# 计算语言学

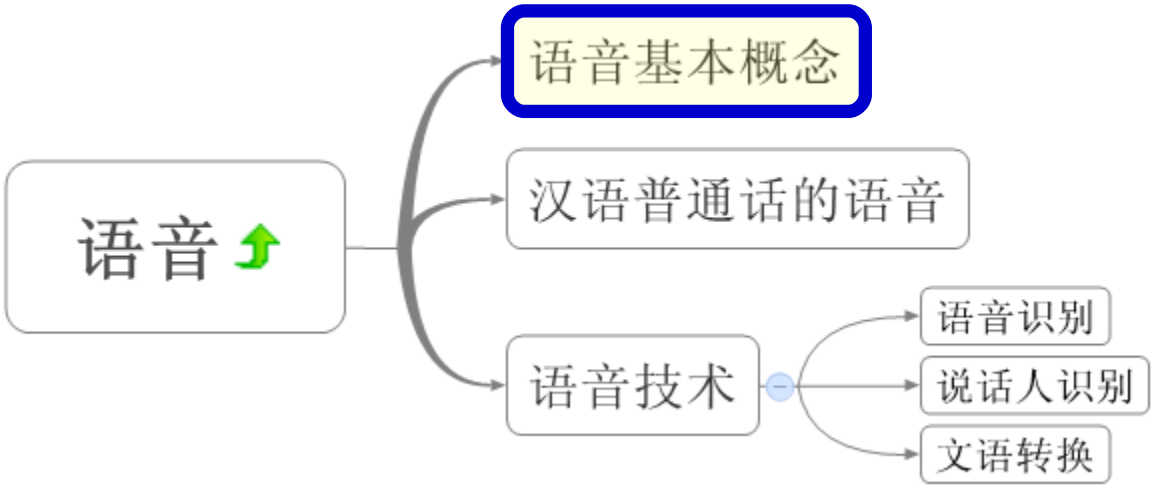
## 第 10 讲　语音、语义与语用

刘群
中国科学院计算技术研究所
liuqun@ict.ac.cn

中国科学院研究生院 2011 年春季课程讲义

# 内容提要

# 内容提要



# 语音基本概念

- 语音的性质
- 音节
- 音素：元音和辅音
  - 清音、浊音
  - 摩擦音、爆破音、塞音、鼻音、后鼻音……
- 音位和音系

# 什么是语音

- 语音是人类发音器官发出的、具有一定意义的、能起社会交际作用的声音。能够代表一定的意义，这是语言的声音同自然界其他一切声音的本质区别。

# 语音的物理属性

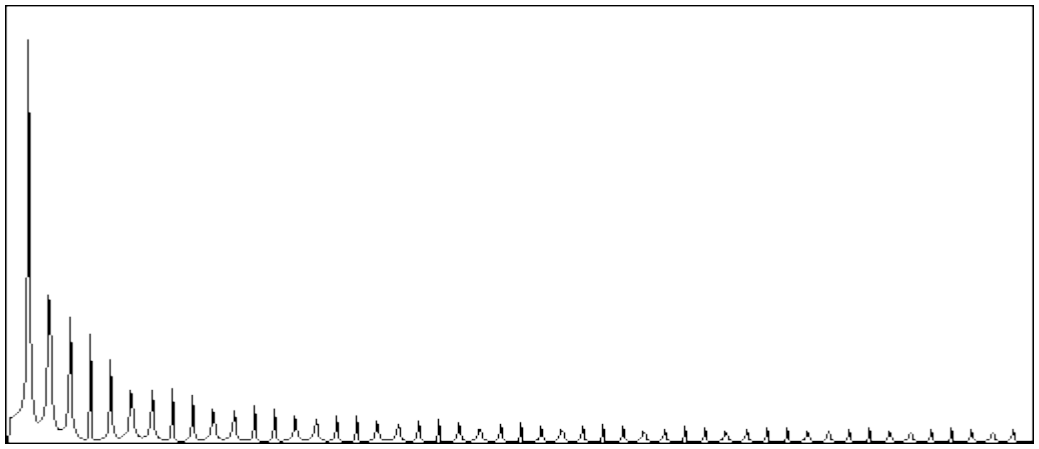
- 音高：频率的高低
- 音强：振幅的大小
- 音长：持续的时间长短
- 音质：又称音色，反映不同频率和振幅的音波的组合形式

# 噪音和乐音

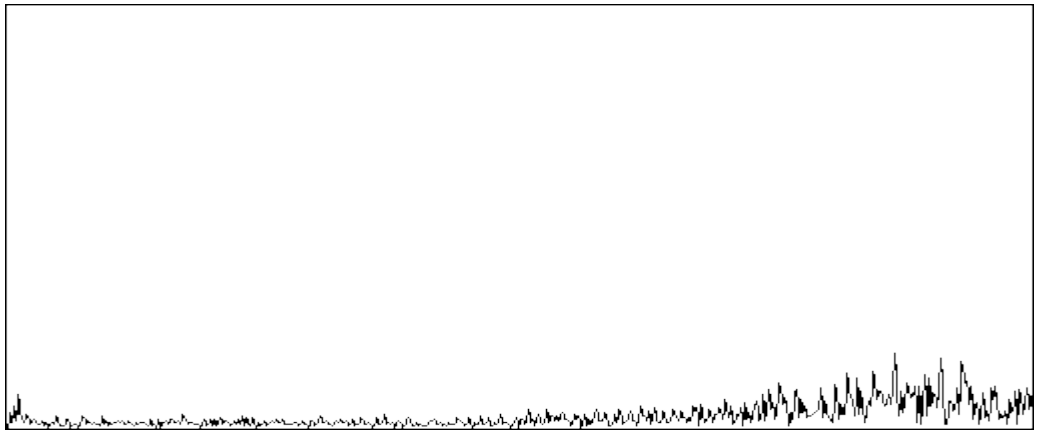
- 噪音与乐音：根据音质不同加以区别
- 噪音：噪音是由许多无规则的音波合成的，它们的音高和强度随时在变化，相互之间没有一定的关系，合成的波形杂乱而无规律。这种声音听起来刺耳、嘈杂，如刹车声，电锯锯木声，马路上车驰笛鸣的喧闹声等等。语音中也有不少噪音成分，如辅声中的塞音、擦音、塞擦音等等。
- 乐音则由若干规则的纯音组成，形成的复合音波有周期性，很有规律，这样的声音听起来和谐、悦耳，歌声、乐声和语音中的元音，都是这样的声音。

# 乐音和噪音频谱样例

乐音



噪音



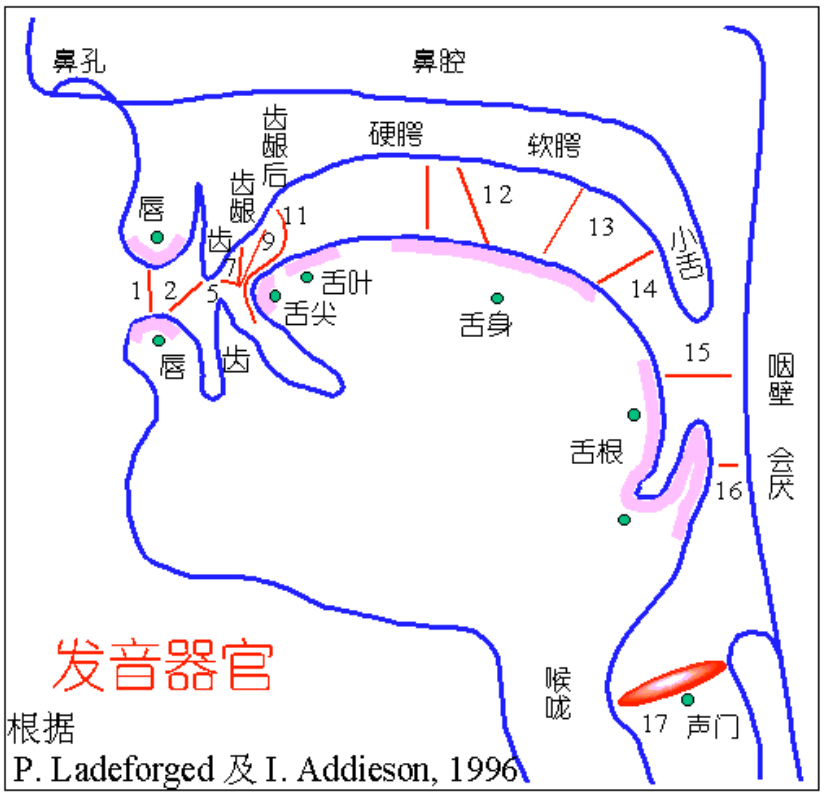
# 语音的生理属性

- 发音器官
  - 肺和气管
  - 喉头和声带
  - 口腔、鼻腔和咽腔

# 发音器官图


1. 上唇　2. 上齿　3. 牙床　4. 硬腭
5. 软腭　6. 小舌　7. 下唇　8. 下齿
9. 舌尖　10. 舌面　11. 舌根　12. 咽头
13. 咽壁　14. 会厌　15. 声带　16. 气管
17. 食道　18. 鼻孔

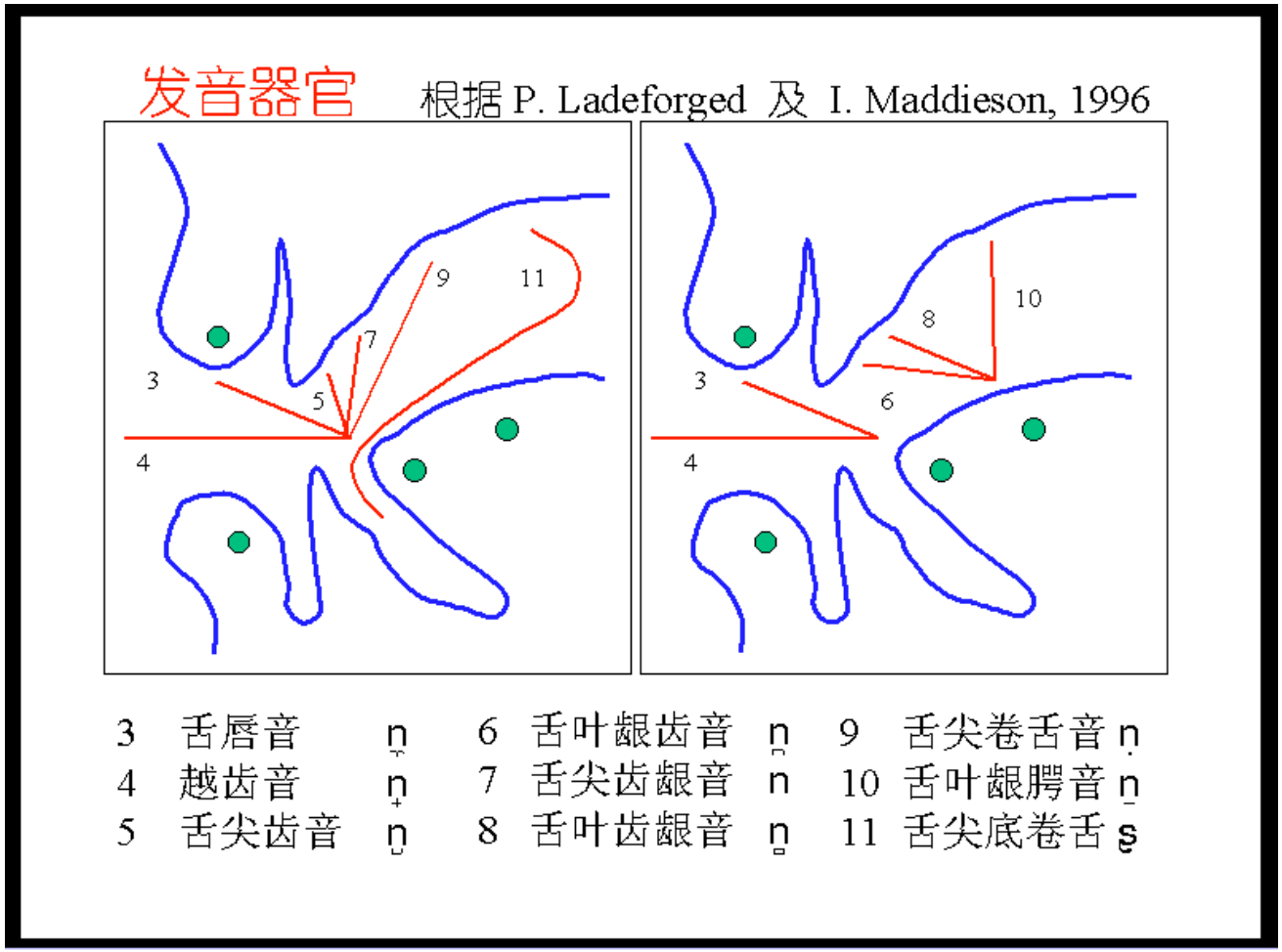
# 辅音发音图 (1)




| | | |
|---|---|---|
| 1 | 双唇音 | m |
| 2 | 唇齿音 | ɱ |
| 5 | 舌尖齿音 | n̪ |
| 7 | 舌尖齿龈音 | n |
| 9 | 舌尖卷舌音 | ɳ |
| 11 | 舌尖底卷舌 | ʂ |
| 12 | 硬腭音 | c |
| 13 | 软腭音 | k |
| 14 | 小舌音 | q |
| 15 | 咽音 | ħ |
| 16 | 会厌音 | ʜ |
| 17 | 声门音 | h |

