

第一讲：补充材料

自然语言 ——数据之外的视角

杨沐昀

语言技术研究中心

哈尔滨工业大学

提 纲

- 语言：进化中的关键优势？
- 语言：无解的起源
- 语言：语言学视角

语言：进化中的关键优势？

■ 从进化观点来看

- 语言是人类区别于动物的关键之一
- 甚至是关键的进化优势

■ 尼人 vs 智人

■ 尼人（13万到3万年前）

- 发达的大脑，脑量达 1300毫升
- 制造比较精致的工具，会人工取火
- 会埋葬同伴，而且用石器陪葬

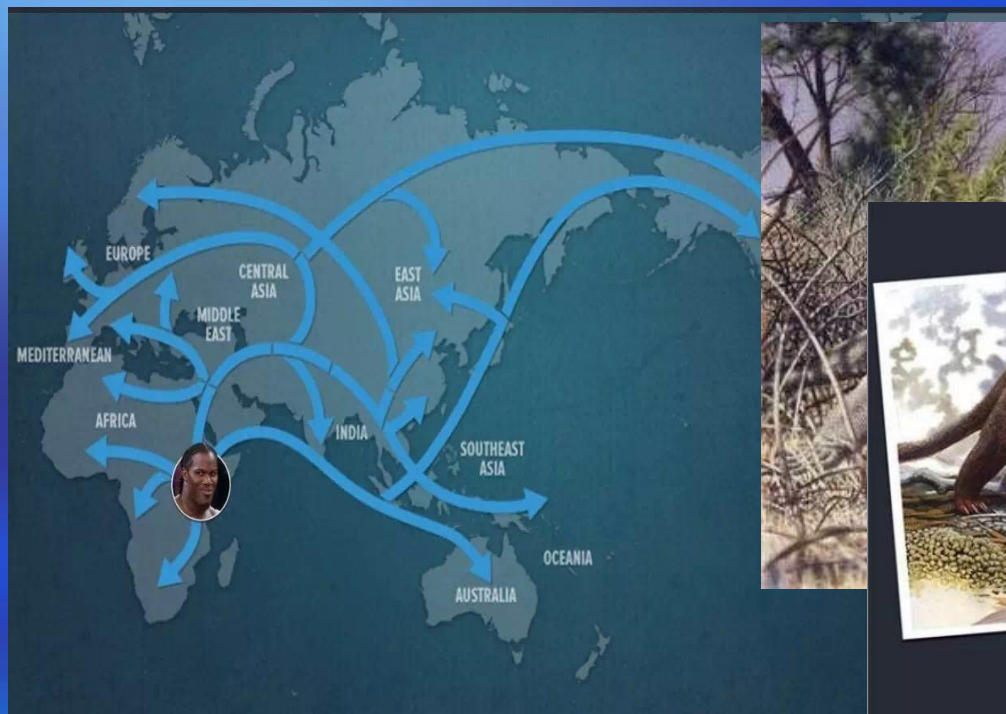
■ 7万年前，智人走上人类舞台，尼人失踪成谜！



根据骨骼重建的尼安德特人

- 尼人声道像猩猩和婴儿一样，发音能力十分有限

语言：进化中的关键优势？

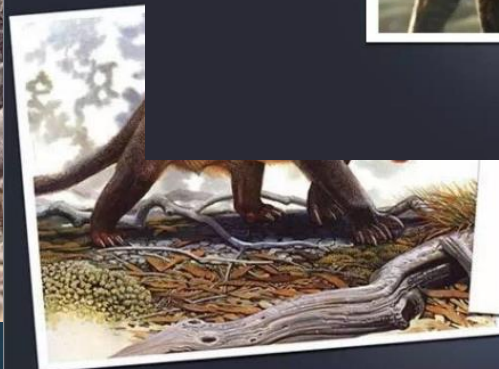


智人征服世界路线图

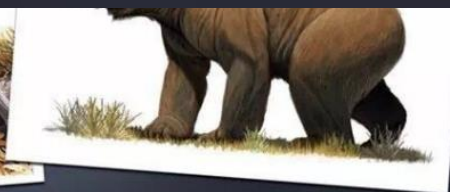


剑齿虎
machairodus

新周刊



袋狮
Marsupial lion



双门齿兽
Diprotodon

新周刊

上述猛兽的尖牙利爪没有任何意义：猛犸象、乳齿象、大地懒、巨型骆驼、拟狮，。

- 语言的协同也许是其中的一个因素，也许传说中咒语的来源

提 纲

- 语言：进化中的关键优势？
- 语言：无解的起源
- 语言：语言学视角

语言：无解的起源

❖ 为何从考虑语言起源

- ❖ 揭示数据天生性质，如分布等等

❖ 1、神授说：认为语言是上帝或神赐予人类

- ❖ 婆罗门教、基督教、神话传说都有类似记载

■ 2、人创说：认为语言是人自己创造的，而非神赐

- ❖ 摹声说：人类对外界声音的摹仿--柏拉图的观点

- ❖ 感叹说：抒发情感的各种叫喊演变来--伊壁鸠鲁的观点

- ❖ 劳动叫喊说：人们劳动时的叫喊声演变来的

- ❖ 社会契约说：彼此约定，规定了事物的名称

语言起源：一个无解的疑问

- 恩格斯提出了劳动创造了语言，语言起源于劳动的观点。
- 普遍认为：人类有声语言的产生大约是在距今四五万年前的旧石器时代晚期，也就是晚期智人时期。
- 回顾有关汉语起源的观点
