

自然话语中语句边界的韵律特征及其交际功能研究

熊子瑜

中国社会科学院语言研究所

xiongzuyu@163.com

摘要

本文基于 973 电话语料库,从话语交际的角度考察自然会话过程中的语句边界特征,运用 SPSS 软件统计语句边界处的时长和音高两个方面的韵律特征,比较不同类型语句边界在韵律表现上的差异性,并在此基础上探讨韵律边界特征的两种话语交际功能:言语行为功能和话轮提示功能。通过研究得出以下结论:在自然语流中,切实存在着可以感知和测量到的语句边界韵律特征;不同的边界韵律特征具有不同的话语交际功能,可以为交际者判断对方的交际意图以及把握话轮行进状态提供必要和有益的线索。

关键词:话轮 言语行为 韵律

1. 引言

话语的韵律结构和韵律表现在话语交际过程中起着很重要的作用,它会影响听话人对话语结构的感知和对意义的理解,正确把握话语的韵律结构,是话语正常交际的前提,同时也是实现人机对话和人工智能的基础。话语的边界韵律特征,为连续语流的切分提供了一个必要的手段。声学实验表明,成年人能够有效地控制时长和基频以表示短语的边界。知觉实验表明,词的时长和基频是划分短语的两个显著的知觉线索。

汉语学界在语句边界的韵律特征方面,已经作了大量的研究工作。林茂灿先生认为语句韵律成分包括韵律词和韵律短语,人们说话时是分别通过简单韵律词,复杂韵律词和复合韵律词三种韵律成分,把词汇词组成韵律短语和语句的。韵律词内部的 F0 变化(韵律词是 F0 变化组)和韵律词和韵律词之间的 F0 重设是韵律词的声学标志;韵律词是口语语调的基石;韵律词对口语的自然悦耳起关键作用。韵律短语不仅有以无声间断为边界的,而且有以有声间断(边界前音节拉长)为边界的。李爱军老师认为汉语的韵律结构是一种层级结构,从小到大是音节、音步、韵律词、次要韵律短语、主要韵律短语和语调组。韵律词是有一个词重音的变调组,韵律短语由一个或几个韵律词组成,如果短语后感知到的停顿较短,它与后面一个短语的连接较紧密,则将此短

语标为次要韵律短语;如果短语后感知到的停顿较长,它与后面一个短语的连接较松散,而且有基频重设等现象,则将此短语标为主要韵律短语。若干个基频曲线依次递降的主要短语组成语调组。熊子瑜的研究表明“语流间断的层级越高,其无声停顿的时间就越长;语流间断的层级越低,其无声停顿的时间就越短。韵律词末尾不出现可以感知的无声停顿。”

熊子瑜还考察了基频重设的程度与韵律间断层级之间的关系,其数据表明,“基频重设的幅度与语流间断的程度二者关系密切,基频的重设幅度越大,听感上的间断程度就越高。”吴洁敏等以句子为中心将汉语的语音层次区分为内三层和外三层,句内包括音节、音步和气群,句外包括句群、段落和篇章,并以此为基础将汉语停延区分为句内停延和句外停延两类,并认为“汉语停延时值往往随着逐级递升的语音层次而不断增加,或者随逐级递降的语音层次而逐渐减小。”

这些实验结果都反映出一个事实,在语流中切实存在着可以感知到的不同等级的韵律边界特征。过去的这些研究大多是基于实验语句和朗读语料进行的,尽管在一定程度上反映了朗读语句或实验语句的韵律边界特征,但不能很好地解释自然口语中出现的某些韵律现象。比如说在电话录音材料中,我们发现在很多情况下语句末尾并不都出现语音停顿,反而在语句内部有时可能会出现较长的语音停顿。这与我们过去的结论很不一致,甚至是矛盾的。那么如何解释这类韵律现象呢?我们认为这可能是因为朗读语句和实验语句是在控制了诸多因素的情况下录音的,它几乎排除了句法结构之外的其他影响因素,因此所得出的结论只能反映出韵律边界特征的静态表现。而在实际的自然交际中,由于受到外在偶然因素和话语交际因素的影响和作用,话语的边界韵律特征会发生一定程度上的变化。

自然口语交际是人类最常用的言语交际形式,其使用的通常是即兴的、没有事先准备的口头交际话语。自然口语交际一般具有实时性、现场性、交互性的特点。受这些特点的作用,使自然口语交际具有以下一些特征:1、自然口语交际中语句单元比较简短,结构比较简单;2、自然口语交际中语句重复罗嗦的现象比较多,信息冗余度较高;3、自然

口语交际中语句简略、不完整的现象比较多；4、自然口语交际中语序安排相对较灵活；5、自然口语交际中话题跳转比较自由；6、自然口语交际中说话人的话语风格和交际策略可能要受到对方的影响而发生变化。自然口语交际的这些特点会对话语的韵律表现产生影响作用。比如说，在自然口语交际中，因交际双方同时存在，交际角色（说者和听者）不断轮换，交际双方有时还存在着话轮竞争，为了能够继续说下去，对于说话人来说，在非终端语句的末尾就不能有太大的语音停顿，否则就容易听话人误认为已经说完。还比如说，当说话人发现听话人心不在焉或者不感兴趣时，往往会通过提高音量的方式吸引对方的注意或以加快语速的方式尽快结束自己的话语。再比如说，有观点认为，语调与标示情感的音高运用直接相关。高而升的音高表示感兴趣、激动和不完整；而低而降的音高表示无兴趣、结束和休息。高和低（升和降）的基本区别在使用音高范围表达感情过程中可以明显地看出来：感情活跃时，如愤怒、惊讶，声音提高；而当厌烦、悲伤之类时声音降低。这表明，自然话语的韵律特征受人际交互因素的影响会产生不同的变化。另一方面也表明，自然话语的韵律特征具有极其重要的交际功能。因此，对自然话语的韵律特征进行研究，就不仅要考虑话语的句法结构这一静态因素对其韵律结构的影响和约束，还要充分考虑自然话语交际的各种动态因素对话语韵律特征的影响作用，在自然话语交际中研究韵律特征的话语交际功能。

