

语法信息与韵律结构的分析与预测*

王永鑫 蔡莲红

清华大学计算机科学与技术系 北京 100084

Email: wangyongxin@mails.tsinghua.edu.cn, clh-dcs@mail.tsinghua.edu.cn

摘要: 韵律结构的自动预测是高自然度文语转换(TTS)系统的关键组成部分,直接影响到合成语音的自然度和表现力。本文建立了一个同时具有语法信息与韵律结构标注的汉语语料库。并在这一语料库的基础上,对汉语的韵律结构组成、韵律结构与语法语义之间的关系进行了分析,并进行了预测试验。研究发现,汉语的韵律结构虽与语法结构不同,但是有着密切的联系,韵律结构可以通过语法结构进行预测。韵律结构除与语法结构有关之外,还要受到语句语义的制约。

关键词: TTS, 韵律结构, 语法结构, 语义

Syntactic Information and the Analysis and Prediction of Prosody Structure

Wang Yongxin, Cai Lianhong

Department of Computer Science and Technology, Tsinghua University, Beijing 100084

Email: wangyongxin@mails.tsinghua.edu.cn, clh-dcs@mail.tsinghua.edu.cn

Abstract: Automatic prosody structure prediction is a very important component of high quality Text-to-Speech systems. It directly affects the naturalness and expressivity of synthesized speech. In this paper, a text corpus with both syntactic and prosodic structures annotated is constructed. Based on the corpus, the composition of prosody structure and the relationship between syntactic and prosodic structures are analyzed, and a prediction experiment is carried out. The result shows that though there are differences between the prosody structure and syntactic structure in Chinese language, they have a close relationship, and prosodic structure can be predicted based on syntactic structure. Prosodic structure is also affected by semantic information of the sentence.

Keywords: TTS, prosodic structure, syntactic structure, semantic information

1. 引言

高自然度和高表现力的语音合成离不开韵律预测。韵律是指语音的超音段特征,是从语音所抽象出的节律、重音和语调特性,在感知上表现为语音的音高、速度和音量随时间的变化,在声学参数上的表现为基频、音段时长和能量随时间的变化。此外韵律还是言语的自然属性和固有特征,隐含了语法、语义、语用等多种语言功能。它不仅体现了说话人对文字的表达和理解,还反映说话人的态度、期望、情绪等高层信息。连续语流中一个重要的现象就是韵律层级结构的存在。韵律结构之间在感知上存在着边界,这种边界对应的声学表现可能是停顿、音高曲线的变化以及边界之前的音段延长。汉语韵律结构主要包含韵律词、韵律短语、语调短语三个层次。

为了实现更加自然的韵律预测,首先要得到更加准确的韵律结构。现在对韵律结构进行预测是,主要是基于语句的语法信息,如词性、词长等[2][3][5][7][8][9],也有研究者使用语法树、语法短语等特征[6]。在这些研究中,有些使用从录音语料得到的韵律标注,语料规模从1000句到6000句不等[1][2][5][6][7][8]。使用从录音标注得到的语料库,可以得到自然语音中的韵律短

*本文的研究工作受到国家高技术研究发展计划(2007AA01Z198)、国家自然科学基金(60805008, 90820304)的资助

语结构,但是由于录音、标注的过程比较复杂,使用录音语料通常语料库的规模比较小。但通过文本标注的方式得到的语料库,由于标注过程可以得到简化,所以可以得到规模比较大的语料库[3][9]。

现在的韵律结构预测多数是有语法信息出发对句子的韵律结构进行预测。在韵律预测的研究中,研究者通常比较关心预测所使用的特征与算法,以提高预测的准确率。如果可以对语法与韵律的关系进行分析,得到汉语韵律结构的生成特点,将可以加深对汉语韵律结构的了解,为韵律预测技术的发展提供支持。

本文通过人工标注建立了一个含有十万句新闻文本的汉语韵律语料库。通过在标注过程中制定详细的标注规范和对标注者长时间的训练保证标注结果的一致性和正确性。在这一语料库的基础上,基于语料库中语法信息的标注对韵律结构的组成进行了分析,并在这一大语料库的基础上进行了韵律预测的试验。通过分析与试验,可以发现通过语法信息进行韵律结构预测可以得到较好的效果。但是同时也发现,韵律结构除了与语法信息有关之外,还会受到语义的影响,要进一步提高韵律预测的准确性,需要对韵律与语义的相互关系进行研究。

