

汉语语篇修辞结构边界韵律表现

杨晓虹^{1,2}, 杨玉芳¹

(1. 中国科学院心理研究所, 北京 100101; 2. 中国科学院研究生院, 北京 100049)

摘要: 本文考察了语篇修辞层级结构边界的韵律表现。对由 20 个语篇构成的语料库进行了修辞结构的层级标注和声学分析。声学参数主要采用小句边界处无声段, 音高重置和边界前音节延长。研究得到以下主要结论: (1) 小句边界处无声段和高音点重置是语篇修辞层级结构的主要声学线索。小句边界在语篇层级结构中的位置越高, 边界处无声段越长, 边界处高音点重置的值越大。但是语篇结构中小句边界处的无声段延长和高音点重置增大也是有限度的。(2) 小句边界处低音点重置和边界前音节延长都不是区分语篇修辞层级结构的有效线索。

关键词: 修辞结构; 无声段; 音高重置; 边界前音节延长

中图分类号: H 013 文献标识码: A

文章编号: 1000-0054(2009)S1-1375-05

Prosodic features of rhetorical structural boundaries in mandarin Chinese text

YANG Xiaohong^{1,2}, YANG Yufang¹

(1. Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China;

2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: This paper investigates the prosodic reflection of rhetorical structures in Chinese texts. A corpus of twenty oral texts that had been read aloud was annotated based on the rhetorical structure theory (RST) and was then acoustically analyzed. The prosodic features included pause duration, pitch reset and pre-boundary lengthening. The results show that both pause duration and top line reset are reliable indicators of the rhetorical structure of texts. Their values increase as the position of the segment climbs up the rhetorical hierarchy, with some limitations. Bottom line reset and pre-boundary lengthening are found to not be valid indicators of the rhetorical structure of texts.

Key words: rhetoric structure; pause duration; pitch reset; pre-boundary lengthening

在语音学的研究领域中, 语篇的韵律特征研究既是热点也是难点。语篇水平上对小句韵律的调节很大程度上影响了语篇语音合成的自然度。语篇既有其表层的结构也有其深层的结构。从表层上来看, 语篇是由段落或句子等信息单元组成的, 是呈线性排列的, 并且信息单元的位置不同或者单元大小都会影响其韵律的表现。Swerts^[1]研究表明, 段首信息单元相对于段末信息单元来说, 段首音节的音高值更高, 音节前无声段更长。王蓓等^[2]在语料库研究的基础上指出, 信息单元越大, 边界处无声段越长且变化的自由度越大。

从深层上看, 由于语义的亲疏远近关系不同, 语

篇中的小句之间会形成复杂的层级结构。国外学者已有研究表明, 语篇的这种深层结构与其韵律表现之间也具有较好的一致性。Noordman 等^[3]最早利用修辞结构理论来分析语篇的层级结构对声学参数的影响, 并得出边界处无声段和小句基频最大值随层级的递减而递减的结论。随后 Ouden 等^[4]利用新闻文体的语料库进行研究, 也得出了相似的结果。本研究拟在语料库研究的基础上, 借鉴修辞结构理论

收稿日期: 2009-03-13

基金项目: 国家自然科学基金资助项目 (60775026)

作者简介: 杨晓虹(1984—), 女(汉), 广东, 博士研究生。

通讯联系人: 杨玉芳, 研究员, E-mail: yangyf@psych.ac.cn

分析语篇中的语义关系,探讨汉语语篇语义层级结构边界韵律表现。

修辞结构理论(rhetorical structure theory, 简称RST)是Mann和Thompson^[5]所创立的理论,目的在于描写自然语篇的结构,为计算机程序提供理论依据。修辞结构理论将语篇的最小信息单元定义为小句(elementary discourse unit, 简称EDU),小句与小句之间依靠修辞关系连接起来。修辞关系总体上可以分成两类:一种是主次关系,如条件关系、阐述关系、让步关系、因果关系等;另一种为并列关系,如对比关系、选择关系、联合关系等。主次关系中比较主要的小句称为核心句,另一个小句称为卫星句(在图1中箭头所指向的句子为核心句)。并列关系中两个句子不分主次,同为核心句。修辞关系把两