2. 语料和标注

本研究材料取自 973 电话录音语料库，该语料库由中国科学院自动化研究所录制，由中国社科院语言所语音室标注。整个语料库录制内容是旅社服务员和客户之间的自然对话，由 94 个独立的语音片段组成，具体内容涉及订房、订票和问讯等方面。

2.1 韵律层面标注

基于 C-TOBI 系统的原则和方法，由六位受过专门语音标注训练的语音工作者对这些电话录音材料进行了音段和韵律的切分、标注、核查和修改工作，主要标注出以下几个方面的信息：1、音节的切分与标注；2、声韵母的切分与标注；3、韵律层级及韵律边界指数的标注；4、重音层级及重音层级指数的标注；5、发音人和话轮转换的标注；6、语音杂类现象的标注；7、语句的语气功能标注。

2.2 话语层面标注

在韵律标注的基础上对这些电话录音材料进行了话语层面的信息标注，以考察韵律特征的话语交际功能，这主要包括以下内容：

2.2.1、区分言语声、副言语声和非言语声

言语声就是指交际过程中交际者所说的与交际主旨有关的话语，并包括话语之间的正常语音停顿。副言语声就是指说话者在交际过程中发出的反映其生理、心理等状态的声音。如笑声、吸气声、哈欠声、哭泣声等。因为这些声音会直接影响话语的韵律表现，同时也反映了说话者的交际状态，并会对另一方的交际行为产生某些直接的影响，因此需要分别标出。非言语声就是指与言语交际活动不发生直接关系的、不是由说话人自己发出的声音。如关门声、翻纸弄出的声音、电话忙音以及其他一些背景噪音等等。尽管这些声音一般不与交际活动发生直接的关系，属于偶然的、不确定的干扰因素，但它同样会影响到说话人的说话方式和交际行为，对话语的韵律表现产生影响和改变，因此需要单独标明。

这些信息中有的已经在韵律标注的语音杂类中表明，在此只是将某些语音现象进一步细分。比如说，根据产生原因的不同将语音上的无声段进一步细分为：1、发塞音和塞擦音时，爆破前的持阻过程留下的无声段；2、因话语的韵律结构需要而产生的无声段；3、因思考或犹豫等语用表达上的需要而产生的无声段；4、因忙于其他活动（如服务员登记客人信息）而顾不上或忘记说话等原因而产生的无声段；5、因录音控制造成的无声段。尽管这些无声段的在语图上都表现为空白段，但背后的产生原因却不同，所传递的信息也就不同，因此需要区分而不能混在一起，否则就会影响数据的一致性。

2.2.2、对言语声区分话轮构造单元和非话轮构造单元

日常会话中有一条会话者自然地、无意识地遵守着的规则：每次至少有一方，但又不多于一方在说话，也就是轮流发话。轮流发话中发话人的话语从开始到结束，就是一个话轮，话轮是会话的基本单位。话轮与话轮之间的转换一般出现在会话的“转换关联位置（transition-relevance place）”上，“转换关联位置”指在句法和语义上可能为听话人识别为话轮结束的位置。在会话过程中，这种由“转换关联位置”断开的语段就是“话轮构造单元（turn-constructive unit）”。构成“话轮构造单元”的语言单位可以是单词、短语、分句或句子。

只有在一个人占据了说话权时，或至少是占据了说话的主动权时，其所说的话才可能构成一个话轮，如果在他说话时，并没有占据主动地位，对方也没有转移说话权的意思而仍旧继续说话，则这个人的话语就不构成一个话轮，而属于话轮外成分。比如在对方说话时，听话人为了表示在听对方的话语而发出的“嗯、啊、好的、行、对、是的、是吗”等等，这些话是对说话人的话语做出的反馈，一方面表示在倾听或明白对方的话语，另一方面还表示示意对方继续说话，而不是要抢占对方的话权，也没有打断对方话头的意图，因此不单独构成话轮，属于话轮外的言语成分，在这里，我

们且称之为“非话轮构造单元”。当然,如果是对方以提问的方式要求作出回答时,此时用作肯定性答语的“嗯、啊、好的、行、对、是的”等又另当别论,因其占据了话权,故应该看作是话轮构造单元。除了这种反馈声之外,非话轮构造单元性成分还包括用来填充停顿的言语成分,如“嗯、哎、啊”等等;还包括一些不完整的言语片段,如说话人有时在一句话还没说完就放弃而重新开始了另一句话,或者还没有说完就被某种外因打断而说不下去,这样形成了一些不完整的言语片段,我们也把它们看作是非话轮构造单元。总之,话轮构造单元和非话轮构造单元的基本区别在于,话轮构造单元在一定的语境中可以作为话轮而独立运用,而非话轮构造单元则在任何条件下都不能作为一个话轮而独立运用。

2.2.3、在话轮构造单元中区分不同的言语行为类型

言语是信息的载体,言语交际过程就是言语信息的传递过程。如用叙述的方式来传递说话人自认为对方不知道的、并期望了解的未知信息,用提问的方式来传递说话人自己不知道的、并期望对方能够给予回答的未知信息,用命令的方式用来传递说话人希望对方能够实施某种行为的信息。受话人在接收到说话人所传递出的信息后,针对不同的信息类型会做出不同的反应。对于第一类信息,通常会将它作为新信息存储起来。对于第二类信息,通常会根据自己所了解的情况做出相应地解答。对于第三类信息,通常会在力所能及的范围内实施说话人所期望的行为活动。根据所负载信息的类型和听话人可能作出的反应,可以将话轮构造单元所实施的言语行为大致区分为三种类型:叙述、提问和命令。

