

# 基于文本信息的韵律结构预测及其在合成系统中的应用

曹剑芬 陈方忻\*

中国社会科学院语言研究所 \*IBM 中国研究中心  
jianfencao@hotmail.com \*chenFX@cn.ibm.com

## 摘要

本文报告一种基于语法信息的韵律结构预测方法以及这种方法在一个文-语转换合成系统中的应用。这种方法的特点在于，能够全面地预测韵律边界的位置分布及其等级层次以及语句重音的位置分布及其等级安排。合成实验表明，这种方法能够有效地改善合成语音的自然度和表现力。

## 1 引言

随着言语处理的不断深入，人们不但越来越需要了解自然话语的韵律结构，而且力图寻找通过文本信息预测韵律结构的方法，以便更加有效地提高合成语音的自然度和语音识别的准确率，同时加深自然语言理解的深度。尤其是文-语转换，如果要实行韵律合成，首先需要在前端为输入的文本确定韵律结构。目前，文-语转换的前端可利用的一般只有来自文本分析的信息，而缺乏韵律的信息。因此，怎样在文本信息的基础上来预测可能的韵律结构，已经成为一个探索的热点。

广义地说，韵律结构应当包括重音、节奏和语调三个方面的结构，例如重音的位置分布及其等级差异，韵律边界的位置分布及其等级差异，语调的基本骨架及其跟声调、节奏和重音的关系，等等。

狭义地说，韵律结构主要指话语节奏的层级组织及其客观标志[4]，包括韵律词的构成以及各级韵律成分边界的界定等等，通常叫做韵律切分。它涉及说话时的组词断句模式，实质上是指语言信息时域分布(或者叫计时 timing)的格局。

本文所说的是广义的韵律结构。

在自然话语中，韵律结构与语法结构往往是不一致的，因此，我们不能完全用语法结构来代替韵律结构。但是，来自自然话语分析的种种迹象又表明，话语的韵律结构还是以句法结构为基础的。尤其是无标记的 (default) 重音分布和韵律边界强度，往往跟句法结构以及句法成分之间关系的松紧程度具有一定的关系。所以，可以利用句法结构关系，再结合其它相关的信息，来预测一般的韵律结构。目前，在这方面已有各种各样的探索，并提出了若干比较有效的方法[1][3]。不过，这方面还有潜力可挖，也有必要寻找更为全面的预测方法。

从表面上看、特别是从书面形式上看，语言信息呈

现线性排列，在时间上先后依次出现；但是，从话语生成的本质上看，语言信息在时间域并不是线性地平均分配，而是以非线性的、层级(hierarchical)的形式分布。每个语句看上去都是一个个词的平等串列，而实质上各个词的轻重地位不同，跟信息时域分布的等级和层次形成有机的统一。因此，预测韵律结构的重点，就是要预测话语的这种组织形式，关键就是预测韵律边界(停延 break)的位置分布及其等级层次以及语句重音的位置分布及其等级安排。

## 2 基于语法信息的韵律结构预测

### 2.1 理论依据

在自然话语中，尽管韵律结构与语法结构不完全一致，但是，语言毕竟是个有机的整体，韵律与语法之间还是具有密切的关系的[2]。在某种意义上甚至可以说，韵律的表达是建立在语法的基础之上的。只是由于说话风格、熟练程度以及其它某些变化因素，同样一个语句，在基本的语法和语义不变的情况下，韵律划分的边界位置和等级也可能会不同的。这就说明，对于同一个语句而言，韵律上的表达可能具有不止一种形式，也就是说，它的韵律结构并不是唯一的。因此，我们不能期望寻找到韵律与句法之间整齐划一的映射关系。但是，也不能因噎废食，而放着有用的信息不去利用。