个小句连成一个结构段(span),结构段之间同样根据修辞关系形成更大的单元,依次类推,直到将整个语篇连接成一个完整的树形结构。以下是一个用修辞结构理论标注的例子(见图1):

- 1) 我国古代就有因材施教的教学思想。
- 2) 孔子对自己的学生就很了解,
- 3) 他能够说出学生的性格特点和智力水平,
- 4) 并且针对不同的特点,
- 5) 用不同的方法进行教育,
- 6) 把学生培养成各种不同的人才。
- 7) 宋代朱熹在《论语》的注解中指出:
- 8) “孔子教人,各因其材。”
- 9) 这就是“因材施教”一词的来源。

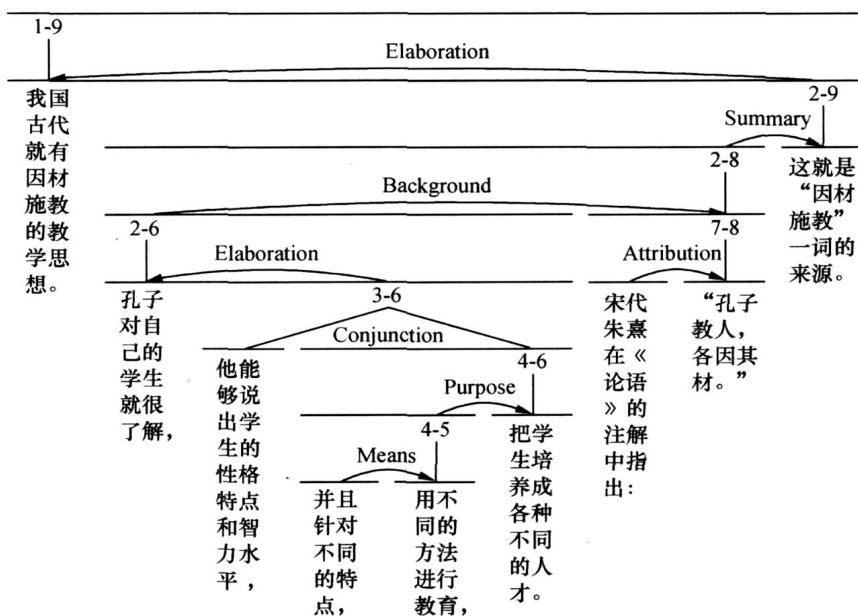


图1 语篇修辞结构的例子

在修辞结构理论的体系下,语篇是由小句间深层语义关系生成的一个层级结构。这种层级结构的边界在汉语的语篇中是否也有相应的韵律表现?制约其韵律表现的因素是什么?本文采用一个由20个自然语篇构成的语料库,用RST理论标注出语篇的修辞结构,考察小句边界处无声段,边界前后音节的音高重置,边界前音节延长这三个声学参数随语义层级结构的变化情况。

1 语料库的录制和标注

1.1 实验语料的录制

实验语料采取的是20个说明文,均为自然语篇,字数在400~500之间。为避免经由RST标注软件标注过后语篇的层级结构差异太大,本研究所采

用的20个说明文语篇是从50个说明文语篇语料库中挑选出来的,挑选标准为标注后的修辞结构层级数目在7至9层之间。一位女性专业的播音员朗读了所有的语篇,年龄22岁。语速适中,在正确理解语篇的基础上以自然的方式朗读。录音在中国传媒大学的录音室完成。

1.2 语料库的声学参数提取

声学参数的提取包括边界前音节发音结束到下一小句首音节发音开始前的无声段的长度、小句边界处前后音节的高音点和低音点、小句边界前音节的时长。声学测量用Multispeech 3700语音分析软件完成。阴平只取高音点,上声只取低音点,阳平和去声分别取高音点和低音点。