• 主动发音器官

# 辅音发音图 (2)


发音器官
根据 P. Ladeforged 及 I. Maddieson, 1996
3 舌唇音 n̪
4 越齿音 n̺
5 舌尖齿音 n̻
6 舌叶龈齿音 n̼
7 舌尖齿龈音 n
8 舌叶齿龈音 n̪
9 舌尖卷舌音 ɳ
10 舌叶龈腭音 n̠
11 舌尖底卷舌 s̠

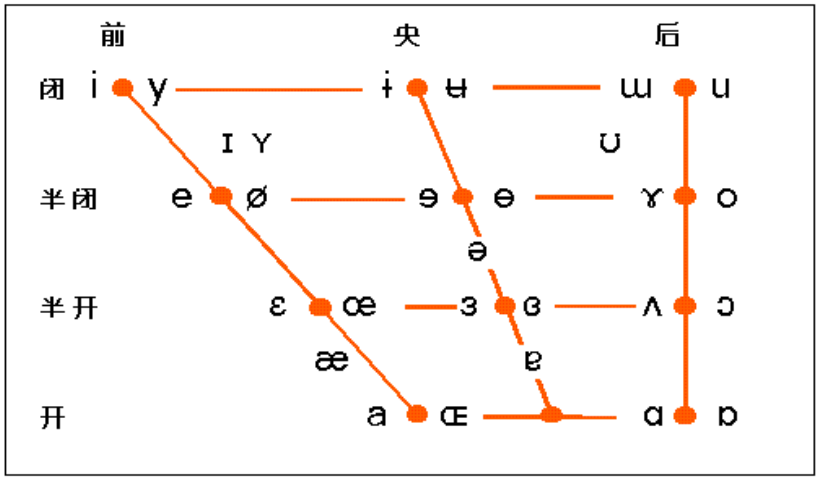
# 元音发音图

舌面元音（1993年版，1996年订正）

舌尖元音

（汉藏通用）

| 舌尖前 | 舌尖后 |
| --- | --- |
| ɿ ʮ | ɿ ʮ |



# 音节

- 音节是语音中最自然的结构单位。人在说话时，发音器官的肌肉总是一松一紧运动，松紧交替一次，就在人们的听觉上形成一个语音段落，这就是音节。因此，人们凭听觉就能自然而然地从语段中分辨出音节来。说话时，人们也总是按音节来发音，肌肉紧张一次，就发出一个音节。

# 音素：元音和辅音

- 音素是最小的语音单位。
- 音素可分为元音和辅音两大类。
  - 气流在咽腔、口腔不受阻碍而形成的音叫元音，如 a.o.e.i.u 等
  - 气流在在咽腔、　口腔受阻碍而形成的音叫辅音，如 p.t.k.ch 等

# 清音与浊音

- 浊音：发音时声带要振动，包括元音和浊辅音
- 清音：发音时声带不振动，包括清辅音
- 英语的 /b/ 是浊音，和 /p/ 是清音，对是否送气不敏感
- 汉语的“波”和“泼”都是清音，区别在于送气和不送气
- 汉语的“波”不等于英语的 /b/，汉语的“波”和“泼”在英语人士中听起来都是 /p/
- 英语中 happened 、 star 中都有“清音不送气”，但在中国人听起来，似乎变成了“ b”和“ d”，这是一种错误的理解（所谓的“清音浊化”谬误）

# 音位和音系 (1)

- 音位是一个语音系统中能够区别意义的最小语音单位
- 在一个语音系统（语言或者方言）中，可能出现的音素通常很多，但有些音素在该语音系统中是没有必要加以区分的，这些音素虽然发音有区别，但这种区别并不用于区别不同的意义，在说该种语言的人听来，这种区别并不敏感，可以把这些音素加以合并，这就是音位

# 音位和音系 (2)

- 要从千差万别的语音中找出哪些是被用来区别意义的单位，一般采用归纳的方法：比较几个音质相近的音，如果它们能区别意义，就分开，如果不能区别意义，就归成一个单位，跟别的单位相区别。音位就是按语音的辨义作用归类出的最小音类。
- 若把几种语言放在一起比较会发现，从物理、生理角度看有所区别的语音现象，在不同语言里地位也不一样，在有的语言中需要加以区别的音，在另一语言里却可以不加区别。

# 音位和音系 (3)

- 如在普通话中， 要是把“ ba”（罢）念成“ pa”（怕），就成另一个词了，“ b[p]” 和“ p[p‘]” 是两个具有辨义作用的音位，要严格加以区别；而在英语中 [p‘] 是字母 p 在一般情况下的读音， [p] 只是 p 在 [s] 这个音素后的读音，如果你还是读成 [p‘] ，也不会引起词义上的误解， 它们在英语中没有区别词义的作用，是同一个音位的两个成员。在不同的语言（方言）中，音位及音位成员的情况是不一样的。

# 音位和音系 (4)

- 音位的划分，可以使得我们对一种语音系统的描述得以简化，不必关注一些对区分语义不发生影响的语素之间的区别
- 一个语言（方言）中所有的音位和音位之间的组合规则，构成了这一语言（方言）的音位系统，简称音系。我们平常所谈的某种语言的语音系统，指的就是这种音位系统，而不是未经归纳的音素系统。

# 内容提要

# 汉语普通话的语音

- 汉语的音节结构
- 声母与辅音
- 韵母与元音
- 声调
- 声韵配合规律
- 音变

# 汉语的音节结构

<table>
<tr><td rowspan="2">音节<br>(Syllable)</td><td>超音段成分<br>(super-segmental)</td><td colspan="4">声调 (Tone) 每个音节的必备成分</td></tr>
<tr><td>音段成分<br>(segmental)</td><td>声母 (Onset/Initial)<br>● 可缺少成分（在汉语音韵学上，没有声母也算是一个单位，称为零声母）<br>● 由辅助性成分组成</td><td colspan="3">韵母(Rime/Final)</td></tr>
<tr><td></td><td></td><td>韵头(Medial)</td><td>韵腹 (Nuclei)</td><td>韵尾 (Coda)</td></tr>
<tr><td></td><td></td><td></td><td>● 可缺少成分<br>● 由高元音组成</td><td>● 必备成分<br>● 一般由元音性成分组成</td><td>● 可缺少成分<br>● 由辅音性成分或高元音组成</td></tr>
</table>

# 声母和辅音 (1)

- 声母指音节开头的辅音。
- 在汉语中，声母都是由辅音充当的，但辅音不一定都是声母。
- 汉语中还存在着不做声母的辅音，如音节“ yáng”（羊）的韵尾 ng, 在普通话中只做韵尾不做声母；还有些既做声母又做韵尾的辅音，如“ nán”（难）这个音节中 , 韵尾和声母是同一个辅音 n 。

# 声母和辅音 (2)

- 普通话有 21 个声母，都是辅音：
  - 双唇音 b p m
  - 唇齿音 f
  - 舌尖中音 d t n l
  - 舌根音 g k h
  - 舌面音 j q x
  - 舌尖后音 zh ch sh r
  - 舌尖前音 z c s
- 普通话还有一个只作韵尾的辅音 ng
- 共有辅音 22 个

# 韵母和元音 (1)

- 韵母指音节中声母后面的部分。

如在“ dà ”（大）这个音节里，声母 d 后面的 a 就是韵母；在“ hóu ”（喉）这个音节里，韵母是 ou ；在“ gōng ”（工）这个音节里，韵母是 ong 。零声母音节整个由韵构成，如“ é ”（鹅），当然也可以理解为是由零声母和韵母构成的。

# 韵母和元音 (2)

- 普通话共有 39 个韵母，主要由元音构成，其中一部分由元音加鼻辅音构成。
- 普通话韵母可由韵头、韵腹、韵尾三部分组成。其中韵腹是韵母的主要元音，声音清晰响亮；韵头韵尾音值不固定，声音轻而短，表示韵母发音的起点或终点的方向。

# 韵母和元音 (3)

- 韵腹可由普通话中任何单元音充当;
- 韵头须由高元音 i 、 u 、 ü 充当;
- 韵尾只由高元音中的 i 、 u 和鼻辅音 n 、 ng 充当。
- 广东话中，还保留有古汉语的 p 、 t 、 k 、 m 四个韵尾。这可以解释 Beckham 普通话音译成“贝克汉姆”，而香港音译为“碧咸”，因为粤语中“碧”和“咸”的韵尾分别就是 k 和 m 。

# 韵母和元音：单韵母

- 由一个元音 ( 单元音 ) 构成的韵母叫单元音韵母。普通话有 10 个： a 、 o 、 e 、 ê 、 i 、 u 、 ü 、 -i (“ 资 ” 、 “ 此 ” 、 “ 思 ” 的韵母 ) 、 -i(“ 知 ” 、 “ 吃 ” 、 “ 诗 ” 、 “ 日 ” 的韵母 ) 、 er(“ 儿 ” )

# 韵母和元音：复韵母

- 前响复韵母：前一个元音清晰响亮，是韵腹；后一个元音音值模糊发音短而轻，是韵尾。前响复韵母有 4 个 :ai 、 ei 、 ao 、 ou 。
- 后响复韵母：后一个元音清晰响亮，是韵腹；前一个元音轻短，是韵头，表示复韵母发音的起点。后响复韵母有 5 个： ia 、 ie 、 ua 、 uo 、 ue
- 中响复韵母：由前响复韵母再加韵头 i 或 u 构成的，中间的韵腹清晰响亮，前后的韵头韵尾音值模糊。中响复韵母有 4 个： iao 、 iou 、 uai 、 uei
- 其中 uei 、 iou 这两个复合韵母，在读阴平声、阳平声时，中响元音常常会减弱，甚至接近消失。汉语拼音方案规定 uei 、 iou 在与声母拼合时省略 e 、 o ，而采用 ui 、 iu 的省写形式。

# 韵母和元音：鼻韵母

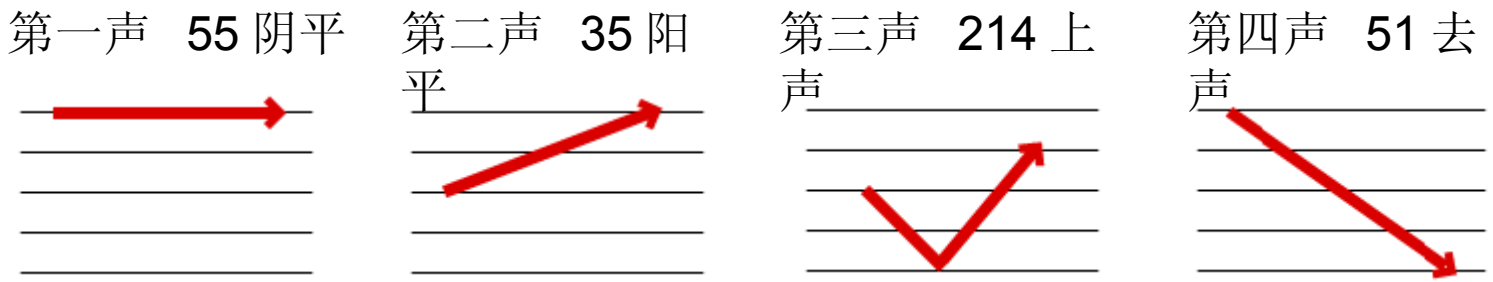
- 前鼻音韵母：由舌尖鼻音 n 作韵尾。发出主要元音后，　舌尖往上齿龈移动，直到抵住；软腭与小舌逐渐下垂，直到鼻腔通道打开；动作完成，发 n ，整个韵母发音就完毕。这类韵母共 8 个： an 、 en 、 in 、 un 是前响型， ian 、 uan 、 üan 、 uen 是中响型，由 an 、 en 加韵头形成。其中 uen 具有跟 uei 、 iou 同一类的语音变化现象。
- 后鼻音韵母：由舌根鼻音 ng 作韵尾。发出主要元音后，舌根上升，软腭下降，鼻腔通道打开；动作完成，发 ng ，韵母的发音就结束。后鼻音韵母也有 8 个： ang 、 eng 、 ing 、 ong 是前响型； iang 、 uang 、 ueng 、 iong 是中响型，分别由 ang 、 eng 、 ong 加韵头 i 、 u 形成

