

◎数据库、数据挖掘、机器学习◎

## 现代汉语“对象格”自动识别研究

汪梦翔<sup>1,2</sup>, 王厚峰<sup>1,2</sup>, 刘 扬<sup>1,2</sup>, 张龙凯<sup>1</sup>WANG Mengxiang<sup>1,2</sup>, WANG Houfeng<sup>1,2</sup>, LIU Yang<sup>1,2</sup>, ZHANG Longkai<sup>1</sup>

1.北京大学 计算语言研究所, 北京 100871

2.北京大学 计算语言学教育部重点实验室, 北京 100871

1.Institute of Computational Linguistics, Peking University, Beijing 100871, China

2.Key Lab of Computational Linguistics, MoE, Peking University, Beijing 100871, China

WANG Mengxiang, WANG Houfeng, LIU Yang, et al. Study for automatic recognition of Chinese object target role. *Computer Engineering and Applications*, 2014, 50(8): 93-98.

**Abstract:** In the Chinese semantic role labeling system, as a “native product” in the Chinese grammar, “Object Target” role has syntactic and semantic features. This paper discusses the semantic features of “Object Target” and the tangled problems in related semantic role labeling. It puts forward a flow chart for the machine recognition of Chinese “Object Target” role and makes an experimental demonstration.

**Key words:** object target; semantic role; automatic tagging

**摘 要:** 在中文语义角色标注体系中, “对象格”作为汉语语法体系的“自产物”, 有着区别于其他语义角色的句法和语义的特征。讨论“对象格”的语义特征及相关语义角色标注间的纠结现象, 提出了面对自然语言处理的“对象”语义角色标注流程图, 并进行了相关实验验证。

**关键词:** 对象; 语义角色; 自动标注

**文献标志码:** A **中图分类号:** TP393 **doi:** 10.3778/j.issn.1002-8331.1309-0427

先看两组动宾搭配:

(1) 学习雷锋

(2) 学习英语

一般情况下, 大多数的语义角色标注系统都会笼统地把“雷锋”和“英语”标注为“受事”。在对语义角色精细度要求不高的情况下, 这种处理无可非议。但是随着语义分析的深入, 已经不能满足将动词所支配的常规论元仅仅处理为“受事”, 而是需要进一步精细化的划分。发现同样是“学习”所带的宾语, 其语义特征有所区别, 这种区别可以通过形式变换来显现, 比如可以说“向雷锋学习”, 而不能说“向英语学习”。相对于“雷锋”来说, “英语”更倾向典型“受事”, 因为可以用“把”字提前(如可以说“把英语学习好”), 具有“处置性”; 而“雷锋”不具

有“处置性”, 只是动词“学习”的“关涉者”, 受到“学习”的作用很小。根据语义角色精细化的要求, 是有必要把像“雷锋”这样的动作作用的“对象”角色从广义的“受事”中剥离出来, 以示区别。而本文就是探讨“对象”语义角色的语义特征, 并提出面向自然语言处理的对象语义角色自动识别和标注的方法。

## 1 “对象格”语义特征

### 1.1 学界对“对象格”的讨论

“对象格”是汉语语义格体系的提法, 目前汉语学界对“对象格”有两种态度, 一种是不承认有“对象格”这一概念, 如文献[1]。因为目前汉语的语义角色描述体系是来源于西方的 Fillmore<sup>[2]</sup>格语法理论, 而 Fillmore 当初

**基金项目:** 中国第53批博士后基金(No.2013M530456); 国家社科基金重大项目(No.12&ZD227); 国家自然科学基金(No.61370117)。

**作者简介:** 汪梦翔(1984—), 博士后, 研究方向为语言理论和中文信息处理; 王厚峰(1965—), 教授, 研究方向为自然语言处理;

刘扬(1971—), 副教授, 研究方向为语义知识库; 张龙凯(1988—), 博士研究生, 研究方向为自然语言处理。

E-mail: wmx1984@pku.edu.cn

**收稿日期:** 2013-09-27 **修回日期:** 2013-11-12 **文章编号:** 1002-8331(2014)08-0093-06

在创立格语法时没有提到“对象格”。另一种承认汉语是有“对象格”,但是其所指比较混乱,而且与本文所说的“对象格”的内涵和外延有所不同。例如北大中文网库在进行语义角色标注时,对于动词所支配的论元进行划分时,就有“对象”一说;再比如鲁川、林杏光<sup>[3-4]</sup>把客体语义成分分为“受事格、结果格、对象格”三类,并且比较明确地指出了“对象格”和其他客体格的区别,他们认为这种区别主要是能不能用“把”进行句法变换,一般能用“把”变换的就是“受事”或“结果”,一般不能用“把”变换的就是“对象格”。比如:“我打了张三”能转换成“我把张三打了”,所以“张三”为“受事”,而“我喜欢张三”不能转换成把字句,所以“张三”为“对象”。而孟琮在《动词用法词典》<sup>[5]</sup>认为“对象”除“爱、恨、喜欢、讨厌、尊敬”等表示心理状态的动词所带的宾语外,双宾语的指人宾语也属“对象”。这里所说的“对象宾语”与本文所说的“对象格”比较相近,但“对象格”不限于宾语,有的甚至不能作宾语。如:“为人民服务”、“向领导汇报”、“我的青春我做主”等。