 - 仓颉造字：黄帝—新石器时代晚期
 - 八卦成文：河图—伏羲—八卦—新石器早期
 - 取火但无文字功绩--燧人氏—旧石器中期
- 目前：学术会议不接受语言起源的论文！

提 纲

- 语言：进化中的关键优势？
- 语言：无解的起源
- 语言：语言学视角

语言：古老的研究对象

- 中国：汉朝时产生了小学---传统语文学
 - 包括文字、音韵和训诂学
 - 文字学：《说文解字》（汉） ... 《康熙字典》（清）
 - 音韵学：研究古汉语各时期声、韵、调系统及发展规律
 - 《切韵》 陆法言（隋）
 - 训诂学：研究古汉语词义；《尔雅》（汉或战国末）
- 印度：4~3 B.C., 出现了梵语语法
 - 形态：将词分成词干和词尾
 - 词类：名词、动词、介词、小品词
 - 语音：区分元音和辅音以及音节等

语言：古老的研究对象

■ 古希腊：关于语言问题的两场论战

- 第一场在“自然发生派”和“约定俗成派”之间展开，焦点集中在语言的形式与意义之间的关系上。
- 第二场在“变则派”和“类推派”之间展开，焦点集中在语法结构是否规则这一问题上。

■ 古罗马

- 当时最著名的语法学家瓦罗的语法巨著《论拉丁语》（《拉丁语研究》），被认为是研究拉丁语的权威性著作
 - 区别了屈折形式和派生形式
 - 详细考察了拉丁语的性、数、格体系和结构功能

语言：古老的研究对象

■ 历史比较语言学：18和19世纪欧洲

- 梳理对比各种语言，探索它们在语音、词汇、语法上的对应关系和异同
- 丹麦的拉斯克（R·Rask）、德国的格里木（J·Grimm）和葆朴（F·Bopp）被称为历史比较语言学的奠基者
- 1786年（英）琼斯证明：古梵语和古希腊语、拉丁语有共同祖先。
- 1799年（匈牙利）沙穆艾尔证明：匈牙利语和芬兰语同出一源。

出现：现代的语言学



■ 现代语言学之父

- 费尔迪南·德·索绪尔
 - Ferdinand de Saussure
 - 1857-1913
- 从1907年始在日内瓦大学讲授“普通语言学”课
- 《普通语言学教程》1916，笔记和讲稿整理出版
- 把语言学塑造成为一门影响巨大的独立学科
- 他认为语言是基于符号及意义的一门科学
 - 现在一般通称为符号学