# 声调

- 声调指音节的高低升降，是音高的变化。它贯通整个音节。
- 汉语的声调包括调值和调类两个方面。
- 普通话有四种基本声调，因而有四个调类。这四个调类的调名按习惯沿用古代声调名称，分别是：阴平、阳平、上声、去声，又简称作一声、二声、三声、四声，统称“四声”。
- 调值就是声调的实际读法，也就是它具体的高低升降曲直长短的变化形式。

# 声调：五度制标调法

- 将音高平分为四格五度，由下向上标出 1.2.3.4.5 ，分别表示低、半低、中、半高、高的声调高度变化。这个高度不是指按频率测得的绝对音高，而是指调与调之间相互比较的相对音高。分别用横线、斜线或曲线来表示具体声调的高低升降变化形式。这些线沿什么高度画，要根据声调的具体情况。也可将每条声调线的高度数字，从五度标记上采下来，作为对调值的数字表示。
- “ 五度制标调法” 只是一种对调值变化的简单抽象。实际语言中声调调值的变化比这几条直线曲线要丰富得多。



# 汉语声调的变迁

- 古汉语的四声为“平上去入”，每一种声调又分阴阳。
- 入声字读音的特点是“入声短促急收藏”。
- 古汉语中的入声在普通话中已经消失，这个过程称为“入派三声”，也就是所有的入声字都变成了其他三种声调。这导致很多入声调古诗词用普通话读的时候不再押韵。
- 一些方言（如广东话）声调比普通话复杂得多，其中很多是保留下来的古汉语声调，如入声。因此有些古诗词用方言读起来比普通话押韵。

# 声韵配合规律

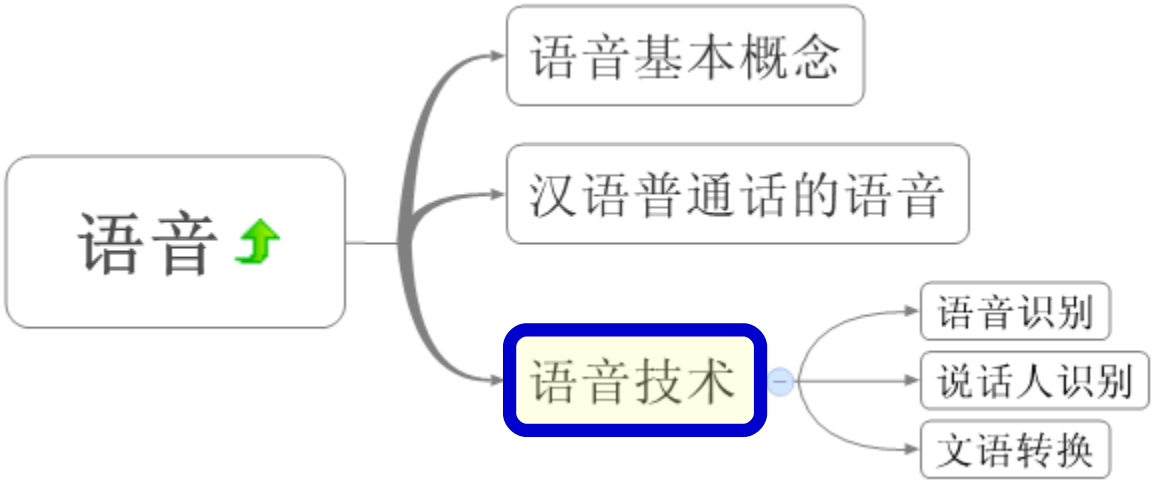
- 声母同韵母相拼就构成基本音节，再加声调便可表示意义。普通话有 21 个声母， 39 个韵母，如果任意搭配，可以拼出八百多个音节。但普通话的基本音节只有四百多个，这说明声母韵母的搭配不是任意的，而是有限制的，有规律的。如果不按规律去拼，组合出的音节可能不是普通话的音。
- 普通话中不会出现的音节如：

  bia biang riong ju zhin ……

# 音变

- 人们说话时，总是把一连串音素、音节、声调连续地说出来，形成语流。在语流中，音素间声调间相互影响，使其中的一些发生了变化，这就叫音变。
- 普通话常见音变形式：
  - 上声的变调：如“指导、搞好”，前一字读成阳平
  - “一”、“不”的变调：如“一样、不对“，前一字读成阳平
  - 轻声：语气词、后缀、重叠词的第二个音节等
  - 儿化：瓜子儿、豆芽儿、角儿、水珠儿、一块儿

# 内容提要



# 语音识别

- 语音识别系统的类型
- 连续语音识别基本原理
  - 声学特征提取
  - 声学模型
  - 语言模型
  - 解码与搜索
  - 自适应鲁棒性
- 语音识别系统的应用

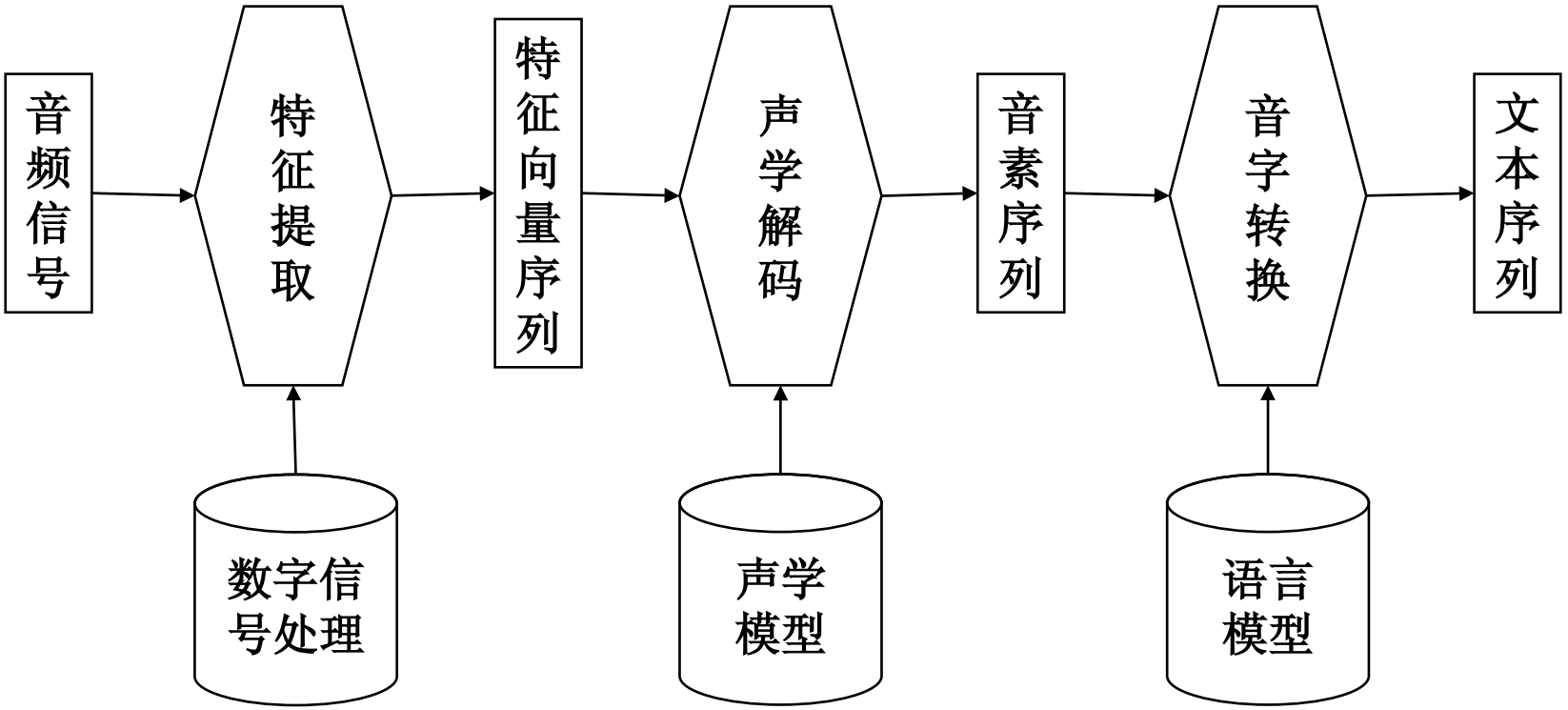
# 语音识别系统的类型 (1)

- 根据所要识别的对象来分：
  - 孤立词识别（字或词间有停顿，用于控制系统）
  - 连接词识别（十个数字连接而成的多位数字识别或由少数指令构成词条的识别，用于数据库查询、电话和控制系统）
  - 关键词检测（从连续的语音中检查出感兴趣的关键词）
  - 连续语音识别和理解（自然的说话方式）
  - 会话语音识别（识别出会话语言）

# 语音识别系统的类型 (2)

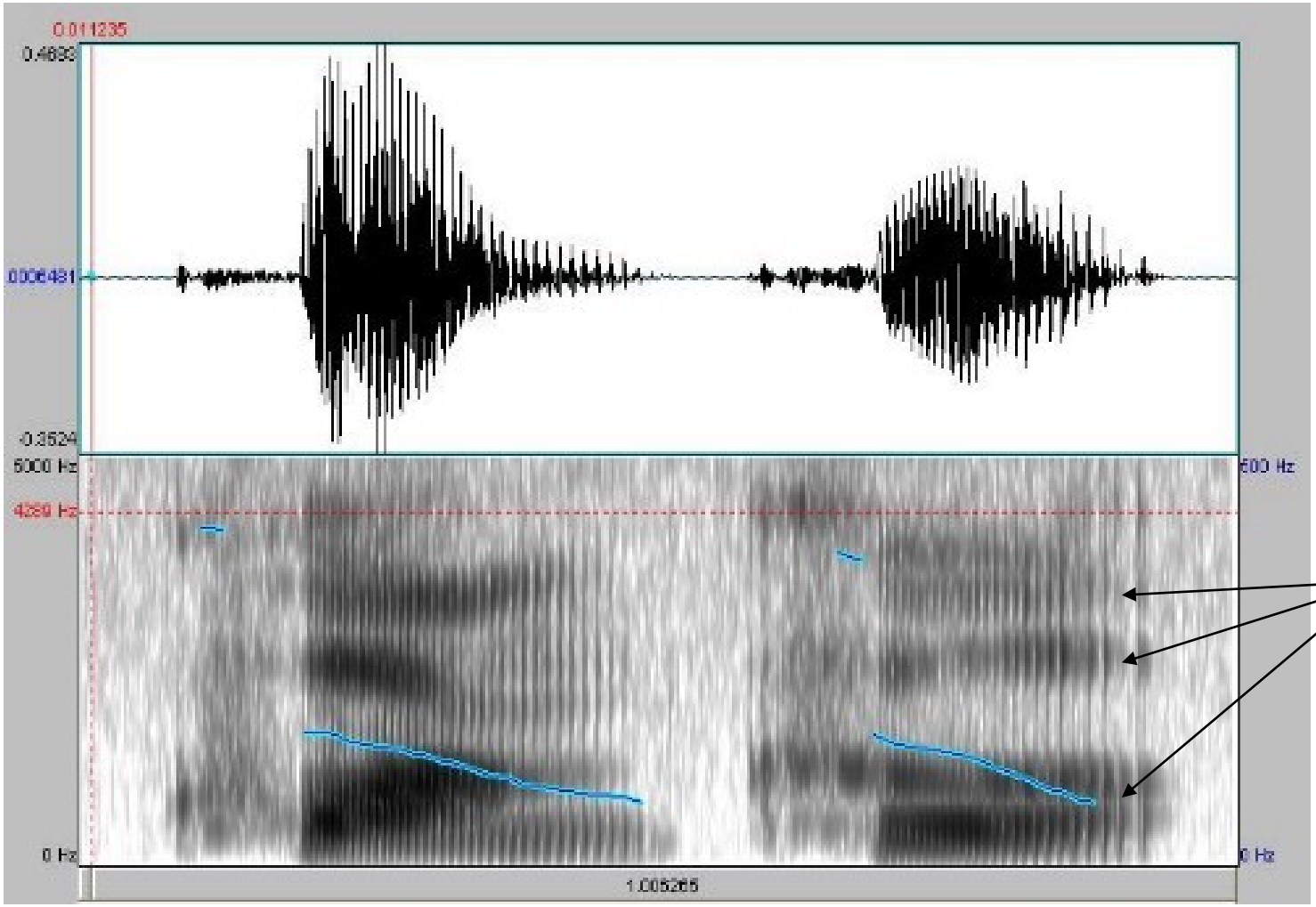
- 根据识别的词汇量来分：
  - 大词汇（1000 个以上的词汇，如会议系统）
  - 中词汇（20 ～ 1000 个词汇，如定票系统）
  - 小词汇（1 ～ 20 个词汇，如语音电话拨号）
- 根据讲话人的范围来分：
  - 单个特定人
  - 多讲话人（有限的讲话人）
  - 与讲话者无关