2. 汉语的韵律结构

本文关注汉语韵律结构中的韵律词、韵律短语、语调短语。较小的韵律成分包含在更大的韵律成分中,由此形成了一个层级结构。

韵律词是语流中节律的基本单位也是语音-语法界面上的基本单位。在韵律词内部没有静音段,在韵律词边界处停顿不是必须的。从语音上来说,韵律词总是作为一个音步出现,即内部只有一个连调域,不会出现韵律分界现象。汉语的标准音步是两个音节,单音节音步是退化音步,三音节音步是超音步,实验观察发现极大多数的韵律词为2-3个音节。包括轻声的韵律词可能延长到4-5个音节。

韵律短语由一个或几个韵律词组成,是介乎韵律词和语调短语之间的韵律单位,边界表现为能感知到的停顿。韵律短语具有相对稳定的短语调模式和短语重音配置模式[4]。韵律短语的长度与语速关系很大,当语速快的时候韵律短语也会相应变成得更长。通常韵律短语为7个音节左右,一般最长为9~10个音节。

语调短语相当于语法上较短的子句,是音系规则作用在句子层面上的辖域[4]。其边界主要服从于语义,通常出现在比较完整的语义成分之后。在实际应用中,往往将句子中的某些标点直接作为语调短语边界的标志。

在语音合成应用中,通常更关心韵律词与韵律短语边界,而使用标点作为语调短语边界的标志。本文中也将对韵律词边界与韵律短语边界进行考察。

3. 具有语法信息与韵律结构标注的语料库

为了可以得到更准确的韵律结构预测结果,从而提高合成语音的质量,现在的韵律结构预测多使用机器学习的方法,从句子的语法信息出发对韵律结构进行预测。建立一个大规模的同时具有语法信息与韵律结构标注的语料库不但有助于对汉语的韵律结构进行进一步的分析,加深对汉语韵律结构组成规律的认识,还可以为提高韵律结构预测的准确性。

本文标注的原始文本来自2001年全年的《人民日报》。北京大学自然语言处理实验室对其进行了语法分词与词性标注[10]。本文随机抽取十万句语料,从文本上直接进行韵律结构的标注。在文本上直接标记韵律结构,需要标注人根据自己对句子读法对句子的韵律结构进行标注。这样可以加快标注速度,并可以通过标注人之间对标注规则的探讨保证标注结果的一致性。

本文选用三位标注人(大学学历)。首先在专家的指导下对标注人进行为期三个月的训练。三名标注人共同标注了一万个语句,在标注过程中反复比较异同,指出错误,讨论问题。共同理

解语法结构与韵律结构在语言表达中的作用与不同；韵律词与韵律短语的生成原则；语法、语义对韵律的约束原则。要求他们排除个人朗读习惯对标注的不良影响。实验表明仅依据少数规则和语法信息，标注人能够很好的把握韵律词划分粒度，三位标注人韵律词边界划分的一致性达到97%。韵律短语边界划分的一致性为86%。经过分析我们发现，韵律短语划分的不一致主要是由于韵律短语本身的灵活性导致，各份标注结果均无明显的错误。可见标注人依据对句子语法、语义和韵律的理解能够较准确的分辨出句子的韵律结构。在训练的基础上，由两位标注一致性比较高的标注人分别标注了其余九万句语料，抽查结果表明，标注的一致性很高。

语料库与标注结果

在标注完成的九万句语料中，语法词的平均长度为1.77个汉字，韵律词的平均长度为2.21个汉字，韵律短语的平均长度为5.61个汉字。语法词、韵律词、韵律短语长度的详细分布情况如图1所示。

从图1可以看到，汉语中存在大量的单音节语法词，但韵律词的长度以两个音节为主，单音节韵律词的数量很少。同时，四字及以下的单音节韵律词的数量也很少。事实上，在由语法词生成韵律词过程中，大多数单音节语法词都会和他前面或后面的语法词合并生成韵律词，而两个音节以上的语法除了可以和单音节语法词合并之外，很少和其它的语法词合并成为韵律词。这完全符合汉语的标准音步为两个音节，单音节词只在极少数情况下成为一个韵律词[4]的规律。

