

汉语名词的内在知识及其表示*

姬东鸿 黄昌宁 赵军

清华大学计算机科学系

北京, 100084

摘要: 本文, 我们详细讨论了汉语名词的内在知识, 即名词区别于其上位概念的语义信息; 探讨了名词内在知识的表示机制。在一个基于继承和缺省机制的汉语概念体系中, 名词的内在知识决定着其组合知识的继承。我们采用的属性结构是一种多阶的属性结构, 它既避免了冗余项的产生, 又能刻划较深层的语义信息。

关键词: 内在知识 属性结构 知识表示 自然语言理解

the Internal Knowledge of Chinese Nouns and Its Representation

Ji Donghong Huang Changning Zhao Jun

Department of Computer Science and Technology

Tsinghua University

ABSTRACT: In this paper, We detailedly discuss the internal knowledge of Chinese nouns, i.e., the semantic information which distinguishes between the nouns and their upper concepts and explore its representation scheme. In a conceptual system of Chinese nouns based on inheritance and default, the internal knowledge determines the inheritance of the compositional knowledge of nouns from their upper concepts. we adopt FS (feature structures) to represent the internal knowledge, the feature system is multi-leveled, it can remove redundant features and depict deeper semantic information.

Keywords: internal knowledge feature structure knowledge representation natural language understanding

1 引言

处理大规模的真实文本和研制高性能的机器翻译系统需要大量的词汇知识 ([1]), 特别是对于汉语这种意合的语言来说, 由于缺乏具体的句法特征, 语义知识、特别是词汇语义知识对于汉语的句法分析和语义分析显得更为重要 ([2])。

一般地, 每个词都表示一个或几个概念, 这些概念在汉语的概念体系中具有一个或几个上位概念。因此词汇知识应包括这些概念区别于其上位概念的知识以及它们与其它概念之间的组合关系, 我们称前者为词汇的内在知识, 后者为词汇的外在知识或组合知识。词汇的内

*本文由国家自然科学基金重点项目资助

在知识往往不同于其上位概念的内在知识，例如“雨”和“毛毛雨”，但词汇的外在知识与其上位概念的外在知识有相同之处，例如，“雨”和“毛毛雨”都可与“下”组合，也有不同之处，例如，“雨”可与“大”组合，但“毛毛雨”却不能与“大”组合。实际上，词汇的外在知识可看作对其上位概念外在知识的一种继承，凡是与其内在知识相容的，则继承，不相容的，则不继承。因此，建立词汇内在知识的表示机制对于建立整个汉语基于继承和缺省机制的概念体系具有重要意义。

名词的语义知识对于确定名词短语的语义，特别是确定复合名词短语的组合层次及其语义是必不可少的资源。J.Pustejovsky 等提出了一种名词语义的表示机制，即 Qulia 结构 ([3, 4])，它大体包括组成，起源，功能，和客观性质四方面的语义信息，其中的客观性质就是指名词的内在知识。但对于这些性质的具体内容，即名词内在知识的语义结构，J.Pustejovsky 并没有给出进一步的探讨。J.Miller 等研制了一个英语词汇数据库 WordNet ([5])，名词的语义信息包括三方面：组成，属性，和功能，但这些性质只涉及名词的组合知识。汉语名词的语义知识通常可用属性及其值来表示，因此，名词的内在知识可表示为由属性及属性值组成的偶对集。由于名词的内在知识可能涉及外貌，组成，功能，性质等方面，名词的属性也可以分为若干种不同的类（参见第 2 节）。

实际上，名词的内在知识可看作一种对其上位概念的语义限制。如何处理这些限制和上位概念之间的关系，有两种不同的策略，一种是将这些限制视为对其上位概念的直接约束，例如，“唐人”中的时间因素“唐朝”直接约束“人”，即“唐人”具有时间属性，其值是“唐朝”；“复杂物”具有复杂性属性，其值是“复杂”。这种策略一方面会导致概念体系中的“属性冗余”：名词具有动词的几乎所有属性，具体名词具有抽象名词的几乎所有属性等。例如，名词具有时间（“唐人”），处所（“空间站”），起源（“对流雨”），对象（“胎教”），范围（“社教”）等，而动词也具有这些属性；具体名词具有复杂性（“复杂物”），重要性（“要人”）等，而抽象名词也具有这些属性。另一方面，这种策略不能确定较深层的语义关系。例如“唐朝”的语义指向是与“人”有关的动词“生”，“复杂”的语义指向是“物”的属性外貌、结构、或功能等。为避免这两方面的问题，本文我们采取另一种策略，采用多阶的属性结构来描述名词语义，在这种属性结构中，名词的属性被分为若干阶，一阶属性是名词本身的属性，二阶属性是一阶属性的属性或属性值的属性，依此类推。例如，时间被看作“人”的二阶属性，同时又是“生”的一阶属性；复杂性是“物”的二阶属性，同时又是结构，外貌，功能等的一阶属性。

