

自然语言理解

(10)

宗成庆

中科院自动化研究所
模式识别国家重点实验室

cqzong@nlpr.ia.ac.cn

<http://www.nlpr.ia.ac.cn/English/cip/cqzong.htm>

No.95, Zhongguancun East Road
Beijing 100080, China



<http://www.ia.ac.cn>
Tel. No.: +86-10-6255 4263

第十章 语义计算

10.1 概述

- 任务：解释自然语言句子各个部分（词、词组及句子）的意义。
- 困难：1) 自然语言句子中大量存在的歧义，涉及指代、同义、量词的辖域，隐涵等。
 - 2) 同一句子不同人有不同的理解。
 - 3) 语义计算的理论、方法很不成熟。



10.1 概述

◆ 例子

例如：1) I bought a car with four wheels.

I bought a car with four dollars.

2) These boys will be dedicated persons.

These boys will be denied license.

3) 我感到很头痛。

4) 他已经先走一步了。

10.1 概述

一个幽默片断：

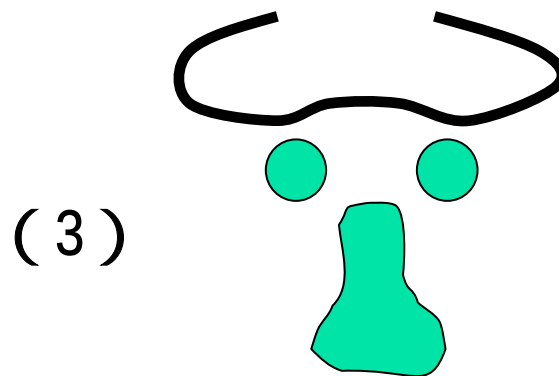
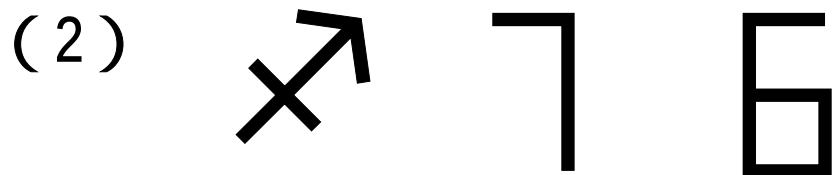
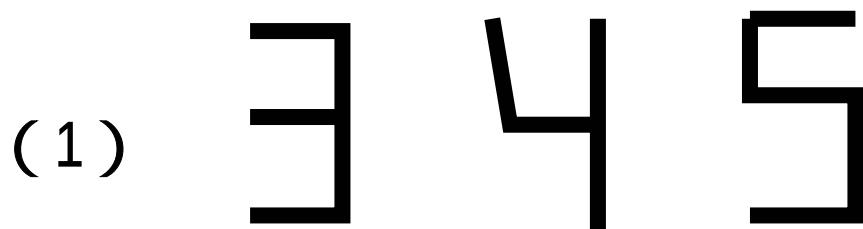
他说：“她这个人真有意思(**funny**)”。她说：“他这个人怪有意思的(funny)”。于是人们以为他们有了意思(**wish**)，并让他向她意思意思(**express**)。他火了：“我根本没有那个意思(**thought**)”！她也生气了：“你们这么说是什么意思(**intention**)”？事后有人说：“真有意思(funny)”。也有人说：“真没意思(**nonsense**)”。

- 《生活报》1994. 11. 13. 第六版



10.1 概述

◆ 上下文的引导与误导



10.2 语义理论简介

□ 词的指称作为意义

该理论认为，词或词组的意义就是它们在现实世界上所指的事物。那么计算语义学的任务就是将词或词组与世界模型中的物体对应起来。

常用的现实世界模型假设世界上存在各种物体，包括人。

缺陷：对于复杂的问题这种定义无法处理。

启明星/暮星 → 金星

神仙？ 鬼？ 妖精？

10.2 语义理论简介

□ 心理图像、大脑图像或思想作为意义

该理论认为，词或词组的意义就是词或词组在人心理上或大脑中所产生的图像。

缺陷：在计算机中把心理图像有效地表示出来并不是一件容易的事情，而且，不一定所有的词义都有清晰的心理图像。

10.2 语义理论简介

□ 说话者的意图作为意义

该理论试图解释语言中一种被成为言语行为（Speech Acts）的现象。

说话者把自己的话语当作行为希望听者理解、作出反应。这种意义被认为是独立于逻辑意义之外的。

10.2 语义理论简介

例如：我想预订明天下午的火车票

c: give-information + reservation + ticket(ticket-spec=(ticket-type=train,time= (relative-time=tomorrow,tod=afternoon)))

