

# 以概念联想的方式完成语句理解

张全

62559371 [zhq@mail.ioa.ac.cn](mailto:zhq@mail.ioa.ac.cn)

中国科学院声学研究所

# 科学问题

- 当今语言技术最严重的缺陷是没有有效方法去获取文本和言语中的真实内容，缺少基本上达到语言理解的系统。因此，计算语言学核心的、也是最宏伟的任务是实现深层语言理解，以便在书面语、口语和有效的语义表达之间建立精确的映射。

— — Hans Uszkoreit (Verbmobil项目)

# 语句的结构意义

- 已经有多种句法分析的策略在计算机中得到实现，已经能够完成语句的句法分析处理。
- 语义角色的标注工作也已经展开，并取得了一定的进展。
- 应用到汉语中还有不尽如人意的地方。
  - 汉语缺乏形态标识，语句的中心谓语动词比较难于确定；
  - 汉语的词汇兼类现象严重，在句子中比较难于根据它的词性属性确定它句法角色，进而语义角色的标注也比较困难。
  - 有些学者甚至用意合来描述汉语的特征。

- 
- 有没有进行句法—语义分析的必要？
  - 如何进行汉语的句法—语义分析呢？
-

- 
- 乔姆斯基和山克进行过大辩论。
  - 要将句法分析和语义分析结合起来。
  - 如何结合呢？
-

# 语言感知处理的模式

- 形成了概念基元体系和句类体系，或者称语义基元空间和句类空间。
- 通过自然语言符号体系映射成为**HNC**概念关联符号体系的方式将自然语言中的语义内容显式的表示出来，完成自然语言语句的理解，为各种语言信息处理提供基础。
- 目前**HNC**已经形成了面向语句理解处理的句类分析处理技术，这一技术基于知识和规则，已经应用于实际的信息处理系统中。

# 概念基元体系——基本载体

- 概念基元符号表示式 (HNC1)

**CP=y|(m//ekm//cmn//dmn//-0|)|(\k//t//i)|**

- 召开 vc39e21a+va01

会议 gc39e21a

回忆 vg6802

会意 v810; (jgwa30\2,jgwa30\4)

## 39 聚散合分（关系效应之三）

39:(e2m,e6m)

39e2m 聚与散

39e21 聚

39e22 散

39e6m 合分独联

39e60 联

39e61 合

39e62 分

39e63 独

<b>39e218</b>	使聚
<b>39e219</b>	集成
<b>39e21a</b>	集中

# 例子

**Zhao kai hui yi**

召开会议

**Zhao kai xue shu hui yi**

召开学术会议

**Hui-yi**

会议 **Meeting**

回忆 **Remember, remembrance**

会意 **Understand, understanding**

- 召开 **vc39e21a+va01**
- 会议 **gc39e21a**
- 回忆 **vg6802**
- 会意 **v810; (jgwa30\2,jgwa30\4)**

- HNC
  - + 语习逻辑概念
  - + 基本概念
  - + 基本逻辑概念
  - + 宇宙
  - + 语法逻辑概念
  - + 人
  - + 人化的基本概念
  - 第二类扩展基元概念
    - + 第一类劳动
    - 表层第二类精神生活(交往与娱乐)
      - + 表层第二类精神生活的基  
本内涵
      - + 交往
      - + 娱乐
      - + 比赛
      - + 行旅
    - + 深层第二类精神生活(想象与信念)
  - + 综合逻辑概念
  - + 物
  - + 物化的基本概念
  - + 属性概念
  - + 基元概念

HNC/第二类扩展基元概念/表层第二类精神生活(交往与娱乐)/比赛 (q73)

生擒  
 田径运动  
 平淡无奇  
 平淡  
 回味无穷  
 观众  
 观看  
 滚翻  
 好看  
 海报  
 滑雪板  
 角斗  
 竞赛  
 戴维斯杯  
 打雪仗  
 打赌  
 蝉联  
 擂台  
 联盟杯  
 对垒  
 观战  
 比赛  
 比武  
 比拼  
 奥林匹克运动会  
 平淡无味  
 炙手可热  
 征程  
 运动会

# 句类体系——联想的框架

- 句类表示式 (HNC2)

$$SCR=SC+fK_m$$

$$SC=GBK_1+\{EK+GBK_m\}(m=2-4)$$

语句: 从大学毕业到现在的六个月间 ~|| , 我 || [都在] 寻找 || 工作。 ↵

句类 ↵

表达式:  $SC = C_{n-1} + TA + T_{19} + TBC$  ↵

概念: ↵

$lq15$  ↵  $152$  ↵  $j110$  ↵  $l41$  ↵  $lh15$  ↵  $l61e51$  ↵  $pea74$  ↵  $j3080c06/wj10-0$  ↵  $p4001$  ↵  $ra00e21$  ↵  
( $v11eb2, gva72$ ) ↵  $v65219$  ↵

# 句类体系——联想的框架

- 打破
- HNC符号：  
(v900#v309a,l10,(gb303,l10,ga00a+gc249a;))
- 句类代码： [1] R3Y0\*21J

[2] XY0\*21J

**R3B + R3Y0 + YC** 谈判双方 || 未能打破 || 僵局。

**A + XY0 + YC** 他 || 打破了 || 谈判的僵局。

- 
- 经过HNC的句类分析实际上得到了句子的句法结构信息和语义角色的信息，还完成了部分词语多义模糊的消解。
-

# HNC的思考

- 语句的理解显然是自然语言理解的基础，但这不等于说，任何一种形式的语句理解处理算法都可以成为自然语言理解的基础。要取得这一资格，就必须把语句理解定位于概念联想脉络运作全过程的激活。
- 自然语言理解的本质是概念联想脉络激活、扩展、浓缩、转换与存储的全过程运作。激活运作的要点是语句的理解；扩展与浓缩运作的要点是段落与篇章的理解；转换与存储的要点是记忆与学习。

——黄曾吻