现代语言学：索绪尔的奠基工作

■ 索绪尔的观点

- 区分了语言和言语这两个不同的概念。
- 区分了语言的“能指”与“所指”
- 区分共时性的研究与历时性的研究
 - 指出语言学重点在共时研究。
 - 语言符号的内在性质而不是演变过程
- 指出语言是一个系统，它有自己的固有秩序。

现代语言学：索绪尔的奠基工作

■ 语言和言语

- **言语：指说话这种行为和说出来的具体的话**
 - a.具有个人特点，丰富多彩（嗓音、用词等）。
 - b.说话所用的词语和规则是全社会共有的、是语言的具体应用
- **语言：是从言语中概括出来的各言语要素的综合，是约定俗成的体系，有统一的语法规则和语音习惯，具有社会性。**
- **两者关系：个别和一般的关系。言语是对语言的具体运用，没有语言也就没有言语；另一方面，语言也不能脱离言语，语言存在与言语之中，而言语是语言的存在形式。**

现代语言学：索绪尔的奠基工作

■ 语言和言语

- 语言是一种工具，是一种规则；言语是人们利用这一工具和规则而产生的结果。
- 语言是言语活动中同一社会群体共同掌握的，约定俗成的，有规律可循而又成系统的部分。
- 言语是言语活动中去除语言部分所剩余的部分，是言语活动中个别的特殊的部分，具有个人特色
- 语言体现在言语中，没有言语就无法体会语言；
- 言语作品是无限的；语言规则是从言语作品中归纳出来的，是有限的。

现代语言学：索绪尔的奠基工作

- 区分语言和言语的意义：
 - 把语言作为语言学的研究对象：从言语中找出规则。
 - 有助于认清语言的客观存在，在语言研究中分清主要和次要；
 - 对语言的规范化也具有重要的意义。

现代语言学：索绪尔的奠基工作

- 研究语言系统的一个性质：符号性

语言的符号性

符号、能指、所指

用甲事物代表乙事物，而甲乙两事物之间没有必然的联系，甲事物就是代表乙事物的符号。甲事物就是符号的能指（形式），乙事物就是符号的所指（内容、意义）。

现代语言学：索绪尔的奠基工作

- **语言符号的性质：语言符号是音义结合的统一体。**
 - 语言是由语音和意义两个方面统一构成
 - 语音是语言的物质外壳，是语言的存在形式；
 - 意义是语言的内容。
 - 语音和意义在具体的语言中统一于一体的
 - 密不可分，二者互为存在条件
 - 语言符号的音义结合是社会约定俗成的。

现代语言学：索绪尔的奠基工作

■ 语言符号的特点

- 任意性：约定俗成
- 线性：书写、口述、理解，都有先后过程
- 稳定性：短期、局部
- 渐变性：长期、全局
- 社会性
 - 后来学者的观点
 - （不是索绪尔关心的，为什么）

现代语言学：索绪尔的奠基工作

■ 语言符号的特点：任意性

- 所指与能指联系具有任意性；
- 很多语言学家认为是语音和语义结合的任意性。
 - Q：为何不说文字和语言语义的联系也是任意的？
- 并不意味着用什么声音表示什么意义，对使用者来说是随意的，不受制约的。
- 事实上，语言符号具有强制性，每个人都要遵守社会的约定俗成，不能任意的改变，否则就无法交际。

现代语言学：索绪尔的奠基工作

■ 语言符号的特点：线性

- 体现在具体的言语（交际过程）中。
- 不管你要表述的思想感情是简单的，还是复杂的，语言符号都只能一个一个地按次序出现
- 使离散的语言符号得以构成大大小小的音义结合体，可以使单个的语言符号组合成连续的语流。
- 用“棋、下、有、一、完、没、的、盘”这八个词组成句子，看能组合成多少句子？
- 答：一盘没有下完的棋、没有下完的一盘棋、没有一盘下完的棋、没下完的棋有一盘、下完的棋一盘没有、下完的棋没有一盘、有一盘没下完的棋、有没下完的一盘棋、棋没下完的有一盘、棋有一盘没下完.....