除了这三种类型之外,还有用来传递友好信息、建立人际关系的“招呼语”,如“你好、再见、谢谢、不客气、没什么”等等表示打招呼或表示礼貌的固定用语,这种言语行为我们称之为“打招呼”。此外将某些确实很难归入上面四类的言语行为类型统统归入“其他”类。尽管在标注说明中可以将言语行为类型区分为几种独立的类型,但在实际标注中,却存在许多交叉的情况,如说话人有时以“提问”的方式来执行“命令”的行为,有时以“命令”的方式来达到“提问”的目的等等,很难一刀切。为了在一定程度上减少这类交叉的现象,有必要对这四种基本的言语行为类型作进一步的细分和标注。因此我们将提问进一步细分为:回声问和非回声问,再在非回声问中区分有特殊疑问点的提问,无特殊疑问点的提问,正反选择方式提问,备选方式提问等不同类型;将叙述进一步细分为两类:客观性的,如对事态的描述、说明等,和主观性的,如表示赞同、反对等等;将命令也进一步细分为两类:以提问方式表达的间接命令和以祈使方式表达的直接命令。

2.2.4、在非话轮构造单元中区分不同的交际功能

非话轮构造单元的言语成分在交际过程中同样具有极其重要的交际作用,能够传递交互信息,表示说话人的交际状态。根据其交际功能的差异,大致分做以下几种类型:背景反馈声、填声停顿的言语成分、不完整的言语片段和其他。这些在 2.2.2 中已有所说明,在此不再解释。

下面是一段根据语音标注和话语标注的内容转写成文本形式的具体例子,标注符号说明请参见本文最后一页:

- T1 C1 A: U1[H-S-K]{有}| ,
C2 U2[H-X-T]您要|什么样的|房间||?
T2 C3 B: U1[H-S-K]单人间||。(S-J .52)
T3 C4 A: U1[H-S-K]单人间有| ,
C5 U2[H-S-K]一百三(LE .35)||。(S-J .20)
T4 C6 B: U1[H-X-H-S]一百三||?
T5 C7 A: U1[H-X-T]什么时候|过来||?(S-J .13)
C8 U2[H-S-Z-T]对|。
C9 U3[H-X-T]您什么时候|过来
||?(S-J .20)
T6 C10 B: U1[H-S-Z-S]我|大概(LE .37)|| (UC)一
个小时|左右吧||。(S-J .85)
T7 C11 A: U1[H-S-Z-M]好的| ,
C12 U2[H-X-T]您贵姓(LE .38)||?
T8 C13 B: U1[H-S-K]我姓(UC)文||。(S-J .49)
T9 C14 A: U1[H-X-H-S]姓文哈|?
C15 U2[H-S-Z-M]好的| ,
C16 U3[H-S-K]您|直接过来|就成。
T10 C17 B: U1[H-S-Z-M]好| ,
C18 U2[H-Z]谢谢啊||。
T11 C19 A: U1[H-S-Z-M]嗯| ,
C20 U2[H-Z]没什么||。

以上是根据第 1 个语音片段的标注数据转写而成的文本文件,从标注转写的数字可以看出,这个言语交际过程发生在客人 B 和服务员 A 之间,话语交谈围绕订房这一活动展开,内容涉及房间的标准、价格,入住时间,客户登记以及人际关系等诸多方面。整个交际过程由 11 个话轮构成,这 11 个话轮总共含有 20 个话轮构造单元,有的话轮由一个话轮构造单元组成,有的由两个或三个话轮构造单元构成。从这些转写的数字还可以看出,有的话轮之间有语音停顿,有的没有语音停顿,还有极少数可能会发生语音叠接的现象。此外,还可以大致看出,有的话轮构造单元末尾有语音停顿,有的末尾没有语音停顿。进一步观察还会发现,处于话轮起始和中间位置的话轮构造单元的末尾几乎很少出现语音停

顿,而在话轮构造单元内部却可能出现较长的语音停顿。这类语音现象在朗读话语或实验语句中几乎不可能出现,有时即便出现了,也往往是被当作例外而不予认真考虑,甚至可能被有意抹去。

3. 统计分析

以下是根据当前已经标注好了的、并做了初步检查的 83 个语音片段数据所做出的一些统计结果:

3.1 语音文件的时长统计

文件数	83
最小值	12.60
最大值	211.96
平均值	75.80
标准差	44.23
总和	6297.80

从上表的数据可以看出,所标注的语音文件总长大约 105 分钟,平均起来一次电话订房(或订票、问讯)活动的时间很短,大约只需要 76 秒,这与其他场合和情境下的言语交际活动有很大的差异,这大概是因为:一、电话交际按时计费,因此闲话、重复罗嗦的话相对较少;二、本语料所涉及的都是公务交际,大多直接进入主题,客套寒暄的话语较少。尽管每场交际活动都比较短小,但个体之间的差异比较大,这表明样本比较离散,如最短的一次言语交际活动总共不到 13 秒,而最长的达到了 212 秒,相差 16 倍之多。观察发现,不同内容的交际活动所需要的时间有一定的差异,就上面所说的三种不同的交际活动来说,问讯所需要的时间最短,而订票所需的时间最长。

3.2 语音文件中的音节数目统计

文件数	83
最小值	61
最大值	1006
平均值	291
标准差	184
总和	24129

从上表的数据可以看出,所标注的语音文件总共含有 24129 个音节,平均起来一次电话订房(或订票、问讯)活动大约需要 291 个音节,但标准差比较大,这表明样本比较离散,个体之间的差异比较大,如最短的一次言语交际活动(问讯)只用了 61 个音节,而最长的(订票)达到了 1006 个音节,相差也有 16 倍之多。

3.3 语音文件中的话语单元数目统计

从下表的数据可以看出,一次电话订房(或订票、问讯)活动平均起来大约由 63 个话段组成,其中包括 47 个话轮构造单元和 16 个非话轮构造单元,这 47 个话轮构造单元组成

了 34 个话轮。另外,从上表的数据还可以看出,已标注的语音文件中总共含有 5263 个话段,其中包括 3937 个话轮构造单元和 1326 个非话轮构造单元,分别占话段总量的 75% 和 25%。