# 连续语音识别的基本原理



# 音频信号：旷课

波形图




频谱图

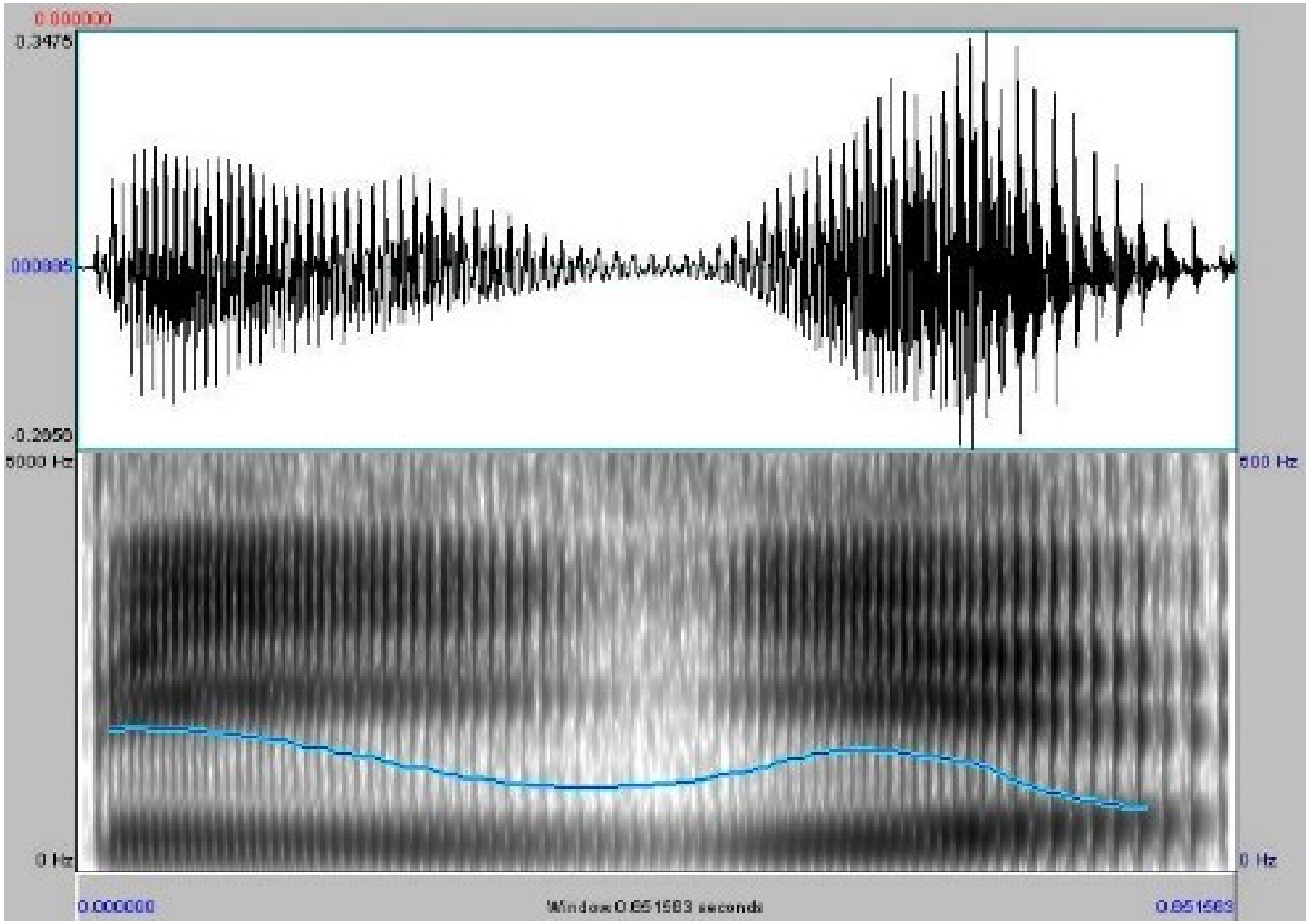
颜色深浅表示该频率处的信号强度

共振峰

# 音频信号：毕业

波形图

频谱图



# 音频信号的解读

- 从语音信号的音图中可以看到一些明显的特征，有经验的人甚至可以看图读出一些语音
- 共振峰：浊音的频谱图上，可以某些频段的强度处于峰值位置，这些峰值称为共振峰。对于元音而言，前三个峰值组合是相当可靠的识别特征。虽然峰值的频率值依基频的不同有所变化，不过其组合关系是很稳定的。
- 爆破音的特征：对于爆破音，可以看到语音信号在一个短暂的停顿后有突然的增强

# 常见的声学特征

- 线性预测系数 LPC
- 倒谱系数 CEP
- 梅尔倒谱系数 MFCC
- 感知线性预测 PLP

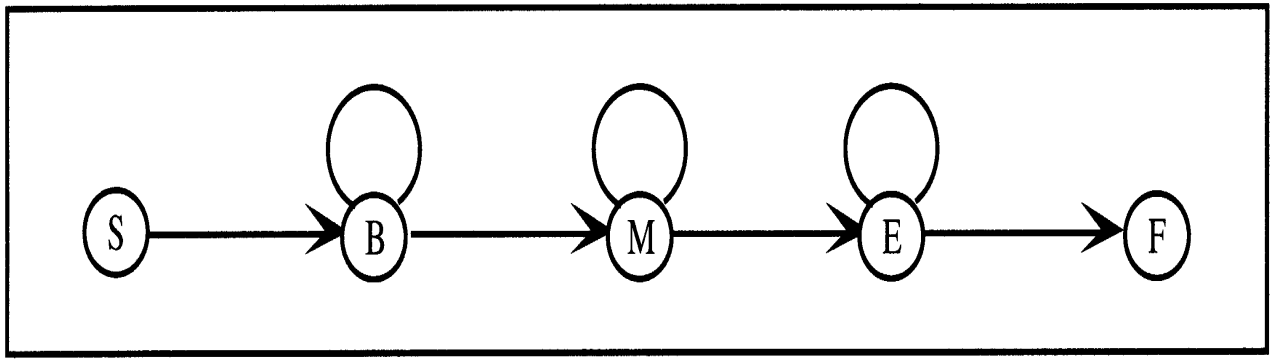
# 梅尔倒谱系数

- 计算方法
  - 首先用 FFT 将时域信号转化成频域
  - 之后对其对数能量谱用依照梅尔刻度分布的三角滤波器组进行卷积，
  - 最后对各个滤波器的输出构成的向量进行离散余弦变换 DCT，取前 N 个系数
- MFCC 特征向量序列
  - 原始的语音数据每一帧为一个表示振幅的实数值
  - 一种常用的 MFCC 特征向量序列为：经过上述的特征提取后得到 12 维向量序列，加上 1 维能量特征，共 13 维，对这 13 维向量再取两次差分，得到一个 39 维的向量序列，作为最终的 MFCC 特征向量序列

# 声学模型：隐马尔科夫模型

- HMM 声学建模：语音识别中使用 HMM 通常是用从左向右单向、带自环的拓扑结构来对识别基元建模，一个识别基元通常是一个音素或音节，用一个三至五状态的 HMM 来表示，一个词就是构成词的多个识别基元的 HMM 串行起来构成的 HMM，而连续语音识别的整个模型就是词和静音组合起来的 HMM。

# 识别基元的隐马尔科夫模型




- 识别基元的 HMM 都采用相同的结构形式，典型的如上图所示
- 这个 HMM 的观察值为一个向量，向量每一维是一个实数
- 这个 HMM 的输出概率用一个混合高斯分布来模拟
- 对于混合高斯分布中的每一个分量，都有一个 n 维均值向量和 n×n 维的协方差矩阵，其中 n 为特征向量的维数

# 声学模型的训练

- 声学模型的训练需要带标注的语音库。
- 语音库可以标注到各个层次。以汉语为例，可以标注到拼音层或音素层。
- 语音库通常还需要进行对齐，也就是给出每个标注在语音中的时间标签。这种对齐也可以对齐到各个层次，如汉语的拼音词或音素层。
- 语料库的标注不可能深入到 HMM 的结点层，因此声学模型的训练都是某种程度的无指导训练。
- 标注与对齐层次越深，训练效果越好，但语料库加工工作量越大。

# 上下文相关建模 (1)

- 协同发音，指的是一个音受前后相邻音的影响而发生变化，从发声机理上看就是人的发声器官在一个音转向另一个音时其特性只能渐变，从而使得后一个音的频谱与其他条件下的频谱产生差异。
- 上下文相关建模方法在建模时考虑了这一影响，把一个音根据其前后音加以细分，从而使模型能更准确地描述语音。
- 只考虑前一音的影响而将一个音细分得到的音的称为 biphone（双音子），考虑前一音和后一音的影响细分得到的音进行的称为 triphone（三音子）。
- 按上述原则细分后的音素又称为“上下文相关音素”（ Phoneme in Context，缩写为 PIC ）

# 上下文相关建模 (2)

- 语音识别系统选择识别基元的要求是，有准确的定义，能得到足够数据进行训练，具有一般性。
- 英语通常采用上下文相关的音素（ PIC ）建模，汉语的协同发音不如英语严重，可以采用音节建模。
- 英语的上下文相关建模通常以音素为基元，由于有些音素对其后音素的影响是相似的，因而可以通过音素解码状态的聚类进行模型参数的共享。聚类的结果称为 senone 。决策树用来实现高效的 triphone 对 senone 的对应，通过回答一系列前后音所属类别（元 / 辅音、清 / 浊音等等）的问题，最终确定其 HMM 状态应使用哪个 senone 。分类回归树 CART 模型用以进行词到音素的发音标注。

# 语言模型：N 元语法

- 语音识别中的语言模型主要采用 N 元语法。常用的是二元语法和三元语法。
- 语言模型的性能通常用交叉熵和困惑度（Perplexity）来衡量。交叉熵的意义是用该模型对文本识别的难度，或者从压缩的角度来看，每个词平均要用几个位来编码。困惑度的意义是用该模型表示这一文本平均的分支数，其倒数可视为每个词的平均概率。
- 常用的平滑技术有 Good-Turing 平滑、删除插值平滑、 Katz 平滑和 Kneser-Ney 平滑。

