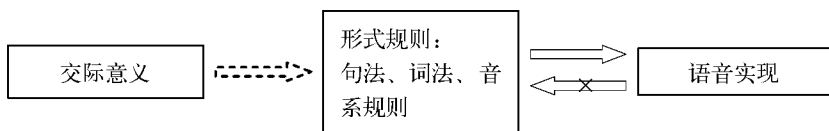


# 从交际功能和编码机制的角度理解语言

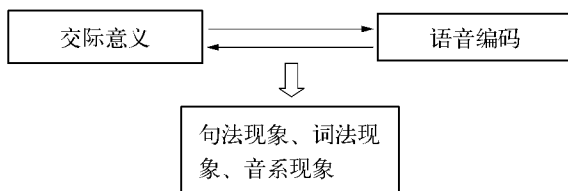
许 毅  
伦敦大学学院

如果我们会说一门外语,或者至少学过一门外语,我们一定会觉得掌握这个语言的语法是非常重要的。这里说的语法不仅仅是句法和词法,还包括语音[音系]规则。语法没有学好,就不能“正确”地使用这个语言。于是我们很容易推论,说自己的母语时,也一定是靠语法的指导,说得正确是因为语法掌握得好。实际上,这正是语言学里相当普遍接受的概念。如图一所示,传统观念认为,说话虽然是要表达意义,但是真正决定语言形式的是一套复杂的句法、词法和音系规则,由它们生成表层的音系形式,最后通过发音生成语音信号。



图一：有关语言的传统模型

这个模型里的重要假设是：① 表层形式的终极解释来自于形式转换的规则；② 交际意义和语音实现对形式规则基本上没有影响；③ 形式规则的解释来自于形式规则的内部规律。在本文里,我要提倡一个跟这种传统观念背道而驰的思路,如图二所示。根据这个思路,说话是通过语音来表达意义的,所以语言是交际意义的语音编码,而语法则是在这个编码过程中产生的种种现象。按照这种思路,要理解语法现象就必须理解语音编码跟交际意义的交互作用。在下面的讨论里,我要先指出传统观念的一些根本性问题,然后再简单阐述为什么只有同时考虑交际功能和语音编码才能深化我们对语言的认识。



图二：有关语言的功能编码模型

## 1. 现象不能用现象本身来解释

无论在科学上或在日常生活里,任何一种现象都不能用自己本身来解释。如果把太阳从东方升起解释为因为有一条规则叫“旭日东升”,这在哲学上叫做循环定义,或者循环式解释、循环论证。循环式解释是无益的,因为它没有告诉我们任何新的东西,没有增进我们对世界的认识。当我们发现人们说某一语言时会遵循一定的规则,我们把这些规则归纳成语法,用它来教人,尤其是成人,学说这个语言,这是完全可以的,而且被一再证明是行之有效的。但是如果试图用这个语法来解释这个语言,我们就走进了循环论证的圈子。同样,当观察到世界上的语言都有一些共性,就解释为所有语言都遵循一个普遍语法(Chomsky, 1975),这并没有增进我们对人类语言本质的认识。对一个现象的解释,必须用该现象之外的因素。譬如,太阳从东方升起,是因为我们是站在由西向东自转的地球上,看位置相对不动的太阳。

## 2. 功能与形式

把关于循环论证的思考放到语言学上就会引出这个学科里长期争论的一个问题,即形式与功能之争。在语法领域里,自从1957年Chomsky出版《句法结构》一书以来,占主导地位的一直是以他为代表的形式学派。形式学派认为语言的核心是一套形式规则,而交际功能则是语言核心之外的。与之相对的是功能学派,他们认为交际功能是语言的核心,语言的形式来自于功能。形式学派在语言学里的主导地位也给现代音系学带来很大优势。从生成音系学,到自主音段音系学,到优选论,音系形式都被视为语音的核心。交际意义与语音实现都被认为对音系形式和规则没有决定性的影响。同样,这种认识也主导了有关语言韵律的理论。例如Goldsmith(1976)提出的自主音段音系学,由Pierrehumbert(1980)、Ladd(1978, 1996)、Guessenhoven(2004)等人发展的语调音系学(intonational phonology),都是以形式为核心的。Pierrehumbert(1980: p. 59)便明确地阐述道:“在已有文献里,判断哪些语调模式在语言学上相互区别,哪些是同一模式的变体,有两种方法:一种是根据观察到的F0曲线特征,归纳出一个音系学代表体系;该体系归纳出来之后,下一步再比较这些音系学区别模式的用途。与之相反的方法是,首先辨识不同语调模式是转达相同还是不同的意义;然后再建立一个音系学体系,其中给相同的意义赋予同一底层代表,给不同的意义赋予不同的代表。…… 本论文采取的是第一种方法,并且只限于第一种方