本文以下的组织是这样的：第 2 节，我们讨论名词的内在知识；第 3 节，我们给出名词内在知识的表示机制，并举出若干实例；第 4 节，我们给出结论并讨论未来的工作。

2 内在知识

名词的内在知识包括它所描述的概念区别于其上位概念的所有属性及其属性值。根据这些属性的所属，它们可以分为三类：

i) 名词本身的属性，例如：(1) 具有属性数量，其值为“小”。再如(2)的意思是“大雨”，作为“雨”的下位概念，它具有属性数量，其值为“大”。

(1) 小雨

(2) 霈

ii) 名词属性的属性, 例如: (3) 和 (4) 都是“人”的下位概念, (3) 区别于“人”的属性是属性才智的属性特殊性, 其值为“奇”; (4) 中, “奇”也是属性特殊性之值, 但此时, 该属性可能属于一阶属性才智、地位、或者外貌等, 因此这种名词具有歧义。它可能的扩展形式是 (5), (6), 或 (7) 等, 它们分别指明了一阶属性。

(3) 奇才

(4) 奇人

(5) 才能奇特的人

(6) 外貌奇特的人

(7) 地位奇特的人

iii) 名词属性值的属性, 例如: (8) 作为“小雨”的下位概念, 它的区别属性是“雨点”的属性体积, 其值为“小”, 而“雨点”又是“毛毛雨”的一阶属性组成的属性值。(9) 的区别属性是持续性, 其值是“间断”, 它是“下”的一阶属性, 而“下”又是“雨”的一阶属性自动之属性值。

(8) 毛毛雨

(9) 阵雨

一般地, 名词的属性名是抽象名词, 属性值是名词, 动词, 或者形容词, 因此名词的内在知识不仅会涉及名词的属性, 而且也会涉及动词和形容词的属性。不同的词类有不同的属性集, 本节, 我们只讨论名词和动词的属性, 形容词的属性将另文讨论。表 1 列出的是名词的属性集, 第二栏的属性类都代表一个属性集, 第三栏列出的是具体的属性。其中, 具体名词的属性对应 1—7, 抽象名词的属性对应 1、2、4、6、7, 人类名词作为一种特殊的具体名词, 对应属性 1—8。

编号	属性类	属性	解释
1	物名	名称 别名 俗称	物的名称
2	物属	上位 所有 组成 1 组成 2	物的名称, 所有者, 组成成分, 作为组成成分的概念。
3	物貌	形状 式样 体态	物的外貌, 内貌, 形象等。
4	物理	密度 数量 范围 年龄 性别 质地 硬度	物的理性性质。
		复杂性 抽象性 正确性 深奥性	
5	物境	环境 经况 政况 关系 心境...	物的环境, 经济状况, 政治状况, 相互关系等以及人的心态。
6	物为	自动 功能 生产...	物的自动作, 功能, 产生或建立。
7	物值	荣誉 地位 意义 价值	物的价值观和审美观。
8	德才	品德 性格 态度 才智	人的品性, 态度和才智。

表 1

表 2 列出了动词的若干属性, 其实动词的属性远不只此, 动词的每一种格关系都可看作一种属性。我们这里只讨论名词的内在知识涉及到的动词属性。

编号	属性	解释
1	动因	动作的起因
2	结果 目的 效果	动作的结果,目的,和效果
3	伴随	动作的伴随物
4	时间	动作发生的时间, 及时性, 持续时间
5	空间	动作发生的空间, 处所
6	对象	动作的对象
7	范围	动作的范围
8	持续性	动作的持续性

表 2

实际上, 每一种属性都可看作由一个词集到另一个词集的映射, 如果两个属性对应的映射互为逆映射, 则我们称它们为逆属性。例如, 组成 1 和组成 2, “桌腿”是“桌子”的属性组成 1 的值, 而“桌子”是“桌腿”的属性组成 2 的值。再如当事和自动, “雨”是“下”的属性当事的值, 而“下”是“雨”的属性自动的值。显然这两种属性并不会造成属性冗余。

3 表示机制

我们采用属性结构描述名词的内在知识, 属性结构应包括名词内在知识涉及到的所有一阶和多阶属性。对于任意名词 w , 假设 f 是其任意属性, v 为其属性值, 则我们用 $f(f: v)$ 表示。因此, 属性结构可看作由属性和属性值偶对组成的嵌套结构。如果某个属性不能确定, 则用变元 x 表示。

考虑例子(1)–(4), (8), (9), 其属性结构分别为(1')–(4'), (8'), (9'),

(1') 小雨 (量: 小)

(2') 霏 (量: 大)

(3') 奇才 (才智 (特殊性: 奇))