来自汉语自然言语分析的种种迹象表明，对于韵律结构的预测来说，语法方面还是存在着许多可资利用的信息的。例如，我们的相关研究[5, 6]发现，韵律词的构成与词性特征存在一定的关联，跟韵律边界相应的停延出现与词性和句法结构特点之间存在一定的关系，重音的位置分布与句法结构之间也存在一定的常规关系(常规重音分布规律)。本文对于汉语基本韵律结构的预测-韵律边界(停延)位置和等级的预测以及重音位置和等级的预测--正是以这些信息为基础的。

### 2.2 来自自然言语分析的相关信息--韵律结构与句法结构及词性的关系

#### 2.2.1 韵律边界的位置分布跟句法结构的关系

话语的韵律切分问题，既涉及听话时听觉上的停延分布，又涉及说话时语音上的组块分合。从对自然话语的分析结果来看，韵律结构层次跟停延的等级层次大体一致。譬如说，句子或者较大的短语边界，必定具有明显的无声停顿，通常在书面上都有标点符号表示，这就

不必多说。而在各个语句的内部,同样也存在着体现韵律边界的停延现象。通常,最大的停延多半出现在语调短语这个层级上,以无声停顿为主;最小的停延主要出现在韵律词层面上,主要表现为边界前音节的延长,一般没有无声停顿。因此,寻找韵律切分的规律首先就是寻找停延分布的规律。我们对朗读语料库的停延分布作过一个粗略的听辨试验,考察结果[5]发现,虽然实验数据比较离散,但是,相对说来,听感上的停延级别的大小,还是跟停延两边的语音单元之间句法关系的松紧程度具有一定的关系。因此,各级停延的分布显示出一定的倾向性,可以用表1(由于篇幅所限,此处表1从略)的数据大致说明。

需要说明的是,就较长的句子内部而言,普通人能够辨别的话语停延一般可能有三级,经过相关训练的人或者通过知觉实验,才能辨别那个细微的第四级停延。至于较短句子的内部,停延的级别没有那么多。不过,那个第四级停延虽然细微,而且在声学上一般只表现为相关音节的延长,而不存在无声间歇,但是,对听觉感知却相当重要,在语音处理上是不可忽视的。

### 2.2.2 停延分布与词性的关系

根据不完整的观察分析,在语句内部,凡是以下情况一般都会出现大小不同的停延,条件是后随多音节结构,或者连接两个复杂的成分:

- (1) 人名、尤其是外国人名之后和地名之后;
- (2) 时间名词和数量词之后;
- (3) 多音节动词之后;
- (4) 方位词(如:前、后、上、下)和趋向词(如:来、去、进、出)之后;
- (5) 并列连词(如:和、及、跟、与、同、以及)之前和/或之后,而且,连词之前的停延常常大于连词之后的;
- (6) 转折连词(如:不但...而且、虽然...但是、由于...因此(因而)、因为...所以、如果...那就、既然...也就、与其...不如)之前和/或之后;
- (7) 助词(如:的、地、得、着、了、过)之后;
- (8) 双音节副词(如:往往、常常、经常;十分、非常;从前、后来、目前、已经;仅仅、多半)之后。

### 2.2.3 自然话语中韵律词的构成规律

通常,双音节或三音节的语法词直接构成韵律词。可是,在实际语言里,还存在着许多单音节的语法词,它们在口语中究竟怎样组合,是确定韵律切分的关键。通过对一个经过韵律词标注的语料库的粗略考察,我们发现[6],在实际的话语中,为了满足韵律上的需要,单音节的语法词都会或者通过延长该音节,或者前附或者后附到另一个标准音步上,或者跟前后其它的单音节词组合,设法凑足一个标准音步或超音步,从而构成一个韵律词。主要原则是:

第一,单音节词或者因为处于重读地位,或者因为左邻右舍已经是完整的音步而无所依附时,就通过本身延长为一个音步,从而构成独立的韵律词;