语音文件中的话语单元	文件中话段的数目	文件中话轮构造单元的数目	文件中话轮的数目
文件数	83	83	83
最小值	12	10	8
最大值	168	118	90
平均值	63	47	34
标准差	34	25	17
总和	5263	3937	2822

3.4 非话轮构造单元的类型分布

在 1326 个非话轮构造单元中,反馈声、填声停顿的言语成分、不完整话段以及其他等类型各自出现的次数和所占的比重如下表所示:

	反馈声	填声停顿成分	不完整话段	其他	总和
次数	597	531	96	102	1326
比率	45%	40%	7%	8%	100%

3.5 话轮构造单元的言语行为类型分布

如前所述,我们根据话轮构造单元所传递的信息内容将话轮构造单元所实施的言语行为大致区分为四种类型:叙述、提问、命令和打招呼。下表给出了这些不同类型的话轮构造单元的分布情况,其中,我们将主观性叙述和客观性叙述分开统计,将直接命令和间接命令归在一起进行统计,将提问合并为三种类型:回声问、是非问和非是非问。次数是指在某种类型的话轮构造单元在所标注的 83 篇材料中总共出现的个数,频数是指该种类型的话轮构造单元在话轮构造单元总量中所占的比重。

	叙述		命令	提问			招呼	合计
	主观叙述	客观叙述		回声问	是非问	非是非问		
次数	846	1384	94	498	424	369	322	3937
比率	21.5%	35.2%	2.4%	12.6%	10.8%	9.4%	8.2%	100%

从上表的数据可以看出,叙述和提问两种言语行为类型的话轮构造单元合起来将近总量的 90%,占绝大多数,而命令和招呼这两种言语行为的出现比率都比较低。究其原因,可能是因为电话交际需要按时计费,因此客套的话比较少,另外由于交际双方都不太熟悉,出于礼貌策略,故一般较少使用命令的方式。

3.6 话轮中所包含的话轮构造单元的数目统计

根据前面的数据可知,在现有的标注材料中,一共有 3937 个话轮构造单元,它们总共组成了 2822 个话轮,平均

起来 1.4 个话轮构造单元组成 1 个话轮。下面统计了话轮中包含话轮构造单元的分布情况：

话轮中所包含的话轮构造单元的数目 统计数据		
话轮所包含的 话轮构造单元 的数目	次数	比率
1	2023	71.7%
2	577	20.4%
3	159	5.6%
4	41	1.5%
5	16	0.6%
6	3	0.1%
7	3	0.1%
合计	2822	100%

从上表的数据可以看出，在电话语料中，一个话轮至少包含 1 个话轮构造单元，至多包含 7 个话轮构造单元。不同数量的话轮构造单元组成的话轮在频数上差异很大，由 1 个话轮构造单元组成话轮的出现几率为 71.7%，而由 6 个或 7 个话轮构造单元组成话轮的出现几率只有千分之一。另外，从上表的数据还可以计算出，97.7%的话轮至多只包括 3 个话轮构造单元。

3.7 不同类型的话轮构造单元的时长统计

下面按照话轮构造单元的言语行为类型及其所处位置的不同统计了话轮构造单元中的音节数目，以分析这两种因素对话轮构造单元的时长的影响。

3.7.1 言语行为类型对话轮构造单元的时长的影响

下表给出了不同言语行为类型的话轮构造单元中所包含的音节数目的统计数据，从中可以看出，音节数目最少的是招呼类，只有 2-3 个音节，而客观性叙述、命令、是非问和是非非问这四种类型相对较长，平均 7-8 个音节，主观性叙述和回声问处在中间，平均 3-4 个音节。从这个结果可以看出，话轮构造单元的音节数目的多少与其言语行为类型密切相关，与其传递的信息类型以及信息量的高低密切相关。

不同言语行为类型的话轮构造单元 所包含的音节的数目			
言语行为类型	平均值	次数	标准差
主观性叙述	3.1	846	2.3
客观性叙述	7.0	1384	2.9
命令	7.4	94	2.1
回声问	4.3	498	2.0
是非问	7.2	424	3.1
是非非问	7.4	369	3.3
招呼	2.4	322	1.1
合计	5.5	3937	2.4

3.7.2 话轮构造单元在话轮中所处的位置对其时长的影响

下表给出了话轮构造单元在话轮中所处位置对其音节

数目的影响数据。从中可以看出，处于不同位置上的话轮构造单元的音节数目也具有明显差异，中间位置的话轮构造单元最长，约 7 个音节左右，起首位置的话轮构造单元最短，约 4 个音节左右。从这些数据可以看出，话轮构造单元的音节数目的多少与其在话轮中所处位置之间有着密切的关系。

不同位置的话轮构造单元所包含的音节的数目			
所处的位置	平均值	次数	标准差
起始位置	4.2	618	2.4
中间位置	7.1	550	4.1
结束位置	5.6	2769	3.3
合计	5.5	3937	3.0

从以上统计数据可以看出，话轮构造单元的音节数目及其时长受其言语行为类型和所处位置制约。因此，根据话轮构造单元的音节数目和时长可以反过来在一定程度上预测该话轮构造单元的言语行为类型及其在话轮中所处的位置。

3.8 不同类型的话轮构造单元的语速统计

下面按照话轮构造单元的言语行为类型及其所处位置的不同统计了话轮构造单元中的语速快慢情况，以分析这两种因素对话轮构造单元的语速的影响。

3.8.1 言语行为类型对话轮构造单元的语速的影响

下表给出了不同言语行为类型的话轮构造单元的语速的统计数据，从中可以看出，命令和是非非问的话轮构造单元的语速相对较快，其内部音节的平均时长只有 160 毫秒，而主观性叙述和回声问的语速相对较慢，其内部音节的平均时长大约有 220-230 毫秒，其他几种类型处于中间状态。从这个结果可以看出，话轮构造单元的语速快慢与其言语行为类型密切相关。

不同言语行为类型的话轮构造单元的 语速统计数据（单位：秒）			
言语行为类型	平均值	次数	标准差
主观性叙述	0.23	846	0.09
客观性叙述	0.19	1384	0.08
命令	0.16	94	0.03
回声问	0.22	498	0.07
是非问	0.17	424	0.04
是非非问	0.16	369	0.05
招呼	0.18	322	0.07
合计	0.19	3937	0.08

