

从语音合成看汉语的重音结构

曹剑芬

jianfencao@hotmail.com

中国社会科学院语言研究所

摘要*

面对言语工程和对外汉语教学等应用方面的迫切需要，语言学工作者越来越感到汉语的重音原来如此重要，而我们现有的认识却又如此不付需要。在同言语工程学界合作的过程中，逐渐对汉语的重音及其分布规则积累了一些新的认识。本文试图结合语音合成试验，简要论述汉语重音的层次结构及其与其它韵律特性的关系。实验结果表明，语句中的重音分布受节奏层次和语调结构的制约，也具有一定的层次和等级结构。而这种结构又跟语法层次存在一定的相关关系：首先，常规的重音等级取决于它所处的句法层次的高低；其次，在同一韵律单元内部，重音的分布又主要取决于该单元的语法结构特点。

1. 前言

重音是语言的一个重要的韵律特征。无论从语言学理论探索的角度看，还是从各种应用的角度看，重音问题向来都是倍受关注的。关于汉语的重音，研究的历史虽然很长，但是议论纷争，许多问题至今尚无定说。

关于汉语重音的主要争议可以概括为两个大的方面：第一，汉语有没有词重音？如果有，有几种类型？第二，汉语的语句重音有几类？语流中的重音到底有多少级别？怎样分布？

就汉语的词重音而言，基本的看法有两种：一种看法认为汉语没有词重音。这是比较早期的看法，现在已经很少有人坚持这种观点了。一种认为汉语有词重音，例如，汉语普通话里具有重音与轻音的对立，轻声就是一种与正常重音相对的弱重音。但是，对于轻声以外的一般所谓正常重音是否还有重与中的对立，则存在不同的看法。例如，殷作炎(1982)认为，重与中的对比同样可以区别意义。而林茂灿等(1990)根据听辨试验结果，认为重与中的对比并不真的具有区别意义的功能。

就汉语的语句重音而言，一是分类上的分歧，有的分类多达四、五类，有的只概括为两大类。不同的分类系统标准不一，彼此之间还存在一些交叉，令人无所适从。同时，关于语流中的重音级别及其分布也没有一致的看法，应用方面无法据以操作。

如果说上述种种对于汉语重音认识上的不足，对于一般人的语言交际还没有什么大碍的话，那么，对于外国人学习汉语或者对于方言区的人学习普通话来说，重音问题就不是一件小事了。尤其是近些年来，人机对话系统的飞速发展，更是迫切需要加深对汉语重音的认识。如今，计算机语音合成、语音识别以及自然语言理解等等，都已经进入处理连续的自然话语的阶段，人们发现，重音问题原来如此重要，而我们现有的认识却又如此不付需要。他们迫切希望语言学能够提供一套比较一致的、便于操作的重音规则来。此外，在近来的应用实践中还发现，还有个词重音跟语句重音的关系问题必须搞清楚。可是，这方面的探索几乎还是个空白。

面对言语工程、对外汉语教学等应用方面的迫切需要，语言学工作者常常感到我们的理论武库“库中羞涩”，非常

尴尬。因此，必须加强这方面的研究与探索，才能适应理论发展和社会应用的需要。

在同言语工程学界合作的过程中，尤其是参与一些语音合成系统的研制过程中，我们对汉语的重音及其分布规律积累了一些新的认识。在这里，试图根据对普通话自然语音的实验分析，结合语音合成试验[1]，对上述几个关键问题发表一点自己的看法。重点讨论以下两个问题：第一，进一步论证轻声以外的一般词重音还要不要再分类？第二，揭示自然连续话语中重音的等级结构及分布规律。

2. 词重音

汉语里存在词重音，这已经是个不争的事实。无论是对自然话语的分析，还是语音合成试验，都已经证明了这个事实。首先，轻声与非轻声的对立自不必说；其次，就是非轻声的普通型，即通常所谓的正常重音型，虽然是中重还是重中似乎不那么固定，但内部各个音节也存在着重度的差异。合成试验表明，如果只是简单地采用并重的两个或三个音节来合成韵律词，合成效果多半不好，往往是其中的某个音节不是听起来音高很不适当地突出，就是觉得它们的长短分布不合适，有的音节好像赶着发出来似的。假如通过音高和时长修正，适当地调节它们的轻重地位，就会马上变得自然了。这说明，普通型词内的音节之间也有轻重之别，只不过还没有真正认识这种轻重差异的实质，不清楚它究竟是发生在词层面上的还是语句层面上的，也没有找到这种差异出现的规律。因此，很有必要加以认真研究。