法的第一步。”<sup>①</sup>这大概是形式优先法最明确的表述了。重要的是,整个语调音系学以及由其衍生的,现已被广泛应用的 ToBI 标注体系,都是建立在用这种形式归纳法得出的描述性语调成分之上的。正如上节所谈到的,由直接观察归纳出的模式,只能是对现象的描述,用来做理论性的解释就难免循环论证,跟用“旭日东升”来解释太阳为何从东方升起没有什么本质上的不同。

事实上,语调音系学的思路甚至违背了语言学上已普遍接受的音位原则。根据这个原则,考察一个语言的语音系统,要先确立哪些语音差别是能分词辨义的,哪些差别无关紧要。那些能分词辨义的差别就叫音位差别。例如,在英语里 [f] / [v] 之差可以将“ferry”和“very”这两个词区别开来,所以这两个辅音在英语里就是不同的音位。相反,在刚才这两个词里的 /r/ 有多个变体,如 [ɹ], [r] 甚至 [R]。它们在形式上的差别是很明显的,但因为不区别词,就被认为是同一音位的不同变体。在这两个例子里,一个音的音位归属决定于它是否分词辨义,所以是以功能而不只是以形式来决定的。而且因为功能在形式之外,所以这样的定义是非循环性的。音位是现代语言学最重要的发现之一。这个发现,给语言学带来了长足进步。但是,现代音系学对形式的强调使人们产生这样的印象:语音形式的变换可以完全独立于意义。这种纯形式的转换在自主音段音系学里比比皆是。语调音系学继承的也是这个传统。Ladd (1996) 干脆把 Pierrehumbert 为代表的这一体系就叫做 autosegmental-metrical phonology of intonation。

也许会有人说,一个以形式为核心的体系即使解释力有限,至少它可以给我们提供系统的描述,这对增进我们的认识还是有帮助的。在下面的讨论里我们会看到,至少在韵律方面,以形式为核心的理论连准确描述都做不到。这其中原因之一,是因为它还忽略了一个跟交际功能同等重要的因素,即发音机制。

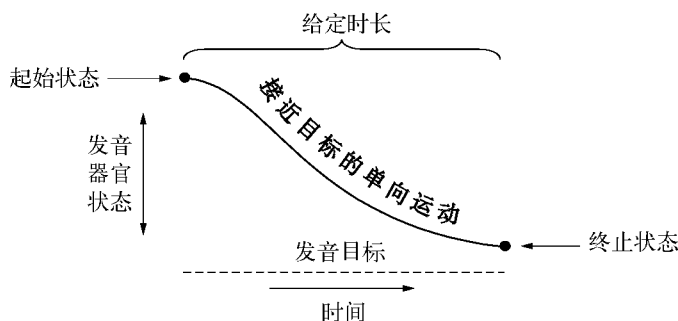
### 3. 发 音 机 制

从交际功能到表层形式,即语音信号,必然有一个编码过程。这个过程应有一套编码机制。不清楚这套机制,就无法理解交际功能与表层形式的关系。有人可能会问,那可不可以把传统的形式规则直接理解为编码机制呢?首先,如上所述,形式学派自己不承认形

<sup>①</sup> “In the literature, one can distinguish two approaches towards the problem of establishing which intonation patterns are linguistically distinct and which count as variants of the same pattern. One approach attacks the problem by attempting to deduce a system of phonological representation for intonation from observed features of F0 contours. After constructing such a system, the next step is to compare the usage of F0 patterns which are phonologically distinct. The contrasting approach is to begin by identifying intonation patterns which seem to convey the same or different nuances. The second step is to construct a phonology which gives the same underlying representation to contours with the same meaning, and different representations to contours with different meanings. ... The work presented here takes the first approach, in fact, it stops at the first step in the first approach.”

