树库在汉语语法辅助教学中的应用初探*
(The application of treebank to assist Chinese grammar instruction: A preliminary investigation)

詹卫东
(Zhan, Weidong)
北京大学
(Peking University)
zwd1972@gmail.com

摘要：本文介绍北京大学中文系近年来所做的树库（Treebank）加工和应用研究的有关工作。全文包括两部分内容：（1）基于树库语料获取汉语句法结构的相关知识；（2）基于树库的汉语句型辅助教学 Web 程序。在汉语语法教学中使用树库资源，可以帮助教师更有针对性地选择语法教学重点，为学生提供句型练习平台，并对学习效果进行自动评估。

Abstract：This article discusses the usage of a large-scale Chinese treebank for Chinese language teaching and learning developed by Peking University. It shows that by specifying the features of a tree, for instance, a structure's immediate constitutes, contexts, tree width and tree depth, one can easily extract the needed language structure from the treebank. It also shows that with the tree structure extracted and the web system developed, one can do structure exercises on the web and the result can be automatically assessed by the treebank.

关键词：中文语法，树库，句法知识获取，人工智能辅助句型教学

Keywords：Chinese sentence structure, Chinese treebank, grammar extraction, intelligent computer-assisted instruction

1. 引言：北大树库简介

树库加工及应用自 1990 年代以来在语料库语言学和自然语言处理领域一直是受到相当重视的研究方向（Marcus, 1993; Abeillé, 2003; Xue Nianwen 等，2000，周强，2004; Huang Chu-Ren 等，2000）。除用于信息处理技术外，还可以利用树库为句法本体研究以及语言教学提供参考。目前计算机辅助语言教学中利用语料库及相

* 本文的研究工作得到霍英东基金项目“大规模中文树库构建及其在对外汉语教学中的应用”（课题号：111098）和国家社科基金项目“语言知识资源的可视化技术研究”（课题号：12BYY061）资助。北京大学计算语言学研究所常宝宝老师对树库加工提供了许多软件技术方面的指导和帮助。匿名审稿人对本文提出了宝贵的修改意见，在此一并致谢。文中尚存错谬，概由本人负责。

关技术已有很多成果（Beatty, 2003; Schmitt, 2007），但把树库资料用于汉语教学的研究和探讨则比较少。本文介绍近年来北京大学中文系树库研究小组在这方面所做的一些工作。北京大学现代汉语树库加工采用的是人机结合的方式，先由程序对原始语料进行断句、分词、词性标注、句法结构分析等处理，然后由人在树库辅助编辑软件（TreeEditor）环境中进行逐句检查，修改程序自动分析的错误，得到最终标注了正确语法信息的树库。下面例 1 是句子“妇女经济地位的提高，是实现男女平等最重要的基础。”的句法标注结果（具体的句法标记符号的含义请见附录，范畴标记前的!号表示该范畴是结构的中心语 Head）。

例 1： (zj (!dj (np (np (np (!n (妇女)) !np (np (!n (经济)) !np (!n (地位)))) ude1 (的) !vp (!v (提高))) wco (，) !vp (!vp (!v (是)) np (vp (!vp (!v (实现)) dj (np (!n (男女)) !ap (!a (平等)))) !np (ap (dp (!d (最)) !ap (!a (重要))) ude1 (的) !np (!n (基础)))))) wfs (。)))