2.1 对自然话语词重音的观察分析

2.1.1 单念的词重音

许多人的观察发现，在单念的情况下，同一个词是念中重还是重中，存在个体差异。有的话者多半念中重，而有的话者多半念重中，没有一定的规律，也并不影响理解。从声学特性分析来看(曹剑芬, 1995)，重中与中重之间也不存在系统一致的本质区别。此外，相关的听辨试验也表明(林茂灿等, 1990)，重与中的对比并不真的具有区别意义的功能。

2.1.2 语句中的词重音

在语流中，一个词究竟念中重还是重中，更是随着具体环境而变。我们近来的研究发现，由于节奏边界上存在着韵律加强的效应，同一个词处于边界前和边界后以及中间位置上时，其内部的音节重度分配存在显著的差异。大致倾向是，念重中的多半出现在界后，少数出现在界前，中间位置上的也以重中的为主；而念中重的既有出现在界前的，也有出现在界后的，而且，界前的中重常常与并重分不清楚。总之，自然话语中实际情况比较复杂，我们正在做进一步的实验分析，具体结果将另文讨论。

上述语流中的这些差异说明，这类词重音并不像重音语言中的那样稳定，它们并不构成严格的对立，这种所谓重与中的差异实际上只是语境中的随机变体，而且可以相互转换，本质上已经属于语句层面的重音了。

2.1.3 自然话语中重音的声学表现

各种语言的研究表明，一般情况下，重音音节的音阶会相对抬高，而非重音音节的音阶会相对压低。普通话里的轻

* 本文是在纪念吕叔湘先生百年诞辰国际学术研讨会上的报告，2004年6月22-23日，北京。

声词就表现出这样的特点,其中轻声音节的音阶就比重读音节的音阶明显降低。照理,如果普通型词内的中重与重中真正也是严格对立的话,那么,就应该可以找到有规律的音阶差异模式,譬如说,中重型的应该是前者音阶略低而后者音阶略高,而重中型的就应该是前者音阶略高而后者音阶略低。可是,从这里对普通话自然语音的初步分析结果来看,尚未发现普通型词内音节之间存在与听感上的重度差异完全一致的音阶差异模式。当然,由于重音还涉及音节的时长效应,情况比较复杂,尚待进一步探讨。

2.2 计算机语音合成的试验结果

从合成语音的效果来看,单念的双音节词的重度分配多半以重中为宜,语流中的则更是如此,多半用重中更为合适,只有少数的以并重为宜,例如,“吃完”、“盛满”无论读哪种类型似乎都可以。可是,有的只能读并重或前重,如果过分后重,听起来就别扭了。譬如“攻击”和“公鸡”虽然这两个词的结构类型不同,但是词重音类型并不对立。合成试验表明,单念的“公鸡”和“攻击”采用重中型和并重型都可以,没有区别,但是,都不能采用太明显的后重型。

此外,在语流中,究竟采用中重型还是重中型,还要视环境而定。例如,在“挑起公鸡斗争”中的“公鸡”用重中型比较合适,而在“调整攻击目标”中的“攻击”则用并重型更为合适。表面上看来,这似乎是“公鸡”和“攻击”在词重音上的不同。实际上,进一步的实验[2]发现,在这里,是因为“公鸡”和“攻击”分别作为韵律词[3],它们在两个语句中的、整词的相对重度不同。根据句法结构特点,在“挑起公鸡斗争”中,常规的句重音落在“斗争”上,而不是在“公鸡”上;而在“调整攻击目标”中,句重音则落在“攻击”上。因此,这里的“公鸡”和“攻击”之间表现出的重中与并重的差异,并不是真正词重音类型的绝对对立,而是因为语句重音的要求而出现的相对轻重变化,是属于语句层面的、随语流环境而变化的轻重对比现象。

3. 语句重音

3.1 对自然语句重音的观察分析

在自然话语里,作为重音的主要声学实现手段,一般表现为相关词或音节的音阶突显。因此,短语内词间的重度差异一般表现为相关词间的音阶差异。但是,从对自然语句重音的观察分析发现,跟语音轻重相关的音阶实现是非常复杂的。例如,根据对某个话者所说的“罗尔斯的恢弘巨著‘正义论’也是从所谓无知之幕出发的”这一句话中“罗尔斯的恢弘巨著‘正义论’”这个韵律短语的测量,其中各个韵律词的音阶(自然音高水平)如图1下部的相关数据所示。

