

# 自然语言处理系统中的事件类名词的耦合处理

陈国华, 赵克, 李亚涛, 易帅

(西安电子科技大学 机电科学与技术研究所, 陕西 西安 710071)

**摘要:**介绍了自然语言处理系统中事件类名词耦合的一种处理方法。事件类名词是汉语名词的一个特殊子类,这类名词兼事物性与动作性于一身。给出了事件类名词的语义认知基础,并采用基于知识的方法,应用概念从属理论,对事件类名词进行详细的语义分析。在此基础上,对事件类名词的耦合情况进行分类,并给出了事件名词解耦处理的具体方法。这种对事件类名词的耦合处理方法在智能辅导领域中得到了较好的应用。

**关键词:**事件类名词;语义分析;自然语言处理;耦合;智能辅导系统

**中图分类号:**TP182

**文献标识码:**A

**文章编号:**1673-629X(2008)06-0060-04

## Coupling Processing of Event Noun in NLP Systems

CHEN Guo-hua, ZHAO Ke, LI Ya-tao, YI Shuai

(Research Institute of Mechatronics, Xidian University, Xi'an 710071, China)

**Abstract:** Primarily introduced a kind of method of the coupling processing of event noun in NLP systems. The event noun is a particular category in Chinese nouns, with its verbal and nominal. Firstly, the semantic cognizing base of the event noun was presented. Secondly, applied the knowledge-based method and the theories of conceptual dependency theory to analyse the semantic of event noun in detail. Finally, on the base of this analysis, a special method of classifying and processing of the coupling of event noun was given. This kind of method of handling event noun has been applied to intelligent tutoring systems.

**Key words:** event noun; semantic analysis; NLP; coupling; intelligent tutoring systems

## 0 引言

文中所涉及的自然语言处理系统为智能辅导系统在中学数学领域内的实现提供了必要的人机交互接口。智能辅导系统是把人工智能应用于计算机辅助教育的一种专家系统,是一种更适用于教与学需求的面向辅导的教育系统。在本领域内,“和、差、延长线、垂线、平行线”等数学名词是普遍存在的,它们作为名词却具有动词的潜在特征。动词的核心意义是“动作”,名词的核心意义是“具体物体”<sup>[1]</sup>。这类兼有动词性的名词在语义上有一个共同点,即都是表示由一系列动作经过抽象的“行为”而得到抽象的“事件结果”。相对“具体物体”来说,抽象的“事件”就是名词比较边缘的意义了。把这种具有边缘意义的名词称作事件类名词。在智能辅导系统中,已经成功地实现了大量事件类名词的语义理解,然而在本领域内,事件类名词与其

它名词,事件类名词与事件类名词之间的耦合的现象大量存在。成功实现事件类名词的解耦,是实现复杂事件类名词语义理解的关键。介绍了事件类名词的语义理论基础、语义特点分析、耦合分类、解耦的实现。

## 1 理论基础

### 1.1 语义基础

汉语使用中,名词短语的功能远远大于单个名词,单个名词只能表达简单的含义,而名词短语可表达复杂的、更为完整的语义<sup>[2]</sup>。当然,名词短语的语义知识表示是基于名词的语义知识表示。名词短语是以名词为核心词的向心结构。对名词短语的自然语言理解实质上是一个确定短语中各个名词与修饰语之间的搭配关系是否合理,并确定核心词的过程<sup>[3]</sup>。因此,对名词短语的分类是正确实现名词短语理解的前提。

文中所阐述的自然语言理解系统是基于产品设计领域,在该系统中,将名词短语分为以下三类:偏正结构、联合结构、事件类名词。

偏正结构由修饰语和中心语构成,偏正结构的功能和中心语相同,中心语是载体,负载整个结构体的功

收稿日期:2007-09-21

基金项目:科技部创新基金(01c26226111002)