式是交际功能的直接编码 (Ladd, 1996)。其次,在操作上也讲不通,因为传统的形式规则全都是像数学公式一样的符号运作,其假设是大脑就是通过符号运作实现编码的。我们只要稍加思索,就会知道这是不可能的,因为编码离不开具体的设备。说话的设备是由神经系统(包括脑和周围神经)、肌肉以及发音器官的骨骼和软组织构成的。编码就是操作这些设备的过程。虽然我们目前对这些设备的认识很有限,但我们至少不难想象编码机制跟编码设备的特性有着密切的关系。不理解这些特性,就不可能理解编码的机制。总的说来,我们对语言的外围设备,即发音器官,比对中央设备了解得更多一些。所以语音学可以为我们认识编码机制提供一个起点。在下面,我集中以声调和语调为例讨论编码的语音机制。

图三显示的是我所理解的发音动作的最基本模式,它不仅适用于声调和语调,也适用于语音的其他方面,也包括音段(Xu & Liu, 2006, 2007)。每个动作都是从发音器官的某个初始状态开始,然后在给定时长,即所分配的时间里向一个理想状态不断接近。理想状态就是当前音的标准状态。初始状态来自于前一个动作的结束状态,或者是发音的起始状态。如果给定时长不够,理想状态就会达不到,如图中所示。这种基本的动作模式实际上不是语音特有的,而是人或者动物的任何动作的基本模式。从这一点上讲,发音动作其实没什么特殊之处。不同的只是理想状态的特性、使用的器官、动作的力度以及给定时长。

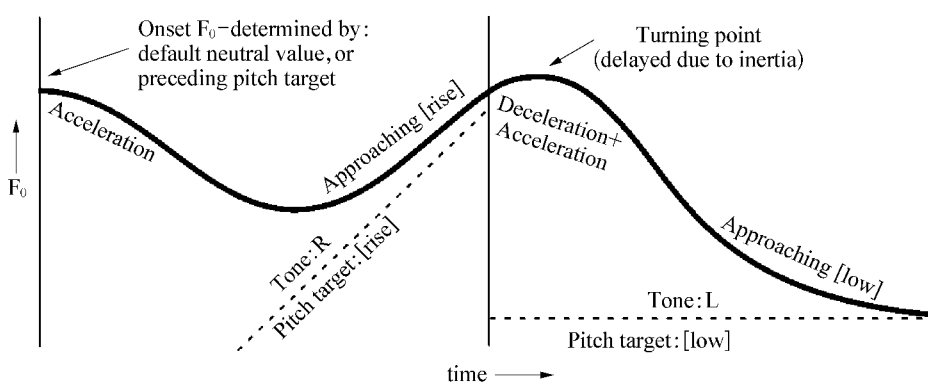


图三: 发音动作的基本模式

图四显示的是建立在发音基本动作模式上的适用于声调和语调的目标接近模型 (Xu & Wang, 2001)。跟发音动作基本模式相比,目标接近模型添加了如下成分:

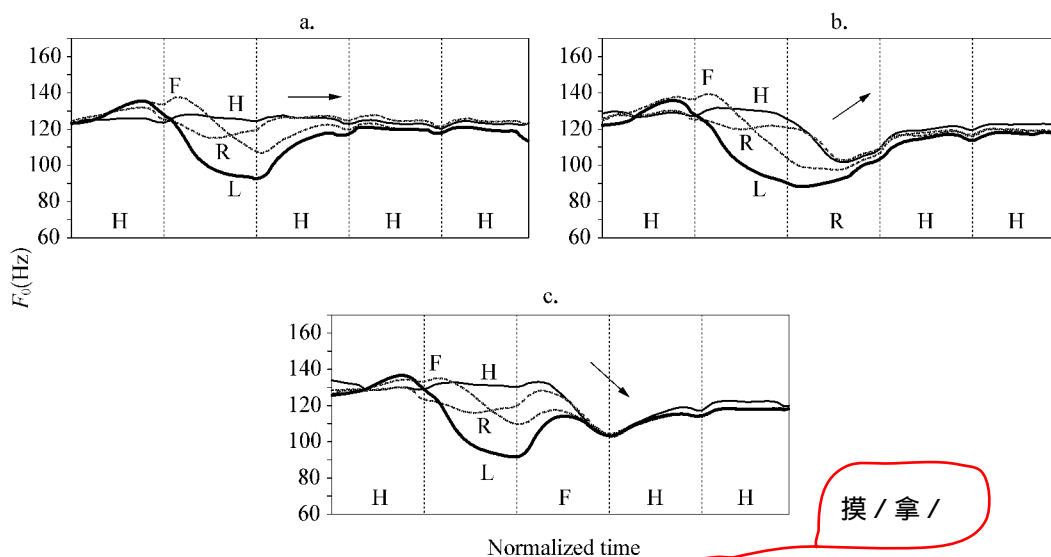
- (1) 音高目标(即模型里的 pitch target)既可以是静态的,也可以是动态的。
- (2) 初始状态不仅包含静止状态,还包含从前一动作继承下来的速率(velocity)和加速度(acceleration)。
- (3) 音高目标所分配的时间与音节所分配的时间等同。

其中前两项其实已经隐含在图三所示的发音动作基本模式里,目标接近模型只是把它们明确化。



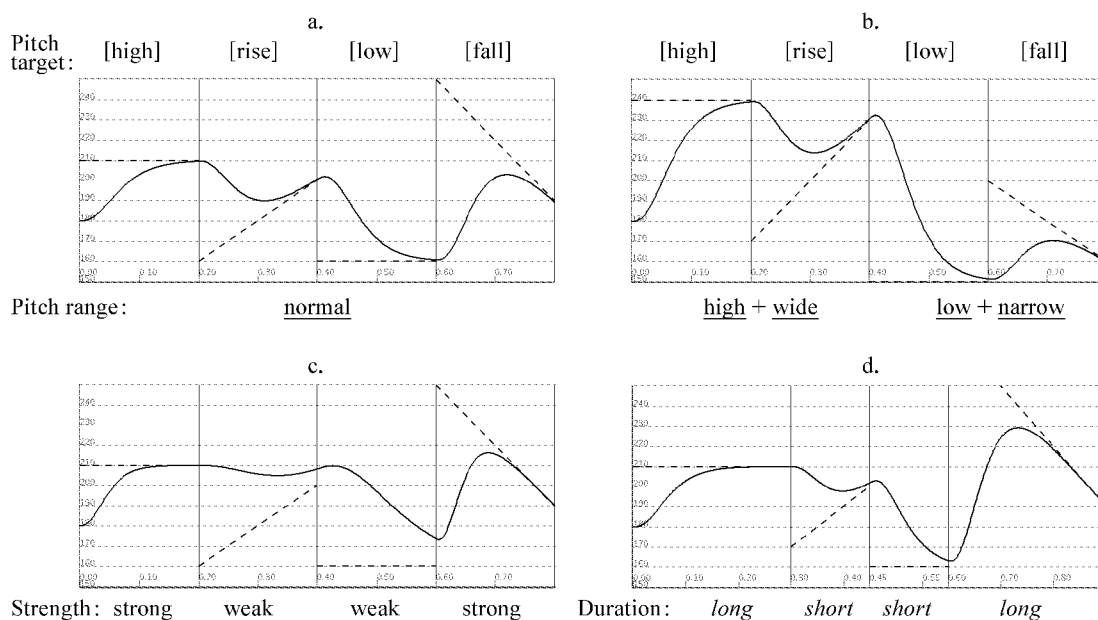
图四：目标接近模型(TA model)。图中的竖线代表音节边界，  
虚线代表底层音高目标，连续的粗实线代表表层基频曲线