说话人
(Speaker)

语句意图
(Speech Act)

概念
(Concept)

具体参数
(Arguments)

10.2 语义理论简介

□ 过程语义

该理论认为，句子的语义定义为接受该句后所执行的程序或者所采取的某种动作。

简单明了，对于计算机智能应用系统来说，这种定义在某种程度上是有效的。

缺陷：对于语言本身缺乏解释，且句子的语义常常和应用连接紧密，缺乏独立性。

10.2 语义理论简介

□ 词汇分解学派

该理论把句子的语义基于它所含有的词和词组的意义之上，而词的意义则基于一组有限特征，这组特征通常称为语义基元。这样，只要给出一组语义基元和一些操作符，就可以把句子的语义描述出来。

类似于化学中的元素学说。

缺陷：语义基元的定义、分解标准等不好把握，基元和组合操作的合理性直接影响句子语义描写的准确性。

10.2 语义理论简介

□ 条件真理模型

该理论以谓词逻辑为基础，句子的语义定义为它所对应的命题或谓词在全体模型（或世界）中的真伪。

例如：“雪是白的”为真当且仅当在这个世界上雪是白的。

优点：对于上下文无关部分的语义描写很有效。

缺陷：对时间、场景有关的语言现象不能很好地描述。不能很好地解释一句多义的问题。

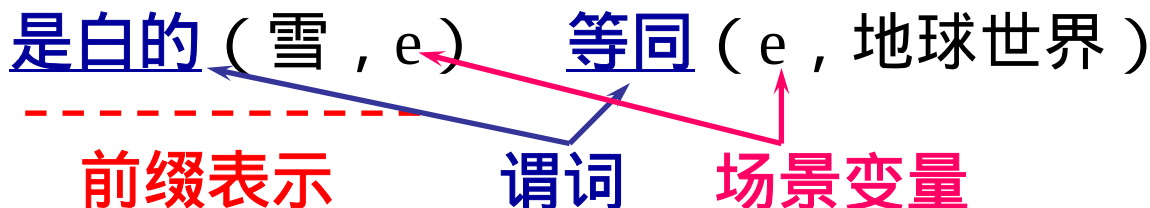
10.2 语义理论简介

□ 情景语义学

该理论认为句子的语义不仅和逻辑意义有关，而且和句子被使用的场景有关。

在语义表达式中引入一些与场景相关的变量，如事件变量、时间变量等，并用逻辑“与”算子对这些变量加以限制。

例如：雪是白的：



10.2 语义理论简介

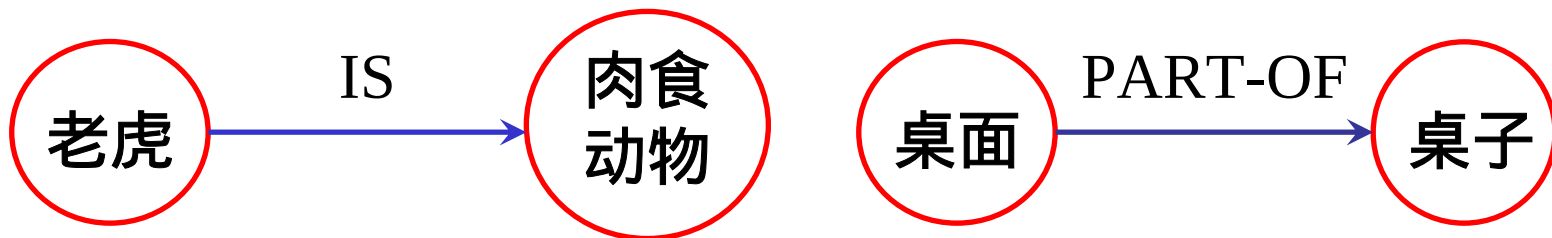
□ 语义网络

有向图：图的结点表示概念，图的边表示概念之间的关系。

边的类型：1) “是一种”：A到B的边表示“A是B的一种特例”

2) “是部分”：A到B的边表示“A是B的一部分”

... ..