句法结构标注在计算机中是以加括号方式标记在原始句子字符串上进行存储的。例 1 对应的直观的树结构图可以在 TreeEditor 软件中显示如下：

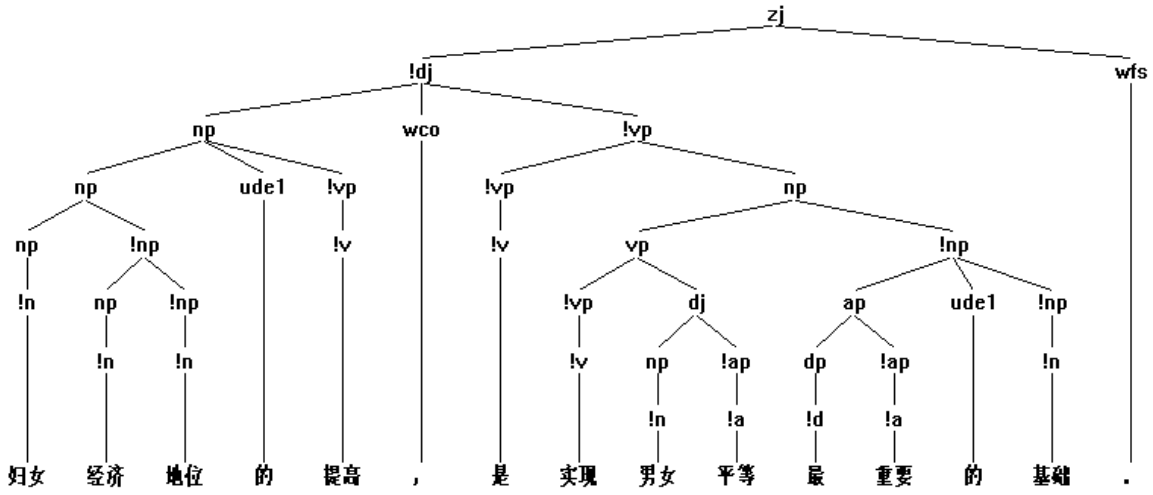
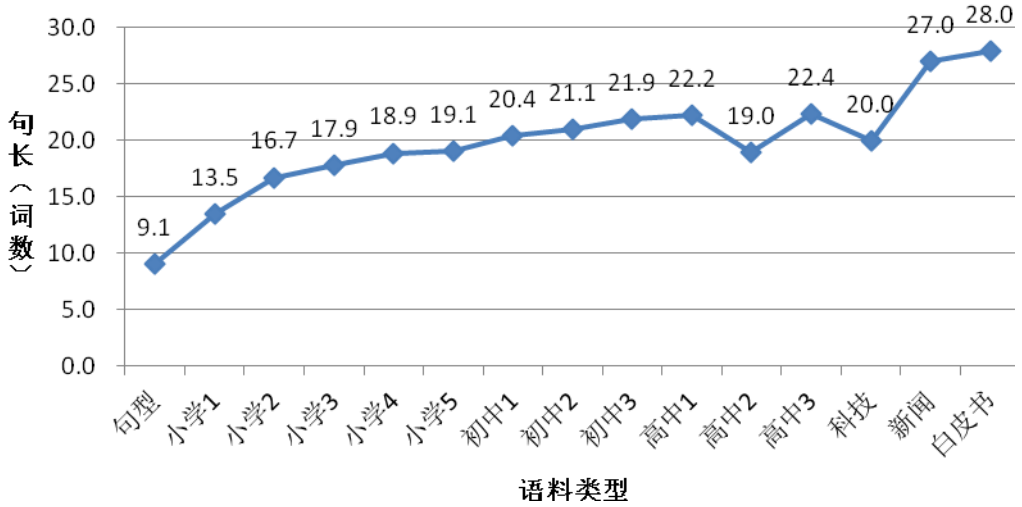


图 1：句子的句法结构树图示例
（语料来源：中国政府白皮书 1994·《中国妇女的状况》）

目前北大中文树库已经标注的语料规模为 55,742 句，1,309,719 字。语料类型包括语文课本（56.85%）、句型例句（13.29%）、新闻语料（13.00%）、科技语料（9.43%）、政府白皮书（7.43%）。各类型语料平均句长（以词数计）的分布如下图所示：

图 2： 树库各类型语料的平均句长



句型语料来自语法文献中的例句，是经过人为裁剪过的句子，因而较短。语文课本语料的句长从小学到高中，整体呈现逐渐加长的规律。只有高二的课文句长出现了下降。原因是高二课文中有 4 个大篇幅的话剧剧本（《雷雨》《茶馆》等，其中的句子大部分是短句。科技语料、新闻语料和政府白皮书代表了当前正式应用文体中的句长情况。

以往加工树库的应用主要集中在自然语言处理（NLP）方面。本文则探讨将树库用于语法辅助教学的途径。本文的思路是，一方面，对于教师和语言学习高级阶段的学习者，可以通过树库查询工具，从树库中提取例句和特定的宏观的句法结构知识，为语法教学提供更有针对性的材料。另一方面，学习者还可以利用画树工具对未标注的自然句子进行结构标注，并将结果与树库中已存贮的参考答案进行自动比对，从而对学习者掌握句法结构的能力进行自动评估。本文接下来就对这两方面的基于树库的应用做具体的介绍。

2. 基于树库获取现代汉语句法结构知识

从树库中可以获取的句法结构知识是多维立体的，包括抽取带词性和频度信息的词表，兼类词的分布统计，短语结构规则及其频次排序，短语分布环境及其频次排序，歧义短语结构及其频次排序，等等。在这些数据基础上，还可以就教师和研究者感兴趣的问题，做专项信息提取和归纳。限于篇幅，本节以三个实例来介绍从树库中获取不同类型的句法结构知识的情况。第一个实例是考察某一类短语在不同句法位置的差异（以名词性短语 np 为例）；第二个是考察汉语中特定句法结构的内部构造特点和外部环境特点（以“把”字结构为例）；第三个是考察汉语中违反中心扩展规约的短语结构的情况（参见沈家煊 2007,2009）。

2.1 名词性短语（np）在主、宾语位置的差异考察

np 在主、宾语位置对应的短语结构规则一共有 6 种情况：np 在主语位置时对应谓语有 4 种：vp，ap，np，dj（规则分别为 dj→np !vp，dj→np !ap，dj→np !np，dj→np !dj）。np 在宾语位置时有 2 种情况：述宾结构（规则为 vp→!vp np）和介宾结构（规则为 pp→!p np）。下面以主语位置为例说明如何用 TreeEditor 工具抽取特定句法位置上的 np 实例。

图 3：TreeEditor 高级查找对话框

在用户指定如图 3 所示的查找条件后，程序可以输出父节点为 dj，右兄弟节点为 vp（即满足 dj → np !vp 规则模式）的所有 np 的实例（或结构模式），并且带有每种结构模式的频次信息。通过像上面这样指定不同树结构作为查找条件，就可以获得 np 分别在主语位置上和宾语位置上的各种内部构造类型及其频次信息。np 的内部构造可以用短语结构范畴的组合来表示，也可以直接用词类组合来表示。下面表 1 显示了在主、宾位置上的高频 np 的内部结构情况，是从短语结构组合的角度来看 np 结构的。表 2 则显示了由 3 个词组成的 np（即 np 词长为 3）在主、宾位置上的内部构成情况的差异。

