语言与理解

周国栋
苏州大学自然语言处理实验室
http://nlp.suda.edu.cn/

认知科学的关键是解决语言机制问题。

– Hans-Georg Gadamer (1900–2002)
Truth and Method (1960)

语言 理解 认知 智能 智慧 …

自然语言理解
是人工智能
皇冠上的明珠





自然语言理解是实现
人工智能的拦路虎



计算智能



感知智能



认知智能

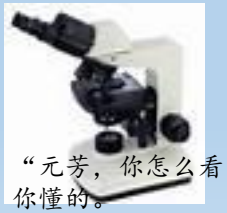
语言是一个哲学问题！



语言：形式主义vs功能主义

形式主义：Aristotle	功能主义：Protagoras、Plato
语言学是哲学的一部分	语言学是人类学的一部分
语法是逻辑学的一部分	语言是文化的一部分
语言是表示肯定与否定的手段	语言是向人谈论事情的手段
语言是一种判断方式	语言是一种活动方式
语言学是规范的	语言学是描写的
语言是规则系统	语言是选择系统
注意规则现象	注意不规则现象
关心语义与真值的关系	关心语义与修辞功能的关系
对句子作形式分析	对话语作语义解释
把合乎语法性作为理想化标准	把可接受型或用途作为理想化标准

形式



“元芳，你怎么看？”
你懂的。
if

意义



“元芳，你怎么看？”
你懂的。
if

功能

什么是自然语言？

语言认知

认知科学的关键是解决语言机制问题。

— Hans-Georg Gadamer
Truth and Method (1960)

- Sapior：人类赖以思维的工具，思维的物质外壳。
- 认知语言学：现实-认知-语言

认知语言学：认知语义学（G. Lakoff）认知语法（R.W. Langacker）
--第三次语言学革命（对Chomsky革命的革命）

语言本体

- 声音、形式和意义相结合的符号系统。
- 变动的结构（整体、成分、关系），意义存在于关系之中。

自然语言是人类籍以思维和交际的一个符号系统，是一个变动的音形义相结合的结构系统，是人类对社会现实的认知反映。

语言功用

- 人类最重要的交际工具和信息载体。
- 社会学、心理学、转播学

功能主义语言学：结构功能学（ V. Mathesius ）
系统功能语法（M.A.K. Halliday）
--第二次语言学革命+++（重点：篇章级）

形式主义语言学：结构主义语言学（F. de Saussure）--第一次语言学革命（Saussure的哥白尼革命）（重点：句子级） 转换生成语法（Noam Chomsky）--第二次语言学革命（Chomsky的心智革命）

