

“自然语言处理导论”课（语言知识部分）要点

詹卫东 北京大学中文系 2018-11

第1讲 第一部分 语言与语言学概述

第二部分 汉语的构词法与中文文本自动分词中的问题

1. 如何认识我们人类的自然语言？

- (1) 语言的功用：**交际工具**、**思维工具**（语言是知识的载体）
- (2) 语言的外在形式：有声语言（**听觉系统**）、书面语言（**视觉系统**）、体态语/**手势语**（视觉系统）
- (3) 语言的内在性质：**符号性**（任意、理据）、**结构性**（离散、线性、层次、组合、聚合）→ **稳定与变异**

2. 语言学：关于语言的知识系统

观念变迁：(1) 看法律的语言学 → (2) 看生物学的语言学 → (3) 看化学的语言学 → (4) 看数学的语言学
核心内容：(1) 句法学 (2) 语义学 (3) 语用学
两个任务：(1) 观察语言现象（知其然） (2) 解释语言现象（知其所以然）

3. 汉语构词法：

基本单位：**语素**（词根、词缀、词尾）

构词模式（17种）：单纯词（7）、合成词（复合6、重叠2、附加2）

造词形式：仿词、缩略、**词语变形/语言游戏**（新词涌现：各类网络新词，合音、译音、谐音、拼音、缩略……）

词的内部层次（语素组）

4. 中文文本自动分词中的问题

词的定义（语法学定义、词典定义 type、语料库定义 token）
分词歧义（交集型歧义，组合型歧义），交集型歧义链长
未登录词（专名、领域词/术语、新词、非汉语词/字母词）

第2讲 汉语的句法结构系统与词类划分

1. 词类划分的目的：词类（**聚合关系**）→ 构建/表述语法系统（**组合关系**）

2. 词类划分的依据：意义、形态、句法功能 ✓ 如何确定句法功能的框架？

3. 现代汉语词类划分的具体操作：

汉语的基本句法结构系统：主谓、述宾、述补、状中、定中、连谓、联合、的、字、地、所、介宾、方位、数量
词类的层级划分与各类的鉴别标准

现代汉语词类系统：**动词、形容词、状态词、名词、方位词、处所词、时间词、量词、数词、数量词、指示词、区别词、副词、**
(20个词类) **代词、介词、连词、语气词、助词、拟声词、叹词**

4. 词类划分中的若干问题

- (1) 词的**同一性**：词形、词音、词义
- (2) 词的**多功能性**
- (3) 词的**临时功能**（活用）
- (4) 词的**兼类**
- (5) 词类的**相对性**

- ⊗ 对内一致性（内部匀质，无特例）
- ⊗ 对外排他性（没有异己分子，无交集）
- ⊗ 系统完备性（子集之合=全集，无遗漏）

* 5. 现代汉语语法信息词典 **词类** + **属性特征描述**（面向人的词类划分体系 vs. 面向计算机的词性标记集+特征描述）

第3讲 句法分析

1. 句法分析任务：句法结构知识（语法模型）+ 分析算法（在语法模型中寻找正确的“树”）
2. 句法结构的形式表示模型：上下文无关文法，依存文法
3. 句法结构歧义：外显型歧义-内含型歧义 真歧义-准歧义-伪歧义
4. 句法分析算法：CYK 算法、Earley 算法、LR 算法、GLR 算法……
5. 现代汉语句法结构系统：常规结构（每类结构的特点，包括成分间关系、形式特点）

主谓、述宾、述补、状中、定中（同位）、连谓、联合、的字、地字、所字、介宾、方位、数量

短语（词组）结构分析的原则：成分独立；功能约束；意义一致；系统普适
非常规结构：易位、插入、省略

第4讲 语义分析

1. 语义分析任务：给出句子的语义表示：谓词逻辑表达式、语义网络、概念依存图、框式图
组性原则（principle of compositionality）：词义+结构义
2. 词汇语义：语义特征集、语义分类树、语义关系网
基于词典释义的语义特征分析、基于句法格式等义变换的语义特征分析
配价/论元结构理论、框架语义、生成词库论
3. 语义知识与句法分析的融合
特征结构、合一运算（Unification）、加入了合一运算的 CYK 算法

HowNet、WordNet、FrameNet、MindNet
Propbank、VerbNet、
Abstract Meaning Representation ...

第5讲 语篇分析

1. 篇章的连贯性 衔接手段：代词回指、零形回指、词语关联、结构关联、语义关联
2. 篇章结构分析 篇章层级结构与篇章单元之间的关系(Rhetorical Structure Theory, RST:23 种)
Penn Chinese Discourse Treebank (11 种)
中文复句内部小句间关系 (14 种)
3. 篇章指代分析 指代语言成分（人称代词、反身代词、零形成分）
指代实物（直指）
面称/背称 通指/专指
语段中心成分理论（Centering Theory）
forward-looking center / backward-looking center / preferred center
Centering transition state: continue / retain / smooth shift / rough shift

附：自然语言理解中的语用因素

- 指示语理论（deixis）
- 会话含义理论（conversational implicature: Grice's theory of implicature）
- 言语行为理论（speech act）
- 话语结构的语用分析（conversational structure）

数量准则、
质量准则、
相关准则、
方式准则