表达，这个“所”可以看作是个冗余成分，应该跟前面的“由……”（pp）结构一道，整体分析为vp。

这里一方面涉及到“由这次停火所带来”的功能定性问题，同时也涉及这个格式内部的层次分析问题，是“由这次停火 + 所带来”二分呢？还是“由这次停火 + 所 + 带来”三分？我们的处理意见是三分。如果二分，“所带来”很难成立，而且一般独立的“所+vp”形成np，为了和独立的“所+vp”形成的np区别开，我们对“xp + 所 + vp”形成的构造做三分处理。

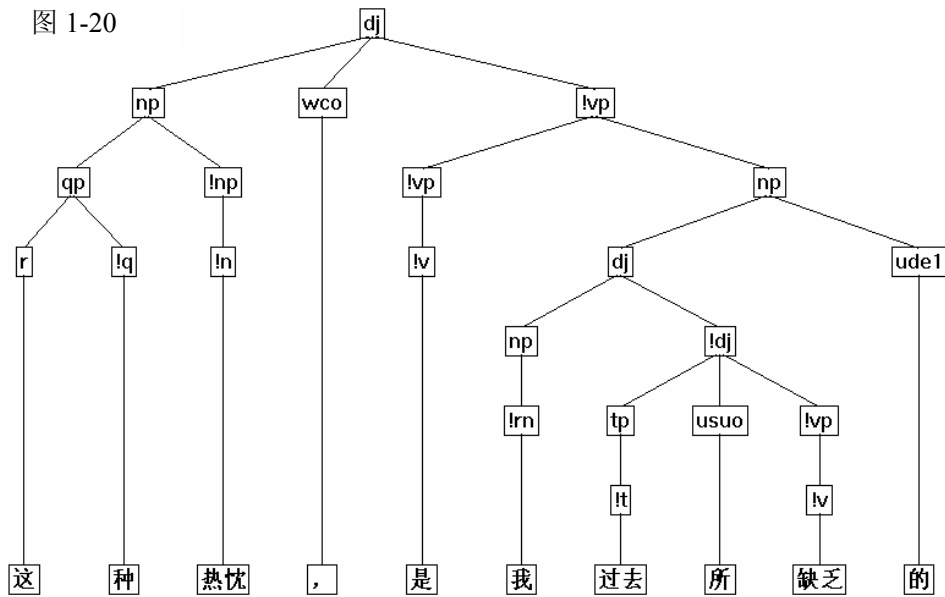
因此，上面的例子如图所示进行分析，得到的组合模式为：

vp → pp usuo !vp

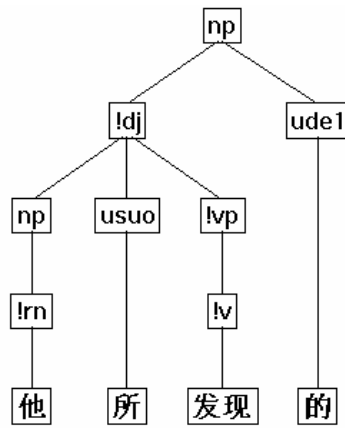
下面这个例子也含有“所”字结构，处理方式同上。“所”字所在的结构作三分处理，得到的组合模式为：

dj → tp usuo !vp

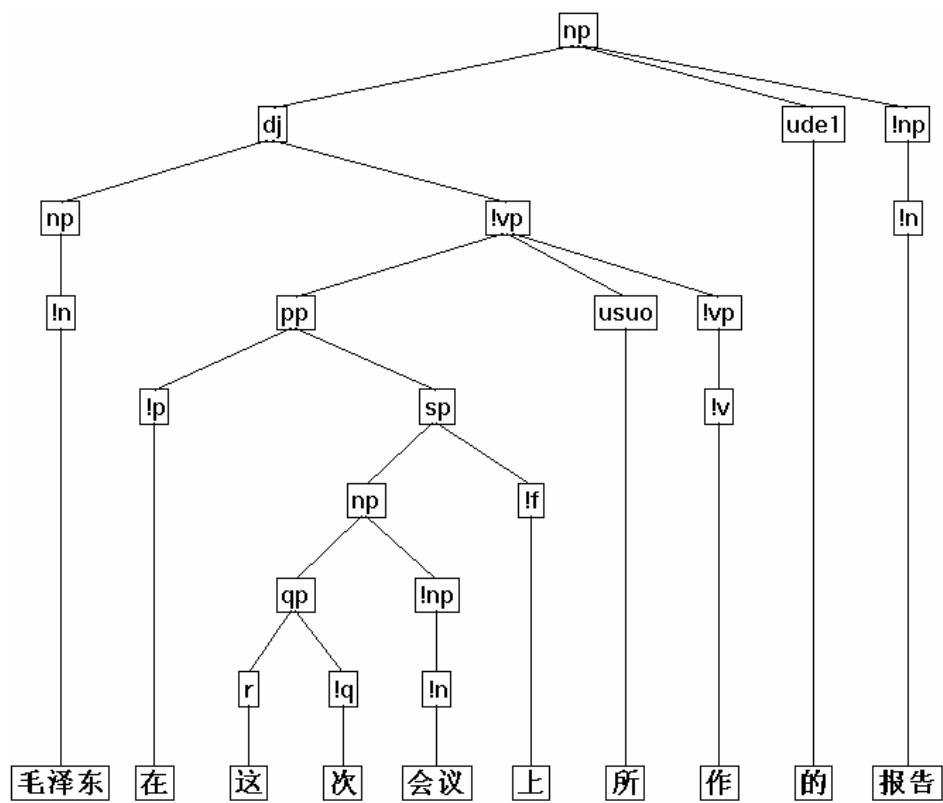
图 1-20



“他所发现的” 分析方式如下：



“毛泽东在这次会议上所作的报告” 分析如下：



二 复句判别问题举例

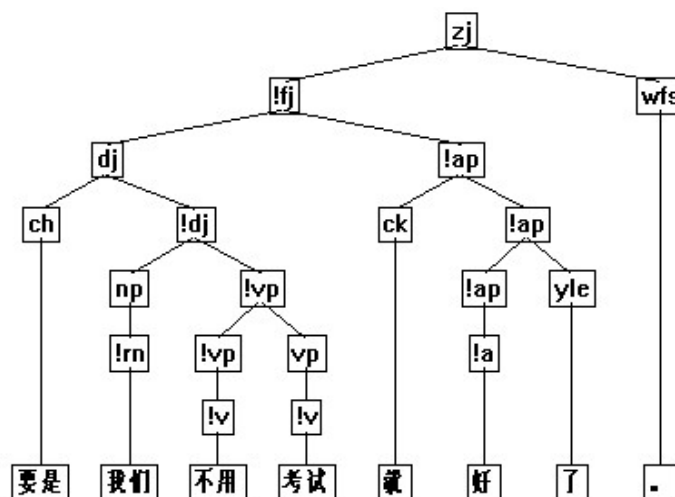
树库校对工作中涉及复句的问题，主要是 dj 和 fj 的判别问题。dj 和 fj 的界限常常难以划清，到底哪些句子定为 dj，哪些定为 fj，有时候有很大的主观性。下面我们举几个例子来说明这个问题。