表 1：主、宾语位置上的 np 的内部结构及其频次

语结构规则	主语位置			介宾位置			动后宾语位置		
	序号	频次	频率	序号	频次	频率	序号	频次	频率
np -> !rn	1	23906	35.01%	2	2436	15.98%	5	4788	7.28%
np -> !n	2	15561	22.79%	1	4218	27.66%	1	18852	28.67%
np -> np !np	3	7638	11.19%	3	2227	14.61%	3	7188	10.93%
np -> np ude1 !np	4	3449	5.05%	4	1180	7.7%	4	5313	8.08%
np -> qp !np	5	2630	3.85%	5	775	5.08%	2	8313	12.64%
…	…	…	…	…	…	…	…	…	…
总计	type 数：172 token 数：68279			type 数：128 token 数：15247			type 数：224 token 数：65751		

表 2：词长为 3 的 np 在主、宾语位置上的内部组成情况及其频次

词结构规则	主语位置			介宾位置			动后宾语位置		
	序号	频次	频率	序号	频次	频率	序号	频次	频率
np -> m q n	3	760	10.05%	4	180	8.26%	1	2771	23.83%
np -> rn ude1 n	1	1091	14.43%	1	260	11.93%	2	1169	10.05%
np -> n ude1 n	4	730	9.66%	3	217	9.96%	3	853	7.33%
np -> rb q n	2	977	12.92%	2	221	10.14%	4	689	5.92%
np -> a ude1 n	7	161	2.13%	5	117	5.37%	5	666	5.73%

可以看到，总体来说主语位置的 np 跟介宾位置的 np 性质更为接近，谓语动词后的宾语位置上的 np 跟二者相差较大。表 1 中“np→!rn”规则表示由人称代词充当的 np，这类 np 在主语和介宾位置都是最常见的，而在动后宾语位置，频次则排在第 5 位。相反，“np → qp !np”规则对应的是汉语中的一般的“数+量+名”结构，这类 np 在主语和介宾位置的出现频率都排在第 5 位，而在动词后宾语位置则居第 2 位。表 2 中结构规则“np → m q n”是“数 + 量 + 名”组合，“np → rb q n”是“指示词 + 量 + 名”组合。前者一般对应语义上的不定指成分，后者则对应定指成分。在动后宾语位置上，不定指性 np 远多于定制性 np。而在动前的主语位置和介宾位置，情况则相反。不过，在动前位置，两类 np 的数量差异没有在动后宾语位置相差得那么大，这主要有两方面的原因，一是汉语允许“无定 np 主语”（范继淹，1985，魏红、储泽祥，2007），二是形式上的无定 np，在语义上也可以表达定指义或者类指义，如“一个人毁坏了别人的东西，应不应该赔偿？”中的“一个人”是无定形式的 np，用于主语位置，语义上并不是表达非定指，而是表达类指。总的来说，从树库中获得的数据实际上印证了以往汉语研究中所观察到的现象，即汉语中旧信息倾向居动词前位置（主语位置 np 和介宾位置 np 都在谓语动词前），新信息倾向居动词后位置（LaPolla，1995）。

2.2 “把”字结构中 vp 的内部构造以及“把”字结构整体分布环境考察

跟考察主、宾语位置上的 np 类似，还可以针对某类句式中的短语抽取其内部组合规则。比如，要考察“把”字结构“把+xp+vp”中的 vp 主要有哪些结构类型，可以在图 4 所示的对话框中指定抽取条件。获取的 vp 结构类型及各自所占比例结果如表 3

所示。图 4 中指定的过滤条件意思是抽取符合“vp → pp !vp”这种形式的短语组合规则中标粗体的 vp（而非根节点 vp）的规则模式，其中 pp 是由介词“把”（标记为 pba）带宾语构成的介词性短语。