3.8.2 话轮构造单元在话轮中所处的位置对其语速的影响

下表给出了话轮构造单元在话轮中所处位置对其语速的影响数据，从中可以看出，处于不同位置上的话轮构造单元的语速也有一定的差异，大致说来，处于话轮中间位置的话轮构造单元的平均语速最快，起首的最慢，但这些差异都不够显著。

不同位置的话轮构造单元的语速统计数据 (单位: 秒)			
话轮构造单元在话轮中所处的位置	平均值	次数	标准差
处于话轮起始位置	0.20	618	0.09
处于话轮中间位置	0.18	550	0.08
处于话轮结束位置	0.19	2769	0.08
合计	0.19	3937	0.08

从这些数据可以看出,话轮构造单元的语速与其言行为类型及其在话轮中所处位置之间有着一定的关联,因此理论上,了解话轮构造单元语速的快慢对于准确识别其言行为类型及其所处位置来说,具有一定的帮助作用。

3.9 不同类型的话轮构造单元的末尾语音停顿统计

下面按照话轮构造单元的言行为类型及其所处位置的不同分析了话轮构造单元末尾语音停顿的具体表现,以考察这两种因素对话轮构造单元末尾语音停顿的出现几率和停顿时长的影响。如果停顿时长大于 100 毫秒,就认为出现了一个可以感知到的语音停顿,才纳入统计,否则,若停顿的时长短于 100 毫秒,因其不可感知,在统计时就不把它看作出现了停顿。

3.9.1 不同位置的话轮构造单元的末尾语音停顿的统计数据

为了考察话轮构造单元在话轮中所处的位置对其末尾是否出现语音停顿的影响作用,下面统计了处于话轮中不同位置上的话轮构造单元末尾的语音停顿数据。在统计时,将那些只含有一个话轮构造单元的话轮中的话轮构造单元看作是处于话轮结束位置,而不把它放到话轮起始位置上进行分析。

3.9.1.1 不同位置的话轮构造单元的末尾语音停顿的出现几率统计

下表是不同位置的话轮构造单元末尾出现语音停顿的几率统计数据,从中可以看出,处于不同位置的话轮构造单元其末尾出现停顿的几率差异很大,处于话轮末尾的话轮构造单元出现停顿的几率是处于话轮中间和起始位置上的话轮构造单元末尾出现停顿的几率的 6-7 倍。对于这类现象的一种可能的解释是,在话轮构造单元的末尾一般都不出现语音停顿,而在话轮的末尾有相对较高的语音停顿出现几率。尽管话轮末尾出现语音停顿的几率相对来说比较高,但实际上并不高,也只有 51.3%,也就是出现语音停顿和不出现语音停顿的几率是一半对一半,没有什么明显的倾向性,而在前两种话轮构造单元的末尾却表现出了明显的倾向性,90%以上的都不出现语音停顿。因此,总得看来,在电话口语交际的话轮构造单元的末尾一般很少出现语音停顿,即便是在话轮结束处,也有一半左右的出现语音停顿。

不同位置的话轮构造单元末尾出现语音停顿的频率				
话轮中的位置	出现语音停顿		不出现语音停顿	
	次数	比率	次数	比率
话轮起始位置	52	8.4%	566	91.6%
话轮中间位置	41	7.5%	509	92.5%
话轮结束位置	1421	51.3%	1348	48.7%
合计	1514	38.4%	2423	61.6%

3.9.1.2 不同位置的话轮构造单元的末尾语音停顿的时长统计

下面我们统计了话轮构造单元在话轮中所处的位置对其末尾语音停顿时长影响。结果表明,处于话轮结束位置的话轮构造单元的末尾语音停顿的平均时长相对较短,而处于话轮中间位置的话轮构造单元的末尾语音停顿的平均时长却相对较长,相差 140 毫秒。F 检验表明这种差异有统计意义的,是显著的。

不同位置的话轮构造单元的末尾语音停顿的时长统计数据 (单位: 秒)			
在话轮中所处的位置	平均值	次数	标准差
起始位置	0.44	52	0.24
中间位置	0.50	41	0.22
结束位置	0.36	1421	0.21
合计	0.36	1514	0.21

从以上关于话轮构造单元末尾语音停顿的分析数据可以看出,话轮构造单元末尾语音停顿的出现几率及其停顿的长短与其言行为类型及其在话轮中所处位置之间有着密切的关联,因此理论上,了解话轮构造单元的末尾语音停顿的出现几率和长短对于准确识别其在话轮中所处位置来说,有一定的帮助作用。

3.9.2 不同言行为类型的话轮构造单元的末尾语音停顿的统计数据

为了考察话轮构造单元的末尾语音停顿与其言行为类型及其在话轮中所处位置之间的关系,下面统计了它们之间的关系数据。

3.9.2.1 不同言行为类型的话轮构造单元末尾语音停顿的出现几率统计

下表是不同言行为类型的话轮构造单元末尾出现语音停顿的几率统计数据,从中可以看出,总共有 1514 个话轮构造单元的末尾出现了语音停顿,占话轮构造单元总量的 38.4%,而不出现语音停顿有 2423 个,占总量的 61.6%,因此可以认为,大部分话轮构造单元的末尾是不出现语音停顿的。另外,从表中的数据还可以看出,不同言行为类型的话轮构造单元末尾出现语音停顿的几率是不同的,其中提问类话轮构造单元末尾出现语音停顿的几率最大,这可能是由于提问类话轮构造单元往往出现在话轮的结束位置上,在其