作者简介:陈国华(1980-),男,广东汕头人,硕士研究生,研究方向为人工智能、知识工程、创新设计;赵克,教授,研究方向为人工智能、知识工程、创新设计。

能。其中,中心语也叫核心词。

联合结构是指直接分成并列的结构。“并列”是就句法(不是语义)而言的,凡并列的成分,在句法上地位平等,可以调换次序而不改变结构关系。联合结构各并列项的语法功能与整个结构的语法功能基本相同,联合结构属多核心结构。

在定义事件类名词之前,先来解释事件。什么是事件?简单地说,一个事件可以表达一个完整的语义,可以说清楚一件事。通常情况下,要有资源和动作才可以完成一个事件<sup>[4]</sup>。动作也就是对对象的操作,动作通常由动词来完成,而上述的对象可以看作是资源。因此,可以将事件类名词看作是动作对对象操作的结果。简单地说,事件类名词就是动词经过推理的结果<sup>[5]</sup>。经过上述分析,可以总结出事件类名词的定义。事件类名词是表示由一系列动作经过抽象的“行为”而得到抽象的“事件结果”。相对“具体物体”来说,抽象的“事件”就是名词比较边缘的意义了。把这类具有边缘意义的名词称作事件类名词。

## 1.2 概念从属理论

任何一种自然语言,无论英语还是汉语,其中所包含的概念都数以万计,对这么多的概念逐一进行理解和处理,无疑是一个非常庞大复杂的工程。倘若从抽象化的角度来处理概念,将使问题得到简化。

概念从属关系是指一个概念的外延包含另一个概念的所有外延。从属关系引入后,知识被划分为一系列结点,知识结点之间呈现树状联系。利用从属关系对特定领域的知识块进行抽象化处理,使其形成具有不同抽象层次的知识结点形成的树状层次结构,这种加入从属关系的树模型就被称为概念从属树(CDT)<sup>[6]</sup>。

在概念从属树上,体现了概念之间的属性继承关系。每个名词概念都是一个抽象程度不同的具体概念,即抽象具体概念树上的一个子节点。上层概念为下层概念的父类概念,下层概念继承上层概念的所有属性,并具有自己特有的属性。概念从属树上的每个节点通过爬树来理解概念的内涵和外延,进行深层语义理解。在向知识库中添加概念时,只需填写与其父概念不同的内涵和外延,这就大大缩减了知识库的容量且便于管理。图1是对三角形的知识结构进行细化之后的概念从属树。

## 2 事件类名词的语义特点分析

在前面的分析中,知道事件类名词是由事件转化而来的,它跟静态事件不一样<sup>[7]</sup>。静态事件是动词的静态描述,每一个动词都对应一个静态事件。而事件

类名词与事件是多对一的关系,如“相交事件”可以产生交点和夹角,所以“交点”和“夹角”这两个事件类名词都是“相交事件”的结果。正因为事件类名词与事件有关,是事件产生的结果,因此在对事件类名词进行处理时,可以先找出该事件类名词中所包含的动态含义,把它转化成相应的事件来处理,得出事件的结果,然后再对这些结果进行相应的处理。在智能辅导系统中,一般处理流程为:

- ①找出事件类名词所对应的动词;
- ②对该事件类名词块进行分析,找出动词事件处理所需的信息;
- ③进行相应的事件处理并找出正确的结果返回;
- ④返回到原事件类名词所在的位置,继续句子的分析。

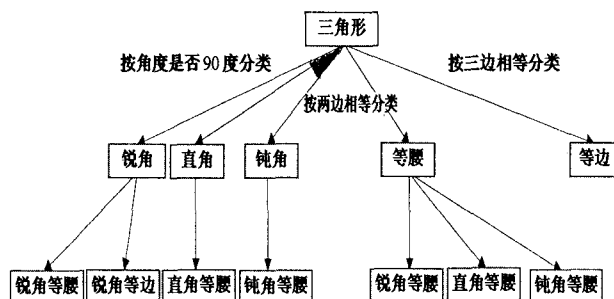


图1 知识细化后的三角形概念从属树

前面说过,事件类名词兼有动词和名词两类词的特性。在事件类名词的语义分析实现中,需要充分考虑其特点,既要对其动词属性即动作性进行处理,又要兼顾其作为名词即事物性的一面。下面分别阐述。