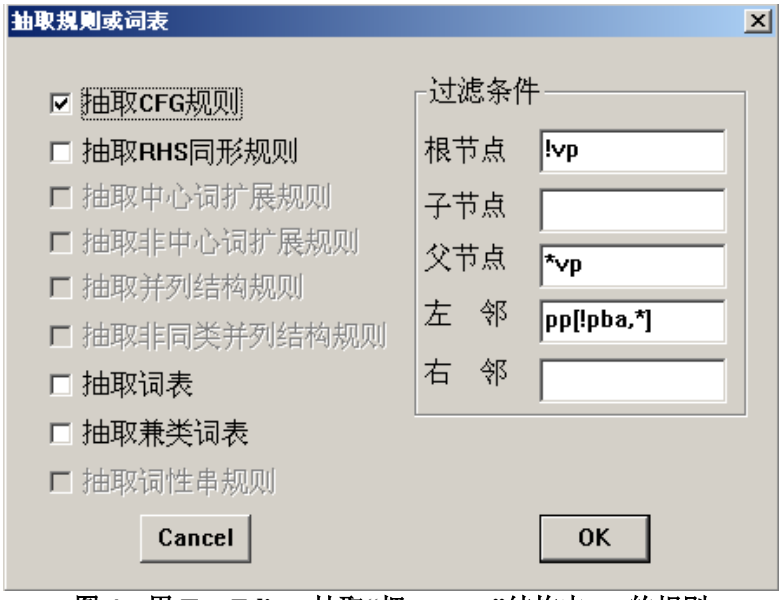


图 4：用 TreeEditor 抽取“把+xp+vp”结构中 vp 的规则

表 3：“把+xp+vp”结构中 vp 的构造类型及示例（vp 各内部构造类型参见詹卫东 2000）

构造类型	实例频次	结构规则	示例
述宾式 vp	999 (39.74%)	vp → !vp np vp → !vp sp vp → !vp qp vp → !vp mp ……	(把 x) 交给 新干部 放在 桌子上 放在 第一位 砍去 一半 …
述补式 vp	907 (36.08%)	vp → !v v vp → !v a vp → !v ude3 ap ….	(把 x) 扔 掉 清理 干净 布置 得 非常漂亮 …
状中式 vp	354 (14.08%)	vp → dp !vp vp → pp !vp vp → ap !vp …	(把 x) 也 抛出来 在几个工作人员中 分配一下 直接 倒到喉咙里去 …
附加式 vp	187 (7.44%)	vp → !vp ule vp → !v uzhe …	(把 x) 摔坏 了 珍藏 着
连谓式 vp	58 (2.31%)	vp → !vp vp vp → !vp wco vp …	(把 x) 带回家 放好 变成电信号，再加以放大 …
其他	9 (0.36%)	vp → !v vp → c !vp …	(把 x) 尽收眼底 一 剥 …
合计	2514 (100%)	48 种	

表 3 反映了“把”字结构的一个特点，即“把”后的 vp 以述宾式构造为最多，这个特点以往语法文献中在描述“把”字句特点时重视不够。以往在提到“把”后 vp 特点时往往是指出 vp 必须是复杂形式。这当然是对的，不过，还应进一步指出各种复杂形式所占比例的多少。表 3 的统计结果更详细地说明了“把”字结构中 vp 的不同类型及频次高低情况。

在树库中还可以进一步考察“把”字结构整体所处的分布环境有什么特点。TreeEditor 程序对分布环境的定义是，一个结构体所处的树结构，即由该结构体的父节点，左邻节点，右邻节点三个项目来描述其分布环境。图 5 显示了 TreeEditor 抽取短语分布环境的对话框。用户在该对话框中指定某种特定类型的短语，包括短语类型（如 vp），以及该短语的内部结构（比如由“pp+vp”构成的 vp，即“把”字结构 vp）。程序根据用户指定的条件，从树库中把符合条件的短语所在的外部树结构环境全部抽取出来，并统计不同环境类型的频次，结果如下面表 4 所示（表中示例粗体文字为“把”字结构 vp，非粗体文字为其外部环境）。

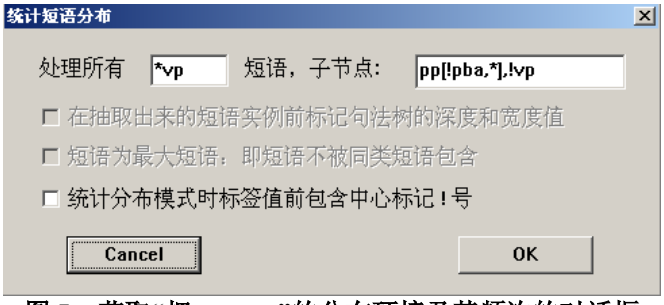


图 5：获取“把+xp+vp”的分布环境及其频次的对话框

表 4：“把+xp+vp”的分布环境的类型统计（按照父节点的类型不同分类）

父节点	左邻	右邻	数量	示例
Vp	dp vp vp wco - …	- - - vp …	1504 (59.86%)	连忙 **把它拾起来** 走过去 **把口琴还给锡海** 爬上树去，**把小鸟放回窝里** **把门打开** 放狗出去 …
Dj	np np wco …	- - …	676 (26.89%)	你 **把它吃了** 古代的埃及人和中国人，**把它用做药物** …
Fj	dj wco …	- …	237 (9.43%)	他一只手抓住绳子，**把另一只手伸给水中的孩子。** …
Zj	- …	wfs …	36 (1.43%)	**把瓶子放在桌上 。** …
np	- …	ude1 …	29 (1.15%)	**把咖啡喝光 的** …
#	-	-	20 (0.80%)	**把桌子拿出去**
Tp	- …	f …	7 (0.28%)	**把羊肉和羊骨粉碎 后** …
Pp	p …	- …	4 (0.16%)	从 **把水放在炉上 到水开** …
合计	82 种		2514 (100%)	

表 4 反映真实语料中“把”字结构的主要用法中，直接做谓语是排在第二位的。排在第一位的是“把”字结构跟其他成分组合成更大的 vp（占到近 60%）。也可以说，现实中的“把”字结构，其前后往往会有其他的谓词性成分共现。这也是把字句教学中应注意的一个特点。

2.3 汉语中不符合中心扩展规约的短语结构考察

通常情况下，一个短语结构的功能类跟其中心成分的功能类是相同的，这样的短语规则是所谓的符合中心扩展规则的组合（记作 HE 规则）。但实际语料中，也有少部分短语结构，整体功能类跟其中心成分的功能类是不同的，本文称之为非中心扩展规则（记作 NHE 规则）。下面是从北大树库中抽取的 HE 规则和 NHE 规则各自所占的比例情况。统计显示，绝大多数规则是符合中心扩展规约的（占 97.20%），只有少部分规则是 NHE 规则（占 2.80%）。