语言的社会性：交际工具

1. 语言是社会现象

- 语言只存在于社会之中，离开了社会就没有语言
- 社会依赖语言而得到维系。
- 语言为人类所独有，既有传承性又有习得性。
- 语言是社会的产物，既要以稳定的形式为社会服务，又要适合社会不断发展的需要。

2. 语言是一种特殊的社会现象。

- 语言是全民的交际工具。
- 语言既不属于经济基础，也不属于上层建筑。

语言的社会性：交际工具

- 朱元璋做了皇帝，他从前相交的一班苦朋友照旧过着很穷的日子。有一天，他从前的一個苦朋友跑到南京求见，准见之后便说：

“我主万岁！当年微臣随驾扫荡芦州府，打破罐州城，汤元帅在逃，拿住豆将军，红孩儿当关，多亏菜将军。”

朱元璋一听，隐约觉得他的话中包含了一些从前的往事，见他说得好听，心里很高兴，所以立刻封他做了御林军的总管。这个消息让另外一个苦朋友听见了，他心想：“同是那时候一起玩的人，他去了有官做，我去当然也不会倒霉吧？”

“我主万岁！还记得吗？从前，你我都替人家看牛，有一天，我们在芦花荡里，把偷来的豆子放在瓦罐里煮着，还没等煮熟，大家就抢着吃，把罐子都打破了，撒下一地的豆子，汤都泼在泥地里。你只顾从地上满把地抓豆子吃，却不小心连红草叶子也送进嘴去。叶子梗在喉咙口，害得你哭笑不得。还是我出的主意，叫你用青菜叶子放在手上一拍吞下去，才把红草叶子带下肚子里去了。……”

朱元璋等不得听完就连声大叫：“推出去斩了！推出去斩了！”