目标接近模型可以有效地解释连续语言里的基频曲线。图五显示的是北京话里声调连续的平均基频曲线。在每个图里，只有第二个音节的声调是不一样的，而其他音节的声调保持不变。我们可以看到，第二音节的终点基频由于声调不同而差别很大，结果第三音节的起点基频也大不相同。但是，在第三音节里，所有的基频曲线都随时间逐渐向一个直线靠拢(如箭头所示)。而这个直线的值跟该音节声调的理想特征相符：第一声为高平，第二声为上升，第四声为下降。所以，图五里的基频曲线跟目标接近模型是一致的。唯一不太一致的地方是第二音节的声调对第一音节也有所影响。但这种影响是逆化的，即第二音节的基频越低，第一音节的基频越高。这种逆化影响的机理还不清楚，有待于今后进一步研究(初步探讨见 Gandour et al. , 1994; Xu, 1997)。



图五：北京话“猫咪 / 迷 / 米 / 蜜 / 卖猫咪”的平均基频曲线。图中 H、R、L、F 分别代表第一、二、三、四声。数据来自 Xu (1999)

通过目标接近模型,我们不仅可以看到实现单个声调的机理,而且可以看出,通过控制该过程的各个方面,就有可能实现多层次的编码。如图六所示,在这个模型里,不光是音高目标([high], [rise], [low], [fall]),还有调域、力度、时长,都可以分别控制。在图六 a 里,只是音高目标有变化。在图六 b 里,第一、二音节的调域被抬高并加宽,第三、四音节的调域被降低并压缩,但每个音节基频曲线的基本形状跟图六 a 里仍大体一致。在图六 c 里,第一、四音节的力度加强,导致目标实现更加充分;第二、三音节的力度减弱,致使目标实现很不充分。在图六 d 里,第一、四音节的时长增加,第二、三音节的时长缩短,其结果跟图六 c 里加强或减弱力度的效果相似。在下一节里我们可以看到,图六所显示的控制手段在交际功能的编码中得到充分应用。



图六: 由 qTA 模型 (Prom-on et al., forthcoming) 模拟的音高目标、调域、力度和时长对表层基频曲线的影响

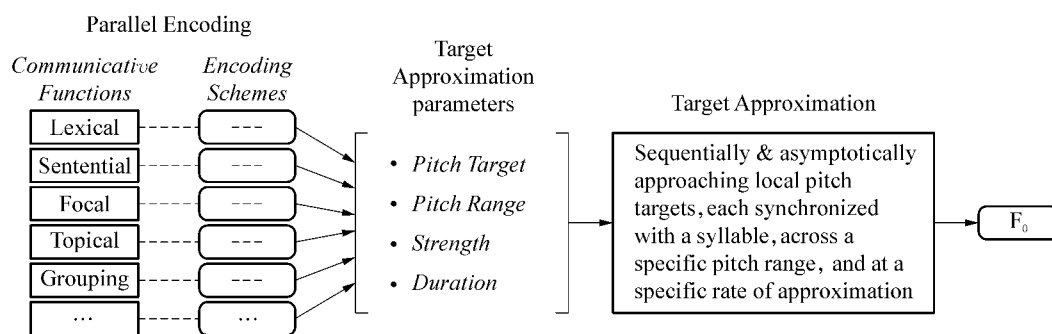
2009

#### 4. 交际功能及其编码

前面已经谈到,从功能的角度来看,我们说话的目的并不是要表现自己的语法是正确的(除非是为了应付口语考试),而是要向听话人传达各种信息。这些信息也并不只是局限于词义和由句法结构体现的命题意义(propositional meaning),还包括很多其他层次的意义。这些意义可以统称为交际功能。下面是一些有基频参与编码的功能:

- ① 分词 (Lexical contrast)
- ② 焦点 (Focus)
- ③ 话题/话轮 (Topic/turn taking)
- ④ 句型 (sentence type)
- ⑤ 分界/组合 (Demarcation/grouping)
- ⑥ 态度 (Attitude)
- ⑦ 情感 (Emotion)