之后通常会发生话轮转换,而在话轮结束位置上出现语音停顿的几率是其他位置的 6-7 倍(数据见前),另外在提问类话轮构造单元之后出现的大多是对问题的回答,在回答问题前人们往往要留点时间来思考以显得慎重(有时面对不需思考即可回答的问题也可能会如此),因此使得提问类话轮构造单元末尾有相对较高的语音停顿几率。在这七类之中,招呼类末尾出现语音停顿的几率最低,这可能是由于礼貌策略的缘故,面对对方的招呼关系语往往要即刻做出反应,否则可能会被认为怠慢或不礼貌,因此其间不应有较长的语音停顿,此外招呼语往往都是惯用语,不需思考,在对方招呼语的刺激下既可脱口而出,因此其间不需有较长的语音停顿。

言语行为类型	出现语音停顿		不出现语音停顿	
	次数	比率	次数	比率
主观性叙述	265	31.3%	581	68.7%
客观性叙述	545	39.4%	839	60.6%
命令	37	39.4%	57	60.6%
回声问	189	38.0%	309	62.0%
是非问	204	48.1%	220	51.9%
非是非问	203	55.0%	166	45.0%
招呼	71	22.0%	251	78.0%
合计	1514	38.4%	2423	61.6%

3.9.2.2 不同言语行为类型的话轮构造单元的末尾语音停顿的时长统计

下面我们统计了不同言语行为类型的话轮构造单元的末尾语音停顿的时长,统计时只计算末尾出现语音停顿的话轮构造单元,末尾没有语音停顿的不放到一起进行平均。从下表的数据可以看出,在这几类话轮构造单元末尾的语音停顿中,叙述类的停顿时间最长,招呼类的停顿时间最短,两者相差 130 毫秒,这表明话轮构造单元末尾语音停顿的时长与其言语行为类型之间密切相关。

言语行为类型	平均值	次数	标准差
主观性叙述	0.40	265	0.22
客观性叙述	0.38	545	0.21
命令	0.33	37	0.16
回声问	0.29	189	0.18
是非问	0.34	204	0.21
非是非问	0.39	203	0.21
招呼	0.27	71	0.13
合计	0.36	1514	0.21

从以上两个数据表中的数据可以看出,不同言语行为类型的话轮构造单元末尾出现语音停顿的几率明显不同,不同

言语行为类型的话轮构造单元末尾语音停顿的时长差异较大,这说明话轮构造单元末尾是否出现停顿及其停顿时长与其言语行为类型之间密切相关。因此理论上,了解话轮构造单元的末尾语音停顿的出现几率和长短对于准确识别其言语行为类型来说,有一定的帮助作用。

另外,为了比较话轮构造单元末尾出现语音停顿的几率,下面统计了话轮构造单元内部出现语音停顿的几率数据。从下表的数据可以看出,在 3937 个话轮构造单元中,有 423 个出现了至少一次的长于 100 毫秒的内部语音停顿,几率为 11%,这个几率甚至高于那些处在话轮起始和中间位置上的话轮构造单元的末尾语音停顿的出现几率(数据见前)。这表明语音停顿既可以出现在话轮构造单元末尾,也可以出现在话轮构造单元的内部,二者几乎没有差异,看不出明显的倾向性。

话轮构造单元的数目	内部出现语音停顿的话轮构造单元的数目	内部出现语音停顿的比率
3937	423	11%

从以上关于话轮构造单元末尾语音停顿的分析数据可以看出,在话轮构造单元末尾和内部出现语音停顿的几率近乎相同,而且都较低,因此仅依据是否出现语音停顿很难准确地识别话轮构造单元的边界位置。但话轮构造单元的末尾语音停顿出现几率及其停顿的长短与话轮构造单元的言语行为类型及其在话轮中所处位置之间有着密切的关联,因此,根据话轮构造单元末尾的语音停顿信息,在一定程度上有助于判断话轮构造单元的言语行为类型以及说话人的话轮行进状态。例如,在话轮结束位置上出现语音停顿的几率相对比较高,相当于其他位置的 6-7 倍,因此语音停顿可以看作是话轮转换的一个重要的提示信息。

观察还表明,语音停顿的长短和出现几率与其后面的语义单元所负载的信息量密切相关,越是重要的语义成分,其前面出现的停顿通常就越长。另外,语音停顿的长短还与话轮进程状态密切相关,这表现为:在那些意义相对完整,可能被理解为话轮结束的位置上,如果说话人的话轮还要继续,则为了防止听话人插话,此时一般很少使用语音停顿,即便使用也很短;而在那些不太可能被理解为话轮结束的位置上,如说话人先使用了某些表示自己在思考或者表示自己接下来还有话要说的话轮控制性词语,诸如“我想、然后呢、呵、那个”等,这时听话人一般会等待说话人继续说下去,因此说话人可以使用语音停顿而不必担心听话人抢走了他的话轮。

3.10 不同类型的话轮构造单元的首尾音节时长统计

下面按照话轮构造单元的言语行为类型及其所处位置的不同分析了话轮构造单元起首和末尾的音节时长的具体

表现,以考察这两种因素对话轮构造单元起首和末尾的音节时长的影响。

3.10.1 不同言语行为类型的话轮构造单元的起首和末尾音节的时长统计数据

下表给出了不同言语行为类型的话轮构造单元中起首和末尾音节的时长的统计数据。从横向上比较来看,不管是哪种类型的话轮构造单元,其末尾音节的时长一般要长于其起首音节的时长,起首音节的时长范围在 150 毫秒至 220 毫秒之间,平均值为 180 毫秒,末尾音节的时长范围在 200 毫秒至 240 毫秒之间,平均值为 230 毫秒,末尾音节和起首音节的时长平均值相差 50 毫秒左右。不同类型的话轮构造单元的末尾音节与其起首音节的时长差异程度却不尽相同,其中客观性叙述的末尾音节与起首音节的时长差异量最大,平均相差 80 毫秒,而主观性叙述的末尾和起首音节的时长差异量最小,相差只有 20 毫秒,其他类型的差异量大多在 40-50 毫秒之间。另外,从纵向上比较来看,这几种类型的话轮构造单元的起首和末尾音节的时长差异也很大。就起首音节时长来看,主观性叙述类型的起首音节时长平均 220 毫秒,而是非问、是非非问和招呼之类的起首音节平均只有 150 毫秒。