在标注结果中，韵律短语的长度一般为四到九个音节。其中，四个字韵律短语出现最多，接近韵律短语总数的四分之一，而七到九个字的长韵律短语则相对比较少。这是因为本文的韵律短语标注粒度较细，同时汉语文本中存在这大量的四音节结构。考虑到汉语的长短语内部大多包含多个小短语，在人工标注中，本文倾向于将它们划分开来，为下阶段的韵律预测提供便利。另外，标注结果中单音节韵律短语的数量极少，因为单音节韵律短语基本只出现与单字句中。十音节及以上的韵律短语也属于特殊情况，多半包含了结构助词“的”或连词“和”之类本身带有分界信息的成分。

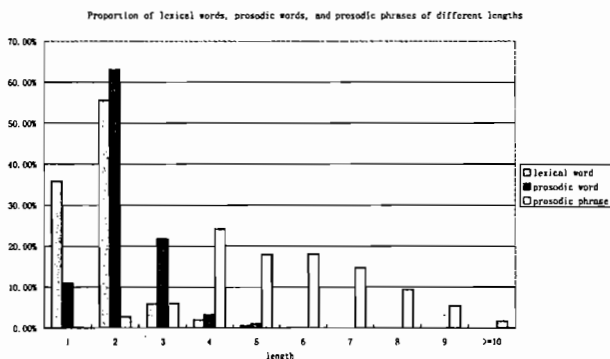


图1 在标注数据中语法词、韵律词与韵律短语的长度分布

4. 韵律结构分析

汉语韵律结构与语法结构虽然不同，但是存在着一定的关系。这也是从语法信息预测韵律结构的基础。通过对标注语料库的分析，可以揭示韵律结构与语法信息的关系。

在进行韵律结构划分的时候，可以认为所有的韵律词边界都出现在语法词边界处。在实际语料中，会出现四个音节以上的语法词。这些语法词应该被切分为多个韵律词。但是，由于这样的语法词数量非常少，仅占语法词总数的0.7%，因而可以暂时认为这些语法词独立成为韵律词，这样可以极大的简化标注与分析过程。

汉语韵律结构划分中韵律组词的过程，实际上是单音节语法词与其它语法词组合生成韵律词的过程。在标注语料库中，单音节韵律词的数量仅为单音节语法词的 24.2%，即有 75.8% 的单音节语法词通过合并生成了韵律词。而任何两个非单音节语法词合并都会导致韵律词长度达到或超过四音节，突破韵律词的词长限制，因而是不可能出现的。连续的单音节语法词出现时，它们总是会互相合并生成韵律词。连续的单音节韵律词出现很少，在九万句语料库中仅出现了 4881 例。

在单音节语法词与其它语法词合并时，根据具体语法词的不同，会出现不同的合并规律。总的来说，但单音节语法词与双音节语法词合并是，倾向与附着在双音节语法词之后。在由两个语法词组成的三音节韵律词中，由双音节+单音节组成的数量，是由单音节+双音节组成数量的 2.5 倍。轻声的单音节语法词总是和它前面的语法词合并生成韵律词。比较典型的如“的”，“着”，“了”等。这些语法词在标注语料库中，与前一语法词合并的比例都为 100%。

在语料中出现的单音节语法词中，助词所占的比例最大，其次是介词、动词和连词。助词、介词与连词由于不是实词，会更倾向与粘附在其他词上。因而本文重点考察了这些词类对韵律结构生成的影响。

助词没有自己的实体意义，在语流中也不会重读，有很多的轻声词。助词用于通常用于协助前词形成一定的语法成分，因而，助词通常都会与前词结合生成韵律词。在标注语料中，助词与前词结合的比例占到了 94.2%。

连词可以分为并列连词与非并列连词。并列连词用于连接前后两个并列的成分，实际上无论在语法与语义上与前后词都没有非常紧密的联系。因而，单音节并列连词只有在前后成分仅含有一个音节时才会与其合并，多数情况下都独立的形成韵律词。但相对来讲，并列连词与后词的联系更加紧密，因而通常在词前会比词后有更高等级的停顿。