现代语言学：索绪尔的奠基工作

❖ 语言系统的2种基本关系：组合聚合

- 1.组合关系

- 若干较小的语言单位组合成较大的语言单位，其构成成分之间的关系就是组合关系，又称线性序列关系。

老师	分析	课文
小王	讲	故事
李明	写	文章
我	看	电视
他	学	英语



现代语言学：索绪尔的奠基工作

❖ 语言系统：2种基本关系

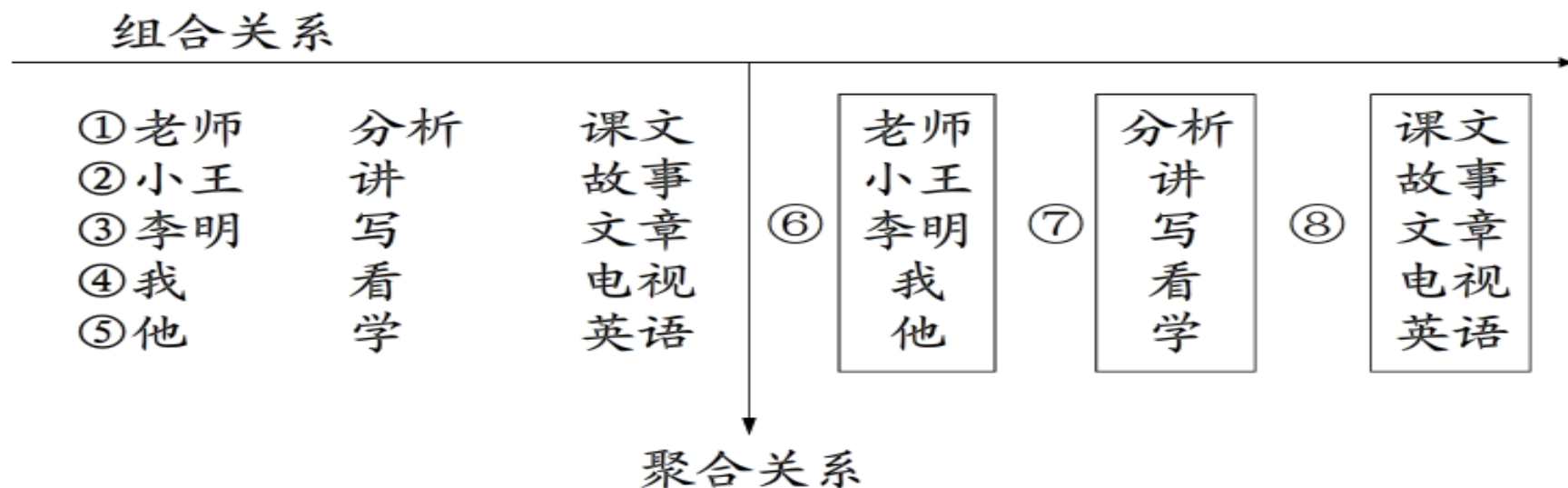
2. 聚合关系

❖ 具有相同组合功能的语言单位之间的关系，就是聚合关系，又称联想关系。聚合关系专指那些具有替换关系的语言单位之间的关系。

	老师	分析	课文
↓	小王	讲	故事
	李明	写	文章
	我	看	电视
	他	学	英语

现代语言学：索绪尔的奠基工作

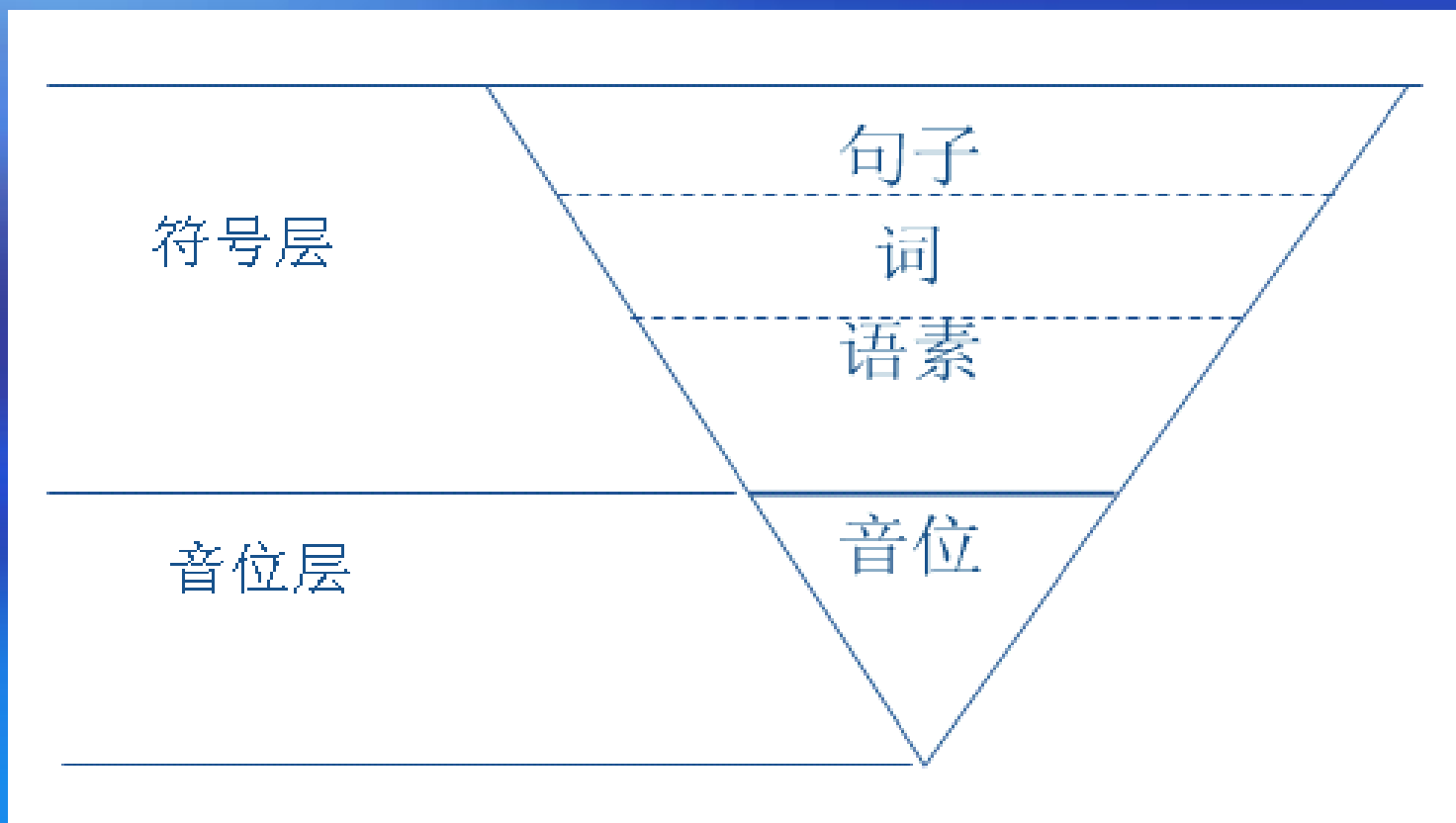
❖ 语言系统：2种基本关系



❖ 语言的组合关系和聚合关系是语言的两种根本关系，是语言系统的纲，把握了这个纲，就基本上把握了语言系统。

现代语言学：索绪尔的奠基工作

❖ 语言系统：层级体系—组合关系带来的结果



现代语言学分支

❖ Micro-linguistics, or Intra-linguistics



语音学

■ 语音的物理性质

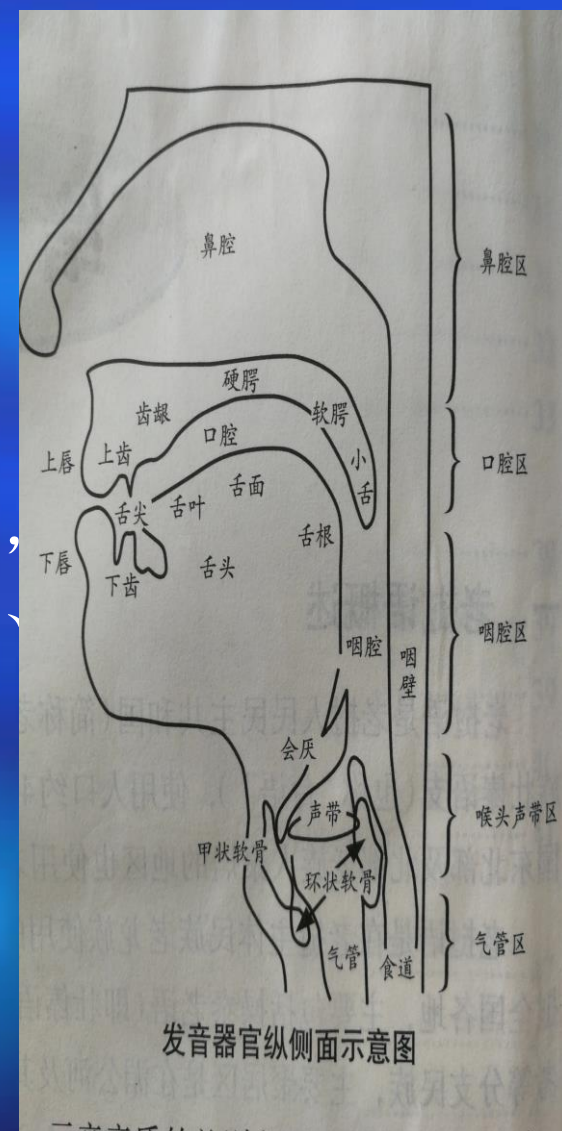
- 音高、音强、音长、音色叫做声音的四要素，也是语音的四要素。

声音要素	含义	决定因素	发音中的影响因素	能够区别意义
音高	声音的高低	频率	声带的长短、厚薄、松紧	声调：妈、麻、马、骂 语调：他不来。他不来？
音强	声音的强弱	振幅	用力大小、呼出气流量大小	轻重音：content(重音在前，名词“内容”，居中为形容词“满意的”)