2.1 有关联词语但没有分句标点的句子

比如“要是我们不用考试就好了。”这个句子处理为 dj 还是 fj 呢？

我们做如下规定：对于有成对关联词出现并且前置关联词 ch 和后置关联词 ck 能够处于树结构的同一层级上的句子，一般处理为 fj。该句符合这一要求，因此应处理为 fj。

图 2-1



2.2 “np + vp₁, + vp₂, + ……, + vp_n” 形式的句子

比如：“广大妇女关心国家大事，关心政府工作，积极参政议政。”

这个句子，是 fj 还是 dj？

对这类在句首的 np 后面并列多个带有排比意味的简短 vp，且没有关联词语连接的句子，宜将这几个 vp 先处理成一个大的 vp，再与前面的 np 构成 dj，如下图所示。

2.3 总分、解说类句子

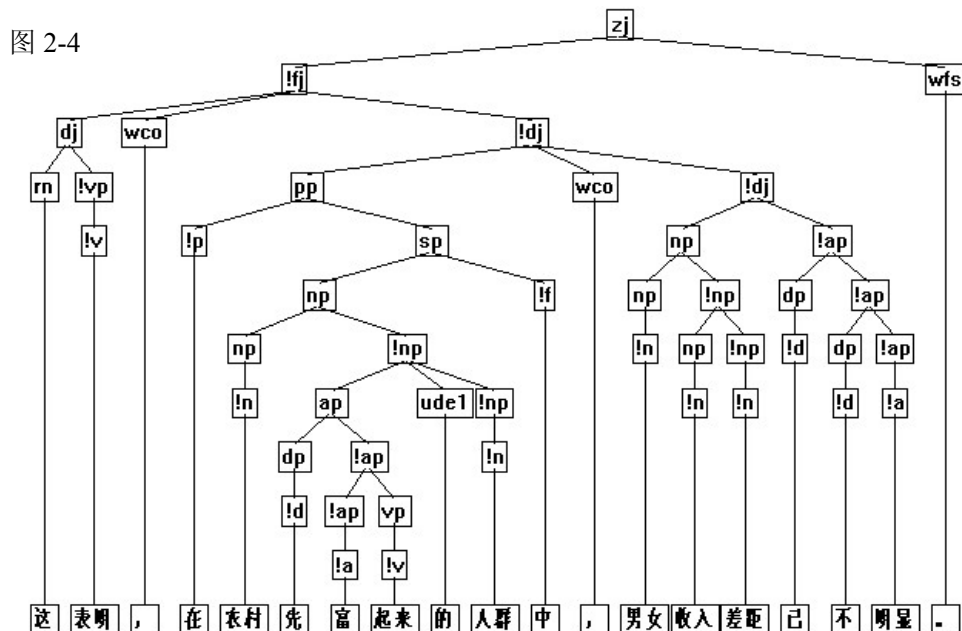
下面这些句子是 dj 还是 fj?

a. 这表明，在农村先富起来的人群中，男女收入差距已不明显。

b. 这主要有：劳动就业的权利，同工同酬的权利和休息的权利，获得安全和卫生保障以及特殊劳动保护的权力，享受社会保险的权利。

c. 法律规定，任何单位在录用职工时不得以性别为理由拒绝录用妇女或者提高对妇女的录用标准。

这类句子虽然可以分析成 dj，中心动词是“表明、有、规定”等，后面一长串都作中心动词的宾语，但很显然，宾语部分都相当长。而且，如果把句子看作 dj，会造成主谓之间没有标点（停顿），述宾之间却有标点（停顿）的情况。虽然一般认为述语和宾语之间要比主语和述语之间的关系更紧密，但由于这类句子句首的 np 和 vp 都比较短小，词数相近，之间没有标点隔开，因此它们倾向于先结合在一起；同时又由于“宾语”太长太复杂，再加上标点符号的分隔，更加疏远了“宾语”与前面 vp 的关系，同时也加强了前面的 np 和 vp 先行结合的趋势。综合考虑，全句不宜分析为 dj。而是分析为 fj。如下图所示：



正如我们在树库加工规范文件中已经指出的，dj，fj 跟 vp，np 等短语功能类不同，基本上是从结构角度对语言单位的定性，而不是从功能角度对语言单位的刻画。一般来说：

- (1) dj 长度相对短一些，fj 长度相对长一些；
- (2) dj 通常内部不含标点，fj 则以内部包含标点为常见；
- (3) dj 通常不含关联成分（包括词关联，结构平行关联等），fj 则包含关联成分。
- (4) dj 的内部成分之间一般形成主谓、状中两种结构关系；fj 的内部成分之间则一般不是这两种关系，而是并列、递进、转折、因果、假设、条件等等关系。

在判别一个语言单位属于 fj，还是 dj 时，应综合考虑上述 4 个因素。这几个因素的重要性顺序⁸是 4 > 3 > 2 > 1

⁸ 从形式角度，或者说可操作性角度来说，这四个因素的醒目程度的顺序大致上是 1 > 2 > 3 > 4。上面在提及这四个因素时，是按照醒目的顺序来描述的。

三 分词问题举例

汉语中‘词’不是一个天然单位，而是一个理论单位，并且没有严格的判别标准来说明什么样的语言单位算作一个词，因而在具体进行语料标注时，会碰到不易决断的情形。就树库标注工作任务来说，多大的一个单位算作‘词’，不是特别重要的问题。举例来说，“北京大学中文系”可以算一个词，也可以分析为“北京大学”+“中文系”两个词，甚至还可以分析为“北京大学”+“中文”+“系”三个词，数清楚这里有几个词，不是树库标注最关心的问题。只要保证，对于这个例子，不分析为“北京大”+“学中文”+“系”，或者“北京大学中”+“文系”，就可以了。

有了上述认识，再来看下面谈到的许多分词问题的例子，不难体会到：一个单位是分开成两个词，还是组合在一起作为一个词，有时是两可的。我们对于分词的要求是：(1) 不要犯严重错误（即分析结果明显不合理）；(2) 同类情况做同样处理（保证一致性）。在树库标注中，更为重要的任务是给一个语言单位确定合理的功能类别。反过来，我们也可以根据是否有利于确定功能类，来决定有些语言成分是合起来作为一个词好，还是分开来作为两个词好⁹。