### 2.1 动作性

事件类名词的动作性分析同动词的语义分析十分相似,最终会生成相应的事件结果。除了词义分析外,还包含了施受动对象分析,这是其结果事件产生的前提。事件类名词的施受动对象的分析可借鉴动词的施受动对象分析。

事件类名词的施受动对象的确定是根据事件类名词所对应的动词的施受动对象来确定的。如“边AB的垂线CD”中的事件类名词“垂线”所对应的标准动词是“垂直”,所以施动对象为“边AB”,受动对象为“线段CD”。找出了施受动对象之后,需要限定事件类名词与其对象的语义搭配关系。因此,将“垂线”的施受动对象进行概念抽象,搜索相关概念从属树,即搜索线的概念从属树,获得“边”和“线段”的抽象概念“线”,判断事件类名词与其对象的语义搭配是否合理。如图2所示事件类名词“垂线”的静态事件知识指出其施受动对象为概念“线”,可见句中的“垂线”与施受格搭配合理,满足语义关系。

分析完施受动对象后,事件类名词也会产生相应的动态事件和结果事件。事件类名词的结果事件就是其相应的动词操作后产生的新结果,是理解事件类名词所需的实效知识。分析“线段  $AB$  垂直线段  $CD$  于点  $O$ ”这个句子,该例中“垂直”为事件类名词“垂线”相应的动词,其施动对象为“线段  $AB$ ”,受动对象为“线段  $CD$ ”,生成了动态事件: $AB \perp CD$ ,结果事件就是垂足  $O$ 。这同样是事件类名词“垂线”的结果事件。图 2 中给出了“垂直事件”的动态事件,而表 1 则比较了“垂直”和“垂线”的处理过程。可见,事件类名词的动作性分析就是对其相应的动词进行语义分析。基于上述分析,构建了事件类名词语义处理的流程如图 3 所示。

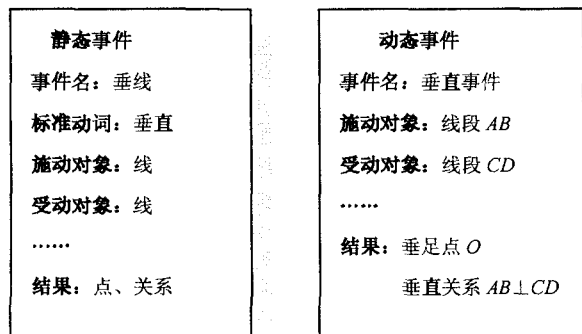


图 2 事件模型示例

表 1 “垂直”与“垂线”处理比较

句子	线段 $AB$ 垂直线段 $CD$	线段 $AB$ 的垂线 $CD$
提取词汇	垂直	垂线
词性	动词	事件类名词
施动对象	线段 $AB$	线段 $AB$
受动对象	线段 $CD$	线段 $CD$
动态事件	$AB \perp CD$	$AB \perp CD$
结果事件	生成垂足	生成垂足