以连词中出现次数最多的“和”为例。“和”所连接的成分通常比较长，因而独立形成韵律词的趋势比较明显。56.9%的“和”前都有一个韵律短语边界，并且独立形成一个韵律词。另有 30.5%前后各有一个韵律词边界。总的来讲，“和”独立形成韵律词的比例达到了 87.5%。“和”字向前合并一共出现了 0.8%，而向后合并的情况也只有 11.8%。从上面的结果中可以看到，并列连词在语音上并不轻读，可以获得独立形成韵律词的资格，在语法上并列连词只起到连接的作用，所以并不倾向与附着在某一方上，在语义上并列连词用于引出下一个并列成分，因而与下一成分的关系要相对更紧密一些。与“和”相比，“与”所连接的成分相对较短，因而其周围韵律短语边界出现比较少，仅有 33.4%，而两边均为韵律词边界的情况达到了 42.1%。它与前后词合并的比例都有所增加，但总的来看，独立形成韵律词的趋势还是比较明显的。

非并列连词的韵律构成上的表现与并列连词就有很大的不同。当连词所连接的两个成分的地位不相同是，连词要引出其后的成分，并指明两者之间的关系。此时连词与其后的成分连接更加紧密。同时，非并列连词所连接的通常并不是词语，而是短句，因而连词之前会有更容易出现韵律短语边界。以表示转折的连词“但”为例，它有 99.5%的情况前面出现了韵律短语边界。

汉语中，介词与动词的在韵律组词的表现上比较类似，也比较复杂。介词用于表示其后的词的语法作用，他总是与其后的词或词组组成介词短语，所以他与其后的词有着自然的联系。因此，当介词的宾语只有一个音节时，介词与其宾语总是会合并形成一个韵律词。

同时，单音节介词与出现在它前面的单音节副词或副词作用的词有很强的合并趋势。副词与介词短语在句子的中的语法地位是相同的，于是在介词没有单音节宾语时，单音节的副词与引导介词短语的单音节的介词紧密的结合在一起，甚至形成了一种副词+介词组成的复合介词的结构。（省委/n /<PW> 宣传部/n /<PPH> /<PW> 近日/t /<PW> 召开/v /<PPH> 新闻界/n /<PW> 座谈会/n）在标注语料中，单音节的副词与介词的连接，合并的比例为 94.5%。

介词前可能存在其它介词，此时两个介词的辖域虽然不同，但它们之间也可以合并成为韵

律词。(如“李/nrf 保东/nrg /<PPH> /<PW> 少数/m 人/n /<PW> 权利/n /<PW> 议题/n /<PPH> 发言/vi 时/Ng /<PW> 说/v”)

当句子的主语是单音节词时，介词还可以附着在句子的主语之后。当介词的宾语比较长时，介词与主语可以形成独立的韵律短语。这种韵律结构组成方式，与动词很相似。

从上面的分析也可以看到，在介词的韵律组词中，也要循序韵律约束（韵律结构长度限制）、语法约束（介词与前后词的语法关系）等多种因素的限制，在它们的共同作用下形成句子的韵律结构。

在由韵律词生成韵律短语时，同样需要同时考虑韵律、语法、语义等多方面的影响。与韵律词的划分相比，在韵律短语划分时语义约束将起到更重要的作用。韵律短语的划分要符合句子的语法结构，因而，韵律短语边界经常出现在句子较高层次的语法边界上，但由于韵律短语的划分还要满足长度的限制，因而不能总是完全按照语法结构划分。同时，韵律短语结构对语义表达有更重要的作用，因而也要受到语义表达的约束。

5. 自动预测试验

分析标注语料库可知韵律结构与语法结构虽然不同，但是存在着较多的联系。因而从语法信息进行韵律结构的预测是可行的。本文在标注语料库的基础上，进行了韵律结构自动标注的试验。

试验中，使用两名标注人分别标注的九万句语料作为试验语料库，并选取其中的一半作为训练集，一半作为测试集。试验使用 wagon 决策树进行训练，使用边界两侧各两个语法词的词长与词性作为特征。本文试验的预测结果与其它一些同类工作的预测结果如表 1 所示