语言是复杂系统：系统之系统

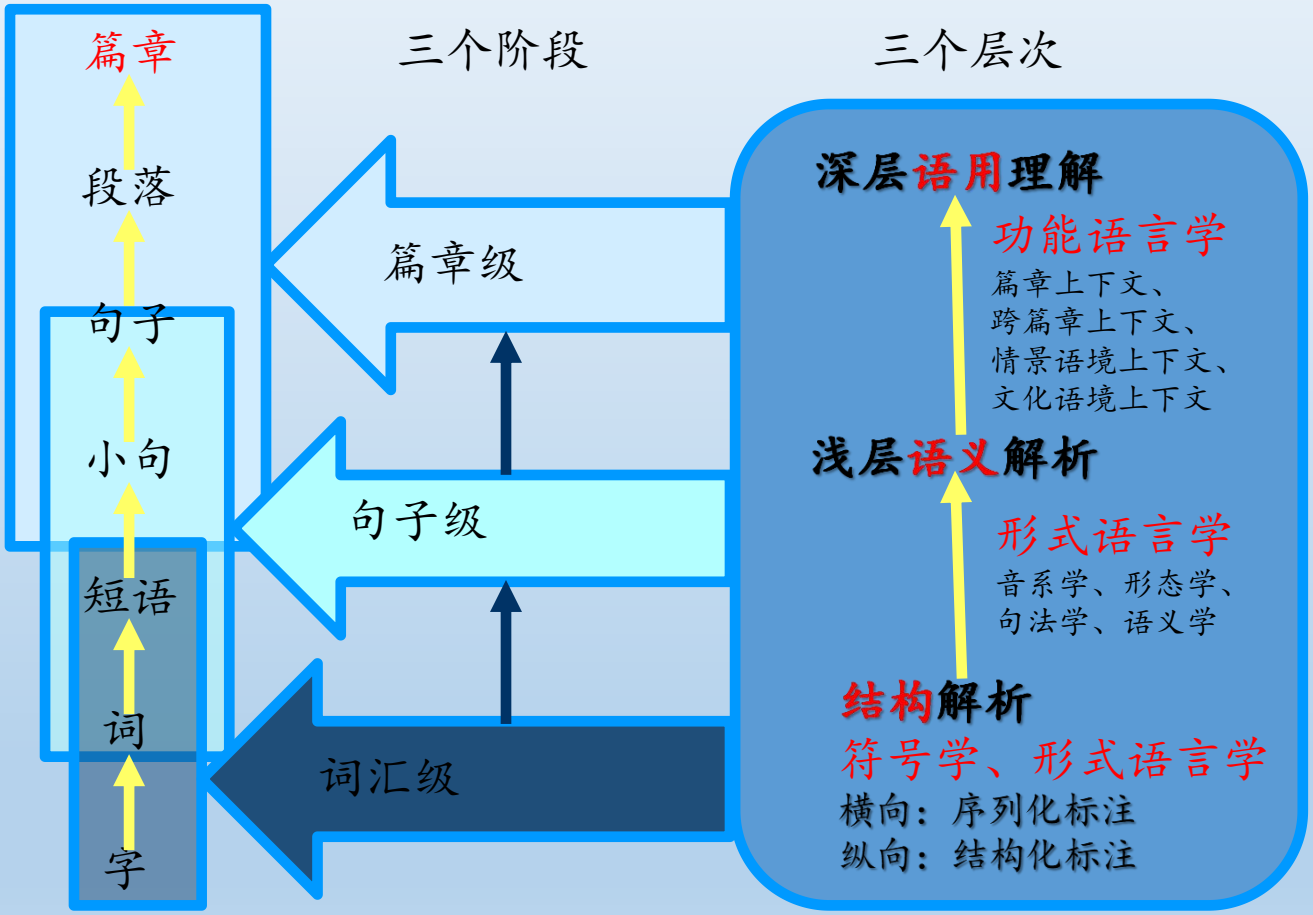


语言学 vs 自然语言理解



- 研究对象：自然语言
- 研究重点：观察、描述、解释=》可计算性

什么是自然语言理解？



自然语言认知（认知语言学）

语言理解==篇章理解

人们理解自然语言通常是在篇章级进行
“不能断章取义”