3.1 组合型歧义字段的切分

组合型歧义字段是指包含至少两个汉字的汉字串，记为 AB。其中 AB 本身是词，切开来，A、B 也分别是词。而合起来和分开来两种处理方式对 AB 的解释是不同的。目前计算机自动分词还不能完全解决组合型歧义字段的切分问题，因而需要对切分不当的进行人工校正。下面是两个例子。

1) “个人”是作为一个词？还是分开成两个词？

“个人”是一个典型的组合型歧义字串，可能是“(我)个人”，也可能是“(三)个人”。对于前者，应合为一个词，标为 n；对于后者，字串的左边相邻词是一个数词，则应该切分开，处理为“个/q 人/n”。

2) “就是”是作为一个词？还是分开成两个词？

“就是”的分合（以及词性）取决于它的语法功能。如果其语法功能是虚词，即作连词、副词，则必须合一。如果是做述宾结构的述语，如“这本书就是他的”，那么允许分开，也允许合在一起，即处理为：

a、这/r 本/q 书/n 就是/v 他/r 的/u

b、这/r 本/q 书/n 就/d 是/v 他/r 的/u

a、b 两种处理方式都可以。“也是”“总是”等也作类似处理。

⁹ 有的分词程序会把“一个”切分为一个单位。主要的考虑是“一个”是个高频成分，认为作为一个单位处理，可能在后续的分析中可以提高效率。但实际上，如果把“一个”合起来，而“两个”是分开的，那么后续处理中就相应地需要规则去判别“一个”跟“两个”属于不同的情况，反而有可能降低效率。我们的认识是，就“一个”这个单位本身来说，切开来，合起来，是两可的。但真正的问题在于，相同的语法现象，应该做同样的处理（强调“一致性”）：如果“一个”合起来，“两个”在分词阶段就也应该合起来。否则在后续分析中因为这种不一致，会造成处理上新的开销。如果从有利于功能定性的角度来考虑，“一个”分开来是更好的分析方式：一/m 个/q 组合为 qp。功能界定很清晰。把“一个”合在一起，词性不好标，一般只能勉强标一个数词 m，但“一个”作为数词，显然跟一般的 m 功能又不一致。这又会在短语组合阶段的处理时带来额外开销。

3.2 “X + 方位词”字段的切分

1) “对内”“对外”是合为一个词？还是分开成两个词？

分开处理比较好，“对/p 内/f”可以组合为一个 pp，这样比较容易对这类成分进行功能定性，如果合为一个词，不太好确定其词性。

这个例子启发我们产生这样一个认识：语言成分（词）在使用过程中，“大小”（长度）会发生变化，也就是在不同层级的语言单位之间发生“跃迁”现象（比如“来信”：从词组跃迁到词）。有时候，这种跃迁就会产生新的“类”，比如语言成分 x 从 A 层级跃迁到 B 层级，x 在 A 层级原有功能分类在 B 层级不适用。于是就需要在 B 层级上重新分类。介词就是这样一种情况，在词层级上，介词 p 的功能是带宾语（除此之外，没有其他功能），在短语层级上，介词短语 pp 的功能是作修饰语或补足语（除此之外，没有其他功能）。显然，p 和 pp 的功能差异及其显著（不妨对比 n 和 np 之间的功能差异，不难发现，n 和 np 的功能一致性要大于其差异）。如果在 p 和 pp 之间发生跨层跃迁现象，就必然造成功能定性上的困难。碰到这种情况，我们应该抓住我们关心的主要问题，即功能定性问题，如何处理分词便于进行功能定性，就如何分词。本着这个原则，“对内”“对外”都适宜分为两个词。

2) “马上”“树上”“墙上”需要切分开吗？

“马上”跟“树上”“墙上”不完全一样，还是个组合型歧义切分。作副词时，切分为一个单位（如“我马上就来”），否则要分开。

“树上”“墙上”一般分开，处理为两个词。当然，“墙上”这样特别常用的方位结构，可以作一个词处理。“桌子上”“理论上”“原则上”都切分为两个词。

“组织上”“国际上”由于常用且有引申义，一般处理为一个词。

重申一点：以上是合是分，都可以。无论分合，对于成分的功能定性不能出错。

3) “班里”可以看成是一个词吗？

可以看成是一个词。这种情况我们按照 ICL《词语切分与词性标注》规范处理。“省、市、县、乡、村、部、局、处、团、营、连、院、系、班”等名词后“里、上”等方位词，仍有组织、机构的意义，作为一个切分单位，标为名词。（ICL“规范”§2.2(14)②c.）

部里/n, 县里/n, 村里/n, 系里/n, 班上/n

“班里的同志”和“支部/n 的/u 同志/n”是平行的，所以“班里/n”。

3.3 重叠式字段的切分

重叠是复合词的一种构词方式，对重叠式字段的切分处理，请参照 ICL《词语切分与词性标注》规范。这里从略，仅举两例。

1) “v—v”中的“—”是什么词性？“v不v”中的“不”是什么词性？

在没有系统地研究这些词的语法性质之前，暂时的处理方式是“—”标为 m，“不”标为 d。“a不a”形式（如“好不好”）也要切开。但是如形成“A不AB”的不完整形式，则不予切分，并分别标以词性 v 或 z（如“相不相信/v”、“容不容易/z”、“漂不漂亮/z”）。

2) “一片片、一个个”等需要切开吗？“一瘸一拐”怎么处理？

“一片片、一个个”等内部构造为“m+q+q”的结构，应该切分开来，处理为三分结

构。

“一瘸一拐”这样的“一 v 一 v”结构形式，也切分开，“一”标为 m。

3.4 离合词的切分

离合词的处理参照 ICL《词语切分与词性标注》规范。这里举三个例子：

出差/v 出/v 过/u 两/m 天/q 差/ng

理发/v 理/v 了/u 一/m 次/q 发/ng

洗澡/v 洗/v 了/u 一/m 个/q 舒舒服服/z 的/u 澡/vg

双音节离合词分开使用，其中一个是语素的，可将它标注为语素。（见 ICL“规范”§2.2(14)①）

3.5 “特别是”“取决于”“倒不如说”等是否要切分开

“特别是”处理为连词 c，对“科学的发展，特别是电子工业的发展”这样的结构形式，“特别是”前后两部分功能类型相同，可处理为一类特殊的联合结构： $xp \rightarrow !x \ c \ y$ 。