表 1 韵律边界预测结果比较

	韵律词边界		韵律短语边界	
	正确率	召回率	正确率	召回率
C4.5[5]	---	---	82.9%	71.2%
C4.5+TBL[5]	---	---	68.4%	85.9%
ME[3]	93.6%	96.3%	79.8%	78.4%
本文的结果	97.5%	97.6%	84.0%	74.3%

从对比结果可以看到，使用语法信息对韵律结构进行预测可以得到比较好的结果，特别是对韵律词边界预测，所得到的预测精度已经比较高。但是在韵律短语预测方面，由于其划分与语法结构及语义有关，因而仅通过词性等语法信息预测的精度还比较低。

虽然本文没有使用复杂的预测算法，但是在一个标注一致性比较好的大规模语料库上，仍然取得了与其它工作类似的结果。一个大规模的语料库可以包含更加丰富的韵律现象，可以更加真实的反映汉语韵律结构的特点，从而从中得到的分析结果也更加真实可信。同时，本语料库在建设过程中，通过对标注人的训练，减小了韵律结构划分灵活性对韵律标注的影响，在保证韵律结构标注正确的基础上，提高了语料库标注的一致性。

6. 语法及语义信息对韵律结构的约束

语音、语义跟语法是人类语言中的几个独立又互相制约的平面。纯语法理论已不再是汉语语法研究的唯一目标，语音、韵律、语法等研究相互渗透。韵律预测是语音、语法和语义的一个非常好的结合点。韵律预测的历程也是一个很好的例证。最初，韵律预测是基于机械式的语法分词和规则处理；进而引入语法信息，基于统计生成规则；目前基于人工标注的大规模语料库，引

入语法和语义信息，再利用机器学习算法实现韵律预测，取得了很好的结果。标注人能根据对文字的理解，很好的平衡语音、语法和语义的关系，准确地表达文字要表达的信息，但机器处理距离之一要求还有不小差距。例如：

语法分词结果	韵律切分结果
走 法治 的 道路	走▲法治的道路
练 体操 的 经历	练体操的▲经历

句中三角形表示韵律词边界。上面的两个短语有相同的词性组合，但由于语义不同，语法词之间的关系也不相同。第一句中，“道路”为“走”的宾语，“法治”与“道路”的结合更为紧密，因而“走”只能单独成为韵律词。但是在后句中，“练体操”却是一个整体，因而“练体操的”形成了一个独立的韵律词。在预测结果中，后句就多出了“练”后的一个韵律词边界。上面两句的区别无论是用语音信息还是词性的信息都难以区分。上例表明语言中韵律要为语义表达服务，仅依据词性特征和语音平衡对韵律结构进行定界和预测还存在很大的不足。

再比如：

荣获本届“**文华大奖**”的剧目**还有**■陕西省**戏曲研究院**的**眉户剧**《迟开的玫瑰》、**重庆市川剧院**的**川剧**《金子》、**湖南省湘剧院**的**湘剧**《马陵道》等**戏曲剧目**。

句中方块表示韵律短语边界。如果仅考虑语音与词性信息，“还有”应该与陕西省合并成一个韵律短语，这样使得韵律短语的长度比较平衡。但是考虑后面出现的排比句式，该韵律短语的边界则应该调整至“还有”之后。

类似的情况还有：

■体操技术□委员会的委员■

■中国体育□代表团的■

上面两个短语，在自动切分长短语时，依照语音平衡原则都被划开。但是两个短语中的词结合比较紧密，朗读时应属于同一个韵律短语。这类长短语是否应该进一步划分很难通过词性信息判断，必须综合考虑句子的语义。

在进行韵律短语划分时，对同样的文本，根据语义表达的不同需要还可能出现多种不同的韵律短语划分方式。如在下面的例子中：

孝义市经济实力■显著增强

孝义市■经济实力显著增强

两句都是合理的短语划分方式，但前一句强调增强的程度，后一句强调增强的是经济实力而不是其它。也就是要，在句子中没有明显的语义重音的时候，韵律短语边界出得的词语由于基频重置会获得一定程度的强调。