10.2 语义理论简介

词义 { 内涵：词本身的意义，是对词代表的概念描述。
外延：词所指代的物体。

如何在语义网络中表示和区分词的内涵和外延？

10.2 语义理论简介

□ 模态逻辑

起源于20世纪80年代初，AI。如：缺省逻辑、时态逻辑、真值维护系统等。

这类逻辑都是试图用一套公理系统来刻画现实世界和自然语言中常见的一些现象。

这类现象从哲学上说就是一般性和特殊性的矛盾。

鸟会飞

企鹅不会飞

10.3 格语法

□ 背景

格语法 (Case Grammar) 是美国语言学家 Charless J. Fillmore 于1966年提出的。代表作：

1966, Towards a modern theory of case

1968, The case for case

1971, Some problems for case grammar

C. J. Fillmore 指出：诸如主语、宾语等语法关系实际上都是表层结构上的概念，在语言的底层，所需要的不是这些表层的语法关系，而是用施事、受事、工具、受益等概念所表示的句法语义关系。这些句法语义关系，经各种变换之后才在表层结构中成为主语或宾语。

10.3 格语法

□ 格的定义

格语法（Case Grammar）中的格是“深层格”，它是指句子中体词（名词、代词等）和谓词（动词、形容词等）之间的及物性关系（transitivity），如：动作和施事者的关系、动作和受事者的关系等，这些关系是语义关系，它是一切语言中普遍存在的现象。

这种格是在底层结构中依据名词与动词之间的句法语义关系确定的，这种关系一经确定就固定不变，不管经什么操作、在表层结构中处于什么位置、与动词形成什么语法关系，底层上的格与任何具体语言中的表层结构上的语法概念，如主语，宾语等，没有对应关系。

10.3 格语法

例如：

- (1) The door opened.
- (2) The key opened the door.
- (3) The boy opened the door.
- (4) The door was opened by the boy.
- (5) The boy opened the door with a key.

- the boy：施事格
- the door：客体格（受事格）
- the key：工具格

10.3 格语法

□ 格语法

三条基本原则：

(1) $S \rightarrow M + P$

句子 S 可以改写成情态 (Modality) 和命题 (Proposition) 两大部分，情态部分包括否定、时、式、体以及其它被理解为全句情态成分的状态。

命题牵涉到动词和名词短语、动词和内嵌小句之间的关系，动词是句子的中心，名词短语按其特定的格属关系依附于该动词。

10.3 格语法

$$(2) P \rightarrow V + C_1 + C_2 + \dots C_n$$

表示命题 P 都可以改写成一个动词 V 和若干个格 C 。动词是广义上的动词，包括：动词、形容词、甚至包括名词、副词和连词。

$$(3) C \rightarrow K + NP$$

K 为格标，是各种格范畴在底层结构中的标记，可以有各种标记形式，如：前置词、后缀词、词缀等。

10.3 格语法

□ 格表

C. J. Fillmore 在1968年的论文中认为，命题中的格包括6种：

- (1) 施事格 (Agentive)：动作的发生者；
- (2) 工具格 (Instrumental)：对动作或状态而言作为某种因素而牵涉到的无生命的力量或客体。
- (3) 承受格 (Dative)：由动词确定的动作或状态所影响的有生物。如，He is tall. 中的 He.
- (4) 使成格 (Factitive)：由动词确定的动作或状态所形成的客体或有生物。或理解为：动词意义的一部分的客体或有生物。如：John dreamed a dream about Mary. 中的“a dream”。