(4') 奇人 (x (特殊性: 奇))

(8') 毛毛雨 (组成 1: 雨点 (体积: 小))

(9') 阵雨 (自动: 下 (持续性: 阵))

(4') 中, 一阶属性可能是地位, 外貌, 或者功能, 因此, 我们用 x 表示。

再如 (10) – (16), 其属性结构分别为 (10') – (16')。

(10) 唐人 (10') 唐人 (建立: 生 (时间: 唐朝))

(11) 东北虎 (11') 东北虎 (建立: 产 (处所: 东北))

- (12) 空间站 (12') 空间站 (建立: 建立 (空间: 空间))
 (13) 雨前茶 (13') 雨前茶 (建立: 采摘 (时间: 雨前))
 (14) 早稻 (14') 早稻 (建立: 种植 (时间: 早))
 (15) 晚茶 (15') 晚茶 (消耗: 喝 (时间: 晚))
 (16) 早餐 (16') 早餐 (消耗: 吃 (时间: 早))

由以上例子可知: 名词内在知识如果含有时间或者空间属性, 它们一般是与该名词有关的建立类动词 (建立, 生产等) 或消耗类动词 (吃, 喝等) 的一阶属性。如果把它们当作该名词的一阶属性, 那么就无法刻划诸如 (13) 和 (15) 中时间信息的差别。

再如 (17) 和 (18), 它们的内在知识可分别表示为两个语义网络 (图 1 和图 2), 这里涉及事件概念及其属性。我们把它们简化为图 1' 和图 2',

(17) 对流雨

(18) 湿雪

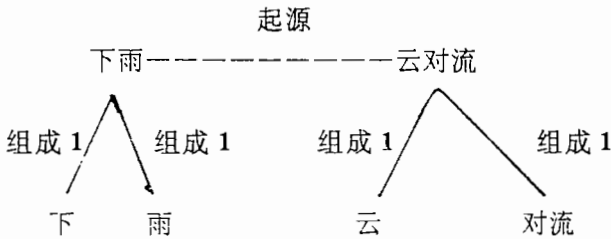


图 1

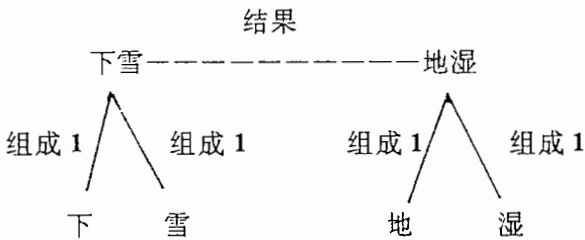


图 2

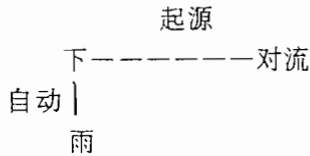


图 1'

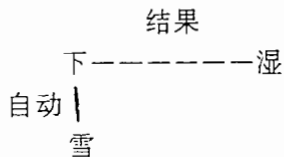


图 2'

这对应属性结构（17'）和（18'）。

（17'）对流雨（自动：下（起源：对流））

（18'）湿雪（自动：下（结果：湿））

由这两个例子可以看出，名词内在知识的表示机制应同时考虑严密性和可操作性，如何处理这两方面之间的矛盾是一个关键问题。

4 结论及展望

本文详细探讨了汉语名词的内在知识及其表示机制，属性结构其实是一种关系模型，属性可表示各种语义关系。我们采用的属性结构是一种多阶的属性结构，这既避免了概念体系中可能出现的属性冗余，又能描述较深层的语义信息。

建立名词内在知识的表示机制是为了建立名词基于继承和缺省机制的概念体系，它对于下位名词对上位名词组合知识的继承起着决定作用。另外，名词的内在知识可能涉及过程或者事件概念，未来的工作主要包括建立过程和事件概念的属性体系，下位名词对上位名词组合知识的继承机制，以及如何从词典文本中自动获取名词及其它词类的内在知识等。名词的内在知识和组合知识一起构成名词的语义知识，未来的工作还包括如何利用名词的语义确定名名短语和形名短语的语义，特别是确定复合名词短语的组合层次及其组合义等。

参考文献

- 1 B.Boguraev, Building a Lexicon, International Journal of Lexicograph, 4(3),1991.
- 2 贾彦德, 汉语语义学, 北京大学出版社, 1990,
- 3 J.Pustejovsky, the Generative Lexicon, Computational Linguistics,17(4),1991.
- 4 J.Pustejovsky, S.Bergler, P.Anick, Lexical Techniques for Corpus Analysis, Computational Linguistics, 19(2),1993.
- 5 G.A.Miller, R.Backwith, C.Fellbaum, D.Gross, K.J.Miller, Introduction to WordNet: An On -line Lexical Database, International Journal of Lexicograph, 3(4),1990.