第二,单音节词处于非重读地位,如遇前后是单音节词或双音节词时,就彼此粘附,构成一个韵律词。

同时,韵律词的组成,还要受各自词性的制约。由于篇幅所限,此处只能略举一、二例说明(详见图1和图2)。

## 2.2.4 语句重音分布跟句法结构的关系

### 2.2.4.1 语句重音的类型和等级

语句重音的分类因具体标准而不同,一般说来,可以大致归纳为语法重音和逻辑重音两个大类。语法重音反映话语的自然节奏,实际上是语句重音最基本的无标记(default)形式。一般说来,它在句子中的位置分布相对稳定,所以,也叫常规重音。逻辑重音是说话者为了强调语句中的某个语言信息而主观安排的重音,它没有固定的出现位置,具体由言语交际的临时需要而定,随着说话者想要表达的语义重心的转移而变化。

关于语句重音的等级,很难精确地说究竟有多少个。因为从理论上说,在自然话语里,语音的轻重差异本质上是一种连续的对比。不过,通常人的耳朵能够感觉到的等级差异还是很有限的,而且,相对说来,人们对比较显著的重轻对比的知觉还是大体一致的。根据一般人群的测听试验结果来看,语句中的轻重大概可以分为特别重的、稍重的和不重(或中的和轻的)的几个基本级别。

### 2.2.4.2 常规重音分布规律

常规重音反映语法结构特点,在语句中出现的位置大致有定。它的分布大致遵循如下的规律[7]:

- (1) 在主谓结构中,语法重音落在谓语部分。
- (2) 在偏正结构中,语法重音常落在修饰名词的定语上或修饰动词或形容词的状语上。
- (3) 在述补结构(由形容词、动词及部分词组充当)中,补语常常念重音。
- (4) 指示代词、疑问代词作主语时,通常带有语法重音。
- (5) 在述宾结构中,宾语常常念语法重音。

此外,常规重音出现的位置往往跟一定的句法成分相联系。例如,下列成分常常重读:

- (1) “是”所强调的成分。
- (2) “连/就/才”所强调的成分。
- (3) 数量成分。
- (4) 否定句中被否定的成分。
- (5) 特指疑问句中的疑问词。

### 2.2.5 语调的基本骨架

从对自然话语的分析来看,汉语语调的基本骨架有两个关键点:第一,语句的总体音高下倾走势及其层级

结构；第二，声调和语调调阶的并存叠加。由于篇幅有限，具体请参阅曹剑芬[8]的相关论述。

## 2.3 汉语韵律结构的预测

这里所说预测的主要任务，就是利用从文本分析获得的句法信息和词性信息，根据已经认识的有关韵律与句法之间的某种相关性，来预测合成话语的总体韵律结构框架，包括停延的位置分布和等级确定，以及重音的位置分布和等级确定。

这种预测的实质，就是首先给每个语法（词）边界确立韵律地位，然后给每个韵律词确立重音地位。至于语调结构，只要韵律边界及其等级一旦确定，它也就迎刃而解了。

### 2.3.1 停延(韵律边界)的位置分布及其等级的预测

这部分预测的核心问题是，如何对文本正确地、自动地实行短语切分。也就是说，通过恰当的停延设置，获得正确的韵律边界及其等级，实质上就是确定哪些词边界只是韵律词的边界，哪些则是韵律短语的边界。

#### 2.3.1.1 基本步骤

- (1) 对目标语句进行分词和词性标注；
- (2) 利用分词和词性信息，实行韵律构词；
- (3) 对目标语句进行句法切分和句法标注；
- (4) 构建韵律结构预测树；
- (5) 确定韵律边界位置，计算其韵律距离的大小；
- (6) 输出语句的韵律组块分合及其层级结构。