1.3 语料库的标注

语料库的 RST 标注由本文第一作者借助 RSTTool345 分析软件, 由下往上根据修辞关系构建语篇树形结构图。为保证标注质量, 在第一遍标注完成以后对所有语篇进行了详细的审查, 并同时由本文第二作者进行校对。

1.4 边界层级计算

对小句边界在树形图中的层级位置做如下计算: 边界的嵌套深度由边界前后小句的嵌套深度叠加而成。从小句在层级中的位置开始向上数, 其结构段(span) 数量就代表小句的嵌套层级。边界前后两个小句的嵌套层级相加得到边界的嵌套层级。按照这种方法计算, 嵌套越深的小句边界, 也就是处于越底层的小句边界, 其层级数越大。按照这种方法计算, 图1中小句2的嵌套深度为5, 小句3的嵌套深度为6, 这样小句2和小句3之间的边界嵌套层级为两者之和11。同理, 小句3和小句4之间的边界嵌套深度则为14。在我们20个语篇的语料库中, 小句边界层级数目处于6~18之间, 为方便统计, 将这些层级数目分成三个大的层级: 6~9为层级一; 10~13为层级二; 14~18为层级三。由于等级数目越大代表在树形结构中嵌套的越深, 因此这里的层级一、二、三实际上代表的是层级在语篇中的位置, 分别对应于树形结构中的上、中、下位置。

2 结果和分析

2.1 三个声学参数随层级的变化情况

由于三个层级间边界的个数不同, 如层级一有97个边界, 层级二有241个边界, 层级三有165个边界, 为了使统计结果更为精确, 将层级二和层级三的边界数目随机删减为97个左右。下面首先考察层级一、二和三的边界处无声段, 音高重置, 和边界前音节延长这三个参数随着修辞结构层级的不同而变化的情况。

2.1.1 RST 层级边界的无声段表现

表1给出了语篇修辞结构的小句边界处无声段长度随三个层级的变化情况。

表1 无声段随层级的变化

层级	平均值/ms	标准差/ms
1	672.91	313.49
2	604.57	283.19
3	502.15	280.23

对表1的结果进行单因素方差分析, 结果是不同层级间无声段有显著差异, $F(2, 281) = 8.150$, $p < 0.001$ 。多重比较的结果是: 层级一和层级二边界处的无声段分别与层级三之间差异显著($p < 0.001$), 层级一和层级二之间差异不显著。这说明, 边界处无声段的值会随着层级的升高而延长。但是无声段的延长也是有一定幅度的, 处于树形结构最底层的层级三的无声段长度明显短于以上两层的无声段长度, 而层级二之后虽然平均数有所增长, 但是不再出现显著的差异。

2.1.2 RST 层级边界的音高重置的表现

根据沈炯^[6]的双线语调模型, 语调短语高音线的变化与语势有关, 而低音线的变化则与节奏的完整性有关。本研究所采用的修辞结构理论标注语篇的结果在深层上是一种语义的层级, 而在表面上又是具有节奏的层级, 因此, 我们分别对高音点的重置和低音点的重置做了考察(结果见表2)。

表2给出了语篇修辞结构的小句边界处高音点重置和低音点重置随三个层级的变化情况。

表2 边界处音高重置随层级的变化

层级	高音点		低音点	
	平均值/Hz	标准差/Hz	平均值/Hz	标准差/Hz
1	142.74	82.93	108.18	56.69
2	131.75	86.78	114.61	61.86
3	87.17	99.89	86.62	55.21

对表2的数据进行单因素方差分析的结果显示, 对于不同的层级来说, 边界处高音点重置($F(2, 188) = 6.781$, $p < 0.005$)差异显著。多重比较结果显示, 对高音点重置来说, 层级一和层级二差异不显著, 层级一和层级二分别与层级三之间差异显著。而边界处低音点的重置差异不显著($F(2, 137) = 2.960$, $p > 0.05$)。