音长	声音的长短	时间		pull (“拖”短音) 与 pool (“水池”长音)
音色	声音的特色	声波形式	发音体、发音方法、共鸣器	f—v g—h

语音学

■ 语音的生理基础

- 人的发音器官是一条由肺部一直延伸到唇、鼻的形状复杂的“管道”。它由三部分组成：发音的动力源、发音体、共鸣腔。
- 主动的发音器官：在人类的发音器官中，有些是能够活动的，如唇、舌头、软腭、小舌、下腭和声带等
- 被动的发音器官：有些不能活动，如上齿、齿龈、硬腭等



语音学

■ 音标

- 目前国际上最通行的是“国际音标”
- 国际语音协会与1888年指定并开始使用的，后经多次修改，最近一次在1989年。
- 大多采用拉丁字母及其大小写、手写体、变形等代表不同的音素，这点跟字母是完全不同的。
- 国际音标中共有20个元音音素和28个辅音音素。

音系学

- **Phonology** studies the rules governing the structure, distribution, and sequencing of speech sounds and the shape of syllables. It deals with the sound system of a language by treating phoneme as the point of departure.
- 语音如何区别和表达意义

音系学

■ 音素 (phone)

- 根据音质不同对一串语音不断加以切分，直到不能再分，这样得到的语音单位就是“音素”。
- 比如汉语“lekailehua（乐开了花）”，往下切分可以得到：l、e、k、a、i、l、e、h、u、a共10个单位，从音质角度看，它们已经是最小的单位了，每一个单位都只体现为一个发音定位。所以，“音素”可以看作是从音质角度划分出来的最小的语音单位。