“取决于”处理为 v，类似的由“双音节动词 + 于”形成的结构，少数使用频繁、内部结合紧密的，可以处理为一个词；否则切分开（构造模式为 $vp \rightarrow !v \ p$ ）。类似的，“致力于”从合处理，标为 v。而“服务于”，内部结合不够紧密，一般从分处理。

“单音节动词 + 于”的形式，“属于”处理为 v，“写于”处理为 v + p。前者内部结合紧密，后者结合不够紧密。

以上关于“v + 于”的处理意见只是建议。都合起来，或这都分开来作为两个词，也都是允许的。树库标注工作者可以根据自己的语感来判断。如果各人对“词”的感觉不同，就没有必要强求相同。“词”在相当程度上，是一个“主观”的语言单位。

“倒不如说”可以处理为一个词。在“与其说……倒不如说……”形成的关联结构中，“倒不如说”充当了关联词的作用。树库加工中可以标为“与其说/ch”、“倒不如说/ck”。类似的习语性质的篇章关联成分，一般都可以整体作为一个切分单位。

3.6 “伸出”“穿上”“激起”“露出”等是否要切分开

这一类“动词 + 趋向动词”的结构形式，按 ICL“规范”意见，分合都可以。当然“伸出”“露出”“激起”等结合紧密、使用稳定的“动趋式”已被收入词典，合在一起自然是更好些。若分开，如“穿/v 上/v”，用 $vp \rightarrow !v \ v$ 规则组合来描述。

3.7 “选举权”“群众性”“正确率”等是否要切分开

这类“词”（可叫做“语法词”）跟一般的分词单位（可叫做“词汇词”）不同，是通过分词单位与“缀”结合的附加式构词法形成的。我们倾向于作简化处理。比较常用的，如“选举权”“被选举权”“群众性”等就作为一个词切分。对于不太常用，有临时组合性质的片断，可以处理为“x + ng(权)”“x + ng(性)”“x + ng(率)”等。

3.8 “外交部长”“课题组长”等是否要切分开

这类字串都是词组的“压缩形式”：“课题组长”是“课题组组长”的压缩形式；“外交部长”是“外交部部长”的压缩形式。这一类语言片断，都可按一个词处理。类似的例子还有“教育局长”“党小组长”。

如果把“课题组长”切开成为“课题/n 组长/n”，问题也不大，因为二者还可以再组合成为一个 np；但如果把“教育局长”切开成为“教育/v 局长/n”，则二者再组合时，就有可能结合为 np，也有可能结合为 vp，造成歧义。这对计算机的自动句法分析会造成困难。因此，在分词和词性标注阶段，就把这类片断作为一个整体单位处理，是有利的。

3.9 “深蓝色”“淡绿色”“粉红色”等是否要切分开

复合形式的颜色词从合处理为一个切分单位。

如果计算机自动分析的结果是“黄/a 白色/n”，则需要修改为“黄白色/n”

3.10 “纸张”“雨滴”“木条”“活字块”等是否要切分开

这一类内部构造是“n+q”的结构，如果比较常用，并且意义固定，收入了词典，那么处理为一个词，如“纸张”“雨滴”“木条”等。否则处理为 np[!n + q]，如“活字块”等。

3.11 “横槛”“人声”“党证”等是否要切分开

这类双音节成分不切开，从合，标为名词。“横槛”整体指称一种事物，横/n+ 槛/n ≠ 横槛/n。类似的，“横线”也是一个名词。

“人/n 声/ng”“党/n 证/n”“月/n 牙/n”“日/ng 食/ng”（“日/j 食/v”）等被切开的，也应合并为一个词，处理为“人声/n”“党证/n”“月牙/n”“日食/n”。

3.12 “6.7%”“上百”等表数成分是否要切分开

数字表示的是一个完整意义，一律从合。

“6.7%”，在计算机自动分词时可能会切做“6/m ./w 7/m %/w”，人工校对时需要合起来作为一个切分单位，“6.7%/m”。“上百”“上千”等也都处理成一个词，标记为 m。

3.13 “米每秒”“英里每小时”等如何处理

看作是术语词，不再切分开，标为量词 q。

3.14 “有的是”

“有的是”有歧义。如果在“有的是，有的不是”中，则“有的”分析为 np，“有的 + 是”分析为 np + vp 形成的 dj。如果是在“地球上的悬崖峭壁有的是”这样的例子中，“有的是”分析为一个单位，标记为动词 v。

3.15 “大不了”

“大不了”跟“有的是”类似，也是有歧义的。如果跟一般的“瘦不了”“胖不了”一样，形成的是“a + 不 + 了”组合，则分析为三个词，整体为 ap。组合模式为 ap → a d v。“了”在这里分析为动词 v。如果在“大不了不去了”这样的例子中，“大不了”合为一个词，词性标为 d，即作副词处理。

3.16 “一整夜”

“一整夜”，意思相当于“一夜”“整夜”。可以合为一个单位，标记为时间词 t。

3.17 “一会儿/一会/这会儿”分别怎么分析？

“一会儿”“一会”都可以处理为一个单位，标记为数词 m。

“这会儿”“那会儿”也可以处理为一个单位，标记为时间词 t。

四 词性判别问题举例

4.1 介词与连词、动词的纠葛问题

1) “为”“为了”是什么词性?

现代汉语中介词和连词之间有一些纠葛，主要是两组：一组是“和、与、同、跟”，一组是“因、以、因为、由于、至于、为了”等。一般来说，充当介词宾语的是体词性短语，但后一组介词也可以带谓词性宾语，这时就与连词发生纠葛。例如：

- a、为了民族的解放，他把自己的一生都献给了革命。 → 为了 np, dj。
- b、为了安置灾民，她忙得整夜没有合眼。 → 为了 vp, dj。
- c、为了儿子能考上大学，她不惜牺牲自己的一切。 → 为了 dj, dj。

我们对此的认识是：连词一般连接表示各种逻辑关系或话题转换等的句子，介词主要用于引介原因和目的短语。“为了”是连词还是介词，我们的区分方法是根据其所引导的成分的结构形式：如果引导的成分是一个 dj 或 fj，那么就标为连词；如果引导的成分是体词性短语或非句子形式的谓词性短语，那么就标为介词。

2) “一心向着党”中的“向着”是什么词性?

现代汉语的介词绝大多数都是从古汉语的动词虚化发展而来的，介词与动词之间存在着交叉、纠葛。对动介兼类的词，具体定为哪一类，要根据词的分布和语法功能来确定。例如“我一心向着党”，“向着党”在句中语法功能明确是谓语，因此“向着”标为 v。当然，在“向着”的功能类明确是 vp 的前提下，把它进一步切分为“向/v 着/u”也是可以的。而在“向着太阳勇敢前进”一类的结构中，短语的功能中心在 vp [勇敢 前进]，“向着太阳”就该标为 pp[向着/p 太阳/n]。“朝着”“对着”等的处理也如此。

3) “像画上画的”中的“像”是什么词性?