基本上是一维的基频怎么能参与传达这么多层次的功能呢？图六已经为我们提供了一个初步的线索。从那里我们看到，目标接近模型的各个控制参数都可以分别调节，而且每种调节对基频曲线都有明显的影响。所以每个交际功能都有可能用某个或某套参数来编码。图七是我在 2004 年提出的“平行编码及目标接近模型”(Parallel Encoding and Target Approximation model), 简称 PENTA, 用来模拟从交际功能到表层语音的编码流程。图左的小方框代表的是各个具体的交际功能，它们每个都有一个对应的编码模式。每个模式由一套目标接近(Target Approximation)参数构成。这些参数控制图四里显示的目标接近模型，由这个模型，像图六显示的那样通过调节各个音高目标来生成表层的基频曲线。在下面的讨论里，我以分词和焦点为例说明交际功能如何通过目标接近实现语音编码。



图七：平行编码及目标接近模型 (Parallel Encoding and Target Approximation model — PENTA, Xu, 2004, 2005)

## 4.1 分词

词是语义和语音的综合体，也是语音和语义的联结点。分词就是将不同的词区分开来。这是语音的最基本功能之一。分词的手段非常多：元音、辅音、声调、发声型、词重音。其中元音和辅音大概是世界上所有语言都用的。声调、发声型、词重音都只是一部分

语言使用。以声调为例,在汉语这样的声调语言里,声调的不同可以让一个声、韵母都相同的音节代表不同的语素。在北京话里,一个音节可能有五个不同的声调:阴、阳、上、去、轻。早期的研究就已经发现每个声调在单念时都有特定的音高模式(白涤洲,1934;赵元任,1922)。这些模式的独特性足以使它们在感知上有效地相互区别开来(Gandour, 1983)。不过在连续语流里,如图五所示,声调的基频曲线会大大偏离单念的形式。但是正如目标接近模型所模拟的,尽管有这些偏离,到了每个音节尾部,基频曲线会接近或达到理想的目标。从感知试验的结果上看,声调的这种实现方式是有效的。即使在连续语流中的声调识别率也可以接近 90% (Prom-on et al., <sup>2009</sup> ~~forthcoming~~)。我们在最近的神经网络模拟实验中发现,一旦把表层基频曲线转换成速率曲线(velocity profile),一个无监督学习系统(unsupervised learning system)就可以通过处理这些曲线自动学会识别连续语流中的声调(Gauthier et al., 2007a, 2007b)。

再来看词重音。在英语里,像 subject、contract 这样的词,当第一个音节重读时是名词,第二个音节重读时是动词。早在 20 世纪 50 年代,Fry (1958) 就发现这样的词的动、名之分可以靠基频、时长和音强来辨别,但是辨别效果最好的是基频。只要两音节的基频之差为 5 Hz,感知上便没有任何混淆。这跟我们考察英语里实际的基频曲线时得出的结论是一致的(Xu & Xu 2005)。那就是,英语里非焦点处的轻、重音节之间的基频之差虽然很小,但总是存在。所以从功能的角度来看,这些很小的起伏载带着分词的信息,因为它们的幅度已足以区别像 'subject/sub'ject 这样的名词和动词。而以形式为主的 AM 理论认为焦点后的基频曲线完全是平的,因为能直接从基频曲线上看到的上下起伏是很小的(Pierrehumbert, 1980)。所以,形式派理论对功能的忽视,导致他们忽略了表层形式上重要的细节。

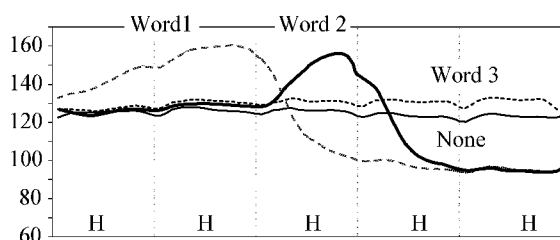
如果对功能和发音同时考察,我们可以看到编码机制里更多的细节。例如,在最近的研究里我们发现,从目标接近模型的角度看,英语像汉语一样每一个音节都有一个音高目标,但这个目标会随音节在词、句里的位置以及在句中焦点的位置而变(Xu & Xu, 2005)。这样看英语和汉语在语调上的差别比一般认为的要小。但跟汉语不同的是,英语音节的音高目标不是完全由分词功能,即词重音决定的。这在下面会进一步讨论。