- 音素可以分为元音和辅音。

音系学

- 音位（phoneme）
 - 区别意义的一组语言特征集合
 - 音系学的基本单位
 - 因素强调语音的差异，音位强调意义的差别
- 音位变体（allophone）
 - 同一音位在不同环境下的实现方式

当g之后以鼻音结尾时，g不发音。
例如： **sign**, **design**, **paradigm**

语法、语义和语用

- 词法：词内部结构，包括语素和构词法
 - 英语的词尾曲折变化
- 句法：句子结构
- 语义：不同语言单元中的意义及其构成
- 语用：在具体场景下语言的使用
 - 固定的语义在不同情况下引起的主观感受

语言学分支

❖ Macro-linguistics, or Inter-linguistics

- ❧ 社会语言学

- ❧ 心理语言学

- ❧ 神经语言学

❖ 应用语言学：另一种考虑维度

- ❧ 计算语言学

- ❧ 语言习得

- ❧ 语言教学与语言测试

❖ 其他：历史语言学、对比语言学....

结语：人们无法不研究语言

信息社会生活严重依赖语言技术

- 从仓颉造字、活字印刷到互联网，人类被语言包围：
 - 对话、广播、电视、书籍、报刊、杂志、邮件、网页、微博
- 互联网：Google索引1,000,000,000,000个网页[2008]
- 一秒钟会发生：
 - 4万多次谷歌搜索
 - 6000条推特、10万条微博被发送、转发或评论
 - 1,00万条朋友圈动态更新，被评论or点赞
 - 超过200万封电子邮件被发送
- 全世界每年科技论文50万篇，中国占1/4以上（2017-）
 - arXiv 论文2014年底超过一百万篇
- 2022年我国专利申请总量261万件