言语行为类型	起始音节			末尾音节		
	均值	次数	标准差	均值	次数	标准差
主观叙述	0.22	846	0.10	0.24	846	0.10
客观叙述	0.16	1384	0.07	0.24	1384	0.24
命令	0.16	94	0.05	0.20	94	0.07
回声问	0.20	498	0.11	0.24	498	0.17
是非问	0.15	424	0.07	0.20	424	0.07
是非非问	0.15	369	0.08	0.20	369	0.09
招呼	0.15	322	0.07	0.20	322	0.08
合计	0.18	3937	0.09	0.23	3937	0.17

从以上关于话轮构造单元起首和末尾音节时长的分析数据可以看出,话轮构造单元的起首和末尾音节时长及其差异值与其所在的话轮构造单元的言语行为类型之间有着密切的关联,因此了解话轮构造单元的起首和末尾音节时长及其差异值对于准确识别其所在的话轮构造单元的言语行为类型来说,具有一定的帮助作用。

3.10.2 不同位置的话轮构造单元的起首和末尾音节的时长统计数据

下表给出了处于不同位置的话轮构造单元中起首和末尾音节的时长的统计数据,从横向上比较来看,不管是什么位置上的话轮构造单元,其末尾音节的时长一般要长于其起

首音节的时长,起首音节的时长范围在 160 毫秒至 190 毫秒之间,平均值为 180 毫秒,末尾音节的时长范围在 210 毫秒至 230 毫秒之间,平均值为 230 毫秒,末尾音节和起首音节的时长平均值相差 50 毫秒左右。不同位置的话轮构造单元的末尾音节与其起首音节的时长差异程度不尽相同,其中处于话轮结束位置上的话轮构造单元的末尾音节与起首音节的时长差异量最大,平均相差 60 毫秒,处于话轮起始位置上的话轮构造单元的末尾音节与起首音节的时长差异量最小,相差只有 30 毫秒。另外,从纵向上比较来看,不同位置的话轮构造单元的起首和末尾音节的时长也有一定的差异,但并不显著。

话轮中所处的不同位置	起始音节			末尾音节		
	均值	次数	标准差	均值	次数	标准差
起始位置	0.19	618	0.11	0.22	618	0.11
中间位置	0.16	550	0.07	0.21	550	0.10
结束位置	0.17	2769	0.08	0.23	2769	0.19
合计	0.18	3937	0.09	0.23	3937	0.17

从以上关于话轮构造单元起首和末尾音节时长的分析数据可以看出,话轮构造单元的起首和末尾音节时长及其差异值与其在话轮中所处位置之间有着密切的关联,因此了解话轮构造单元的起首和末尾音节时长及其差异值对于准确识别其在话轮中所处的位置来说,有一定的帮助作用。

3.11 不同类型的话轮构造单元的调阶统计

由于录音条件的原因,本语料库中有些语音片段的音质较差,很难进行频谱分析,也很难获得有效的基频数据(所谓有效是指话轮构造单元中的每一个音节都具有相对较完整的基频数据),获得基频数据的方式是,先用 Praat 软件提取语音的基频数据,然后根据语音的窄带频谱分析的结果对提取的结果进行人工核查和修改。下面是对 3937 个话轮构造单元中能够获得有效的基频数据的话轮构造单元的统计。从中可以看出,大约有 40%左右的话轮构造单元不能获得有效的基频数据,对于这一部分语料,在分析基频时需要剔除。

	能够提取基频的话轮构造单元	不能提取基频的话轮构造单元	合计
次数	2460	1477	3937
比率	62.5%	37.5%	100%

3.11.1 不同位置的话轮构造单元的调阶统计数据

下面对能够获得有效基频数据的 2460 个话轮构造单元进行了统计。下表给出了处于不同位置的话轮构造单元的基频最大值、最小值和平均值的数据。

不同位置的话轮构造单元的调阶统计数据 (单位:赫兹)							
位置	单元基频最大值		单元基频最小值		单元基频平均值		
	均值	标准差	均值	标准差	均值	标准差	次数
起始	232	85	145	61	187	74	379
中间	219	81	135	53	183	71	347
结束	211	77	126	49	166	62	1734
合计	215	82	130	52	172	63	2460

从上表的数据可以看出,处于话轮结束位置上的话轮构造单元的调阶相对较低,而处于非结束位置上的话轮构造单元的调阶则相对较高。这表明,话轮构造单元调阶的高低在一定程度上反映了话轮构造单元在话轮中所处的位置,通过考察话轮构造单元的调阶,能够在一定程度上揭示说话人的话轮行进状态。

3.11.2 不同言语行为的话轮构造单元的调阶统计数据

为了比较不同言语行为类型对话轮构造单元调阶的影响,下面分析了不同言语行为类型的话轮构造单元的调阶的平均情况,数据如下表所示。从中可以看出,客观性叙述的调阶最高,而招呼类的最低。这表明话轮构造单元的调阶的高低可能与其所传递的信息量有关,信息量越多,调阶就可能越高,信息量越低,调阶就越低。招呼类的话轮构造单元通常是由比较短的惯用语组成,而且其出现位置比较固定,容易预测,因此信息量比较低。从下表的数据还可以看出,是非问和回声问之间在调阶上的差异也比较大,尽管回声问在疑问形式上与是非问相同,都可用“是”或“非”来回答,但它们在语音形式上是有差别的。由此可以看出,话轮构造单元的调阶在一定程度上具有提示其言语行为类型的功能。

不同言语行为类型的话轮构造单元的调阶统计数据 (单位:赫兹)							
言语行为类型	单元基频最大值		单元基频最小值		单元基频平均值		
	均值	标准差	均值	标准差	均值	标准差	次数
主观叙述	195	75	135	56	164	63	548
客观叙述	230	82	130	50	180	63	830
命令	234	87	136	56	184	69	65
回声问	200	84	121	49	159	61	315
是非问	226	77	132	49	177	59	290
非是非问	225	80	124	53	170	61	238
招呼	196	81	139	57	167	65	174
合计	215	82	130	52	172	63	2460