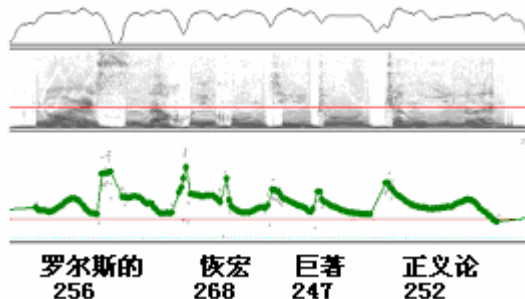


图1 语句中韵律词的音阶实现举例

初看起来,这些数据之间的差异似乎跟听感上的轻重差异不太一致。因为在听感上,这段话中的“恢弘”和“正义论”是最重的。可是“正义论”的音阶却不如“恢弘”的高,甚至不如相对较轻的“罗尔斯的”高。而且,四个话者表现出同样的分布规律。从表1的数据可以看出,这段话里各个韵律词四个话者的音阶平均值分别为 183.2、195.3、164.1 和 178.0,与图1下部的音阶分布趋势一致,可见图1里这个

话者所表现的并不是个别的偶然现象。

表2.不同话者所说的“探讨一下权力制约的制度问题”中“探”、“权”、“制”的音阶实现

音节 话者	探	权	制
F1	302.65	221.8	269.05
F2	228.05	153.7	192.3
M1	175.5	166.5	167.95
M2	148.4	117.75	121.6
Av.	213.65	164.9375	187.725
Sd.	67.92482	43.16143	61.63942

仔细分析表明,这是由两个因素决定的:首先,作为韵律短语的语调,其内部要求一定程度的音阶下倾,由于“罗尔斯的”处于韵律节奏短语的开头,尽管听起来不那么重,其音阶总是相对地高一些;而“正义论”处于韵律节奏短语的末尾,所以它的音阶必定会被相对压低。第二,由于这是一个持续短语,也就是赵元任先生所说的悬念子句,它要求一个“不低性”边界调作为话语持续的标志,所以它的音阶又略高于“巨著”的,从而在听感上造成略微上挑、话语持续的印象。由此可见,语句中的重音实现必定要受语调和节奏的制约。同时,还跟它在韵律单元中的位置前后具有一定关系,相对来说,音阶突显作为重音的主要声学实现手段,它在语句首尾(也就是韵律边界前后)位置上的变化不如在内部中间位置上的来得自由。

再如,在“探讨一下权力制约的制度问题”一句中,“探”、“权”、“制”分别处于三个韵律短语的开头,从听感上讲,“探”与“权”差不多重,而“制”显得更重一些。可是,从声学表现上看,“探”的音阶最高,“权”的音阶略低于“制”的,如图2所示。而且不同话者的语音表现一致,如表2的数据所示,说明图2表现的也不是个别的偶然现象。

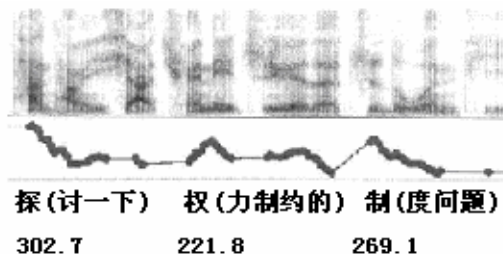


图2 语句中音节的音阶实现举例

表面上看来,图1和表2所反映的这种分布现象,似乎既不符合语调的下倾走势规律,也不符合重音突显的音高突出规律:若论下倾走势,“权”的音高应该略高于、而不是略低于“制”的音高;若论重音突显,也是“权”的音高应该是最高的,然而,它却又低于“探”的音高。实质上,如果再深究一下,就会发现,这是由于语调和重音突显共同作用的结果:第一,由于语调下倾走势的需要,所以“探”的自然音高高于“权”和“制”的,以满足整句总体音阶自前向后逐渐下落的要求;第二,此句是由三个韵律短语构成的,根据由句法制约决定的常规重音分布规律(郑波,2000),整句的句重音应该落在句子宾语部分的定语“制度”上,由于“制度”的“度”通常念轻声,因此实际重音就落在“制”上,它的音高应当突显于周围的音节,所以,它的音高略高于“权”的也就顺理成章了。由此可见,“制”的音高实现实际上是语调的下倾走势跟重音突显两种制约并存叠加的结果。