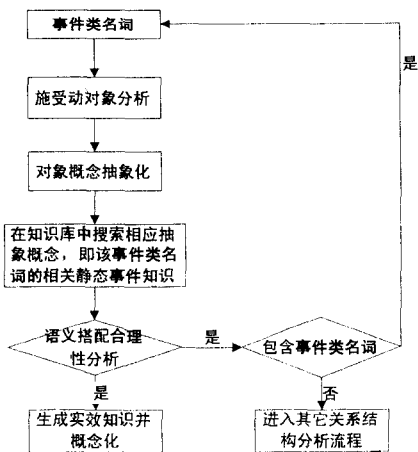


图 3 事件类名词语义实现流程

## 2.2 事物性

作为名词,事件类名词可与其他名词构成偏正或联合结构的名词短语,如:“3 加 2 的和”与“两个数的积与商”。前者是由事件类名词和名词构成的偏正结

构,后者虽然两个事件类名词“积”与“商”形成了联合结构,但实际上该短语仍是偏正的结构。所以,事件类名词与其他名词搭配时只会构成偏正这种向心结构<sup>[7]</sup>。偏正结构中的事件类名词与别的名词构成名词间的操作关系,其中心语为事件类名词。操作关系是指一个名词是与另一名词相关的操作类名词。如上例中,“和”就是与“3”和“2”相关的操作类名词。在这类关系中,操作类名词是描述的重点,且一般位于另一个名词的后边,而另一名词是它的修饰限制成分。而判断这种操作关系是否成立,这需要在名词概念的概念从属树中查找,判断两名词是否相关概念。事件类名词的偏正结构名词短语的处理流程如图 4 所示。

以“三角形  $ABC$  的边  $AB$  的延长线”为例,从左到右识别该名词短语。首先“边  $AB$ ”与“三角形  $ABC$ ”满足外延关系,故“边  $AB$ ”为中心名词;然后搜索 CDT 得到“边”的父概念为“线段”,且“延长线”是“线段”的一个外延。由此可得“延长线”是第一步分析所得的中心名词“边”的事件名词,而所得到的“延长线”,即为整个短语的中心名词。

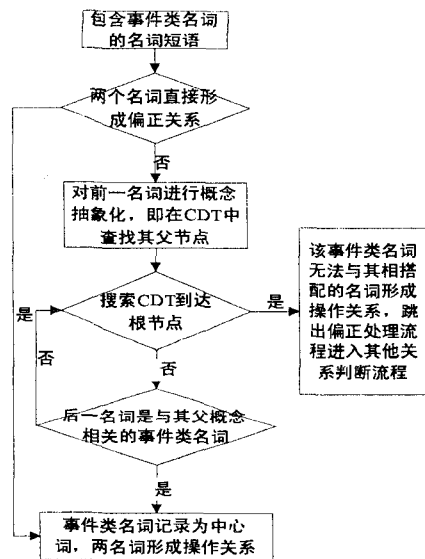


图 4 事件类名词的偏正结构名词短语处理流程

## 3 事件类名词耦合分类及处理

### 3.1 事件类名词与并列关系的耦合

- ①  $AC$  与  $BC$  的交点。
- ② 5 与 3 的和与差。
- ③ 求  $AB$  的长和角  $B$  的正弦值。
- ④ 求 6 与 8 的积与商和  $A$  的值。

该类型的耦合关系主要包括以下三个特点:

(1) 事件类名词的受动对象之间构成并列关系,主要是由于事件类名词对受动对象的数量有一定的限制,如题①中的“交点”要求受动对象要两个或两个以

上。

(2)由相同受动对象的事件类名词之间构成并列关系,如题②中“和”与“差”构成并列关系。

(3)事件类名词与非事件类名词之间构成并列关系,如题③中“AB的长”与“角B的正弦值”。

在对事件类名词的语义分析中知道,事件类名词具有动作性和事物性。动作性体现在事件类名词对其受动对象的操作关系上,而事物性体现在事件类名词与其它名词的偏正关系或并列关系上。处理事件类名词时,要先处理其动作性,再处理其事物性。基于上面的分析,结合事件类名词与并列关系的耦合的特点,在处理事件类名词与并列关系的耦合时,应该依照以下原则处理:首先处理事件类名词的受动对象之间的并列关系;其次处理事件类名词之间的并列关系;最后处理事件类名词与非事件类名词之间的并列关系。如题④中,两个受动对象“6”和“8”先构成第一层并列关系,事件类名词“积”和“商”构成第二层并列关系,最后“6与8的积与商”和“A的值”构成第三层并列关系。

### 3.2 事件类名词与偏正关系的耦合

①求三角形的中线的长度。

②求平行四边形的高的长度。

事件类名词的事物性体现了事件类名词可以与其它名词构成偏正关系或并列关系。事件类名词与偏正关系耦合的特点是事件类名词后面有核心词,且核心词与事件类名词构成外延关系,整个名词块呈偏正结构,事件类名词作为修饰语修饰核心词。所以该类型的耦合处理主要是判断事件类名词与其后面的核心词是否构成偏正关系,若是,则事件类名词后面的核心词为整个名词块的核心词。如题①中“中线”与“长度”构成偏正关系,核心词为“长度”;题②中“高”与“长度”构成偏正关系,核心词为“长度”。