“像”处理为动词。

4.2 量词与名词、动词的纠葛问题

1) “一个人”“两碗饭”都是数量名结构的短语吗?

从整体功能来看，二者都是 np，也可以说都是数量名结构的名词性短语。但二者“数量”的“量”不同。“一个”的“个”，是“专职量词”，而“一碗”的“碗”，只是“临时量词”。一般语法著作和文章在给量词分类时都有临时量词这一类。如“一盆水”“两杯酒”“一脸皱纹”“一走廊人”等。是把临时量词处理成名词兼量词呢，还是把临时量词处理成名词的一种用法，词类不变，仍然是名词呢？如果是前者，在词典中标注词性时既要标出名词也要标出量词。如果是后者，只需要标出名词，不必再标出量词。事实上两种处理方法都有。和“个”“种”“篇”“斤”等专职量词相比，临时量词有三个特点：可类推性，数量很多、是开放的类，词义无明显变化¹⁰。我们主张把临时量词看作是名词的一种用法，仍

¹⁰ 关于临时量词及其词性标注，参看谭景春《从临时量词看词类的转变与词性标注》，沈家煊主编《现代汉语语法的功能、语用、认知研究》，商务印书馆，2005。

然处理成名词。“一个人”分析为 $qp[m + q] + np$ ，“一碗饭”则分析为 $qp[m + n] + np$ 。

除上面的量、名纠缠外，还有其他的问题需要注意，例如“堆了一床书”和“两床棉被”，这两个例子中“床”的功能是不同的，前者是名词作临时量词，跟一般的临时量词一样，充当“容器量词”的功能，后者则是普通的个体量词。对这种多义的名量兼类词，需要仔细分辨。

2) “一弯月牙”“一弯胳膊”有何不同？

对人来说，二者区分非常明显，根本不是问题。但对于计算机自动分词和标注来说则有一定困难。二者的切分与标注结果可能是一样的，

一/d 弯/v 月牙/n

一/d 弯/v 胳膊/n

这反映的就是量词和动词的纠葛问题。很显然，“一弯月牙”中的“弯”是作为量词来用的，而“一弯胳膊”中的“弯”则是作为动词来用的。这两例应该标注为

一/m 弯/q 月牙/n，整体功能是 np。

一/d 弯/v 胳膊/n，整体功能是 vp；

4.3 助词与其他词类的纠葛问题

1) “着”“了”“过”的词性

“着”“了”“过”有时候不是助词，而是动词，应标记为 v。校对过程中要注意分别。例如：

- a. 这次你₁还是没猜₂着。 b. 这么点饭我₁吃得₂了。 c. 他₁回₂过头看了₃看。

其中“着”“了”分别读成“zháo”“liǎo”。在这几个句子里，“着”表示达到目的，有了结果（“着”还可以用于述补结构中的补语位置，例如“找不着”中的“着”就应该分析为动词 v）；“了”跟“得”连用，表示一种能力或可能；“过”表示掉转方向。它们作为动词，还有其他的意义，这里就不一一列举了。

2) “极了”是什么词性？

“极了”只能附在形容词后面，如“美丽极了”、“好极了”，因为有的时候“了”可以不说，只说“好极”，因此“极了”分开为两个单位，“极”标为副词 d，“了”标为语气词 yle。“a 极了”的组合模式为： $ap \rightarrow !ap[!a d] yle$ ，或 $ap \rightarrow !a d yle$ 。也就是说，形容词跟“极”可以先捆绑成 ap 再跟“了”捆绑，也可以由“a、极、了”三项平列组合。

3) “一样”是什么词性？

“一样”的词性有两种不同情况：

(i) 标为 a。这时候“一样”作定语（如“一样的梦想”），作状中结构的中心语（如“他跟你不一样”）。

(ii) 标为 u。这时候“一样”与“像/v……似的/u”平行，形成“像……一样”这样的结构形式。如“日子过得像神仙一样”。

4) “以来”是什么词性？

“x + 以来”整体功能是 tp。考虑到时间是空间的隐喻这样一种关系，把“以来”标为

方位词 f。

5) “近”的词性

“近”有三种用法：

(1) 接近，将近，大约。 例：近两百人参加了这次活动

(2) 最近。 例：近两年他很活跃

(3) 距离近。 例：他们两家离得很近

第三种用法的“近”分析为形容词应该没有争议。前两种用法的“近”是区别为不同词性，还是统一为一个词性，可能有不同看法。如果区别为不同词性，可能第一个“近”会参照“接近”，定为动词 v，第二个“近”会参照“最近”定为时间词 t。如果定为一个词性，那归入哪一类，又是一个问题。

我们的看法是，表示“将近、接近”义的“近”后面跟非时间数量成分的可能更多一些，表示“最近”义的“近”后面一般跟时间数量成分。无论是哪种情形，“近”后面都是跟“数量成分”，从句法功能上看，“近”还是处理为一个词性标记比较好。这个“近”是粘着性的词，对这种词，有两种处理方式，一种是放在数词里面，作为“概数词”“助数词”；一种是放在助词里面，作为“数助词”。朱德熙先生采用前一种方式，郭锐老师采用后一种方式（不过他们都没有谈到“近”这个词，而是讨论“多、来”等例子）。两种方式似乎很难分出好坏。本规范暂定“近”为数词 m。

4.4 与代词相关的词性判别问题

1) “之一”是什么词性？“之后”呢？

“之一”处理为代词 r。“之后”处理为方位词 f。“之前”也处理为方位词 f。

“之类”处理为代词 r。

2) “各二斤”里的“各”是什么词性？整个结构如何标注？

“各”处理为 r，“各二斤”处理为主谓结构，标为 dj。

同样，“各顾各”切分为“各/r 顾/v 各/r”，整个结构是 dj。

3) “有些”是什么词性？

“有些”处理成 r，可视其功能细分为 rd（如“有些害怕”）、rm（如“有些人”）。

4) “为什么”是词吗？什么词性？

“为什么”可分可合。如果处理为一个词，词性标为 r，还可视具体功能细分，比如单独作谓语，可标为 rv，上升为 vp 再参与组合。如果是在连谓结构前项，可分开，结构分析为：pp[为/p 什么/r]。