这说明, 边界处高音点的重置较之低音点重置来说更能够反应这种语义的层次性结构。高音点的重置会随着层级的升高而重置程度加大, 但是, 高音点的重置的程度也是有限度的, 从层级二到层级一虽然平均数上有所增长, 但是差异没达到显著性水平。由此可以看出, 边界处音高重置的程度增大到一定程度以后就不会出现显著的增大。

2.1.3 RST 层级边界的边界前音节延长的表现

表3给出了语篇修辞结构的小句边界前音节延长随三个层级的变化情况。

表3 边界前音节延长随层级的变化

层级	平均值/ms	标准差/ms
1	223.87	45.76
2	223.02	48.63
3	236.58	58.54

对表3的数据进行单因素方差分析显示,各层级的边界前音节延长程度差异不显著($F(2, 296) = 2.183, p > 0.05$)。可见,对于以小句为基本单元的语篇修辞结构来说,边界前音节延长不是一个的有效线索。

2.2 三种层级边界所对应的边界类型

在以上的分析中可以看到,虽然小句边界处无声段和高音点重置的值会随着边界层级的递增而递增。但是,这种线性的趋势在到达层级二之后就不再出现特别显著的增加,是什么原因造成这种现象呢?以下我们试图从RST各层级结构所对应的边界类型来给出解释。我们将在RST层级结构中的边界分成三类: a类为句群与句群之间的边界; b类为句群与小句之间的边界; c类为小句与小句之间的边界。从边界的强弱程度来看,句群与句群之间的边界应该是最强的边界,而小句与小句之间的边界程度是最弱的,句群与小句之间的边界居中。我们将这20个语篇中的503个边界分别按照a、b、c三种标准进行归类,对这三种边界在语篇RST层级一、二、三(也就是树形结构中位置上、中、下)中的分布百分比情况做一统计,结果如图2所示:

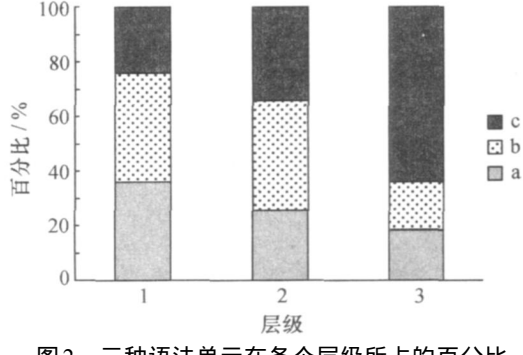


图2 三种语法单元在各个层级所占的百分比

从图2可以看出,对层级三,也就是嵌套在RST结构中最底层的层级来说, c类边界所占的百分比最多。随着层级的递增, c类边界所占的百分比逐渐减少,而a类和b类边界所占的百分比逐渐增多。而c类边界是比较弱的边界,这也可以解释为什么第三层级的边界处无声段长度和高音点重置的值远远低于上面两个层级。从图2还可以看出,对层级一

和层级二来说,三种边界分布的百分比比较接近,这也正是边界处无声段长度和高音点重置在这两个层级上没有显著差异的一个重要原因。

3 讨论

本文利用修辞结构理论探索语篇层级边界的韵律表现,实验结果是随着边界在语篇中层级的递增,边界处无声段和高音点重置的值也递增;同时,边界处低音点重置和边界前音节延长不是区分语篇修辞结构的有效线索。进一步分析显示,边界处无声段和高音点重置的值并不是无限制增长的。这一点发现也弥补了Noordman等^[3]和Ouden等^[4]的研究。这两位学者的研究结果都是认为边界处无声段长度和小句的基频最大值会随着修辞结构层级的渐增而渐增,但并没有对层级内部的变化进行细化研究,因此也未能说明语篇中这种渐增趋势是有限的还是无限的。同时,本文还结合汉语语言的特性采用双线模型进行分析,从而从实验的角度证实了边界处高音点重置和低音点重置在区分语篇层级上的功能不同。