10.3 格语法

(5) 方位格 (Locative) : 由动词确定的动作或状态的处所或空间方位。

如 : He is in the house. 中的“house”。

(6) 客体格 (Objective) : 由动词确定的动作或状态所影响的事物。

如 : He bought a book. 中的“a book”。

10.3 格语法

后来 C. J. Fillmore 在语言分析时又增加了一些格：

（7）受益格（Benefactive）：由动词确定的动作为之服务的有生命的对象。

如：He sang a song for Mary. 中的“Mary”。

（8）源点格（Source）：由动词确定的动作所作用到的事物的来源或发生位置变化过程中的起始位置。

如：He bought a book from Mary. 中的“Mary”。

10.3 格语法

(9) 终点格 (Goal) : 由动词确定的动作所作用到的事物的终点或发生位置变化过程中的终端位置。

如 : I sold a car to Mary. 中的“Mary”。

(10) 伴随格 (Comitative) : 由动词确定的与施事共同完成动作的伴随者。

如 : He sang a song with Mary. 中的“Mary”。

*** 格的数目和名称并不是确定的。**

10.3 格语法

□ 用格语法进行语义分析：格框架约束分析

◆ 格框架表示

格框架中可以有语法信息，也可以有语义信息，语义信息是整个格框架的最基本的部分。

一个格框架可以由一个主要概念和一组辅助概念组成，这些辅助概念以一种适当定义的方式与主要概念相联系。

一般地，在实际应用中，主要概念可以理解为动词，辅助概念理解为施事格、受事格、处所格、工具格等语义深层格。

10.3 格语法

例如 : In the room, he broke a window with a hammer.

[BREAK

[case-frame:

[agentive: HE

objective: WINDOW

instrumental: HAMMER

locative: ROOM]

[MODALs:

time: past

voice: active]]]

10.3 格语法

◆ 分析的基础

词典中记录动词的格框架和名词的语义信息。

对于动词：规定它们所属的必备格、可选格或禁用格，同时填充这些格的名词的语义条件。

- 《动词用法词典》把名词按其与动词格的关系分为14类：受事、结果、对象、工具、方式、处所、时间、目的、原因、致使、施事、同源、等同、杂类。
- 林杏光《词汇语义和计算语言学》
- 符淮青《词义的分析和描写》等

对于名词：填充语义信息，建立名词的语义分类体系。

10.3 格语法

◆ 分析步骤

(1) 判断待分析词序列中主要动词，在动词词典中找出该词的格框架；

(2) 识别必备格

如果格带有位置标志，则从指定位置查找格的填充物；如果格带有语法标志，则在这个分析的词序列中查找语法标志，进入相应的填充；如果格框架还需要其它必备格，查找其它名词的语义信息，按格框架的语义信息要求进行相应的填充。

(3) 识别可选格

(4) 判断句子的情态 Modal

10.3 格语法

格框架分析可以和句法分析结合起来：

- (1) 句法分析：判断出句子的动词、名词短语、介词短语等；
- (2) 查找动词的格框架与名词短语、介词短语的格关系，并进行相应的填充。

必须首先找到动词，从而获得格框架。

10.3 格语法

The young athlete will be running in Los Angeles next week.

从词典中查找 run 的格框架，比如：

Verb: run

Case-Frame [

Neutral

-required (中性格)

Dative

-not allowed

Locative

-optional

Instrumental

-not allowed

Agentive

-required]

run 的中性格象一个物理实体或组织，如：

John ran the machine.

He ran the corporation.

10.3 格语法

CASE

[Agentive: the young athlete

Locative: in Los Angeles

Neutral: the young athlete

[Modal

[Tense: Future

MOOD: Declarative

Time: next week]]]

10.3 格语法

□ 格语法描写汉语的局限性

汉语的一些无动句、流水句、连动句、紧缩、动补、省略等结构，无法或不必用一个统率一个句子的模式来描述，其中连动句和兼语句尤为突出。

例如：1) 他拿了书就上楼去了。

2) 我们选他当班长。

10.4 语义网络

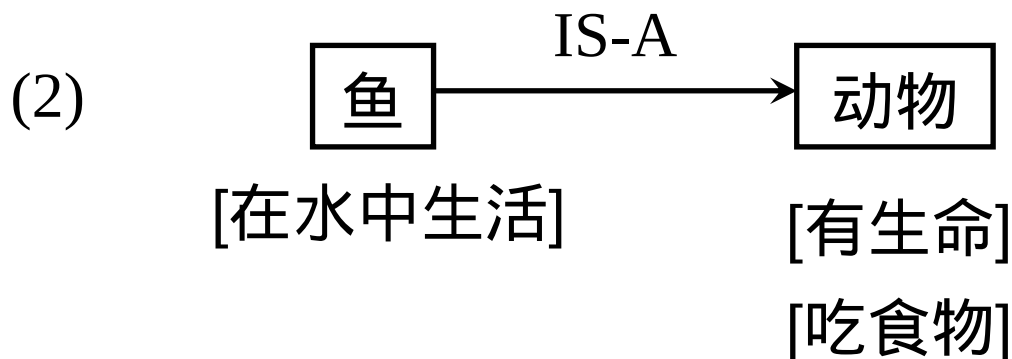
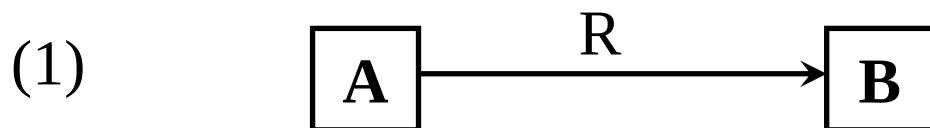
□ 背景