上述例子还进一步说明,语句中的重音分布是跟句法上的层次关系十分密切的。事实上,句法上的层次是底层语义结构的外部表现,所以,从这个意义上来说,语句中的重音

分布最终是受语义结构的制约的。

3.2 语句重音合成试验结果

合成试验同样表明，语句重音跟其它韵律特征具有密切的关系，它在词重音的基础上受节奏和语调的制约和调节，既要符合节奏上的层次结构，又要受语调下倾走势的制约，从而形成一定的层次结构。

3.2.1 语音合成反映出短语内重音存在着层级关系

合成试验表明，语音合成必须处理好重音的层级结构。例如，在合成“恢弘巨著正义论”这个短语时，考虑到“恢弘”和“巨著”都是修饰“正义论”的，根据常规重音规律，它们都应该比“正义论”读得重一些，音阶就应该设得相对高一些。可是，合成结果表明，仅仅这样设置语句重音还不够，这样合成的结果不自然，“巨著”一词听起来特别突出，显得太重、太高。究其原因，就是因为对短语内部的层级关系没有处理好。因为一方面，“恢弘巨著”是上一级短语“恢弘巨著正义论”中的一个下位短语，它除了修饰另一个与它并行的下位短语“正义论”以外，本身也应该有它独立的短语语调模式，也就是应该有它独立的音高下倾走势。因此，“巨著”不应该与“恢弘”并重，它的总体音高应该低于“恢弘”的。另一方面，在这个层级上，“恢弘”跟“巨著”之间还存在着修饰关系，“巨著”受“恢弘”的修饰，它们之间本来就必须分出相对的轻重。根据基于语法信息的韵律结构预测方法(曹剑芬, 2003)，整个短语的重音等级应该是“罗尔斯的 1.0 恢弘 2.5 巨著 2.0 ‘正义论’ 1.0”（具体预测方法如末页附图所示）。根据这个预测，我们对“巨著”的重音等级做了适当降低，结果，合成话语的轻重效果就很好。因为这样的处理既符合节奏上的层次分别，又满足了语调上的音高下倾的需要。这个事实进一步证实了我们在 3.1 节中观察到的自然语句重音的分布规律和结构特点。

又如，在合成“逐步引入多个变量”这个短语时，假如只是把相关的韵律词简单地并重排列，合成的语音听起来别扭，其中的“引入”和“变量”显得特别突出，究其原因，也还是因为没有处理好重音的层次关系。首先，从节奏层级上看，这个短语是由“逐步引入”和“多个变量”两个下位短语构成的，因此，短语重音设置起码要分两个不同的层次：根据常规重音分布规律，从整个短语看，“多个变量”充当宾语，应该比“逐步引入”更重一些；从各个下位短语看，“逐步”是状语，应该比“引入”重，“多个”为定语，应该比“变量”重。同时，从语调结构来看，也应该是整个短语有一个短语调，两个下位短语又应该有各自独立的短语调，而且，彼此之间应该存在音高重置。根据这些韵律结构特点进行处理以后，整个短语的重音层次就分别出来了，结果合成的话语相当自然。

为了进一步深化上述认识，我们对同一句话的自然语音跟合成语音的音阶运动做了一个对比分析，这个合成语音是未经上述重音调节的，听起来不太自然。具体的对比情况可以用图 3 和表 3 来概括。首先，从图 3 中可以观察到，自然语句的每个韵律短语都以节奏强音开始，起始音阶较高，接着逐步下倾，等到下一个韵律短语开始，又重复出现这样的过程，而且，较大的韵律短语之间的这种重置比较小的韵律短语之间的更为显著，形成了明显的层次结构；而合成语音虽然也有类似的结构，但层次不够分明。同时，还可以通过下表韵律词的音阶差异来说明。表 3 所列的数据是自然朗读的和合成的“世间的问题原来极复杂的，可以用极简单的事例加以说明”的音阶比较。首先，从韵律词之间的音阶差异来看，在句首的“世间的问题”和句尾的“加以说明”两个韵律短语中，“世间的”跟“问题”之间和“加以”跟“说明”之间的音阶差异，显然都是自然语音的明显高于合成语音的；至于在句子中部的“原来极复杂的”和“可以用极简单的事例”两个韵律短语中，各个韵律词之间的音阶差异虽