一般来说，HE 规则更符合逻辑上的整体与部分之间的一致性要求，NHE 则违反了这种一致性要求，因而在不同语言的对比中更有可能突显出差异。汉语中的这些 NHE 组合规则，其对应实例更应成为对外汉语教学中重点考虑的教学对象。通过树库的查询功能，可以尽可能多地从实际语料中发现 NHE 规则的类型和实例。为教学中选择重点对象提供线索。在从树库中自动抽取 NHE 结构时，本文定义的操作（判定）标准为：一个短语组合规则为 NHE 规则，当且仅当根节点的短语类与中心语的短语类属非同类范畴。比如“np → np 的 !vp”是 NHE 规则，因为根节点 np 跟中心语 vp 属非同类范畴。而“dj → np !vp”是 HE 规则，因为根节点 dj 和中心语 vp 都属谓词性结构，属广义同类范畴。下面表 6 展示了一些常见 NHE 结构的实例（示例中粗体文字为短语的中心成分）。

表 5：树库中 HE 规则与 NHE 规则频次及示例

规则类别	规则种数（type）	规则例数（token）	结构示例	示例
全部规则	1930	1,318,488	dj → np !ap	人 多
HE 规则	1672(86.63%)	1,048,669 (97.20%)	ap → dp !ap	最 冷
NHE 规则	258(13.37%)	30,252 (2.80%)	np → sp !vp	体内 分布

大体上 NHE 规则可以分为两类，一类是通过标记成分，如结构助词“的”“地”“似的”等，系统地改变结构的性质，使得结构整体的功能不同于其中中心成分的功能。比如上表中例 1、2 的情况。另一类是不需要标记成分的帮助，直接发生功能转换的情况，比如上表中例 3、4、5、6 都是这类情况。其中陈述性成分（如 ap、dj）、修饰性成分（如 qp）都直接转为指称性成分（如 np、sp）。

表 6：从树库中抽取的 NHE 规则及其示例

序号	内部构成/中心成分	NHE 规则	示例
	跟“的”相关的 NHE	np → np 的 !vp np → pp 的 !vp np → sp 的 !qp	唯一 的 消遣 在电子产品可靠性方面 的 应用 他们中 的 三个
	跟其他助词（似的、地）相关的 NHE	ap → !np 似的 ap → !dj 似的 dp → !qp 地 dp → !vp 地	一窝蜂 似的 雪片 似的 他是这个地方的主人 似的 一寸一寸 地 有秩序 地
	ap 扩展为 np	np → qp !ap np → vp !ap	一点 清凉 （联系） 教学 实际
	ap 扩展为 sp	sp → vp !ap	过桥 不远 （有一座寺庙）
	dj 扩展为 np	np → qp !dj	这种 再狭窄发生率降低
	qp 扩展为 np	np → np !qp	这 三本

3. 基于 Web 树库的汉语句型辅助教学原型系统

上面第 2 节介绍的树库应用主要是面向教师的，通过树库的丰富查询功能获得的有关汉语句法结构的知识，可以直接作为教学材料，也可以在设计教案时参考。此外，基于树库数据，还可以进一步设计和开发帮助学生自学汉语句型的计算机辅助教学系统。

在本文设计的汉语句型辅助教学系统[1]中，用户可以通过网页浏览器，对系统给出的句子进行结构标注，然后由系统比对标准答案，给出一个评分反馈。这里假设了语言学习者对句型的学习效果可以通过他分析句子结构的好坏反映出来。教师也可以通过这个系统评估学生的知识掌握情况，从而更有针对性的组织教学内容。系统的总体结构如下页图 6 所示。

该系统除将 TreeEditor 程序的树库检索和编辑功能移植过来之外，还在已有树库基础上，将北京语言学院句型研究小组归纳的汉语句型系统及其例句的句法结构树一起加入到数据库中，这样可以方便教师在汉语句型系统的整体框架中根据需要选择适当的句型作为教学对象。此外，系统还设计了一个主要的功能模块：通过树结构编辑距离计算，对用户的句法结构标注结果跟库中标准答案进行比对，给用户的标注结果进行自动评分。这个功能可以帮助汉语学习者进行句型练习，帮助教师收集学生容易犯的错误并进行分析，从而改善句型教学的教学方案，提高教学效果。本节通过一个例子演示系统对用户手工标注句子的结构进行自动评分的功能。

例句：你应该清醒、冷静。

该例句在“北京语言学院句型系统中所属句型大类为“动词谓语句”，次类为 “主 || 能愿动词 + 宾语”

[1] 网址：http://ccl.pku.edu.cn:8080/WebTreebank 程序代码的编写工作主要是由北京大学软件与微电子学院 2006 级硕士研究生谭大伟、王更生、邓结慧、吴桐等几位完成的。