功能主义

形式主义

可接受性

意图性

信息性

连贯性
（纵聚合）

衔接性
（横组合）

语境性（文化）
（文化图谱：知识图谱++）

语境性（情景）
（情景图谱：知识图谱+）

跨篇章性
（知识图谱：认知图谱+）

语言认知
（认知图谱）

篇章理解vs连贯性/衔接性

篇章：一系列连续的语段或句子构成的语言整体单位

——Beaugrande R.D. and Dressler W. Introduction to Text Linguistics（1981）

自动问答

a 李四比较年轻，
b 既没有张三的工作经验丰富，
c 又没有王五的学历高，
d 但是不论做啥事情，
e 他都比那两人认真负责，
f 所以，领导非常器重他。

太难了！



人问1：领导非常器重谁啊？
人答：李四。
机器答1：王五。
人问2：为啥呢？
人答2：不论做啥事情，李四都比张三王五认真负责。
机器答2：他都比那两人认真负责。

- 问答1涉及上下文的篇章话题结构信息（衔接性）
- 问答2涉及上下文的篇章修辞结构（连贯性）和话题结构（衔接性），两者配合才能生成需要的答案

2017年国家自然科学基金人工智能应急重点项目

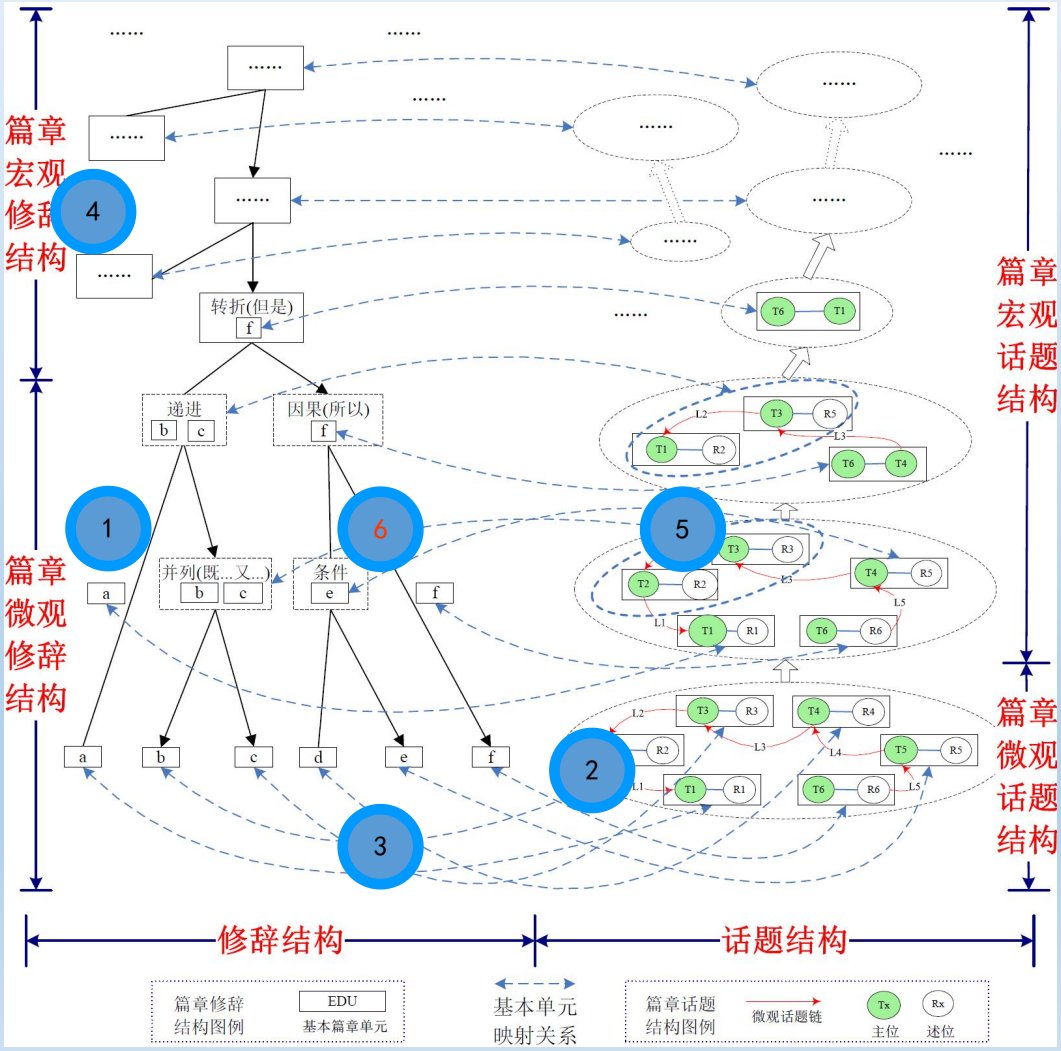
话题驱动的汉语篇章机器阅读理解（NSFC 61751206）

周国栋 李培峰 陈文亮 孔芳

苏州大学 自然语言处理实验室

http://nlp.suda.edu.cn/

微观与宏观融合的
篇章修辞结构和话题结构一体化表示体系



① 篇章微观修辞结构（层次化）
② 篇章微观话题结构（序列化）
③ 微观层面对应关系
④ 篇章宏观修辞结构（层次化）
⑤ 篇章宏观话题结构（层次化+序列化）
⑥ 宏观层面对应关系

篇章理解vs跨篇章性/语境性

篇章存在信息缺省，需要词语所激活的相关概念和信息

跨篇章性
（知识图谱）

（上例）领导器重李四的主要原因：李四是领导的小舅子！

语境性（情景）
（情景图谱：知识图谱+）

She didn’t want to talk about the operation but she couldn’t think about anything else. Maybe it would turned out to be benign; on the other hand, maybe they would open her up and find that she was permutated, riddled, rotting away from inside.
（operation可以激活一个ECM，包括主刀医生、护士、病人、病灶、手术方式、手术部位…）