4.5 兼属副词和区别词的修饰词

“先后”可以作定语（比如“先后次序”），也可以做状语（比如“先后离开”）。跟“先后”语法性质类似的词还有一批，比如“长期”“临时”“真正”，等等。这些词同时又都不能受“很”的修饰，因而不符合一般形容词的鉴别标准。这一类词的词性是什么呢？

可以处理为兼属副词和区别词的兼类词：

在定语位置为区别词 (b)，在状语位置为副词 (d)。

4.6 “同时”是什么词性

“同时”可以修饰谓词性成分，可以在 vp 前，也可以在 dj 前位置出现，还可以在体词性短语的中心语位置出现（“教师在上课的同时”，“他被打倒的同时”）。换句话说，尽管大多数时候“同时”起状语作用，但它的功能并限于作状语，因此不适合看作副词。如果按照检验时间词的严格标准来判别，“同时”也不能直接作“在、到”的宾语，不能用“什么时候”提问，似乎也不适合看作时间词。在这种情况下，采取“两害相权取其轻”的策略，将“同时”定为“时间词”比定为“副词”更好一些，因此定为“时间词”。

“上课的同时”整体功能是 tp，结构内部按定中式“x 的 y”三支结构进行分析。

4.7 “好多年”“二十多年”中的“多”是什么词性

“好多年”切分为三个单位：

好/d 多/a 年/q

“好”处理为 d（相当于“好坏”“好臭”“好美”里的“好”），“多”处理为 a，“好多”构成一个 ap，“年”是量词，ap + q 构成 qp（数量结构）。

“二十多年”中的“多”应看成 m（不定数词）。这个词组的组合模式为：

qp → mp q 其中 mp 的组合模式为 mp → !mp m

“二十多”组合为一个数词短语（mp），然后再跟“年”组合为一个数量短语（qp）。

对两个词组中“多”的词性的不同处理，也可反映“多”在两个词组中的含义差异，“好多年”中的“多”相当于 many，“二十多年”的“多”相当于 more。

类似的，“要记住这么多笔账”中的“多”处理为形容词 a；“四点多”中的“多”处理为数词 m。

4.8 “一个又一个”中的“又”如何标注

这里的“又”分析为连词 c。

4.9 “爬呀爬”中的“呀”是 y，还是 o？

“呀”用在“v 呀 v”格式中，词性如何标注。下面也是一个“v 呀 v”的例子：

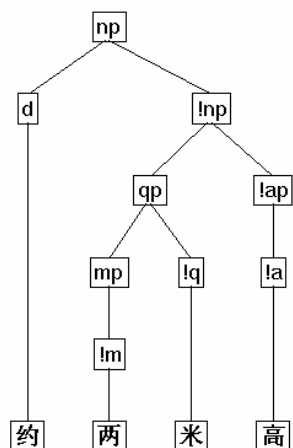
“野花遍地是：杂样儿，有名字的，没名字的，散在草丛里像眼睛，像星星，还眨呀眨的。”

这种格式中的“呀”应分析为语气词，类似的书面上还有“啊”，也可以用在“v 啊 v 啊”格式中。这类格式的组合规则均为：vp → !v y v。

4.10 “约两米高”中的“约”

“约”分析为副词 d。“约两米高”分析为：np → d !np。

“两米高”分析为：np → qp !ap



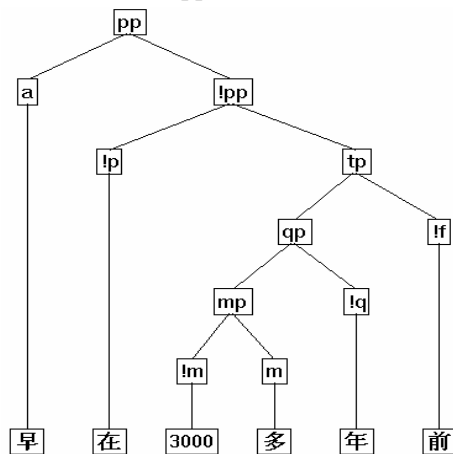
需要说明的是，这里“约”以副词身份参与组合，而不上升为 **dp**，因为在“约”所在的位置上一般只能出现单个副词，而不是 **dp**。同样的，“高”所在的位置还可以出现“那么高”这样的 **ap**，因此“高”需要从 **a** 上升为 **ap**，再参与组合。

4.11 “一般五寸至二尺长”中的“一般”

这里的“一般”分析为副词 **d**。

4.12 “早在 3000 多年前”中的“早”

这个“早”分析为形容词 **a**，修饰 **pp** 结构。类似的例子如“直到两年前”。



4.13 “三十年来”和“三十来年”中的“来”

“三十年来”整体分析为 **tp**，其中“来”标记为 **f**。

“三十来年”整体分析为 **qp**，其中“来”标记为 **m**。



4.14 拟声词的功能定性

拟声词可以在句法结构中占据不同的位置，应根据其所能占据的位置判别其功能类。比如：

- i. 哗，哗，哗，大雨真的下起来了
- ii. 大雨哗哗哗下起来了

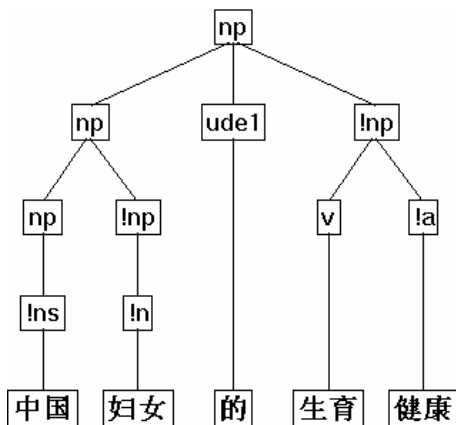
这两例中的“哗，哗，哗”应该分析为 vp。 $vp \rightarrow o \ o \ o$ 。

4.15 “那就比一比吧”中的“那”

这里的“那”标记为连词 c。

4.16 “生育健康”中的“生育”和“健康”分别是什么词性？

“中国妇女的生育健康”



“生育健康”的组合模式为 $np \rightarrow v \ !a$ 即处理为由 动词加形容词形成的 np 结构。在词性层面，维持“生育”“健康”一般常见的词性，在短语层面，则根据功能将“生育健康”定性为 np。