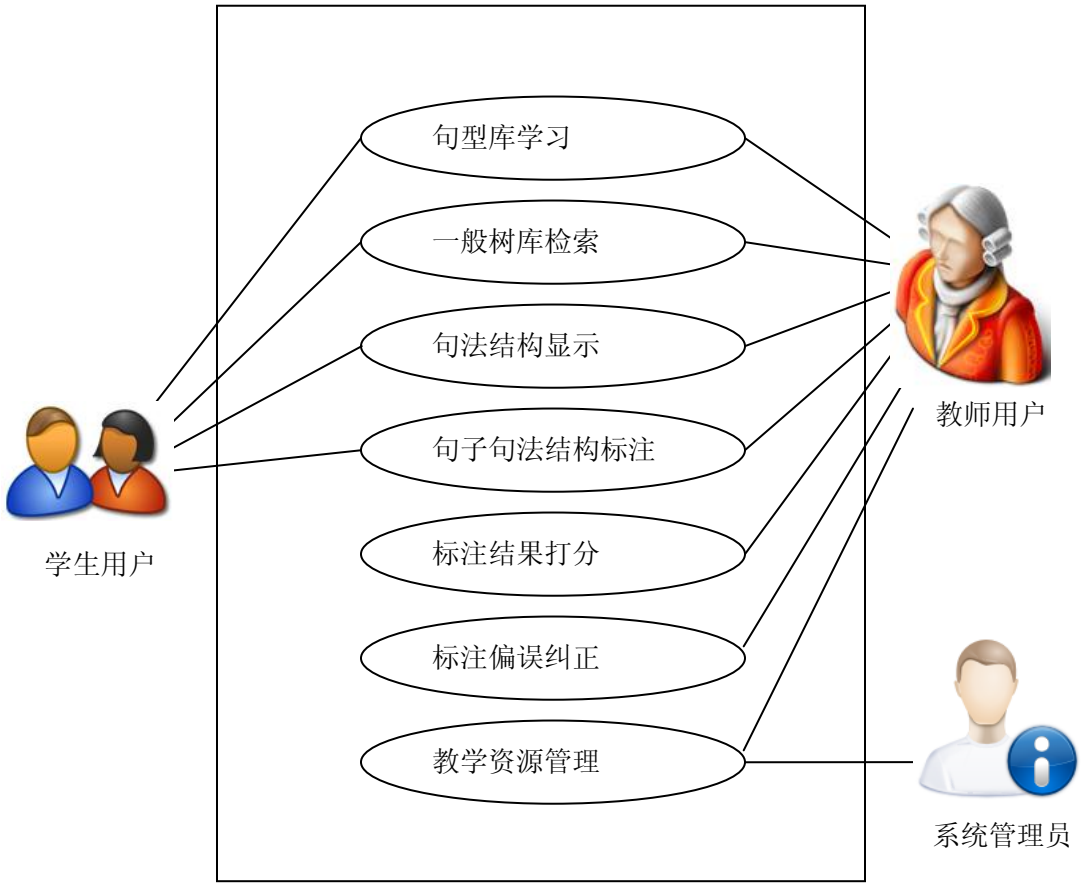


图 6：基于 Web 的汉语句型辅助教学系统

学生可以通过系统的句型树状分类指引定位到该例句。在练习开始时，网页上呈现出原始句子。学生练习的顺序则是：把句子切分为若干个词语（即分词处理），对每个词语，标注其词类，划分出句中合法的短语（直接成分），并为每个短语标注其功能类别，最终得到一个句子的完整的短语结构树。在整个过程中，系统后台的编辑距离程序一直在实时计算学生标注的中间结果跟数据库中存放的标准答案之间的差距，并给出评分。通过下面截图显示的程序界面，可以对这个句型辅助学习系统的工作方式有一个直观的了解。