### 3.3 事件类名词与事件类名词的耦合

①三角形ABC边AC与BC的夹角的外角平分线。

②求线段AB、BC的延长线的交点。

③方程两个根的最小公倍数的平方。

事件类名词与事件类名词的耦合的主要特点是事件类名词之间的嵌套,即前面的事件类名词作为后面的事件类名词的受动对象,事件类名词之间构成操作关系。后面的事件类名词作为其前面部分结构的核心词,最后的事件类名词作为整个名词块的核心词,整个名词块呈偏正结构。处理该类型的耦合关系,主要是从前向后判断紧挨着的两个事件类名词之间是否构成操作关系,即前面的事件类名词为后面的事件类名词的受动对象。若是,则后面的事件类名词为核心词,接着往下判断,直到最后一个事件类名词,则最后的事件

类名词为整个名词块的核心词。如题①中事件类名词“外角平分线”与事件类名词“夹角”构成操作关系,核心词为“外角平分线”;题②中事件类名词“交点”与事件类名词“延长线”构成操作关系,核心词为“交点”;题③中事件类名词“平方”与事件类名词“最小公倍数”构成操作关系,核心词为“平方”。

### 3.4 事件类名词耦合处理流程图

事件类名词耦合处理流程图如图5所示。

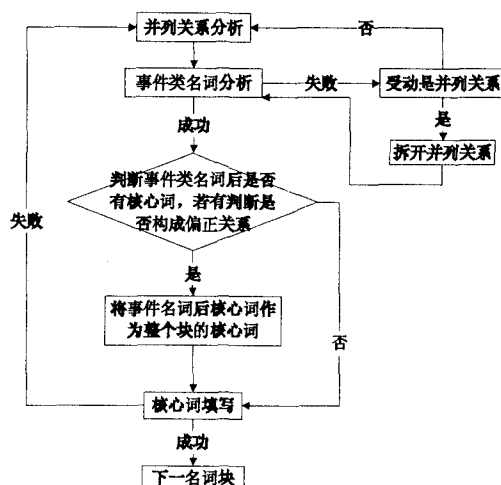


图5 事件类名词耦合处理流程

## 4 结束语

事件类名词是自然语言理解中较为特殊的名词短语,而事件类名词的解耦是自然语言处理系统中名词处理模块的重要组成部分,也是较为复杂的部分。已经将这种处理模式应用到了智能辅导系统中,经过反复的、大量的句子测试,基本达到了预期的目的。但对一些有歧义的耦合的处理并不尽如人意,如“求角A和角B的余弦值。”,该题可以理解为分别求两个角的余弦值,也可以理解为分别求角A的度数和求角B的余弦值。对于这种类型的耦合,系统还不能很好地处理,还需要进一步的研究和完善。下一步的工作将着重于这一方面的研究,相信不久将有突破性的进展。

### 参考文献:

- [1] 陆炳甫. 核心推导语法[M]. 上海:上海教育出版社,1993.
- [2] 石纯一, 黄昌宁, 王家钦. 人工智能原理[M]. 北京:清华大学出版社,1993.
- [3] Allen J. 自然语言理解[M]. 第2版. 刘群,等译. 北京:电子工业出版社,2005.
- [4] Jurafsky D. 自然语言处理综论[M]. 冯志伟,孙乐译. 北京:电子工业出版社,2005.
- [5] 韩蕾. 现代汉语事件名词分析[J]. 华东师范大学学报:

(下转第66页)

个个功能模块构成的程序结构称为模块化结构。

C 语言是一种结构化的程序设计语言。它直接提供了三种基本结构的语句,并提供了定义“函数”的功能,在 C 语言中,它是用函数来完成上述子程序的所有功能,也就是函数真正实现了上述一个个“模块”的具体功能;而且 C 语言允许对函数独立进行编译,从而可以实现模块化。