文[6]中将预测句子的语法树加入韵律结构预测的特征集，虽然在一定程度上提高了预测精度，但其利用的是人工标注的语法结构信息，还很难在大语料库及自动预测中应用。

可见，在韵律结构的划分中，语法和语义信息起着至关重要的作用。如何将语法和语义信息融入韵律结构的预测过程中，将成为进一步提高韵律结构预测精度的关键问题。

7. 结论

本文在一个已完成语法分词与词性标注的语料库上，进一步进行了韵律结构的标注，得到了一个含有十万句文本语料的人工标注的同时具有语法信息与韵律结构信息标注的语料库。语料库标注过程中通过对标注人训练，保证了语料库内部标注高度的一致性。

在这一语料库的基础上，本文进行了对汉语韵律结构的分析。从分析中可以看到，韵律结构虽然与语法结构不同，但是韵律结构与语法信息还存在着比较紧密的关系。文中通过对汉语文

本中出现较多的单音节语法词的韵律组成规律进行分析发现,助词总是和前一个语法词组成韵律词,并列连词除和单音节并列成分合并之外,倾向与独立形成语法词,非并列连词倾向于与后词组合形成韵律词,介词则通常与其单音节宾语或词前的单音节副词合并。通过对语料库的分析,可以得到韵律词与韵律短语的划分规律。因而使用语法信息进行韵律结构的预测是可行的,从文中进行的预测试验的结果也可以看出这一点。

从预测实验的结果可以看到,现在的韵律结构预测中,对韵律词边界的预测结果更好,但对韵律短语边界的预测结果稍差。这主要是由于韵律短语划分不但与韵律约束相关,还与句子的语法、语义等多种因素相关。只使用句子中的语法信息就不能完全描述韵律短语划分的全部约束条件。

本文还在这一语料库的基础上,分析了韵律结构与句子语法、语义表达的关系。韵律结构的划分是为准确的语义表达服务的,因而韵律结构的划分必须符合语义表达的需要。在韵律预测算法中加入语义的约束将成为提高韵律结构预测,特别是韵律短语边界预测精度的关键。

参考文献

- [1] Min Chu, Yunjia Wang. Rhythmic Organization of Mandarin Utterances – A Two-Stage Process. Proc. of International Symposium of Chinese Spoken Language Processing, Singapore, 2006, pp 138-148.
- [2] Min Chu, Yao Qian. Locating Boundaries for Prosodic Constituents in Unrestricted Mandarin Texts. Computational Linguistics and Chinese Language Processing, Vol.6, No.1, February 2001, pp.61-82
- [3] Xiaonan Zhang, Jun Xu, Lianhong Cai. Prosodic Boundary Prediction Based on Maximum Entropy Model with Error-Driven Modification. Proc. of International Symposium of Chinese Spoken Language Processing, Singapore, 2006, pp 149-160.
- [4] 曹剑芬. 基于语法信息的汉语韵律结构预测. 中文信息学报, Vol.17, No.3, 2003, pp.41-46
- [5] 赵晟, 陶建华, 蔡莲红. 基于规则学习的韵律结构预测. 中文信息学报, Vol.16, No.5, 2002, pp.30-37
- [6] Zhao Sheng, Tao Jianhua, Jiang DanLing. Chinese prosodic phrasing with extended features. Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, 2003, pp. 492-495.
- [7] Honghui Dong, Jianhua Tao, Bo Xu. Prosodic Word Prediction Using the Lexical Information. Proc. of IEEE International Conference on Natural Language Processing and Knowledge Engineering. 2005, pp.189-193.
- [8] Honghui Dong, Jianhua Tao, Bo Xu. Prosodic Word Prediction Using a Maximum Entropy Approach. Proc. of International Symposium of Chinese Spoken Language Processing, Singapore, 2006, pp 169-178.
- [9] Heng Kang, Wenju Liu. Prosodic Words Prediction from Lexicon Words with CRF and TBL Joint Method. Proc. of International Symposium of Chinese Spoken Language Processing, Singapore, 2006, pp 161-168.
- [10] 俞士汶, 段慧明, 朱学锋 and 孙斌. 北京大学现代汉语语料库基本加工规范. 中文信息学报, 2002, Vol.16, No.5, 49-64