# 解码与搜索 (1)

- 解码就是在所有可能的结果中，找到一个最优的结果，也就是一个搜索过程
- 常见的搜索算法有 Viterbi 算法和 Beam Search 等，目前最常用的算法是 Beam-Viterbi 算法
- N-best 搜索和多遍搜索：为在搜索中利用各种知识源，通常要进行多遍搜索，第一遍使用代价低的知识源，产生一个候选列表或词候选网格，在此基础上进行使用代价高的知识源的第二遍搜索得到最佳路径。第一遍搜索使用的知识源通常有声学模型、语言模型和音标词典。为实现更高级的语音识别或口语理解，往往要利用一些代价更高的知识源，如 4 阶或 5 阶的 N-Gram 、 4 阶或更高的上下文相关模型、词间相关模型、分段模型或语法分析，进行重新打分。

# 解码与搜索 (2)

- N-best 搜索产生一个候选列表，在每个节点要保留 N 条最好的路径，会使计算复杂度增加到 N 倍。简化的做法是只保留每个节点的若干词候选，但可能丢失次优候选。一个折衷办法是只考虑两个词长的路径，保留 k 条。
- 词候选网格以一种更紧凑的方式给出多候选，对 N-best 搜索算法作相应改动后可以得到生成候选网格的算法。
- 前向后向搜索算法是一个应用多遍搜索的例子。当应用简单知识源进行了前向的 Viterbi 搜索后，搜索过程中得到的前向概率恰恰可以用在后向搜索的目标函数的计算中，因而可以使用启发式的 A* 算法进行后向搜索，经济地搜索出 N 条候选。

# 自适应与鲁棒性

- 语音识别系统的性能受许多因素的影响，包括不同的说话人、说话方式、环境噪音、传输信道等等。提高系统鲁棒性，是要提高系统克服这些因素影响的能力，使系统在不同的应用环境、条件下性能稳定；自适应的目的，是根据不同的影响来源，自动地、有针对性地对系统进行调整，在使用中逐步提高性能。具体包括说话人自适应、方言自适应、噪声自适应等等。
- 解决办法按针对语音特征的方法（以下称特征方法）和模型调整的方法（以下称模型方法）分为两类。前者需要寻找更好的、高鲁棒性的特征参数，或是在现有的特征参数基础上，加入一些特定的处理方法。后者是利用少量的自适应语料来修正或变换原有的说话人无关（SI）模型，从而使其成为说话人自适应（SA）模型。

# 语音识别系统的应用

- 语音命令系统：通过语音给计算机发出指令。可以用于语音遥控器、手机语音拨号等场合。
- 听写机：大词汇量、非特定人、连续语音识别系统通常称为听写机。可用于语音录入、会议记录、监听等场合。
- 对话系统：用于实现人机口语对话的系统称为对话系统。受目前技术所限，对话系统往往是面向一个狭窄领域、词汇量有限的系统，其题材有旅游查询、订票、数据库检索等等。由于目前的系统往往词汇量有限，也可以用提取关键词的方法来获取语义信息。

# 语音识别技术的发展前景

- 虽然语音识别技术得到了很多的应用，但与人们的预期相比，还有较大的距离
- 连续语音识别应用中面临的主要问题有：
  - 噪音问题：强噪音环境下如何达到高识别率
  - 方言：如何适应各种不同的方言口音
  - 自然（Spontaneous）语音：自然语音往往包含了多变的语速、语气、韵律和真实的情绪，以及严重的协同发音，这就会造成大量的音素级的插入、删除和替换现象。此外，不同的人具有不同的口音背景和发音习惯，这些都会严重影响语音识别的准确率

# 说话人识别 (1)

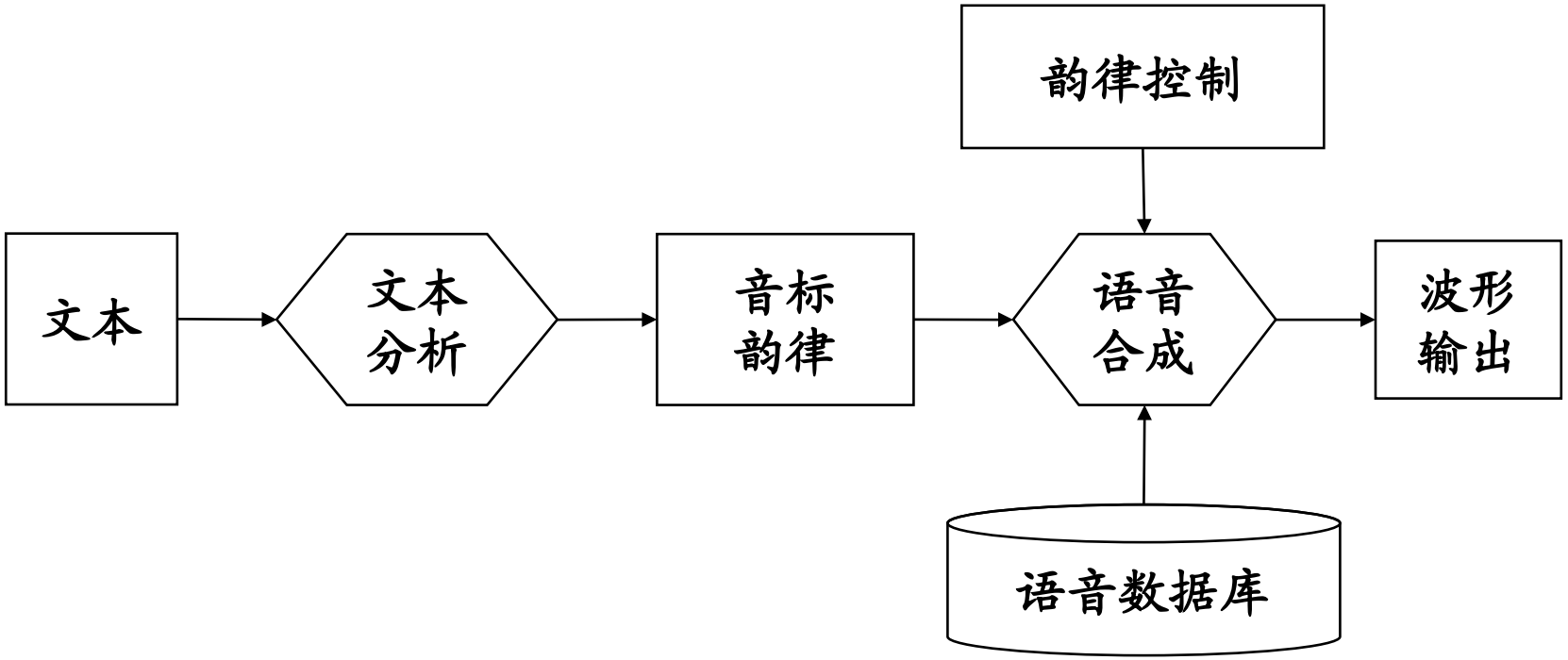
- 说话人识别，又称声纹识别，分为说话人确认和说话人辨认。
- 说话人确认要求判断某一段语音的说话人是否为某一特定的人
- 说话人辨认要求找出某一段语音的说话人是一个已知集合中的哪一个人（或者都不是）。
- 与其他生物识别技术相比，声纹识别具有更为简便、准确、经济及可扩展性良好等众多优势，可广泛应用于安全验证、控制等各方面，特别是基于电信网络的身份识别。

# 说话人识别 (2)

- 说话人识别的基本原理是将语音输入语音特征与特定人的语音特征进行比对并进行判断。
- 说话人识别包括特征提取和模式匹配两个过程
- 声音中包含的个人特征信息有两种，一种是声道长度、声带等先天性发音器官的个人差别所产生；另一种是由方言、语调等后天性讲话习惯产生。前者是以共振峰频率的高低、带宽的大小、基频的平均值和频谱的基本形状来表现；后者以单词的时间长等特征来表现。
- 常见的模式匹配方法包括：统计分类方法、动态时间规整方法、矢量量化方法、隐马尔科夫模型方法、神经网络方法等等。

# 文语转换的基本原理

- 文语转换： Text to Speech ， TTS



# 文本分析

- 文本分析包括语音排岐和韵律分析等过程。
- 语音排岐要确定多音字词的读音。
- 人类的语音中含有丰富的韵律信息语音合成时如果不考虑韵律信息，发出的语音将是单调而机械的机器声，自然度很差，也不易理解
- 韵律分析，就是为了给待合成的文本加上韵律信息，以便语音合成模块能够产生自然的语音
- 韵律分析的好坏很大程度决定了语音的自然度，糟糕的韵律分析会导致误解，还不如没有
- 完美的韵律分析涉及对语言句法、语义、语用层面的深入理解，实际上是非常困难的。

# 韵律

- 语音中的韵律信息包括：
  - 语调：语音基频的上升或下降
    - 英语中句尾语调的上升通常表示疑问，而下降表示断言
    - 汉语中由于协同发音现象汉字的音调通常会发生变化
  - 语速：音节的长短变化、停顿等
    - 一个词内各音节长短不尽相同，需要加以区分
    - 适当的停顿可以表示句子或者短语的开始和结束，缺乏停顿的语音会使听者感到难以理解。
  - 强度：语音的强弱变化、重音、弱音等。
    - 重音可以指示句子的焦点，焦点不同会导致句子完全不同的理解

# 语音合成

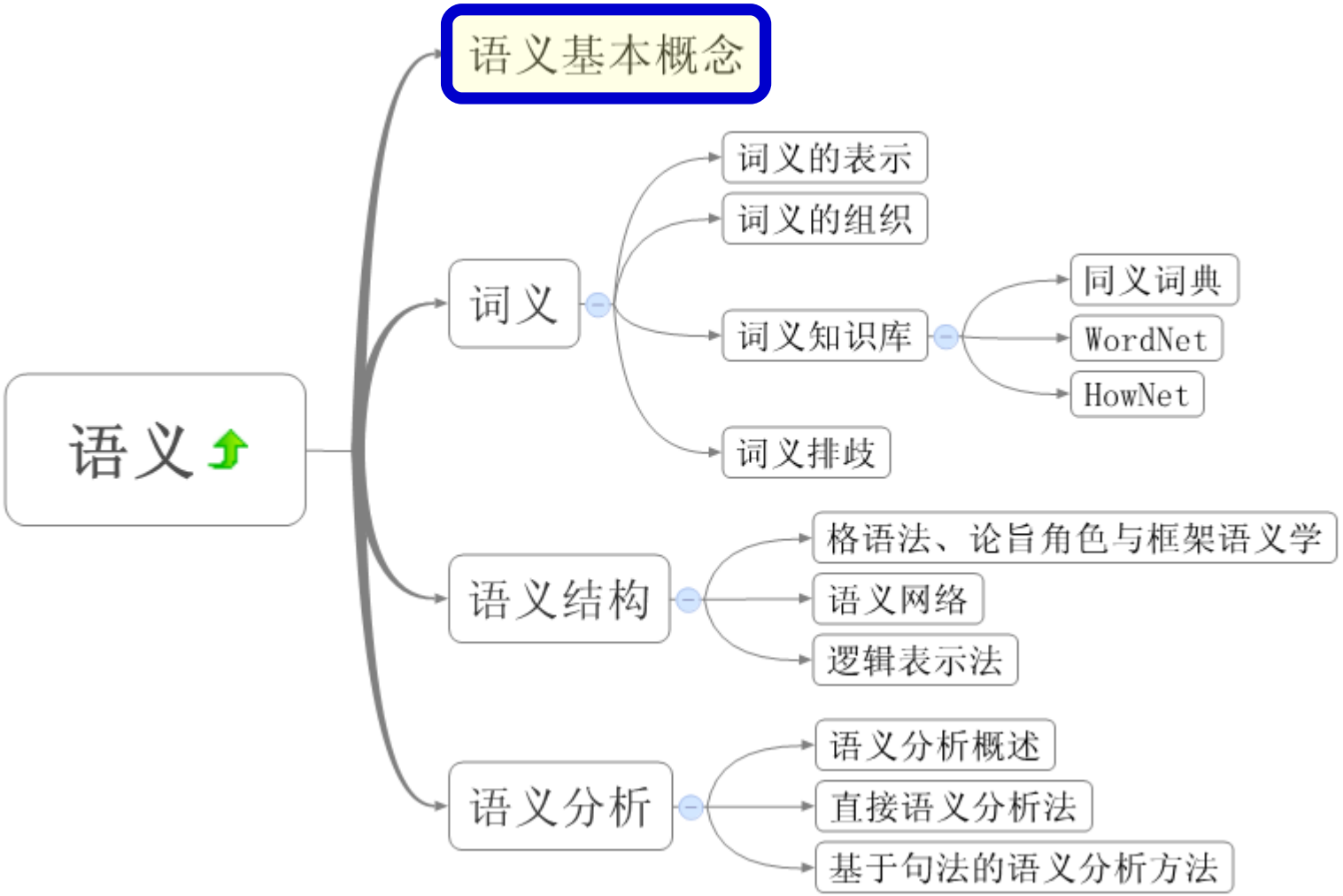
- 语音合成主要有两种方法：参数合成法和波形拼接法
- 参数合成法是通过计算机根据给定的参数产生一种特定的组合波形来模拟各种语音
- 波形拼接法是事先在系统中存储大量人工发音片段，合成的时候用这些片段组合成要合成的语音
- 波形拼接法自然度较好，尤其是语音数据库规模大质量高的时候。但其对于韵律参数的调整能力较弱，处理片段间的协同发音不太方便。因此二者相结合的方法是今后的发展趋势。