#### 2.3.1.2 预测树的构建方法：

这里的预测树(详见图3)是一种综合利用句法、词性和韵律信息构建出来的。它的上半部是个倒置的二叉树(binary tree)，它是在语法分词和韵律构词的基础上，根据通常韵律上偶分的倾向，首先把句法上距离最近的韵律词两两连接为节点，然后再依此法建立更高层次的节点。因此，其中的每个节点就代表各自所包含的两个下位树杈(即韵律词或韵律短语)之间的一个韵律停延。根据这种二叉树结构，就可以判断这个句子的韵律层级：越是接近根部的节点，所代表的停延就越明显，也就是韵律边界越强。预测树的下半部代表通过句法切分获得的句法结构，这里采用直接成分分析法，把整个句子逐层划分成短语和词，并且加上了句法和词性标注，以供韵律构词和构建预测树参考。

### 2.3.2 语句重音位置和等级的预测

语句重音位置和等级预测的基本步骤是：

- (1) 利用2.3.1的多重二叉预测树，根据韵律组块内部两个分叉之间的句法关系，按照常规重音分布规律，给每个二叉韵律成分赋予相对的重轻(S/W)地位；
- (2) 计算每个韵律词最后负载的重度指标，确定每个韵律词的重度；
- (3) 根据各个韵律词的重度，输出语句中重音分布的层

级结构。

### 2.3.3 语调结构的预测

基本步骤：

- 1) 利用2.3.1.1输出的语句韵律组块分合确定语调单元；
- 2) 根据文本信息，标注各个语调单元的语语气类型；
- 3) 根据语语气类型，预测各个语调单元的基本调阶走势。

### 2.3.4 韵律预测小结

初看起来，这里预测出来的结构似乎就是句法的结构，但事实上不一样。因为我们首先进行了韵律构词，剔除了不是韵律边界的那些词边界。这就等于首先确定了全部的韵律边界位置，剩下的任务就只要确定这些边界的等级，从而形成一定的韵律层级结构。由此可见，韵律构词是预测的关键，它保证了预测过程既建立在句法结构的基础上，又不完全受句法的限制。这是充分考虑了韵律与句法的实际关系：(1) 韵律上的边界基本上都是句法的边界(局部韵律词边界除外)；但句法上的边界不一定都成为韵律的边界。(2) 韵律结构与句法结构的不一致，主要是它们在层级结构上的不一致。

这里预测出来的结构是一种无标记的韵律结构，它反映了平常的自然话语的一种韵律结构。根据这种结构生成的合成语音的韵律必定是自然的，但并不是唯一的。不同的人说同一句话，采用的韵律结构可能略有不同，但都是可以接受的。此外，人在说话时，还常常需要采用特别的强调或对比等手段，来突出所要传达的信息。在那种情况下，就可能要改变平常的、无标记的韵律结构，而采用不同标记的韵律结构，才能满足不同的特殊表达的需要。

## 3 韵律结构预测在 TTSSPY 合成系统中的应用

### 3.1 输入文本韵律结构的自动预测

根据本文提出的韵律结构预测理论，我们在语音合成系统平台上归纳和实现了文本分析中韵律结构自动切分的基本规则。并在此韵律结构基础上，确定韵律词、韵律短语及语调短语的声调和语调模式。

#### 3.1.1 韵律词切分和重音模式标记

语音合成系统的文本分析首先是自动切分韵律词。根据自然言语的韵律词构成规律，文本分析将输入文本中的语法词、尤其是那些单音节词自动组合成韵律词，基本上由1-3音节组成。合成系统为此制定了具体实现韵律构词的规则，例如，行为动词后跟动态助词或趋向动词等组成韵律词；程度副词与形容词组成韵律词；名词与类别词组成韵律词；数词与量词组成韵律词，等等。这样形成的韵律词串列，就为输入文本确定了基本的韵律边界。