3.12 不同言语行为类型的话轮构造单元的调域统计

3.12.1 不同言语行为类型的话轮构造单元的调域统计数据

下面是不同言语行为类型的话轮构造单元的调域统计数据,从中可以看出,不同言语行为类型的话轮构造单元的调域差异较大,其中招呼 and 主观性叙述类的调域比较窄,而客观性叙述、命令以及是非问和非是非问的调域都比较宽,回声问的调域处于中间状态。由此可以看出,话轮构造单元的调域与其锁传递的信息类型和信息量的多少密切相关,在一定程度上具有提示其言语行为类型的功能。

不同言语行为类型的话轮构造单元的调域统计数据 (单位:赫兹)			
言语行为类型	平均值	次数	标准差
主观性叙述	60.50	548	39.51
客观性叙述	99.35	830	51.75
命令	97.91	65	46.58
回声问	78.42	315	49.71
是非问	94.36	290	46.86
非是非问	101.68	238	49.21
招呼	56.76	174	40.45
合计	84.82	2460	50.30

3.12.2 不同位置的话轮构造单元的调域统计数据

下表给出了处于不同位置的话轮构造单元的调域统计数据。从中可以看出,它们之间没有太大的差异。

不同位置的话轮构造单元的调域统计数据 (单位:赫兹)			
在话轮中所处的位置	均值	次数	标准差
起始位置	87.47	379	53.12
中间位置	84.44	347	51.24
结束位置	84.31	1734	49.36
合计	84.82	2460	50.30

4. 结论

通过以上数据统计分析,可以看出,自然口语的韵律表现与话语的言语行为类型和话轮行进状态密切相关,其边界韵律特征具有一定的话轮提示功能和言语行为功能,通过分析不同类型的话轮构造单元的边界韵律表现有助于准确地分析说话人的交际意图及其话轮行进状态,这对于实现真正意义上的人机对话来说其重要性是显而易见的。

话语标注的符号说明:

1. 第一列“T”表示话轮,其后的数字表示该话轮的序号;
2. 第二列“C”表示话轮构造单元,其后的数字表示该话

- 轮构造单元的序号；
3. 第三列“A、B、C”等表示不同的说话人；
 4. 第四列“U”表示话段，其后的数字表示该话段在话轮中所处的序号；
 5. “|”表示该位置有较小的韵律间断；
 6. “||”表示该位置有较大的韵律间断；
 7. 字体的加黑加大表示该音节在其所在的话轮构造单元中是最重的音节；
 8. 斜体字表示语速较快（说明：以韵律词为计算单元，其内部各音节的平均时长若小于130毫秒，则认为该韵律词的语速较快）；
 9. 用花括号“{}”圈住表示语速较慢（说明：以韵律词为计算单元，其内部各音节的平均时长若大于280毫秒，则认为该韵律词的语速较慢）；
 10. “LE”表示音节被拉长，其后的数字表示该音节的实际时长值，单位为秒（说明：若某音节的时长大于350毫秒则认为是被拉长了）；
 11. “S-J”表示话轮构造单元末尾的正常语音停顿，其后的数字表示实际停顿的时长，单位是秒（说明：停顿时长大于100毫秒才计算）；
 12. “SIL”表示话段内部的正常语音停顿，其后的数字表示实际停顿的时长，单位是秒（说明：停顿时长大于100毫秒才计算）；
 13. “UC”表示模糊、听不清楚的声音；
 14. 下划线表示语音叠接；
 15. 话段前方括号内的符号表示话段的交际功能或言语行为类型，如叙述、询问、命令等。

16. “[NON-PitchTier]”表示该话段由于音质较差，无法进行频谱分析。

参考文献

- 1、Katz, W. F., Beach, C., Jenouri, K., and Verma, S. (1996). “Duration and fundamental frequency correlates of phrase boundaries in productions by children and adults,” *J. Acoust. Soc. Am.* 99, 3179-3191.
- 2、Beach, C., Katz, W., and Skowronski, A. (1996). “Children’s processing of prosodic cues for phrasal interpretation,” *J. Acoust. Soc. Am.* 99, 1148-1160.
- 3、林茂灿(2000),“普通话语句中间断和语句韵律短语”,《当代语言学》,2000年第四期。
- 4、李爱军(1999),“普通话对话的韵律短语和语句重音的声学分析”,第四届全国现代语音学学术会议《现代语音学论文集》,吕士楠等主编,金城出版社。
- 5、熊子瑜 林茂灿,语流间断处的韵律表现,《第六届全国人机语音通讯会议论文集》2001年11月。
- 6、熊子瑜,基频重设与语流间断,载于蔡莲红 周同春 陶建华主编《新世纪的现代语音学——第五届现代语音学学术会议文集》,清华大学出版社2001年版。
- 7、吴洁敏 朱宏达著《汉语节律学》第62页,语文出版社2001年版。
- 8、Aijun, L., (2002), “Chinese Prosody and Prosodic Labeling of Spontaneous Speech,” In *Phonology and Prosody 2002*.

A study on the prosodic features of the utterance boundary in natural speech and the interactive function of these features

Xiong ZIYU

Institute of Linguistics, CASS

xiongziyu@163.com

Abstract

The paper mainly studies on the prosodic features of the utterance boundary in nature speech in term of speech intercommunication based on 973 telephone dialogue corpus. A painstaking investigation has been paid to discover the duration and pitch feature near the utterance boundary, using the SPSS program to measure and compare the difference of the prosodic features of the utterance boundary among different kind utterances. Grounded on the study results, I try to investigate two kinds of interactive function of the

prosody features near the utterance boundary: the indicator of speech acts and the cue of turn state in progress. The results are: there indeed exist some perceptible and measurable prosodic features of the utterance boundary. Different boundary prosodic features may have different interactive function, and they are the important and necessary cues to help the listener catch the speaker’s intention and detect the speaker’s turn state in progress.

Key words: Turn Speech act Prosody