语境性（文化）
（文化图谱：知识图谱++）

你什么时候过门啊？（转喻）

Men are verbs, but not nouns. 男人应当干活，而不是让人看的物品。

My wife always says ‘you shall’, and I always say ‘I will’. 老婆总是下命令，丈夫总是要服从。

2018年国家自然科学基金重点项目

面向领域大数据的事件知识图谱构建（NSFC 61836007 ）

朱巧明 郑凯 李培峰

苏州大学自然语言处理实验室

电子科技大学

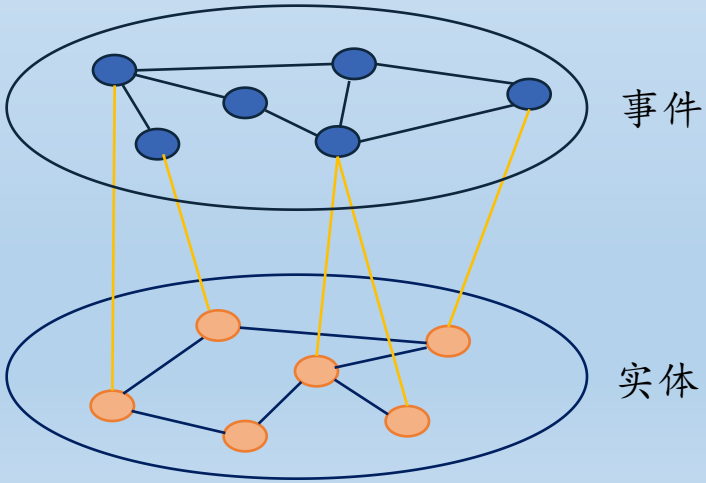
面向领域大数据的事件知识图谱构建

面向特定领域

- 国防军工
  军委科技委，140万/年
- 电子商务
  阿里巴巴，200万/年
- 金融服务
  科沃斯（服务机器人），300万
- 舆情信息处理
  NSFC重点项目，61331011
- ……