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。1977年美国 AI 学者 G. Hendrix 提出了分块语义网络的思想，把语义的逻辑表示与“格语法”结合起来，把复杂问题分解为几个较为简单的子问题，每个子问题以一个语义网络表示，把自然语言理解的研究向前推进了一步。

10.4 语义网络

□ 语义网络的概念

语义网络通过由概念和语义关系组成的有向图来表达知识、描述语义。

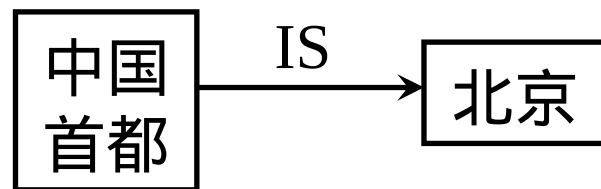


10.4 语义网络

□ 语义网络的概念关系

语义网络各个概念之间的关系，主要由 IS-A, PART-OF, IS, COMPOSED-OF, HAVE, BEFORE, LOCATED-ON 等谓词表示。

- IS-A：表示“具体 - 抽象”关系
- PART-OF：表示“整体 - 构件”关系
- IS：一个结点是另一个结点的属性
- HAVE：表示“占有、具有”关系
- BEFORE/AFTER/AT：表示事物间的先后次序关系
- LOCATED-ON/UNDER/AT：表示事物之间的位置关系