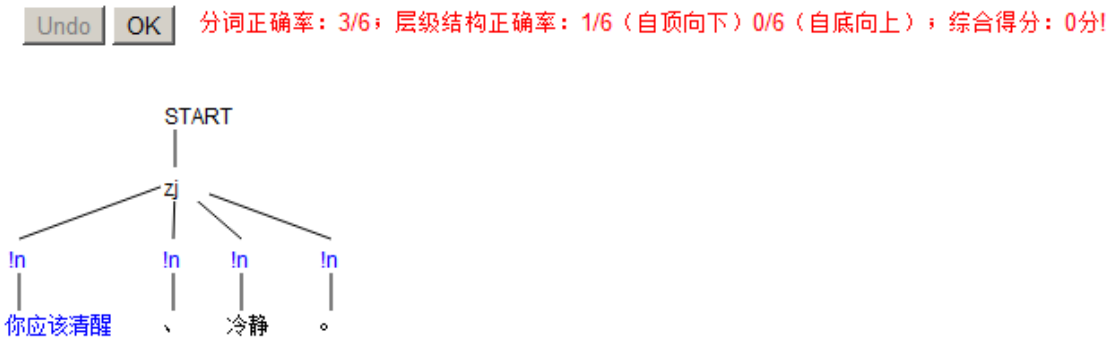


图 7-1 用户点击“练习”后的初始页面

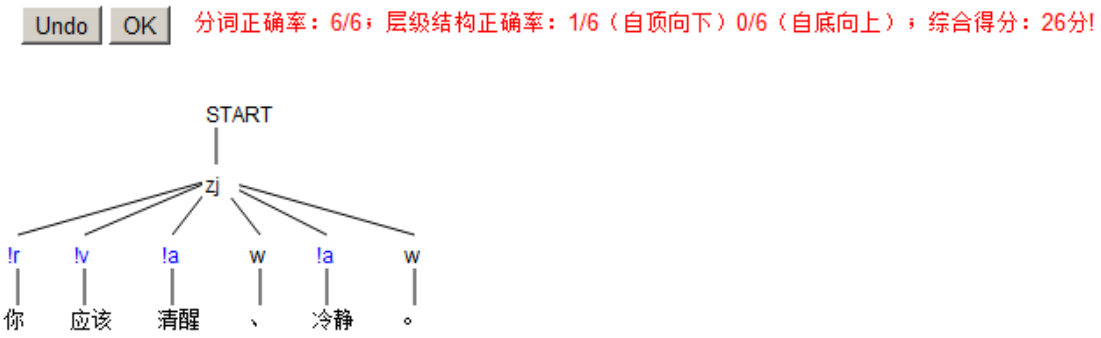


图 7-2 例句中的词语得到正确切分和词类标注后的截图

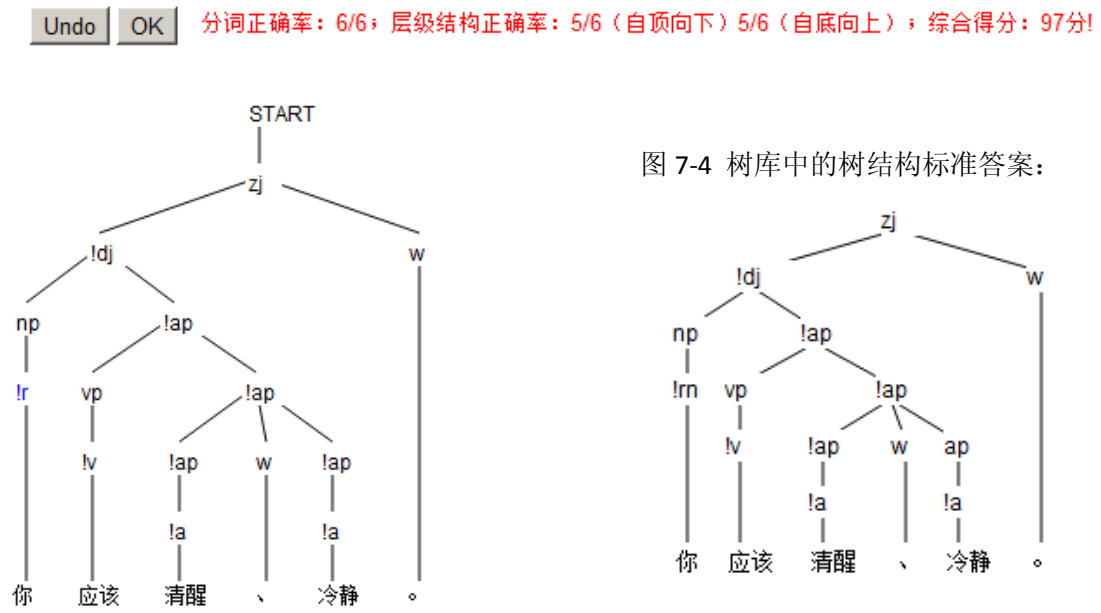


图 7-3 整个树结构分析基本正确之后的截图：

图 7-4 树库中的树结构标准答案：

上面整个操作过程基本都可以通过鼠标点击选择完成。在标注过程中，系统会将不正确的树节点以蓝色高亮显示，以提醒用户进行修改。比如图 7-3 跟图 7-4 标准答案相比还有细微的差异，用户练习中“你”标记为“r”（代词），跟标准答案中“你”标记为“rn”（人称代词）还不完全一样，因此，用户练习中的“r”以蓝色高亮显示。此外，练习过程中，可以通过点击“undo”按钮撤销当前操作。操作完成后点击“ok”按钮保存结果。

需要说明的是，该系统目前只做过少量内部测试，还有待在真实教学过程中做一定规模的并发请求环境下的测试，以观察系统运行效率如何。并在用户界面的友好性方面、系统的响应时间等方面做更多改进。

4. 结语

树库标注是一项成本较高的语料库加工工作。由于标注任务的复杂性，树库内部标注一致性问题也比其他浅层加工的语料库更为突出，北京大学中文树库是一个

仍在继续发展的项目。树库标注质量还在进一步提高，此外，TreeEditor 软件工具和 WebTreebank 系统的功能，以及界面友好性，操作便利性等问题都还在进一步完善当中。

本文从树库语料出发，提出了将树库中获取的语言知识用于语言教学的可能性，同时在树库资源基础上，设计了一个网页程序，可以通过互联网访问树库数据，在网页上进行句子分词和结构标注的操作，来模拟人对句子的理解过程。这种方式对于汉语的句型教学是否有效，还有待跟从事对外汉语教学工作的老师合作，在真实的课堂环境中开展实验，加以验证。

参考文献

Abeillé, A. (Ed.). (2003). *Treebanks: Building and using parsed corpora* (Text, Speech and Language Technology Volume 20). Dordrecht, Netherlands: Kluwer Academic Publishers.

Beatty, K. (2003). *Teaching and researching computer-assisted language learning*. Essex, England: Pearson Education.