处于RST层级最低的边界无声段和高音点重置的值远远低于上面两个层级边界的值,这可能是由于嵌套在RST结构中最底层的往往是小句与小句之间的边界,而这种边界一般比较弱,其声学参数的值也因此较低缘故。而层级结构越往上,则更多的出现句群和句群之间的边界以及句群和小句之间的边界,并且这两种边界在中层和上层的分布都较为均匀,因此也导致上面两层的声学参数的值不会出现显著变化。

值得探讨的是,根据修辞结构理论所产生的韵律边界是由语义生成的,与根据语法单元大小而产生的韵律词、韵律短语、语调短语和句子这样的韵律边界存在本质上的不同。以往对根据语法单元大小划分出来的韵律边界进行声学线索探索的研究不少,如王蓓等^[7]的研究得出,韵律词边界的声学线索是低音线的不连续性和边界前音节的延长,一般没有无声段。韵律短语和语调短语边界的声学线索是低音线重置和无声段。并且,边界等级越高,低音线重置程度越大,无声段的长度也越长。知觉等级与无声段长度成对数增长关系。

而将本研究结果与前人的研究结果进行对比,基于语义关系划分的层级关系和基于语法单元大小划分的层级关系的声学表现之间既存在相同之处也存在不同之处。相同之处在于,边界处无声段对区分

两种层级结构来说都是有效的线索。不同之处在于: 首先, 对基于语义划分的语篇结构来说, 高音点重置比低音点重置更能体现边界的层级性, 这可能是因为高音点比低音点变化的幅度更大, 也可能是因为高音线与语义更相关的缘故; 再者, 在语篇结构中, 无声段和高音点重置的增长是有限的, 不会随着层级的增长而无限制增长; 最后, 边界前音节延长对区分语篇结构来说也不是一个有效的区分线索。

本文对边界前音节延长的统计表明, 边界前音节延长量不会随着层级的升高而出现显著的增加。这可能是因为以修辞结构理论来标注的语篇时把语篇的最小单元定义为小句(EDU), 而小句一般等同于语调短语, 这也就使得最后标注出来的层级变化是以小句为最小信息单元通过修辞关系连接逐渐变化为更大的信息单元的。而边界前音节的延长主要是弱韵律边界的声学线索, 如杨玉芳^[8]的研究表明, 边界前音节时长随边界等级的变化是双向的, 在短语边界处达到最大。而以小句为最小信息单元的层级结构显然超出了边界前音节所能起作用的范围。

4 结 论

1) 对于以小句为最小信息单元的语篇修辞结构来说, 其韵律边界是通过边界处无声段长度和边界处高音点重置来体现的。但是, 边界处无声段和边界处高音点重置在语篇水平上的增长也是有限度的。边界类型的分布对语篇树形修辞结构的韵律表现起到很大的影响作用。

2) 低音点重置和边界前音节延长不是区分语篇水平上语义层级结构的有效线索。

参考文献 (References)