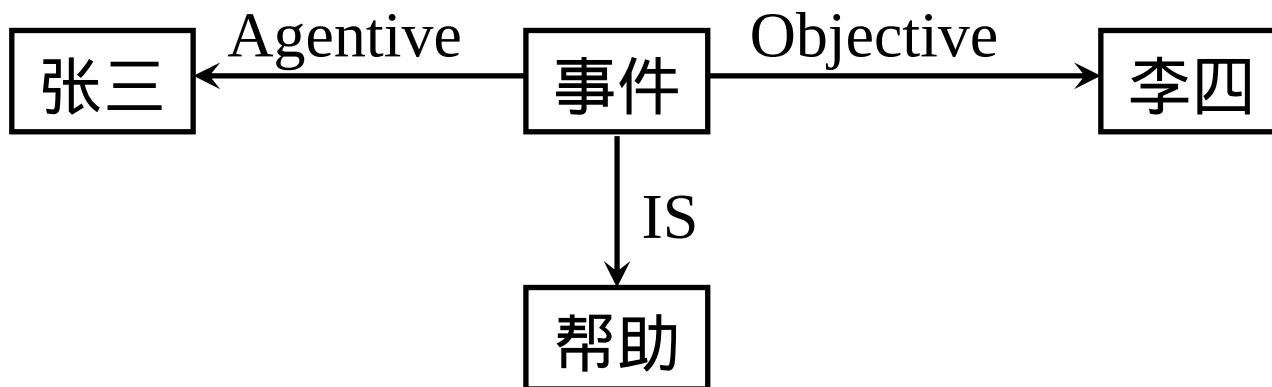


10.4 语义网络

□ 事件的语义网络表示

当语义网络表示事件时，结点之间的关系可以是施事、受事、时间等。

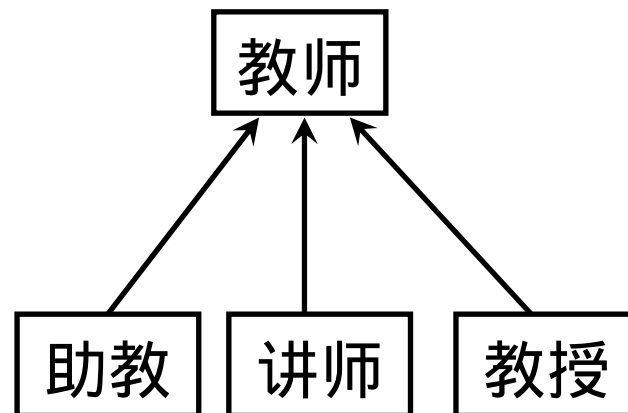
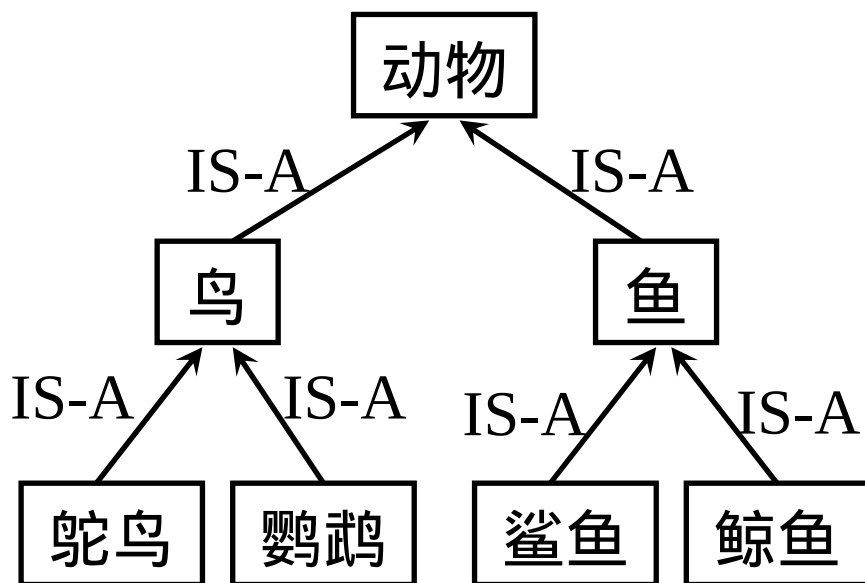
例如：张三帮助李四。