LaPolla, R., J. (1995). Pragmatic relations and word order in Chinese. In P. Downing & M. Noonam (Eds.), *Word Order in Discourse*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Schmitt, N. (Ed.). (2007). *An Introduction to Applied Linguistics*. London: Edward Arnold Publishers.

Xue, N., Xia, F., Chiou, F. D., & Palmer, M. (2005). The Penn Chinese treebank: Phrase structure annotation of a large corpus. *Natural Language Engineering*, *11*(2), 207-238.

北京语言学院句型研究小组 (1989)，现代汉语基本句型，《世界汉语教学》, 1989 年第 1-4 期。

范继淹 (1985)，无定 NP 主语句，《中国语文》1985 年第 5 期。

陆丙甫 (2006)，不同学派的“核心”概念比较，《当代语言学》2006 年第 4 期，289-310 页。

沈家煊 (2007)，汉语里的名词和动词，《汉藏语学报》2007 年第 1 期，27-47 页。

沈家煊 (2009)，我只是接着向前跨了半步——再谈汉语的名词和动词，《语言学论丛》第 40 辑， 3-22 页

魏红、储泽祥 (2007)，“有定居后”与现实性的无定 NP 主语句，《世界汉语教学》2007 年第 3 期， 38-51 页。

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》，清华大学出版社。

詹卫东 (2011)，《从语言工程的角度看“中心扩展条件”与“并列条件”》，第三届两岸三地现代汉语句法语义研讨会，2011 年 8 月 12-14 日，北京，中国社会科学院语言研究所。

附录

树库短语标记（17 个）

短语功能类标记	内部结构	实例
zj　整句	暂缺	三点钟全体集会。 你怎么了？ 快跑！ 他走了……
yj　引句	暂缺	“我买的东西，为什么让给你？”小芬不满地说。
ypc 语篇插入成分	暂缺	平均每百户居民拥有彩电 86 台（比 1990 年增加 27 台）
yph 语篇呼语成分	暂缺	喂，喂，你是谁？
fj　复句	暂缺	只要他在，你就过不去； 风也停了，雨也住了
dj　单句	主谓 状中	爱夸张事实的孩子往往喜欢喜剧；三点钟全体集合；今天星期一；他二十来岁；长两米；重三斤；
np 名词性短语 npr（指人 np） npz（机构 np）	定中结构 “的”字结构 并列结构 同位结构	粒子碰撞噪声检测仪；计算机在国外应用的现状；世界名牌服装；新问题；自己的； 桌椅门窗；理想与现实；支持总统的群众；给孩子们的；服装设计；两国之间的合作；几十年的努力；他们两位； 录象带两百盘；最善良的一个；三斤重；两米宽；
vp 动词性短语	述宾结构 述补结构 连谓结构 兼语结构 并列结构 状中结构	把杂志放进抽屉里；进行多方面的经济结构的调整；从暴风雪中救出了一群羊；来了；请客人吃饭；去外婆家玩；烧毁证物并袭击警察；跑得我累死了；
ap 形容词性短语	状中结构 并列结构 “的”字结构 述宾结构	很不高兴；冷得发抖；比他们房间冷得多；干干净净的；通红通红的；亮了；干净不了三天；不礼貌而且不诚实； 长三米； 小两岁；
dp 副词性短语	并列结构 “地”字结构	飞快地； 轻松而愉快地； 波浪式地；
pp 介词性短语	并列结构 介宾结构	关于专家系统； 从桌子上； 被我们； 在后面； 比这里； 从北京到那里； 除他之外；
sp 处所短语	并列结构 定中结构	报纸上； 我前面； 我们班里；
tp 时间短语	并列结构 定中结构	一个秋天的早晨； 下星期一； 吃饭前；
qp 数量短语	并列结构 定中结构	两百张； 三十岁； 三场； 多少斤；
mp 数词短语	并列结构 定中结构	六七百； 三万两千零五十； 四又二分之一； 五点三二； 大多数； 不少； 几；

词类标记（上位标记 26 个，下位标记 69 个，合计 95 个）

词类标记	含义	下位分类	词类标记	含义	下位分类
a	形容词	ad 形用作状 an 形用作名	N	名词	nr 人名 ns 处所专名 nt 机构专名 nx 非中文字符串 nz 其他专名
b	区别词		O	拟声词	
c	连词	ch 前句连接词 ck 后句连接词	P	介词	pba 把 pbei 被
d	副词		Q	量词	
e	叹词		R	代词	rb 区别性代词 rd 副词性代词 rm 数词性代词 rn 名词性代词 rs 处所词性代词 rt 时间词性代词 rv 动词性代词
f	方位词		S	处所词	
g	语素	ag 形容词性语素 bg 区别词性语素 dg 副词性语素 ng 名词性语素 sg 处所词性语素 tg 时间词性语素 vg 动词性语素	T	时间词	
h	前缀		U	助词	udh 的话 uetc 等，等等 uguo 过 ule 了 uzhe 着 ude1 的 ude2 地 ude3 得 usd 似的
i	成语	ia 形容词性成语 in 名词性成语 iv 动词性成语	V	动词	vd 动用作状 vn 动用作名
j	缩略语	ja, jn, jv（参考 i）	W	标点	22 个下位标点标记
k	后缀		X	非语素字	
l	习用语	la, ln, lv（参考 i）	Y	语气词	yde 的 yle 了
m	数词		Z	状态词	