4.17 “等我醒过来”中的“等”是什么词性？

这里的“等”分析为动词 v。

4.18 “这一发现”中的“发现”是什么词性？

这里的“发现”分析为名词 n。

4.19 “以图”、“报经”、“诸如”等紧缩型词语如何标注词性？

类似的例子还有“上述”、“与会”等紧缩而成的词语。

按照一般的功能判别标准进行判别。比如“报经”v “以图”v “在内”v “诸如”v “上述”b（和“下列”对举）“与会”v

4.20 “两尺见方”中的“见方”是什么词性？

“见方”标为名词 n，语义相当于“正方形”，“两尺见方”即“边长为两尺的正方形”。

4.21 “总的来说”“综上所述”是什么词性？

这些成分都分析为连词 c。这些词的作用是篇章关联作用，在句法结构分析中跟一般连词 c 一样进行组合分析。

五 文言及书面特殊用法的处理

真实文本语料中时常会有“突破规范”的地方，如文言表达、作者“个性化”写作造成的词语特殊用法，等等。这些“特殊情况”，会对自动分词和词性标注造成影响，而目前计算机还不善于应付这些问题。在树库加工中需要视具体情况进行人工校正。

5.1 文言

1) 双字组“x+然”的切分标注

双字组的“x+然”，绝大多数都是情态副词，其中像“骤然”“悄然”“飘然”“傲然”等较常用的都已经作为词收入了词典。然而，有相当一部分在词典中没有记录。考虑到“x+然”式的情态副词在书面语中使用较多、数量也大、功能明确，可以把它们作为一个切分单位看待。比如下面三个例子：

a、一定能够傲然自立于世界民族之林

机器切分结果：一定/d 能够/v 傲然/d 自立/v 于/p 世界/n 民族之林/n

b、对这种充满血泪的怪诞与孤傲，我们怎能不**悚然**面对，肃然起敬，**油然**生爱？

机器切分结果：对/p 这种/r 充满/v 血泪/n 的/u 怪诞/a 与/c 孤傲/a ， /w 我们/r 怎/r 能/v 不/d 悚/ag 然/k 面对/v ， /w 肃然起敬/i ， /w 油/v 然/r 生/v 爱/v ？ /w

c、听的人都**肃然**了。

机器切分结果：听/v 的/u 人/n 都/d 肃/vg 然/r 了/y 。 /w

d、其时几个旁听人倒也**肃然**的有些起敬了。

机器切分结果：其/r 时/dg 几/m 个/q 旁听/v 人/n 倒/v 也/d 肃/vg 然/k 的/u 有些/r 起/v 敬/v 了/y 。 /w

从例子中我们看到，计算机对“x+然”的切分标注很不理想。例a中“傲然”正确切分标注为“傲然/d”；例b中“肃然起敬”是成语，切分标注没有问题，例c、d中的“肃然”就都切分标注错了。例b中“悚然”“油然”都切错、标错了。这几个“x+然”，根据其语法功能不同，应分别处理：

傲然/d 悚然/d 油然/d 肃然/z

2) “发诸真情”中的“诸”如何标注？

“诸”标为介词p。

3) “先生知其事否”中的“否”如何标注？

“否”标为副词d。

4) “迥乎不同”该如何分析？

原文是“自然条件迥乎不同”

对于这个例子，如果要分析其组合结构，则应处理为 $ap \rightarrow a \ u \ !a$ 。或者选择不作为结构处理，而作为一个词处理，a(迥乎不同)。

5) “X 之于 Y”应如何分析？

分为两种情况：

I “公之于众” “取之于民” “拒之于门外” “(把原因) 归之于领导的失误”

$vp \rightarrow !v \ u(\text{之}) \ p(\text{于}) \ np(\text{众})$

II “秋 之于 人” “未庄的人们 之于 阿Q” “父母 之于 子女”

“年龄 之于 每个人 (都是不可抗拒的)”

$np \rightarrow !np \ u(\text{之}) \ p(\text{于}) \ np$

对于涉及到古代汉语表达形式的结构，一般应该在词语层面描述其组合模式。不宜在短语层面描述其组合模式，除非该结构在现代汉语中具有普遍性。对于短小的古代汉语表达形式，可以从合，作为一个单词处理。此外，对这类结构的功能类定性，也应根据分布情况来确定。