10.4 语义网络

□ 事件的语义关系

- (1) 分类关系：事物之间的类属关系。
- (2) 聚焦关系：多个下位概念构成一个上位概念。



10.4 语义网络

(3) 推论关系：由一个概念推出另一个概念。

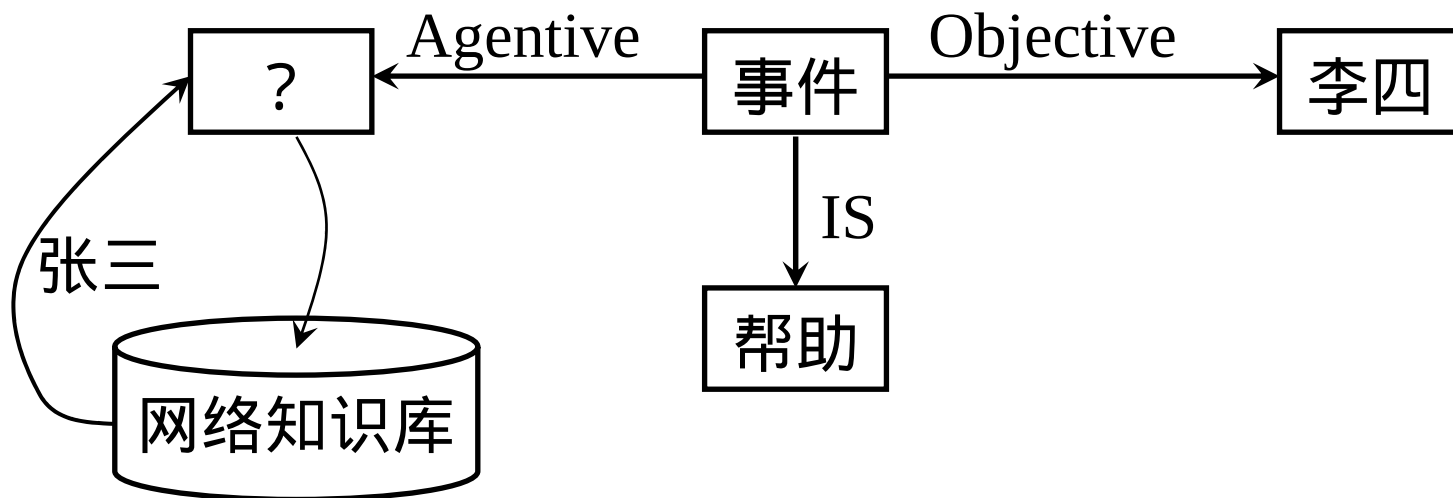
(4) 时间、位置关系：事实发生或存在的时间、位置。



10.4 语义网络

□ 基于语义网络的推理、分析

- (1) 根据提出的问题构成局部网络；
- (2) 用变量代表待求的客体。



10.5 概念依存理论

□ 背景

Schank 和他的同事在70年代提出概念依存理论（ Concept Dependence, CD ）。

1975, Conceptual Information

1977, Scripts, Plans, Goals and Understanding

10.5 概念依存理论

□ CD 理论的组成

1) 三个层次，之一：动作基元

(1) 在概念依存层次，规定了一组动作基元，其它动作是由这些动作基元组合而成的。

抓(Grasp)、移动(Move)、传送(Trans)、去(Go)、推(Propel)、吸收(Ingest)、撞击(Hit)等。

(2) 关于精神世界的概念：心传(MTrans)、概念化(Conceptualize)、心建(MBuild)。

(3) 关于手段或工具：闻(Smell)、看(Look-at)、听(Listen-to)、说(Speak)。

10.5 概念依存理论

2) 三个层次，之二：剧本

用来描写遇到一些常见场景或场合时所采取的一些固定的成套的动作。如：

- a) A推购物车或拿购物筐；
- b) A根据购物单或随意选购一些物品B；
- c) A把选购好的B给收帐员算帐、付款。

10.5 概念依存理论

3) 三个层次，之三：计划

计划中的每一步都是一个剧本，如，外出旅游的安排：

- a) 出门前的准备；
- b) 搭乘交通工具到目的地；
- c) 找住宿地点安顿下来；
- d) 在旅游地游玩；
- e) 若还未尽兴，转 b)，否则，转 f)；
- f) 搭乘交通工具回家。

10.5 概念依存理论

□ 依据CD 理论理解语言

一般文章中一些动作的细节被忽略，计算机难以发现事件、人物、地点等各种指代之间的联系，而 CD 理论试图建立这种联系，正确描述常识，并利用基本动作推理。

该理论对限定领域内的特定应用比较有效。

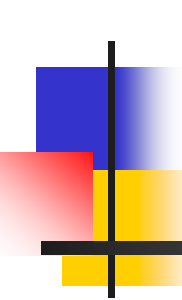
缺陷：对常识的描写过于刻板和定式。

10.5 汉语中的语义计算研究

- ❑ 人大林杏光、北大符淮青等专家的词汇语义学研究；
- ❑ 鲁川教授的句模（sentence model）理论；
- ❑ 东北大学姚天顺等利用语义网络实现汉语的分析与翻译；
- ❑ 上海交大陆汝占教授的内涵逻辑；
- ❑ 中科院声学所黄曾阳教授的概念层次语义网络（HNC）
- ❑ 董振东教授的 HowNet
- ❑ 梅家驹《同义词词林》，上海辞书出版社，1996
- ❑ 《现代汉语分类词典》汉语大词典出版社，1998
-

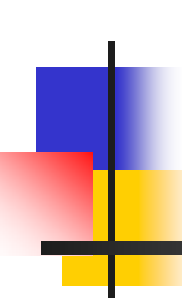
小结

- 语义分析的基本任务及其面临的困难
- 语义计算研究概括及常见的语义理论
- 格语法（定义、格语法、格表、格框架约束分析）
- 语义网络（概念、概念关系、事件语义网络表示、事件的语义关系、基于语义网络的推理分析）
- CD 理论（三个层次：基本动作、剧本、计划）
- 汉语语义计算研究概括



习题

1. 了解汉语 HowNet、HNC、《同义词词林》及其应用。
2. 了解蒙塔格语法（Montague Grammar）。
3. 阅读有关词汇歧义消歧（Word Sense Disambiguity）的论文。



Thanks

谢谢!