# 语音合成的应用与前景

- 语音合成目前达到了相当的自然度，在大量场合得到了广泛的应用，如电话查询、信息广播、教学、游戏等等。
- 为了研究更加自然的语音，需要对语音中含有的语义、语用等信息进行更深入的研究，包括语音所包涵的情感等因素

# 内容提要

# 内容提要


语义
语义基本概念
词义
词义的表示
词义的组织
词义知识库
同义词典
WordNet
HowNet
词义排歧
语义结构
格语法、论旨角色与框架语义学
语义网络
逻辑表示法
语义分析
语义分析概述
直接语义分析法
基于句法的语义分析方法

# 语义基本概念

- 简单的说，语义就是语言的意义。
- “ 意义”的含义非常复杂：
  - 意义可以是某种客观的事实
  - 意义也可以是说话人的思维状态
- 语义研究的方法通常是：
  - 首先，给语言所描述的对象（客观事实或者思维状态）建模
  - 其次，解释语言和对象模型直接的对应关系

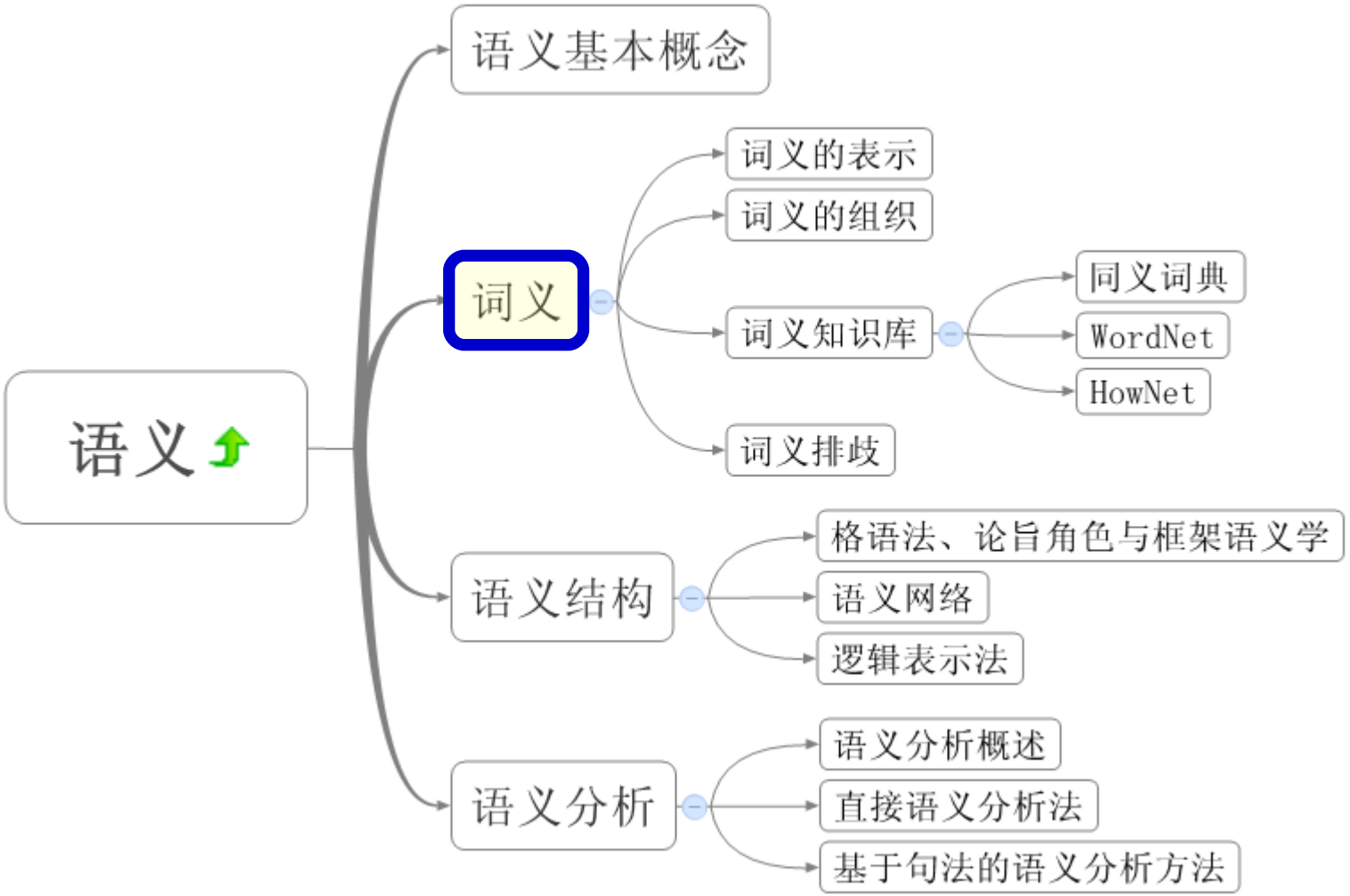
# 语言描述的对象 (1)

- 如前所述，语言所描述的对象主要包括客观事实和思维状态
- 本质上，语言所描述的都是说话者的思维状态，所谓客观事实，都是通过思维间接反映出来的
- 为了研究的方便起见，我们有时可以忽略思维这个中间过程，直接研究语言和客观事实之间的关系

# 语言描述的对象 (2)

- 客观世界和思维世界都是非常复杂的，我们不可能建立一个包罗万象的模型
- 任何一种语义理论，实际上都是建立一个简化的世界模型，然后研究这个模型与实际语言的某一个子集之间的对应关系

# 内容提要


语义
语义基本概念
词义
词义的表示
词义的组织
词义知识库
同义词典
WordNet
HowNet
词义排歧
语义结构
格语法、论旨角色与框架语义学
语义网络
逻辑表示法
语义分析
语义分析概述
直接语义分析法
基于句法的语义分析方法

# 词义

- 词义分析
  - 义位与义素
  - 语义场和义素分析法
- 词义组织：知识本体
- 词义排岐（WSD）

# 义位

- 义位：义位是从具体语言或方言中归纳出来的、能够独立运用的、具有独立形式标志的最小意义单位。
- 义位是语义系统中最基本、最自然、最现成的语义单位。
- “义位”的概念与“词义”、“义项”有重叠，但又有区别：
  - 一个词的“词义”可能包含多个“义项”，每个“义项”都是一个“义位”
  - 一个词或者一个语素都可以有“义项”，但只有词的“义项”才称为“义位”，语素的“义项”不是“义位”，因为“语素”的“义项”不能独立使用

# 义素

- 义素：义素是构成义位的最小意义单位，即义位的区别性特征。它是由分解义位得到的比义位低一级的语义单位。在语义体系中它是无法被直接观察到的，所以属于语义的微观层次。
- 义素属于语义的微观层次，它没有对应的语言形式，不能直接观察到，只有组合起来才能形成现实的语义。义素是义位的组成成分，也是最小的语义单位。

# 语义场

- 在词义上具有类属关系的词集合在一起所形成的聚合体即是语义场。
- 在同一语义场中，根据词义上的类属关系，词义可分为上位意义和下位意义（简称上位义和下位义），如A是B的一种，则B是A的上位词，B词义是A词义的上位义；A是B的下位词，A词义是B词义的下位义。
- 语义场根据上、下位意义关系形成层次结构，一种语言中的词汇根据词义的类属关系，可划分出若干个大语义场，每个场下面可以再分出若干个“子场”，“子场”下面可再分出“次子场”等等。这样就可以使词汇体系及词义系统有次序地展现出来。

# 义素分析法

- 义素分析法指从义素的角度分析义位的方法。是现代语义学的重要分析方法和重要范畴。它借助于结构语言学的对比性原则，将一组义位放在一起进行对比分析，从中寻找出共性义素和互有差别的义素，这样既可以看到同组义位之间的联系，也可以看到它们之间的区别。
- 这种方法类似于数学中提取公因式的方法，也类似于音位学中寻找音位的区别性特征的方法。
- 义素分析法的步骤：
  - 确定语义场
  - 比较语义场中的各个义位，找出共同特征和区别特征（义素）
  - 用得到的义素对各个义位进行描写

# 义素分析法：例 (1)

| 义素 \ 对比词 | 鞋 | 靴子 | 袜子 |
|---|---|---|---|
| 服饰 | + | + | + |
| 穿在脚上 | + | + | + |
| 走路时着地 | + | + | − |
| 有筒 | − | + | + |

# 义素分析法：例 (2)

| 义素 \ 对比词 | 父亲 | 伯父 | 舅舅 | 弟弟 |
|---|---|---|---|---|
| 人 | + | + | + | + |
| 男性 | + | + | + | + |
| 长辈 | + | + | + | − |
| 有血统关系 | + | + | − | + |
| 直系亲属 | + | − | − | + |

# 词义的组织

- 词义的组织有两种常见的方法
  - 分类法：将词义按同义关系组织成类，类与类之间按照上下文和其他关系组成分类体系
  - 分解法：将词义用更基本的单位（如义素或义原）来表示
- 要将一种语言中的所有词义组织成一个完整的体系，是一件浩大的工程，典型的工作有 WordNet， HowNet 等
- 词义体系的构建，是一件主观性很强的工作，面临实际应用的时候，很多问题很难有圆满的解决办法，但这件工作又是不得不做的，只能尽量做好

# 知识本体 (1)

- 知识本体是对概念体系的明确的、形式化的、可共享的规范（ An ontology is a formal explicit specification of a shared conceptualization, Studer, 1997 ）
- 具体地说，如果我们把每一个知识领域抽象成一个概念体系，再采用一个词表来表示这个概念体系，在这个词表中，要明确地描述词的涵义、词与词之间的关系、并在该领域的专家之间达成共识，使得大家能够共享这个词表，那么，这个词表就构成了该领域的一个知识本体。
- 知识本体已经成为了提取、理解和处理领域知识的工具，它可以被应用于任何具体的学科和专业领域，知识本体经过严格的形式化之后，借助于计算机强大的处理能力，可以对于人类的全部知识进行整理和组织，使之成为一个有序的知识网络

# 知识本体 (2)

- 知识本体的本质是对概念和概念之间的关系进行明确的、形式化的描述
- 知识本体中常见的关系描述
  - 上下位关系
  - 整体部分关系
  - 同义、近义、反义、对义关系
- 知识本体可以用于描述一种语言的通用的词汇语义知识（如 WordNet），也可以用于描述某一专业的术语，甚至用于一个特定的小领域
- 由于知识本体描述的是概念，因此很多研究工作开始试图利用知识本体作为语言之间、领域之间知识交流的工具和平台。如现在很多种语言都在构造其相应版本的 WordNet，这就为各种语言的词义之间的交流奠定了良好的基础

# 知识本体 (3)

- 语义网（ Semantic Web ）的出现，对知识本体的研究起到了很大的促进作用
- 语义网的初衷，是实现语义的精确表达，以便用户对网络知识进行准确的查询和推理。
- 而要做到这一点，就需要借助于知识本体，以确定在语义描述中每一个概念的准确含义，以及与其他概念之间的关系。
- 在语义网中，每一个网页都要指明自己所采用的知识本体，以便确定网页中每一个概念的准确含义。如果整个 Internet 上所有的知识本体都能够互相关联、互相共享，就可以实现基于整个网络的知识推理，这是一种理想的境界。