完成韵律词切分以后，就根据各个韵律词的构词特

点, 确定韵律词重音模式, 除了标出轻声词以外, 同时确定普通型词内前后音节的相对重音度。合成系统参考词性特点, 为韵律词内的每个音节标记相对重音度等级, 共分 0-3 四级, 由 0 至 3 级渐增。例如, 由单音节行为动词和趋向动词组成的韵律词, 行为动词音节的重音度为 3 级, 趋向性动词音节的重音度为 1 级; 而由单音节行为动词与动态助词组成的韵律词, 行为动词音节的重音度为 3, 动态助词音节重音度为 0, 等等。系统根据每个音节的重音度调节该音节的能量、音高变化和时长的设置。具体实施参数不属本文讨论范围。

### 3.1.2 建立韵律短语层级和语句重音等级

在韵律词边界切分的基础上, 文本分析进一步进行韵律短语切分。根据基于韵律构词和句法结构预测出来的韵律边界指标, 同时参考韵律短语长度限制[4], 实施不同层级的韵律短语切分。实际上就是参照预测的韵律边界指标, 在各个韵律词边界上设置相应的停延等级, 从而把一个个语句内的韵律词串列, 组合为不同层级的韵律短语。

接着, 根据合成系统中的韵律短语的语法结构类型, 例如定中结构、状中结构、述宾结构、述补结构、联合结构等等, 按照常规重音分布规则, 为语句中各韵律词预测相应的重音等级。合成系统将韵律词的重音度分为 0-4 五级, 由 0 至 4 级渐增。系统根据韵律词的重音度调整词的能量、音高变化和时长伸缩。无标记的词重音度为 2 级。在韵律短语中, 韵律词重音度随其所处的语法位置变化。例如, 在定中结构的韵律短语中, 充当中心词的韵律词重音将比修饰它的韵律词重音减弱一级, 而在述补结构中, 充当补语的韵律词重音比充当述语的韵律词重音增加一级, 等等。

### 3.1.3 建立韵律层级的语调结构

合成系统通过韵律短语切分, 再建立韵律层级的语调结构。语调结构包括语调宏观结构(macro prosody)和语调微观结构 (micro prosody)。语调宏观结构以语调短语 (Intonation Phrase) 为基本单位。宏观结构根据语调短语句型确定音高下倾曲线和音高下倾的重置, 并根据语气变化和语法重音确定语句的总体音高走势和音节时长。例如陈述语调在句尾下降速度加快, 一般疑问语调趋平或趋升, 而感叹句总体调阶提升, 总体音高走势可升可降; 语法重音或语义强调所处的音节, 声调调域扩大, 并且时长增加, 等等。语调的微观结构 (macro prosody) 则主要通过韵律短语来实现。韵律短语的音调依附于语调宏观结构。但也有其相对独立的音高起伏规则。韵律短语的音高模式主要通过其内部各个韵律词的重音分配得以实现。如前所述, 韵律词重音级别由其在韵律短语中所处的语法地位而定。而重音又跟不同程度的音高、时长和强度相关连。这样, 词重音的变化便自

然构成韵律短语的音调起伏。

## 3.2 合成检验

### 3.2.1 合成语音的改进

对本文提出的韵律结构预测理论的合成检验, 是在一个参数规则合成系统上进行的。非正式的听辨试验结果表明, 应用韵律预测的合成语音比没有应用韵律预测的合成语音, 在以下几个方面有了比较明显的改进:

- (1) 改进了词重音自然度;
- (2) 改进了停顿自然度, 在一定程度上避免了原合成系统出现在韵律词或韵律短语中的不适当停顿;
- (3) 为较长的合成语句增加了层次感, 从而改进了长句的理解度;
- (4) 合成话语的整句语调及轻重对比听起来更加接近自然话语。

### 3.2.2 存在问题

汉语韵律结构自动预测在很大程度上依赖于完善的汉语自动词法和句法分析。由于目前汉语自动词法和句法分析中存在的问题, 在韵律边界层级以及重音位置和等级的预测方面, 难免出现相关的错误。因此, 怎样完善汉语自动词法和句法分析系统, 将是一个亟待解决的相关问题。