还有一个跟词重音有关的问题也是从功能的角度更容易看清。那就是,有分词功能的词重音只在某些语言里有,如英语、德语、俄语。另外有许多语言不用重音分词,如法语、波兰语、西班牙语。在某种程度上,北京话里的轻声有点类似于词重音(Chen & Xu, 2006)。在不用重音分词的语言里,也常常可以观察到一些貌似词重音的语音现象。人们往往把它跟词重音相提并论,并试图找到一种跨越不同语言的、以形式为定义的抽象的重音(prominence)。从功能的角度看,这种抽象的重音没有多大意义。因为不同功能的编码必须是不一样的,不然它们就会相互混淆。这在下面关于焦点的讨论里可以清楚地看到。



## 4.2 焦点

我们说一句话,有可能特别强调其中某一部分。这种强调现在一般叫做焦点(focus)。例如在“是我请客”这个句子里,强调的是“我”。在这里强调的手段是用“是……”这种句型。但是也可以说“我请客”,只是在语气上额外加重“我”。所以焦点既可以用句法也可以用语音来实现。那么语音上如何实现焦点呢?以往的研究发现,音高变化是焦点的主要特征之一。可是在汉语里,音高已经用于声调,怎么能又用来表现焦点



图八:北京话“猫咪摸猫咪”四种焦点的平均基频曲线。数据来自 Xu (1999)。

呢?在图六 b 里我们已经看到,在音高目标基本不变的条件下,调域是可以作调整的。根据系统实验的发现(Xu, 1999; Liu & Xu, 2005),在北京话里,跟焦点有关的调域控制分三个区域:焦点前、焦点处、焦点后。焦点前的调域不变,也就是跟没有焦点的句子相似,焦点处的调域加宽,焦点

后的调域降低并压缩。图八显示的是一个全是阴平字构成的句子如何通过三区段调域调整实现焦点。

进一步的研究发现,英语里焦点的实现方式跟汉语的很相像,也用三区段调域调整,而且这些调整并不抵消词重音对音高的影响,就跟北京话里焦点并不抵消声调对音高的影响一样(Xu & Xu, 2005)。但英语有一点跟北京话不同的是,焦点处的词如果最后一个音节是重读,该音节的音高目标会由[high]变成[fall]。不过这种从功能角度看英语语调的观点跟当前流行的观点是不一样的。按照流行的语调音系学或 AM 理论,词重音只是在载带 pitch accent 时才会有音高上的表现,而确定是否载带 pitch accent 完全取决于基频曲线有无一目了然的运动(Pierrehumbert, 1980)。由于焦点后的基频运动幅度很小,于是就认定不值得考虑(Beckman & Pierrehumbert, 1986)。所以,完全从形式出发,有可能连形式本身的描述都会不准确。因为描述必然要抽象,抽象就免不了取舍。只有考虑了功能之后才能知道哪些该取哪些该舍。

英语和汉语以及其他一些语言里的焦点还有一个重要特点,就是句末焦点的调域扩展幅度很小(Eady & Cooper, 1985; Xu, 1999)。其后果之一是句末焦点的感知识别率大大低于非句末焦点(Botinis, 1999; Liu & Xu, 2005)。这一发现很可能不仅有语音学的意义,还有可能帮助解释一些句法现象。例如英语、汉语里的强调句型都是把被强调成分挪到靠近句首的位置,而那样正好可以充分实现焦点后的调域压缩,从而保证焦点的感知识别。同样,在汉语里特殊疑问词一般是被强调的。而英语里的特殊疑问句里则把疑