# 知识本体 (4)

- 语义网的相关研究发展很快，已经形成了一些相关的国际标准，包括网络知识本体的描述标准 OWL （ Web Ontology Language ）
- 语义网和知识本体的相关资料：
  - http://www.w3.org/2001/sw
  - http://www.semanticweb.org
- 《 Scientific American 》（科学美国人） 2001 年 5 月出版了 Semantic Web 专辑
  - http://www.sciam.com/

# 词义排歧

- 一词多义是自然语言中最常见的岐义现象，越是常见的词岐义现象往往越严重
  - 打：打人、打仗、打饭、打毛衣、打渔……
  - play： play football， play piano， play game …
- 词义排歧是很多自然语言处理工作的重要基础，如机器翻译、信息检索、自动问答、语音合成等等
- 词义排歧是研究得比较充分的一个问题，早期的一些相关研究可见 Computational Linguistics 的词义排歧专辑（ 1998 年）：

  http://acl.ldc.upenn.edu/J/J98/

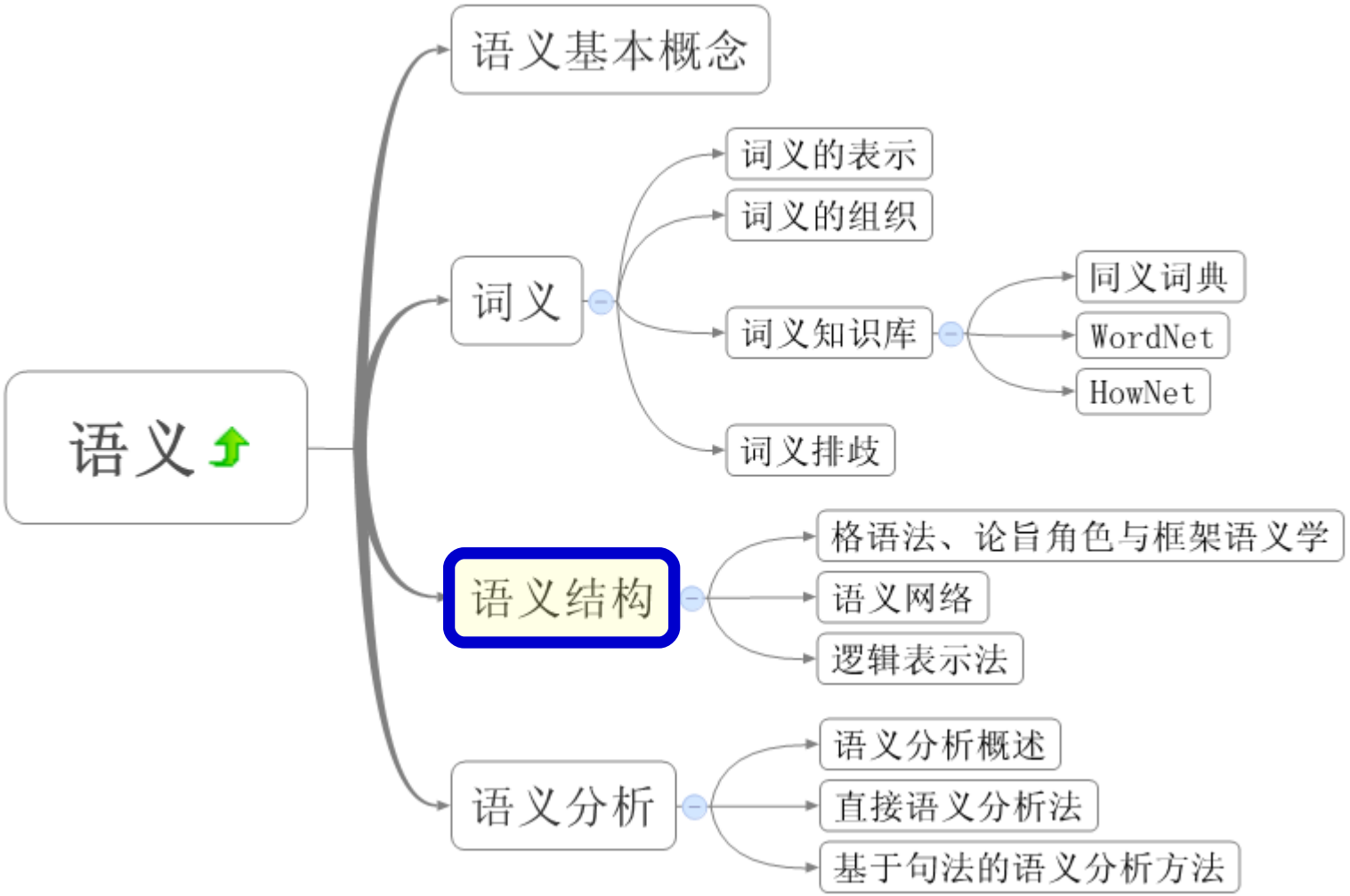
# 词义排歧方法的分类

- 基于词典的方法
- 基于语料库的方法
  - 有指导的方法
    - 基于实例的方法
    - 统计机器学习方法： Bayes 、互信息、决策树、最大熵、支持向量机、粗糙集……
  - 无指导的方法
    - 词义聚类
    - 语料增强方法： Bootstraping
    - 双语语料增强方法

# 词义的泛指与特指

- 在具体的句子中，任何一个概念都有泛指和特指之分，在语义研究中，这二者应严格加以区分
  - 贾宝玉爱林黛玉：这里“爱”是一个特指的概念，就是指贾宝玉对林黛玉的爱
  - 爱是没有国界的：这里“爱”是泛指，所有的“爱”都是不分没有国界的
  - 贾宝玉看书：这里“书”是特指，指贾宝玉正在看的某一本书
  - 贾宝玉爱书：这里“书”是泛指，只要是书，贾宝玉都爱
- 汉语中，由于没有必须出现的限定词（如英语的冠词），所以一个汉语名词短语的泛指和特指是比较难判定的，往往必须结合上下文中才能判定。而且对于汉语说话人而言，往往不容易意识到这二者之间的区别。

# 内容提要


语义
语义基本概念
词义
词义的表示
词义的组织
词义知识库
同义词典
WordNet
HowNet
词义排歧
语义结构
格语法、论旨角色与框架语义学
语义网络
逻辑表示法
语义分析
语义分析概述
直接语义分析法
基于句法的语义分析方法

# 语义结构

- 格语法、论旨角色与框架语义学
- 语义网络
- 逻辑表示法

# 格语法 (1)

- 句法分析不足以刻划句子中词语之间的语义关系：
  - 动宾关系：
    - 吃饭、吃食堂、吃大碗
    - 救人、救火、救灾
  - 主谓关系：
    - 房子烧起来了，火烧起来了，他烧饭
- 要准确刻划词语之间的语义关系，需要定义另外一套体系，这就是格

# 格语法 (2)

- 考虑下面句子：
  - The door opened.
  - The key opened the door.
  - The boy opened the door.
  - The door was opened by the boy.
  - The boy opened the door with a key.
- 在上面的句子中，有一个动词 open 和三个定指的名词短语： the door 、 the boy 、 the key ，从语义上看，这个动词和这三个名词短语的关系都是相同的，我们把这些关系定义如下：
  - 施事格：该动作的发出者，动词 open 和 the boy 的关系
  - 客体格：该动作所影响的事物，动词 open 和 the door 的关系
  - 工具格：该动作所凭借的工具，动词 open 和 the key 的关系

# 格语法 (3)

- 格语法是美国语言学家 Charles J. Fillmore 于 1966 年提出的一种理论。其经典著作为：
  - Towards a modern Theory of case, 1966
  - The case for case （格辨）， 1968
  - Some Problems for Case Grammar， 1971
- 格语法源自转换生成语法，现在已被更新的理论所取代，不过其核心思想已被普遍接受，如管辖约束理论中的论旨理论
- 格（ case ）这个概念来源于语法理论，很多语言中的名词都具有明显的格标记。但在格语法中，“格”指的是“深层格”，描述的是某种语义关系，而不是语法关系。不管语言本身有没有句法格，深层格都是存在的，而且深层格与句法格不一定一致。

# 格语法 (4)

- 基本规则
  - S→M+P （M：情态，P：命题）
  - P→V+C1+C2+…+Cn （V：动词，C：格）
  - C→K+NP （K：格标，NP：名词短语）
- 每一个动词的格框架是固定的：包括必备格和可选格。必备格是必须出现的，可选格可有可无，除此之外的其他格是不允许出现的
- 除了主语和谓语，很多格出现在介词短语中，介词可以认为就是某种格标志
- 格语法还规定了转换生成的方法
- 采用格语法可以用特定的算法进行语义分析

# 格语法 (5)

- Fillmore 定义的格：
  - 施事格（ Agentive ）
  - 受事格（ Dative ）
  - 使成格（ Factitive ）
  - 工具格（ Instrumental ）
  - 方位格（ Locative ）
  - 客体格（ Objective ）
  - 受益格（ benefactive ）
  - 来源格（ Source ）
  - 目标格（ Goal ）
  - 伴随格（ Comitative ）

# 格语法 (6)

- 格语法的缺陷
  - 转换的思想随着语言学的发展已被更完善的理论所取代
  - 格集合的确定没有客观的标准，主观性太强，每一个具体应用都有自己的格集合
  - 不管怎样定义格集合，在实际的应用中总是会遇到各种问题，例如无法区分句子中的两个名词短语的情况（总是不够细），但太细了又会导致其他问题

# 格语法的发展

- 在管辖约束理论中的题元理论源自于格语法，在题元理论中“格”被称为“论旨角色”
- Fillmore 将“格语法”发展成为“框架语义学”，并试图用框架语义学对主要的英语的动词进行描述，构造了一个 FrameNet

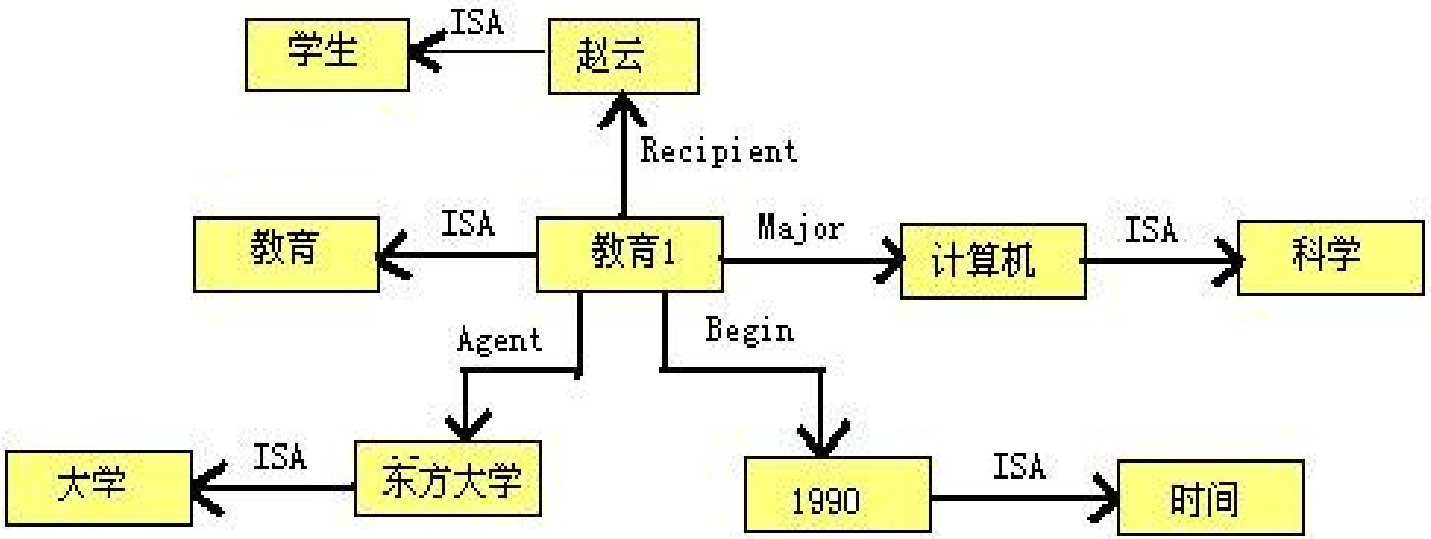
# 语义网络 (1)