### 3 C 程序设计语言的优缺点和生命力

在当今众多的计算机高级程序设计语言中,C 语言表现出如下优点。

#### 3.1 功能强大

C 语言不仅提供了大量的标准库函数实现常规通用的功能,而且,它还允许用户根据自己实际的需要编写自定义函数,满足了不同用户特定的需要。在 C 语言中,函数就是一个个功能模块,实现了具体特定的功能,用户在程序中通过函数之间不同层次的调用完成了强大的功能。

#### 3.2 运算丰富

C 语言提供了极其丰富的运算符,主要包括:算术运算、赋值运算、自加自减、关系运算、逻辑运算、位运算等等,每种运算都完成了一定的功能,从而极大地增强了 C 语言的编程功能。

#### 3.3 结构精练

正如上所述,任何 C 程序都可由三种基本的结构而构成,使得程序结构整齐,也使得 C 程序形成了一个模块结构。标准的基本结构使得 C 程序结构显得十分精练。

#### 3.4 语句简洁

在 C 语言程序中,语句种类仅仅包含了赋值语句、函数调用语句、三种基本结构语句等,即可构成了功能强大的程序。可见,C 语言中的语句十分简洁,也便于用户掌握和使用。

#### 3.5 强类型语言

在 C 语言中,要求运算符两侧数据类型必须一致,为此,它提供了如下三种类型之间的转换。

##### 3.5.1 算术类型转换

即在算术运算过程中所进行的数据类型转换,使得算术运算符两侧的数据类型一致,并且这种转换是由系统自动完成的。

##### 3.5.2 强制类型转换表达式

一般形式为:(类型名)(表达式)。其作用:可利用强制类型转换运算,将一个表达式的值转换成指定的类型,并且这种转换是人为设定的。例如:(int)3.14 将 3.14 转换为 3。

##### 3.5.3 赋值运算中的类型转换

如果赋值运算符两侧的数据类型不一致时,在赋值前系统将自动将右侧表达式所求的数值按赋值号左边变量的类型进行转换,然后再赋值给左边的变量,这称之为赋值类型转换(赋值兼容)。

C 语言是一种典型的面向过程的高级程序设计语言,同时,它通过位运算等操作也具有了低级语言的强大功能,使 C 程序的运行效率相对较高。为此,C 程序设计语言不仅可以用来开发大型的系统软件,如 UNIX 等,而且还可适用于应用软件的开发,因此,它具有很强的生命力。

当然,在 C 语言中由于指针的引入给某些操作带来了不安全的隐患,但是,同时指针的存在也为对字符串处理等操作提供了极大的方便。

### 4 结束语

从三个不同侧面透视了 C 高级程序设计语言的主要特性,分析了其基本编程思想和用其编程解决实际问题的思路,同时,总结了上述三个方面的内在联系和它们的统一性。并论述了 C 程序设计语言面向过程的模块化编程思想和具体程序中的实现形式,显示了其优点和其强大的生命力,具有较高的概括性和理论意义。

#### 参考文献:

- [1] 谭浩强.C 语言程序设计[M].第 2 版.北京:清华大学出版社,1999.
- [2] 谭浩强.C 语言程序设计[M].第 3 版.北京:清华大学出版社,2005.
- [3] 张 艳.基于 VC++ OpenGL 的分形山三维模拟生成技术[J].计算机技术与发展,2007,17(8):243-244.
- [4] Alsuwaiyel M H.算法设计技巧与分析[M].北京:电子工业出版社,2005.
- [5] 范 华,秦茂玲,张新法.典型“稳定婚姻问题”的简明矩阵算法实现[J].山东师范大学学报:自然科学版,2007,22(1):29-32.

(上接第 63 页)

哲学社会科学版,2004,36(5):106-112.

- [6] 吴中兴,赵 克,胡钢伟,等.概念从属树——一种新的树模型设计[J].计算机应用,2004,24:99-100.

- [7] 张小林.机械产品设计领域的自然语言理解中名词性短语的语义分析[D].西安:西安电子科技大学,2005.