问词挪到句首,也就是实现焦点的有效位置。当然这些都只是初步的观察和推理,有待于进一步的系统研究。但是我们已经可以看到,也许相当一部分句法现象有语音的来源。这跟冯胜利关于韵律句法学的思路是一致的(冯,2000)。当然,从语音机制到语法现象,是要有一个像王士元先生文中强调的演化过程的。要用演化论来解释语言现象,对交际功能和发音机制的理解就更重要了。因为只有它们才能帮我们认识演化的单元以及演化中的选择压力。

## 5. 结 语

我希望通过本文的讨论说明,要解释语言里能够直接观察到的句法、词法和语音现象,不能仅仅靠归纳法得出描述性的规则,而是需要找到背后更基本的机制。本文考察了交际功能和语音编码这两种可能的机制,探讨了多重交际功能如何通过语音编码转换成听者可以破译的表层语音信号。由于时间和篇幅的缘故,这里对两者的讨论都很简略。更详细一点的探讨散见于别处(Xu & Wang, 2001; Xu, 2005)。当然这方面的研究才刚刚起步,还需要大量的实证性研究。同时,编码的设施并不是只是发音器官,还有人的神经系统,包括大脑。尤其是命题意义方面的编码很可能受到大脑的各种特性的限制,例如短时记忆的容量(见陆丙甫的初步探讨:陆,1983)。虽然我们对神经系统的编码机制仍然知之甚少,但近年来大脑研究的手段有很大进展。可以想象,随着大脑研究的进一步发展,随着我们更多地了解语音跟语言其他方面的密切关系,未来的句法学、语义学、语用学也会像语音学一样逐步变成全面的实证科学。

### 参考文献

- 白涤洲(1934): 关中声调实验录,《史语所集刊》4: 355 - 361。
- Beckman, M. E. and Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook* 3: 255 - 309.
- Botinis, A., Fourakis, M. and Gawronska, B. (1999). Focus identification in English, Greek and Swedish. In *Proceedings of The 14th International Congress of Phonetic Sciences*, San Francisco: 1557 - 1560.
- Chen, Y. and Xu, Y. (2006). Production of weak elements in speech — Evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica* 63: 47 - 75.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Chomsky, N. (1975). *Reflexions on language*. New Jersey: Pantheon.
- 冯胜利(2000): 《汉语韵律句法学》,上海教育出版社。
- Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech* 1: 126 - 152.

- Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics* **11**: 149 - 175.
- Gandour, J. , Potisuk, S. and Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics* **22**: 477 - 492.
- Gauthier, B. , Shi, R. and Xu, Y. (2007). Learning phonetic categories by tracking movements. *Cognition* **103**: 80 - 106.
- Gauthier, B. , Shi, R. and Xu, Y. (2007). Simulating the acquisition of lexical tones from continuous dynamic input. *JASA Express Letters* **121**: EL190 - 195.
- Goldsmith, J. A. (1976). *Autosegmental Phonology*, Ph. D. dissertation, MIT. [Published in 1979 by Garland Press in New York] .
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*: Cambridge University Press.
- Ladd, D. R. (1978). *The Structure of Intonational Meaning*. Ph. D. dissertation, Cornell University.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- Liu, F. and Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* **62**: 70 - 87.
- 陆丙甫(1983): 人脑短时记忆机制和人类语言结构的关系,《世界科学》1983. 9: 21 - 30。
- Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Ph. D. dissertation, MIT, Cambridge, MA. [Published in 1987 by Indiana University Linguistics Club, Bloomington] .
- Prom-on, S. , Xu, Y. and Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *Journal of the Acoustical Society of America* **125**: 405-424.
- Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics* **25**: 61 - 83.
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* **27**: 55 - 105.
- Xu, Y. (2004). Transmitting Tone and Intonation Simultaneously — The Parallel Encoding and Target Approximation (PENTA) Model. In *Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, Beijing: 215 - 220.
- Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication* **46**: 220 - 251.
- Xu, Y. and Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* **33**: 319 - 337.
- Xu, Y. and Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* **33**: 159 - 197.
- 赵元任(1922): 中国言语字调底实验研究法,《科学》7: 871 - 882。