然都相对缩小，但总体上看，仍然是自然语音的大于合成语音的。至于韵律短语内部或句子内部最高和最低音阶差异，也都是自然语音的绝对地高于合成语音的。

上述实例再次说明，适当的重音结构设置是提高合成话语自然度的一个关键。

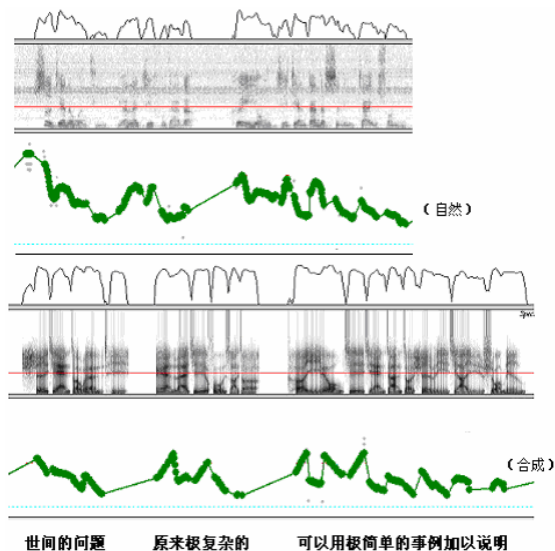


图 3 “世间的问题原来极复杂的，可以用极简单的事例加以说明”自然音和合成音之比较

表 3 自然和合成语音的音阶比较

韵律词语	音阶		韵律词间的音阶差		短语内最高/最低音阶差		话语内最高/最低音阶差	
	自然	合成	自然	合成	自然	合成	自然	合成
世间的	238.6	123.5						
问题	157.9	100.7	80.7	22.8	80.7	22.8		
原来	209.8	126.0						
极	159.4	129.8	14.4	-3.2				
复杂的	150.0	85.6	45.4	44.2	59.8	40.4		
可以用	221.4	132.9						
极	236.6	145.2	-15.2	-12.3				
简单的	212.7	94.7	23.9	50.5				
事例	163.7	100.8	49.5	-5.2	73.4	50.5		
加以	173.0	105.2						
说明	122.8	103.5	50.2	1.7	50.2	1.7	98.6	59.6

3.2.2 语音合成反映了处理好语流轻音的重要性

我们最新的观测分析表明，现在的语句合成效果之所以不理想，重要原因之一就是忽视了语流轻音问题。也就是说，合成语音的轻重差异不够明显。实际上，在自然话语里，除了重读的词或音节以外，存在大量非重读的词或音节，它们有时读得很轻，近似于轻声，但却对重音起着不可忽视的烘托作用，使得重音更加突显。可是，在合成系统中，往往只注意词重音和语句主要重音的设置，而忽略了大量存在的非重读词或音节的轻音处理。

试比较图 3 中自然语音跟合成语音的声学表现，你会发现它们之间具有明显的不同：首先，自然朗读时，语句的音高起伏幅度明显大于合成语音的；第二，自然语句中音节的时长伸缩变化大于合成语句的；第三，自然语音音强的起伏也明显地大于合成语音的，而且具有类似于音高的下倾走势及其重置，并形成一定的层级结构，而合成音的音强则缺乏明显的层次。

从上述几个参量的比较可以看出，在自然语句中，语音的轻重变化很大；而在合成语句中，轻重变化则不够大。例如，在“可以用极简单的事例加以说明”这个短语中，自然

发出的语音符合常规重音分布规律：“极”和“简单的”携带短语重音，而其余的韵律词则相对轻读，尤其是“可以用”中的“以”和“用”，以及“事例”中的“例”都是语流轻音。然而，在合成语音中，相比之下，“简单的”(尤其是“单”)不够重，而“可以用”中的“以”(注意：“以”作为上声，其音高越低，听起来就越重)和“用”，以及“事例”中的“例”又都不够轻。因而听起来总是不那么自然。

3.2.3 语音合成说明了处理好四声固有音区差异的必要性

此外，尽管音节的声调音阶在语流中的变化看起来非常随意，但是，还是需要注意汉语的四声之间固有的音区差异问题。例如，在合成“探讨一下权力制约的制度问题”时，虽然“权力”的音阶应该比“制约”的略高一些，但是，“制约”的“制”的高音点一定要设得略高于“权”的，否则合成语音听起来就别扭。同样，“制”又要略高于“约”的。这是因为，普通话的去声和阴平都具有高音区特征，而且，去声又比阴平略高。