## 参考文献

- [1] 应宏, 蔡莲红, “基于结构助词驱动的韵律短语界定的研究”, 《中文信息学报》, Vol. 13. No. 6, 1999.
- [2] Gee, J. P., and Grosjean, F., (1983). Performance structures: A psycholinguistic and linguistic appraisal, *Cognitive Psychology*, 15, 411-458.
- [3] Niu Zhengyu, Chai Peiqi (2000), Segmentation of Prosodic Phrase for Improving the Naturalness of Synthesized Chinese Speech, (In The Proceedings of ICSLP'2000, Vol. III).
- [4] 曹剑芬, “普通话节奏的声学语音学特性”(见 吕士楠等主编《现代语音学论文集》, 北京, 金城出版社, 1999 年)。
- [5] 曹剑芬, 2001, “汉语韵律切分的语音学和语言学线索”, 蔡莲红等编《新世纪的现代语音学》, 清华大学出版社。
- [6] Jianfen Cao & Weibin Zhu (2002), Syntactic and Lexical Constraint in Prosodic Segmentation and Grouping. (In The Proceedings of Speech Prosody 2002, Aix-en-Provence, France).
- [7] Zheng, B., Wang, B., Yang, Y., Lu, S. & Cao, J. (2000), The regular accent in Chinese sentences. (In The Proceedings of ICSLP'2000, Vol. I).
- [8] 曹剑芬, 2002, “汉语声调与语调的关系”, 《中国语

文》2002年第3期。

## THE PREDICTION OF PROSODIC STRUCTURE BASED ON TEXT INFORMATION AND THE APPLICATION OF THIS MODEL IN A TTS SYNTHESIS SYSTEM

Cao Jianfen \*Chen fangxin

Institute of Linguistics, CASS \*IBM China research centre

### ABSTRACT

This paper reports an automatic prediction method of prosodic structure based on text information and the application of this method in a TTS synthesis system. The main steps of this method includes (1) prosodic word formation based on grammatical word segmentation; (2) syntactic parsing and annotation; (3) prediction-tree building and (4) predicted index calculation both for prosodic boundary distribution and accent distribution. The synthesis inspection to this prediction theory was conducted on a parameters synthesizer by rule. The results of informal perception test to the synthesized speech show that comparing to that of without using this method, the naturalness of synthesized speech has improved in

following aspects: (1) the word stress sounds more natural; (2) the accuracy of prosodic segmentation is increased; (3) the hierarchical organization of longer sentence become more clear, thus the comprehension degree to long sentences is improved and (4) the global intonation and accent contrast of whole utterance sounds more close to that of natural speech. The advantage of this method is able to predict prosodic structure of an utterance as a package but not only a individual adjustment. The shortcoming of this method is too much dependent on syntactic analysis, while the automatic syntactic analysis is far not perfect up to date. Therefore, some error in the prediction of prosodic structure is hard to avoid.

前加:

后附:

单音节形容词/数词+名词+“的”、“里、外”(旧书的、九折的、新房外)