- 语义网络 (Semantic Network) 由美国心理学家 M. R. Quilian 于 1968 年在研究人类联想记忆时提出。
- 1972 年，美国人工智能专家 R. F. Simmons 和 J. Slocum 首先将语义网络用于自然语言理解系统中。
- 1977 年，美国人工智能学者 G. Hendrix 提出了分块语义网络的思想。

# 语义网络 (2)

- 语义网络的基本思想是用有向图来表示语义结构：
  - 结点：表示概念，或者实体，或者事件
  - 弧：表示概念之间的关系，这些关系包括：
    - ISA ：具体－抽象关系
    - KINDOF ：类属关系
    - PARTOF ：整体－部分关系
    - BEFORE 、 AFTER 、 AT ：时间关系
    - LOCATED-ON 、 LOCATED-AT 等：位置关系
    - 格关系

# 语义网络一例




- 赵云是一个学生。
- 他在东方大学主修计算机课程。
- 他入校的时间是 1990 年。

# 语义网络 (3)

- 语义网络可以表达各种逻辑关系（与、或、非、蕴含、量词）
- 语义网络也可以进行一些逻辑推理（属性继承、匹配等）
- 语义网络具有表达能力强、方便、直观等特定，并且具有一定的推理能力，在很多场合得到了应用

# 语义的逻辑表示

- 形式逻辑是一门研究比较充分的学科，其语义有严格的定义（如模型论语义学）
- 形式逻辑的命题是无歧义的
- 人们很容易联想到要用逻辑形式来表示自然语言的意义，以消除自然语言的歧义性
- 简单的一阶谓词逻辑不足以表达自然语言的语义，因此人们研究了各种更加复杂的逻辑，如模态逻辑、缺省逻辑、时态逻辑、真值维护系统等等
- 由于逻辑具有表达精确无歧义的特点，因此受到了很多理论研究工作者的青睐，发展出了蒙太格语法、情境语义学等语义理论，但这些形式由于过于复杂，很难在工程实践中得到应用，大多都处于理论研究阶段

# 一种逻辑形式表示法 (1)

- 概念的表示：
  - 抽象概念：表示为一阶谓词（ Unary Predicate ）
  - 具体名词概念：表示为项（ Term ）
  - 具体动词概念：表示为谓词（ Predicate ）
- 命题的表示：
  - 贾母是主人： (Master 贾母 )
  - 贾宝玉爱林黛玉： (Love1 贾宝玉 林黛玉 )
  - 贾宝玉看书： (And (Read1 贾宝玉 书 1) (Book 书 1) )

    注意：这里“书 1”不能用泛指的“ Book”或者“书”表示

# 一种逻辑形式表示法 (2)

- 逻辑关系的表示：
  - 贾宝玉爱林黛玉，或者贾宝玉爱薛宝钗

    (Or (Love1 贾宝玉 林黛玉 )
    (Love1 贾宝玉 薛宝钗 ) )
  - 贾宝玉爱林黛玉，也爱薛宝钗

    (And (Love1 贾宝玉 林黛玉 )
    (Love1 贾宝玉 薛宝钗 ) )

# 一种逻辑形式表示法 (3)

- 量词的表示
  - 贾宝玉爱每个女孩

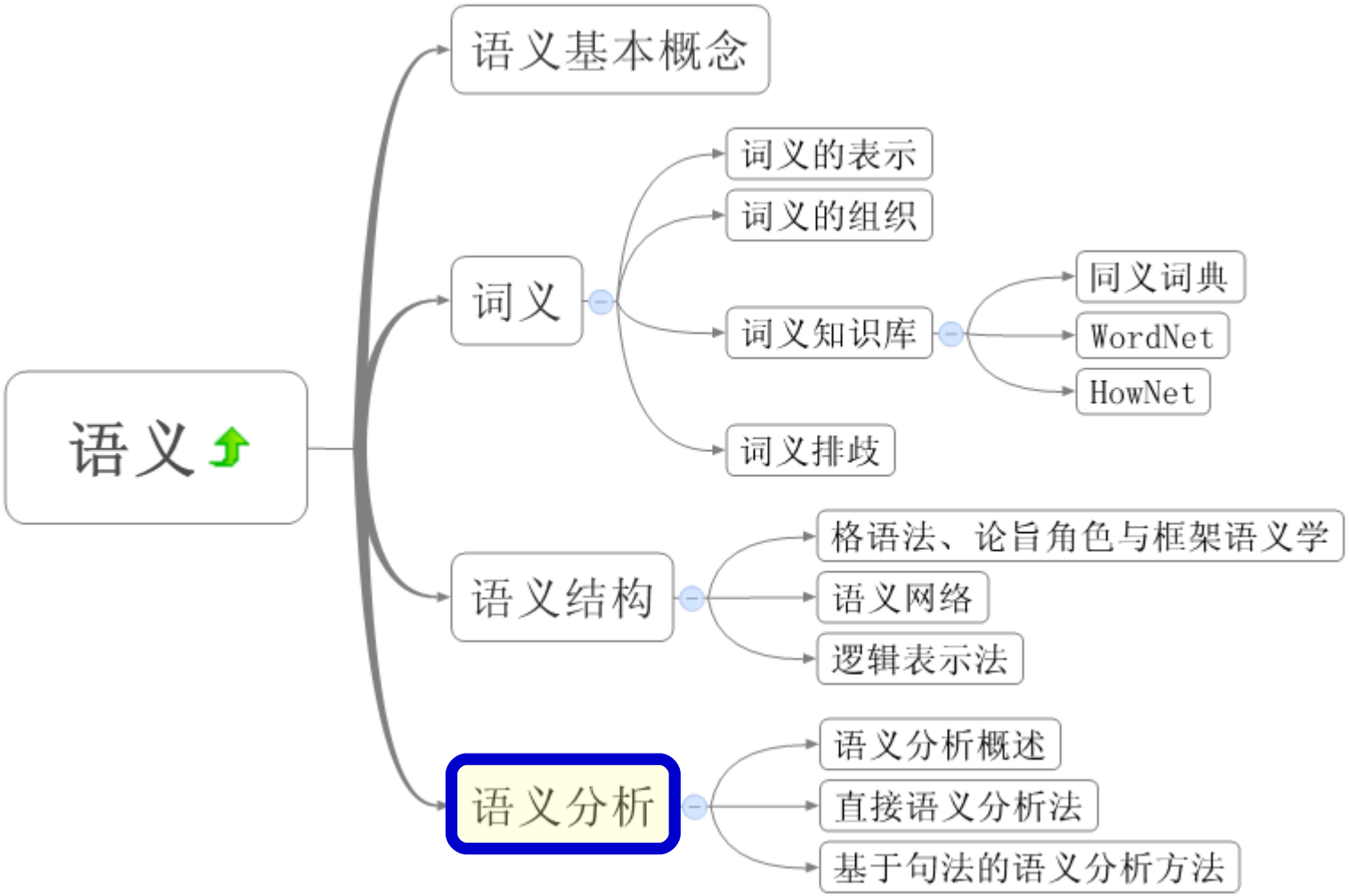
    (Every x: (Girl x) (Love1 贾宝玉 x) )
  - 有个女孩爱贾宝玉

    (Exist x: (Girl x) (Love1 x 贾宝玉 ) )
  - 有些女孩爱贾宝玉

    (Some x: (Girl x) (Love1 x 贾宝玉 ) )
  - 大多数好女孩爱贾宝玉

    (Most x: (And (Girl x) (Kind1 x)) (Love1 x 贾宝玉 ) )

# 内容提要


语义
语义基本概念
词义
词义的表示
词义的组织
词义知识库
同义词典
WordNet
HowNet
词义排歧
语义结构
格语法、论旨角色与框架语义学
语义网络
逻辑表示法
语义分析
语义分析概述
直接语义分析法
基于句法的语义分析方法

# 语义分析

- 语义分析概述
  - 语义分析要达到的目标
  - 语义分析要解决的问题
- 直接语义分析法
  - 语义语法
  - 统计语义分析法
- 基于句法的语义分析法
  - 选择限制与语义优选
  - 统计语义排岐

# 语义分析的目标

- 生成一个与上下文无关的语义表示形式
  - 语义网
  - 格框架
  - 逻辑表达式
  - ……
- 上下文相关的分析是篇章分析和语用分析要解决的问题，如指代消解、话语分析等等

# 语义分析要解决的问题

- 词义排歧
- 语义关系排歧
- 更复杂的问题
  - 量词辖域判定
  - 名词定指判定
  - 指代消解

# 语义分析的两种做法

- 直接语义分析：不通过句法分析阶段，直接对源语言进行语义分析
- 基于句法的语义分析：先通过句法分析得到句子的句法结构，在此基础上再进行语义分析

# 直接语义分析：语义语法

- 采用通常的句法分析方法，但在语法规则定义中，直接采用语义标记作为非终结符，而不是用句法标记作为非终结符
- 这样，句法分析完成的时候，得到的句法结构可以直接对应到其语义表达
- 语义语法通常可以用在一些比较狭窄的应用领域中，在比较宽泛的领域很难使用

# 航空定票领域的语义语法

- 非终结符：
  - Loc-From, Loc-To, Time-From, Time-To, Location, Time, Flight, Query-Time, Query-Loc, Query-Flight……
- 规则

  S → Flight 航班 Query-Time 起飞 ？

  Flight → CA Dight | MU Dight

  Query-Time → 什么时候 | 几点
- 句子：

  CA175 航班几点起飞？

# 句法与语义的关系

- 很多句法结构的岐义必须依赖于语义才能解决
  - 英语 PP-Attachment 问题：
    - I bought a table with three legs / dollors.
  - 汉语结构岐义：
    - 他吃鱼的样子 / 他吃海里的鱼
- 而句法结构对语义结构有很强的约束作用，也就是说，给定一个句法结构，其对应的语义结构是固定的、有限的，不能违反。
  - 汉语的 N+V+N 结构：
    - ✓ 实施＋谓词＋受事：鸟吃虫
    - ✓ 工具＋谓词＋受事：枪打出头鸟
    - ✓ 施事＋谓词＋工具：我吃大碗
    - ✓ 施事＋谓词＋处所：我吃食堂
    - X 受事＋谓词＋施事：虫吃鸟

# 基于句法的语义分析

- 完全隔离的句法和语义分析不是好的做法
- 直接语义分析对于大规模应用不适用
- 合理的做法还是将句法分析和语义分析分开，但二者之间又有交互：
  - 句法分析阶段利用语义信息进行排歧
  - 语义分析利用句法分析的结果作为起点

# 选择限制

- 选择限制：词义之间的搭配不是任意的，而是有选择性的。
- 选择限制被广泛用于语义岐义的消解：
  - “ 油”可能是指“燃油”，也可以指“食油”，其具体含义，根据其搭配的词语可以做出选择
    - “ 飞机的油耗尽了。” ——指“燃油” /“ 飞机”的选择
    - “ 油多不坏菜。” ——指“食油” /“ 菜”的选择限制
  - “mouse” 可以指“老鼠”，也可以指“ 鼠标”
    - The mouse stole the egg. ---- 指“老鼠”，“ stole” 的选择
    - Click the mouse. ---- 指“ 鼠标”，“ click” 的选择

# 统计语义消歧 (1)

- 很多情况下，选择限制难以解决问题
  - 很多的选择限制都不是绝对的。比如说，“吃”的客体通常只能是“食物”，但经常有报道说发现了某怪人喜欢吃“泥土”
  - 在选择限制都被满足时难以区分。比如说：“ He saw a girl with a telescope”，这里，“ with a telescope” 既可以修饰“ a girl”，也可以修饰“ saw”，但显然后者更合理。这时用选择限制就很难区分。

# 统计语义消歧 (2)

- 在语义消歧中引入统计方法是很自然的选择，可以解决单纯使用选择限制所面临的问题
- 统计语义消歧中研究得比较多的问题是英语的 PP-Attachment 问题
- 汉语句法分析中也可以采用这种策略，清华大学周强博士在这方面做了很多工作

# 内容提要

# 语用（略）

- 语用（略）

# 复习思考题

- 试比较作为声学模型的 HMM 和作为语言模型的 HMM 的相同和不同之处
- 试分析汉语语音合成中韵律和情感所起的作用
- 研究 FrameNet，写出研究报告，并对构造汉语 FrameNet 可行性及实施方案提出自己的建议
- 用统计语义排歧方法解决一种具体的汉语语义岐义问题（如介词的辖域岐义）