面向事件

事件知识图谱



实体知识图谱

面向非结构化大数据

本课题　　传统

非结构化（文本）

半结构化（维基百科）

结构化（数据库）

知识图谱

实例（以“金正男遇刺”为样例的事件知识图谱）

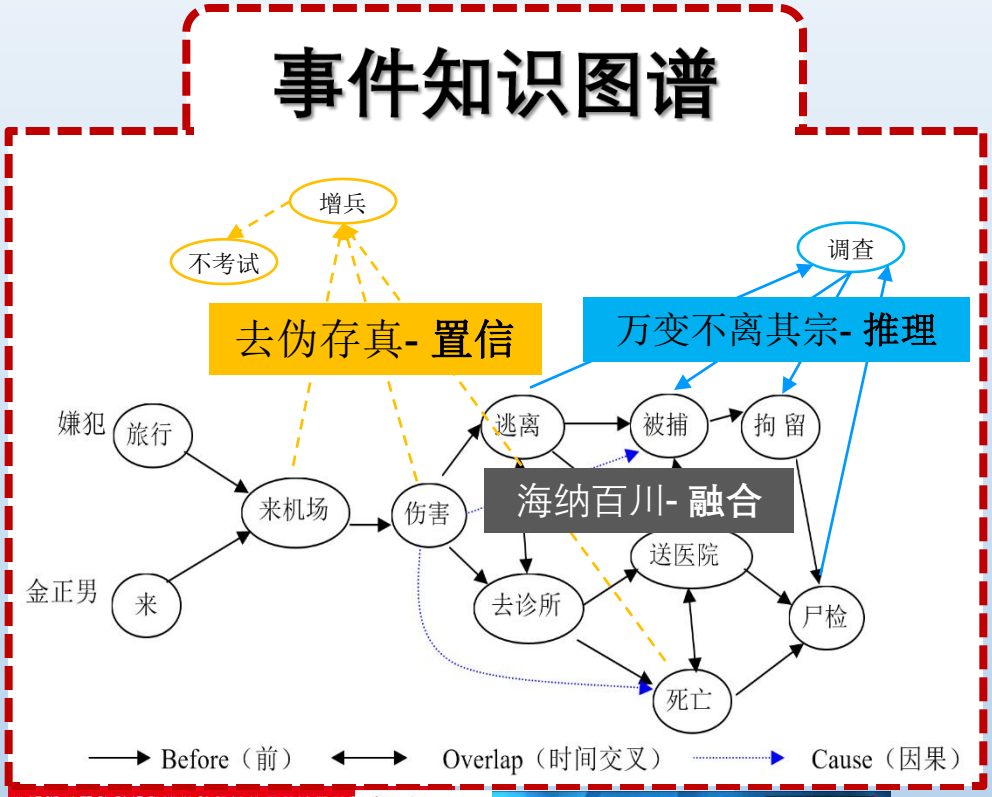


答：我们也注意到媒体有关报道。据我们了解，有关事件发生在马来西亚，目前马方正在对事件进行调查。

问：关于金正男在马来西亚遇害，中方对此有何评论？









在逃嫌疑人在吉隆坡国际机场实施了对金正男的蓄意谋杀

金正男向机场柜台职员求助，后被送入布城医院求治，但途中死亡

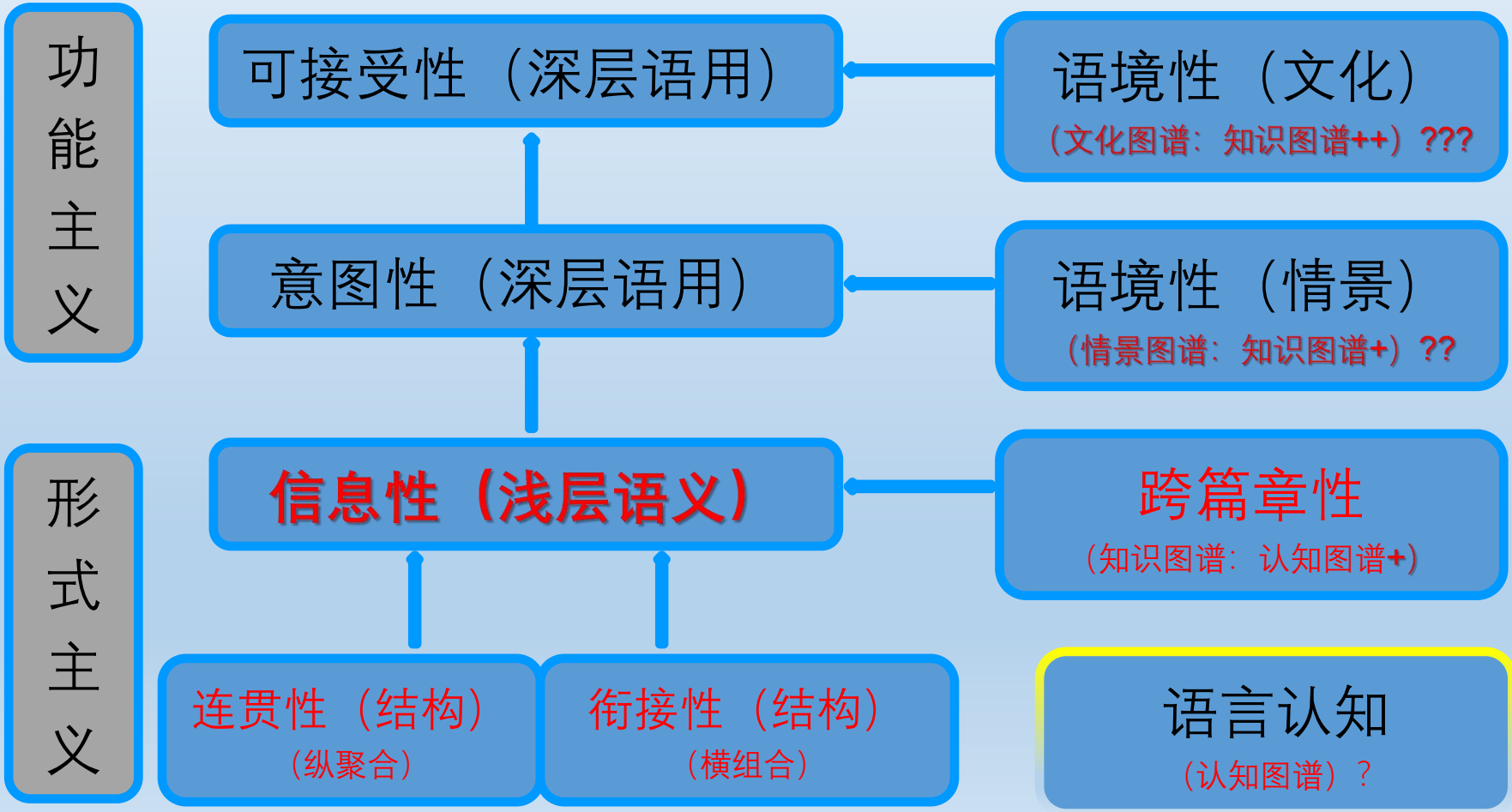
警方周六（18日）逮捕涉及金正男遇刺案的朝鲜男子



语言理解==篇章理解

人们理解自然语言通常是在篇章级进行

“不能断章取义”







谢谢大家！