5.2 特殊用法

5.2.1 “印（任）务” 如何处理

这是“任务”一词的特殊使用方式。在真实文本中这类现象是少数，目前只能作个案处理。

“印（任）务”出现在如下语句中：

妈妈说，我的印（任）务是看妹妹。

机器切分结果：妈妈/n 说/v ，/w 我/r 的/u 印/j (/w 任/rq)/w 务/dq 是/v 看/v 妹妹/n 。/w

计算机自动标注的结果显然是不对的。应修改为：

妈妈/n 说/v ，/w 我/r 的/u 印/x (/w 任/vg)/w 务/ng 是/v 看/v 妹妹/n 。/w

在进行结构标注时，“印（任）务”捆绑成一个整体，标为 np。组合模式为：

np → x w vg w ng

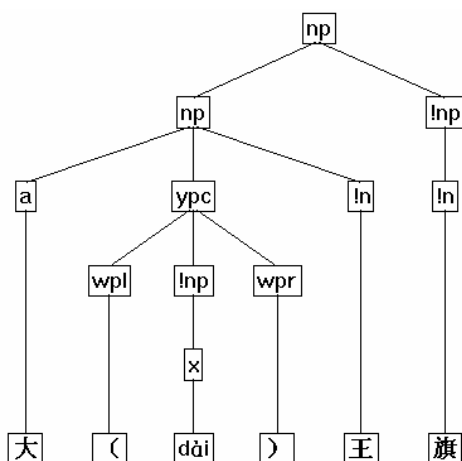
5.2.2 被注音分开的词如何处理

例 1： 城头变换大（dài）王旗

分词和词性标注结果： 大/a （/w dài/nx ）/w 王/n 旗/n

例 2： （台（tai1）州的宁海）

分词和词性标注结果：（/w 台/ns （/w tai1/nx ）/w 州/ns 的/u 宁海/ns ）/w



5.2.3 “好——笑” 这类被破折号分开的词如何处理

“好——笑”这样的情况，属于词的特殊用法。这种情况应该分开来标注：

好/a ——/wda 笑/v

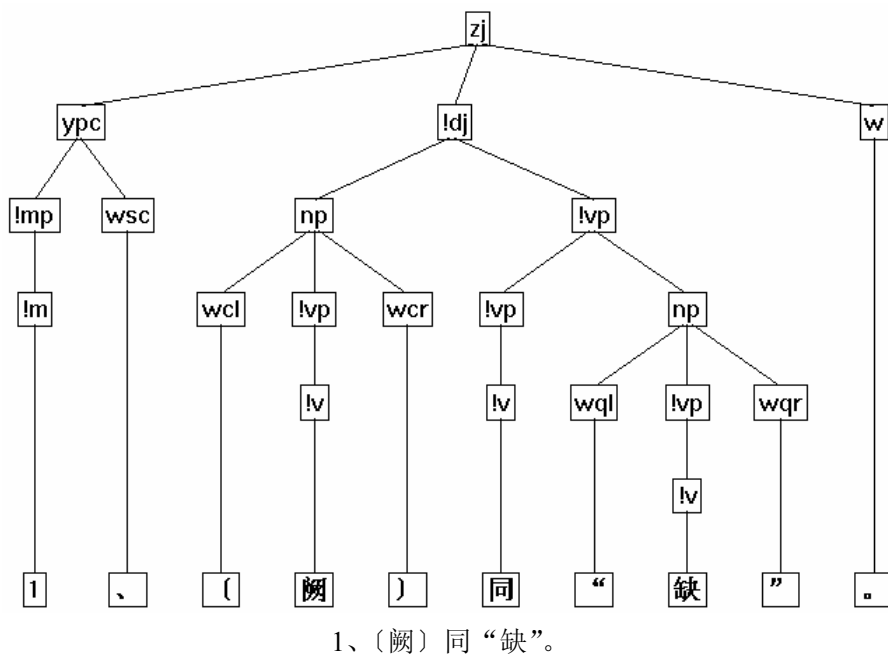
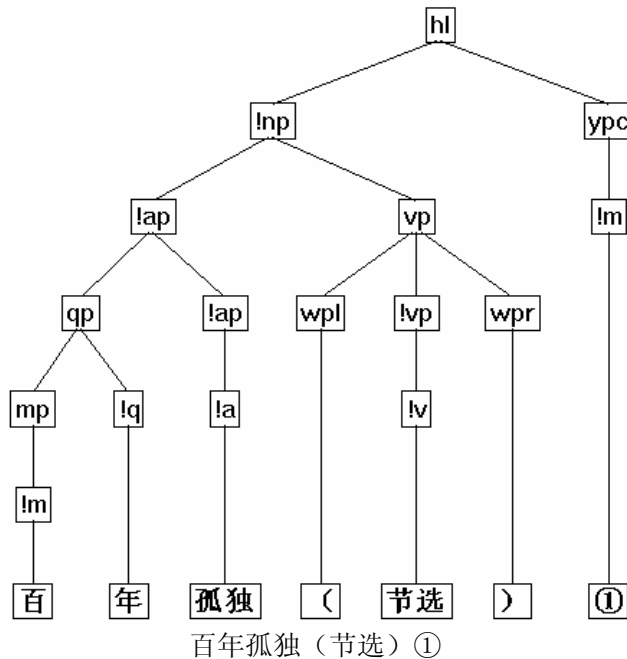
整个片断标注为 ap。组合模式为 ap → !a wda v

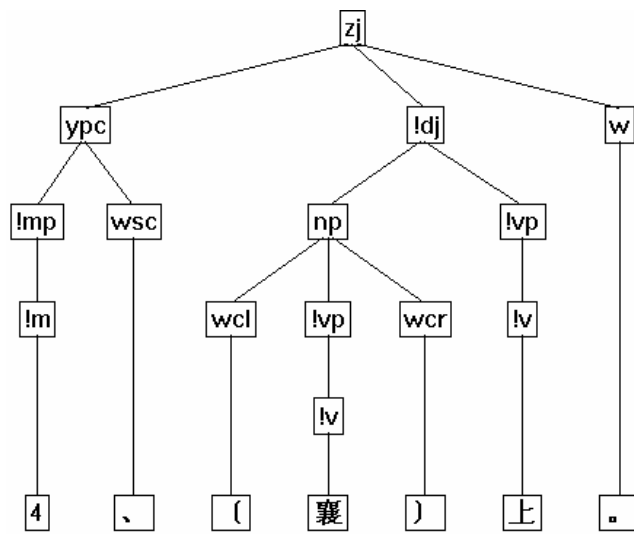
这样处理的目的是便于今后提取这一类特殊用法（这种组合模式很少见，是低频规则模式）。

5.2.4 文章中的脚注或尾注如何处理

- (1) 注释编号一律标为数词 m；
- (2) 注释编号 m 一律上升为“语篇插入成分” ypc，即不参与句法结构组合。

下面是对文章中注释相关部分进行标注的例子：





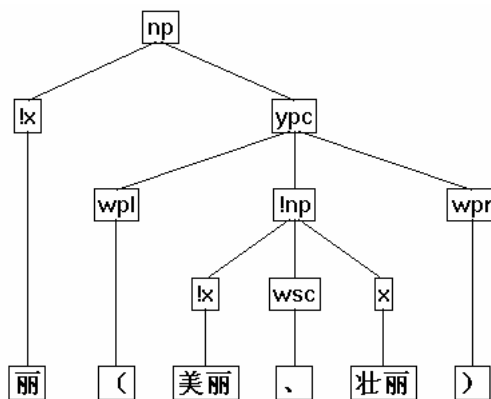
4、(襄)上。

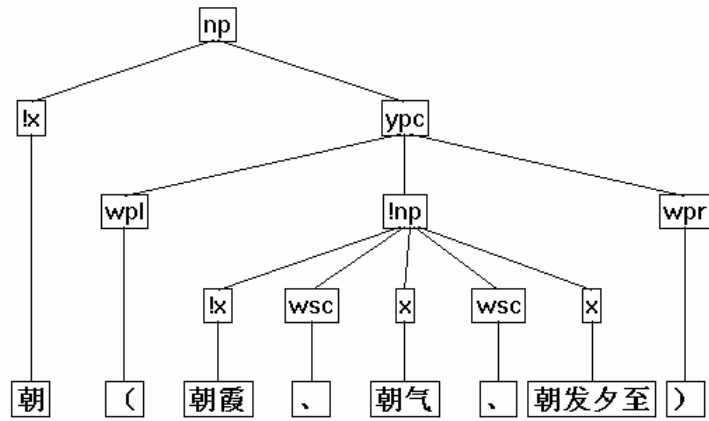
5.2.5 正文中的随文注释如何处理

注释的部分一律处理为语篇插入成分 ypc。例如：

第一种情形是意义没有改变，但是现在不能单用，只能作为复音词或者成语的一个成分。有的构词的能力还比较强，如：形、貌、衣、镜、北、何、自、信、日、外；有的只在极少数词语里出现，如：丽（美丽、壮丽）、朝（朝霞、朝气、朝发夕至）、窥（窥探、窥测）、妻（夫妻、妻子）、甚（欺人太甚）。

其中“丽（美丽、壮丽）、朝（朝霞、朝气、朝发夕至）…”中包含了随文注释的内容。标注可按如下方式处理：





在上述语境中，“丽、朝、美丽、……”等，都是符号的自指用法，因而在词性层次上均标为 x。

---- 正文完 ----