- [1] Swerts M. Prosodic features at discourse boundaries of different strength [J]. *Journal of the Acoustical Society of America*, 1997, 101: 514 - 521.
- [2] 王蓓, 杨玉芳, 吕士楠. 语篇中大尺度信息单元边界的声学线索 [J]. *声学学报*, 2005, 30(2): 177 - 183.
WANG Bei, YANG Yufang, LU Shinan. The acoustic characteristics of large information units' boundaries in monologue discourse [J]. *Acta Acustica*, 2005, 30(2): 177 - 183. (in Chinese)
- [3] Noordman L, Dassen I, Swerts M, et al. Prosodic markers of text structure [C]//van Hoek K, Kibrik A, Noordman L. *Discourse Studies in Cognitive Linguistics*. Amsterdam: Benjamins, 1999: 133 - 145.
- [4] den Ouden H, Noordman L, Terken J. Prosodic realizations of global and local structure and rhetoric relations in read aloud news reports [J]. *Speech Communication*, 2009, 51(2): 116 - 129.
- [5] Mann W C, Thompson S A. Rhetorical structure theory: Towards a functional theory of text organization. *Text*, 8(3): 243 - 281.
- [6] 沈炯. 北京话声调的音域和语调 [C]//林焘, 王理嘉. *北京语音实验录*. 北京: 北京大学出版社, 1985: 73 - 130.
SHEN Jiong. The Pitch range and intonation of Beijing tones [C]//LIN Tao, WANG Lijia. *Recordings of Beijing Phonetic Experiment*. Beijing: Beijing University Press, 1985: 73 - 130. (in Chinese)
- [7] 王蓓, 杨玉芳, 吕士楠. 汉语韵律层级结构边界的声学分析 [J]. *声学学报*, 2004, 29(1): 29 - 36.
WANG Bei, YANG Yufang, LU Shinan. A acoustic analysis on prosodic hierarchical boundaries of Chinese [J]. *Acta Acustica*, 2004, 29(1): 29 - 36. (in Chinese)
- [8] 杨玉芳. 句法边界的韵律学表现 [J]. *声学学报*, 1997, 22(5): 414 - 421.
YANG Yufang. Prosodic cues to syntactic boundaries [J]. *Acta Acustica*, 1997, 22(5): 414 - 421. (in Chinese)

(上接第1374页)

参考文献 (References)

- [1] Casey M, Veltkamp M, Goto M, et al. Content-based music information retrieval: Current directions and future challenges [C]//*Proceedings of IEEE*. 2008, 96(4): 668 - 696.
- [2] Deutsch D. *The Psychology of Music*, Second Edition [M]. San Diego, CA: Academic, 1999.
- [3] 李重光. *音乐理论基础* [M]. 北京: 人民音乐出版社, 1979.
LI Chongguang. *Basic Theory of Music* [M]. Beijing: People's Music Publishing House, 1979. (in Chinese)
- [4] Fujishima T. Realtime chord recognition of musical sound: A system using common lisp music [C]//*Proceedings of the International Computer Music Conference*. Beijing, 1999.
- [5] Harte C, Sandler M. Automatic chord identification using a quantised chromagram [C]//*Proc. Audio Eng. Soc. Spain*, 2005, paper Nr. 6412.
- [6] Sheh A, Ellis D. Chord segmentation and recognition using EM-trained hidden markov models [C]//*Proc. Int. Conf. Music Inf. Retrieval*. Baltimore, 2003: 185 - 191.
- [7] Temperley D. *The Cognition of Basic Musical Structures* [M]. Cambridge, MA: MIT Press, 2001.
- [8] Brown J. Calculation of a constant Q spectral transform [J]. *Acoust Soc Amer*, 1991, 92(1): 425 - 434.
- [9] Zhu Y, Kankanhalli M. Precise pitch profile feature extraction from musical audio for key detection [J]. *IEEE Transactions on Multimedia*, 2006, 8(3): 575 - 584.
- [10] Krumhansl C. *Cognitive Foundations of Musical Pitch* [M]. New York: Oxford University Press, 1990.
- [11] Chew E. *Towards a Mathematical Model of Tonality* [D]. Cambridge, MA: MIT, 2000.
- [12] Bello J, Daudet L, Abdallah S, et al. A tutorial on onset detection in musical signals [J]. *IEEE Transactions on Speech and Audio Processing*, 2005, 13(5): 1035 - 1047.
- [13] Goto M. A predominant-F0 estimation method for CD recordings: MAP estimation using EM algorithm for adaptive tone models [C]//*Proc. ICASSP 2001*. 2001: 3365 - 3368.
- [14] Duda R, Hart P, Stork D. *Pattern Classification*, Second Edition [M]. John Wiley & Sons, Inc., 2001.