又如，在合成“内在关系”时，不管是简单的并重排列还是简单的音高下倾，合成结果都不自然。后来经过适当的修改，既考虑了音高下倾，又考虑了层次关系，合成结果就自然了。因为在这个例子里，虽然有三个去声和一个阴平，但是它们的高音点不能简单地随意下倾，而需要考虑这个短语的层次结构：首先，在“内在”一词内部，要遵守两去相连的变调规则，第一去不全降，第二去高点比第一去低点略高；其次，要考虑“关系”一词内部“系”的高点要略高于“关”的；第三，也是最重要的一点，既要设置两词之间的下倾走势，以构成短语调；又要考虑两者之间的音高重置，一定要把“关”的高音点定得略高于或接近于“在”的、但必须低于“内”的才行。由此可见，音高下倾既要遵守一定的层次结构，又要处理好四声之间固有的音区差异，合成音听起来才会有适当的轻重缓急。

4. 结语

(1) 汉语里不但有词重音，而且是必不可少的，它还是构成语句重音的基础。

(2) 汉语的词重音类型除了普通(正常)重音型(即它的常规 default 形式)跟轻声型(即弱重音 weak stress)的对立以外，并不存在所谓“中重”与“重中”绝对的区别性对立。至于那些普通重音型音节，在连续话语中的确也存在着重度上的差异，不仅可以分为重与中，甚至还可以分出次轻和轻等更多的层次。但是，这些差异只不过是普通型词重音的语境变体，本质上属于语句层面上的相对对比性重音差异，而不是词层面上的对立性重音区别。

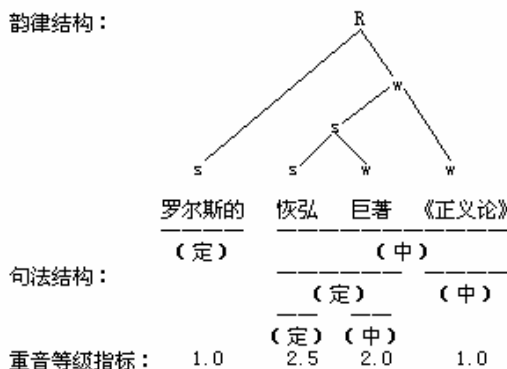
(3) 自然语句中的各个韵律词，虽然表面上看似线性的排列，但并不意味着各个词都是并重的，也不是简单的轻重相间，而是受节奏层次和语调结构的制约，也具有一定的层次和等级结构。这充分说明，轻重跟语调看似相互独立，实质上密切相关。它们既彼此关联，又相互制约：没有轻重层次的下倾，合成的语调不自然；没有语调下倾的合成，也不会产生自然的轻重音。

(4) 语句重音的层次结构跟句法结构具有一定的相关关系：首先，常规的重音等级取决于它所处的句法层次的高低；其次，在同一韵律单元内部，重音的分布又主要取决于该单元的语法结构特点，同时，还跟它在韵律单元中的位置前后具有一定关系。事实上，句法上的层次是底层语义结构的外部表现，所以，从这个意义上来说，语句中的重音分布本质上是受语义结构的制约。

参考文献

- [1] 曹剑芬, 1995, 连读变调与轻重对立, 《中国语文》, 第4期。
- [2] 曹剑芬, 2003, 基于语法和语音信息的汉语韵律结构预测, 《中文信息学报》, 第十七卷第三期。
- [3] 林茂灿、颜景助, 1990, 普通话轻声与轻重音, 《语言教学语言就》, 第3期。
- [4] 殷作炎, 1982, 关于普通话双音常用词轻重音的初步考察, 《中国语文》, 3期。
- [5] Zheng Bo, et al., 2000. The regular accent in Chinese sentences. The Proceedings of ICSLP'2000, Beijing.

附图：基于语法结构的语句重音结构预测图示



注解：

- [1] 本文相关的合成试验是在 IBM 中国研究中心语音部的一个规则合成系统 TTSSPY 平台上进行的，特此致谢！
- [2] 在这两句中，如果“公鸡”和“攻击”都采用重中型，但是分别相对加重“斗争”和“攻击”，合成效果大致相当。
- [3] 本文所说的词都是指韵律词，或者叫语音词，而不是严格的句法意义上的词。