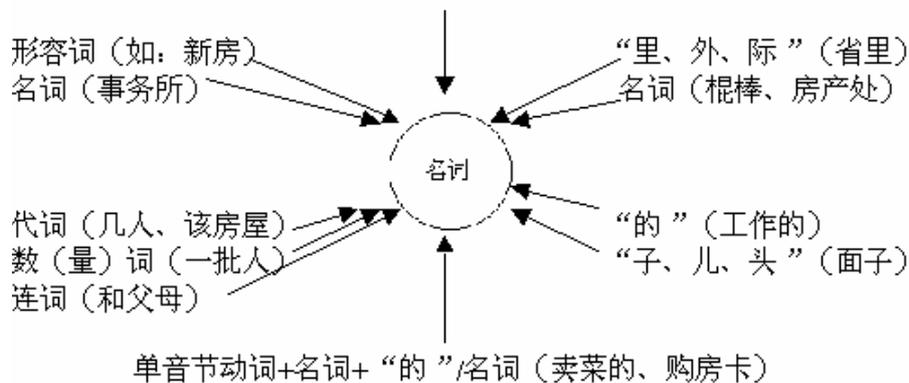


图1 以名词为核心的韵律词构词规律

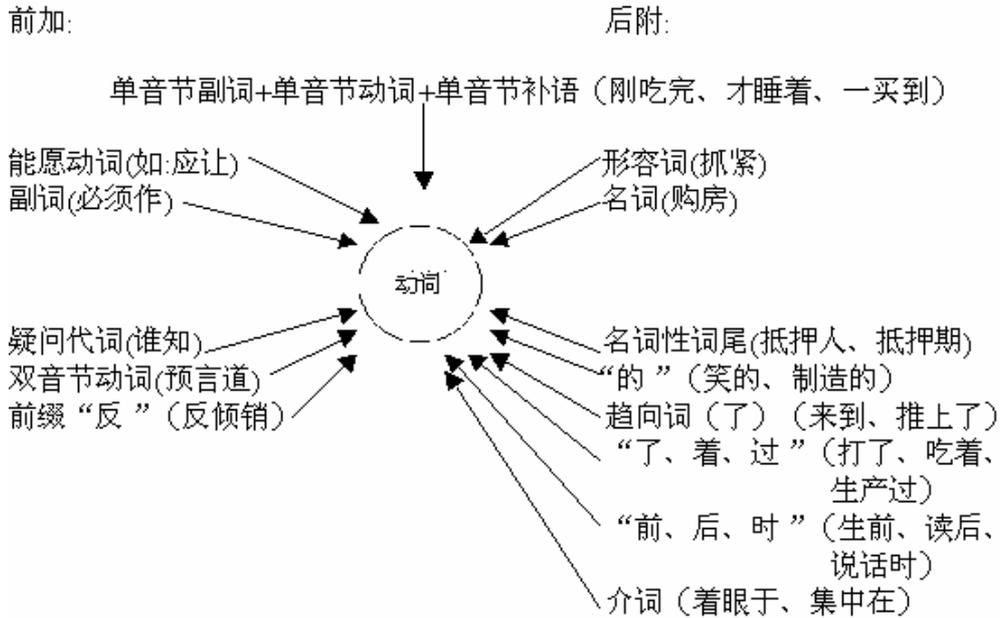


图2 以动词为核心的韵律词构词规律

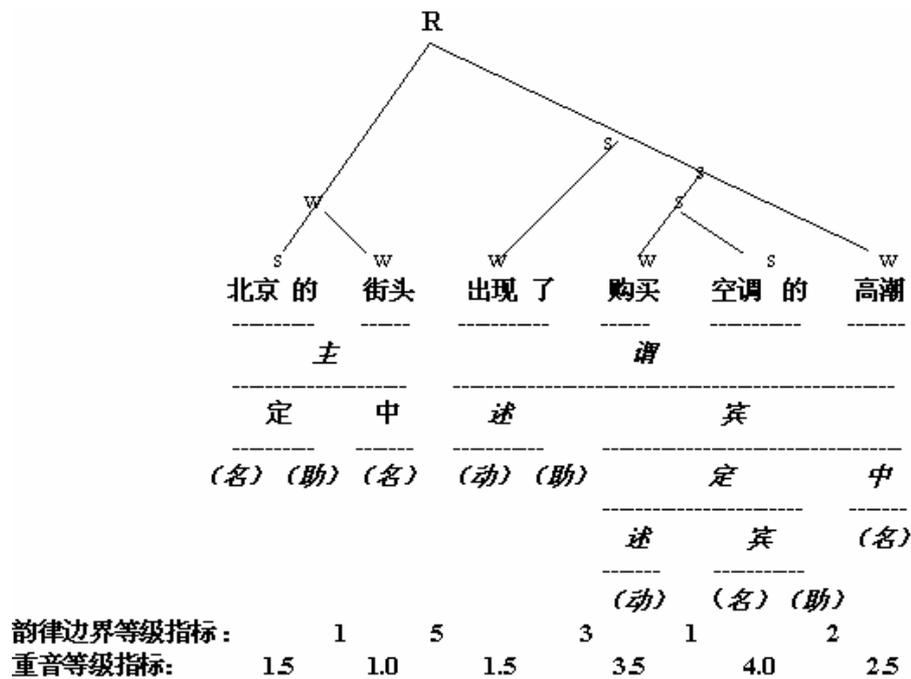


图3 无标记韵律边界和重音等级预测示例

补充：有标记情况下的韵律预测

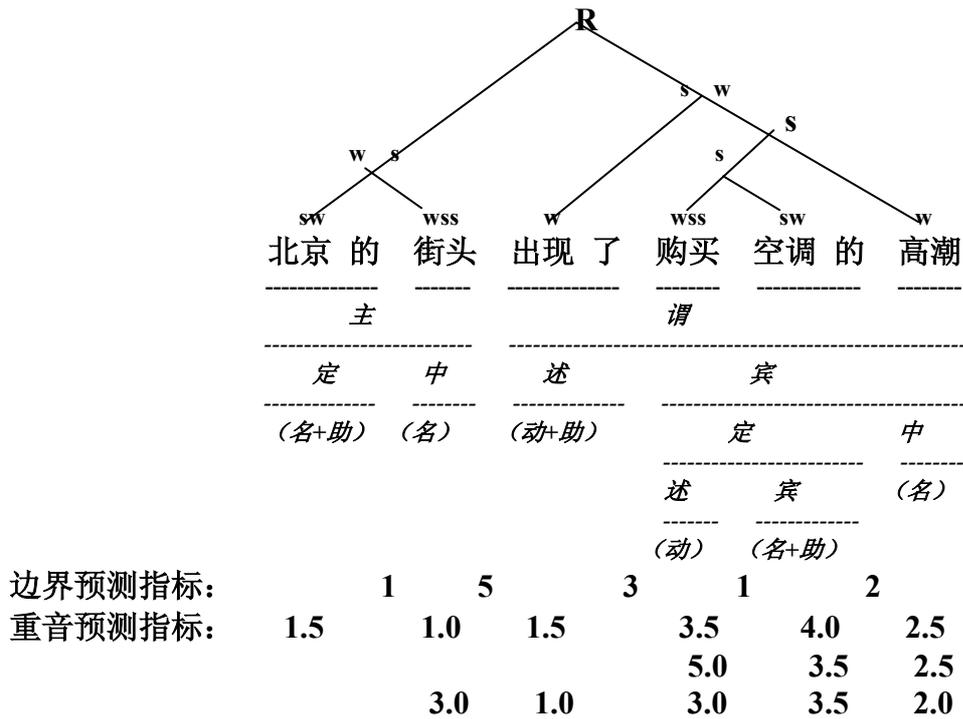


图4 有标记韵律边界和重音等级预测示例

根据图4，可以看出以下两点：

- 1) 当强调“购买”时，句中两个短语的重轻地位未变，只是改变了所在短语内部的重音分布，因而不影响和改变另一个短语内部的重音分布。
- 2) 当强调“街头”时，句中两个短语的重轻地位改变了，不但改变了所在短语内部的重音分布，同时也影响到另一个短语内部的重音分布，这种影响主要表现为重音程度的相对降低，而重轻分布结构不变，仍然跟无标记情况下大体一致。

以上情况表明，即使在有标记的情况下，也还是以无标记情况下的基本韵律结构（这里主要以重音结构为例）为基础的，只是由于标记成分的实际辖域不同，它对基本韵律结构的影响和改变也不同。但是，无论如何，有标记情况下的具体韵律结构同样是可以预测的。