## 1.2 本文对“对象格”的判定

单纯从语义上笼统地界定“对象格”这一复杂的类是比较困难的,因为“对象格”所涉及的并不只有单独的一类动词,所以必须借助动词的分类来探讨现代汉语“对象格”的语义标准。本文认为“对象格”指的是动作行为的关涉对象,一般不是动作行为的发出者,也不是动作行为的内容,常和心理动词、交际动词搭配,简单地说,典型的“对象格”是指交际对象和感知对象,语义上具有[+关涉性]、[+致动性]。

所谓“关涉性”,是相对于“处置性”来说的,对象名词并不是动作的直接作用者,不具有处置性,而是参与者或关涉者。例如“借你一支笔”和“喜欢你”中的“你”,几乎都没有受到动词“借”或“喜欢”的直接作用,但是都参与到了这个动作的发生过程中来,是动作所涉及的对象,而非动作的直接处置对象。

所谓“致动性”就是对象的存在可以导致心理动作或者是交际动作的发生。例如“我喜欢你”中的“你”可以理解为,是对象名词“你”的存在导致心理动作“喜欢”的发生;“我告诉他一件事”的“他”可以理解为,是对象名词“他”的存在导致主体要发生“告诉”这个交际行为。

形式上,一般可用“对/向/与……相”等介词来介引,而不能通过“把”提前。例如:

(3)他非常思念故乡→他对故乡非常思念→他把故乡非常思念\*

(4)他要询问下医生→他要向医生询问下→他把医生询问下\*

(5)我要结合具体情况→我要与具体情况相结合→我把具体情况结合\*

(6)他要抱小孩→他要对/向/小孩抱\*→他要把小孩抱一下

在这四个例子中,“思念”是心理动词,“询问”是言语交际动词,“结合”是行为交际动词,都属于交际动词,“抱”是最常见的行为动词。“故乡”、“医生”、“具体情况”都参与了事件的发生,受到动词的影响也不会变化太大,只是动作关涉的对象,它们的存在都导致了动作的发生,形式上可以用“对/向/与……相”等介词来介引,不能通过添加“把”来提前。而第四例中“小孩”的位置变化会随着“抱”这个动作而改变,受动性较强,不能用“对/向/与……相”等介词来介引,但可以用“把”来突显“小孩”的被处置义。按照前文所提供的语义及形式的标准,前三例动词所接的都是“对象格”,第四例中的“小孩”应该属于“受事格”。

当然“关涉性”和“致动性”只是“对象格”相对于其他语义格的语义特征来说的,其实“对象格”内部也有分化,比如感知对象和交际对象比较的话,在“受动性”上就具有一定的差别。感知对象几乎不受动词的直接作用或影响,因此不具有“受动性”,或为[+弱受动性];交际对象常和交际动词搭配,而这类动词必须双方进行互动,所以这类对象格语义范畴可以受动词的影响,有时也不一定受到动词的影响,可以标示为[±受动性]。不过这种“受动性”如果相对于其他类的语义格,比如说对于“受事格”来说,“对象格”的“受动性”还是要弱得多。例如:

(7)我向你询问一个问题。

(8)我非常讨厌你。

(9)我要揍你。

这里有三个“你”,但是其受动性程度和被动词的影响程度是有差别的。第一个句子的“你”是交际类言说动词“询问”的对象,较于第二句心理动词“讨厌”所接的“你”其受动性要略强。但是这两个句子的“你”和第三个句子比起来,其受动性和受影响程度都不如第三句,“我要揍你”的“你”属于受事格,是动词“揍”的直接承受者,直接受“揍”的影响。

## 2 “对象”和“与事”的区别及融合

### 2.1 “对象”和“与事”的区别

“与事”是“事件中有利害关系的间接客体”(《动词大词典》<sup>[4]</sup>解释),一般出现在双宾语句中,如“厂长奖了劳模一千元”中的“劳模”。

汪梦翔<sup>[6]</sup>从语义关系、形式标记、名词语义特征、所涉及动词的类别以及是否受时态影响几个角度对典型“与事格”和典型“对象格”的特征进行了分析:

对象:[+受动]、[±格标必现]、[±省略]、[±生命性]、[±交际类动词]、[-主受动关系受时态制约]

与事:[±受动]、[+格标必现]、[-省略]、[+生命性]、[+交际类动词]、[±主受动关系受时态制约]<sup>①</sup>

发现“与事”和“对象”的这些区别性特征只是标明了差异性,却没有对立的特征,大多数特征都有蕴含或交叉,所以这两种语义角色不容易完全区分。不过它们和“受事”有着较强的对立,那就是“处置性”,一般“受事”是动作行为的直接影响者,是可“处置”的,而“对象”和“与事”只是一个动作发生的关涉者,是不可“处置”的,因此“对象”和“与事”具备融合的可能,可以“打包”,然后和“受事”相区别。

## 2.2 “对象”和“与事”的融合

在现实语义角色标注系统中有把“对象”和“与事”进行分离标注,这种做法是对“对象”的语义所指进行限制的基础上进行的。比如北大中文网库在进行语义角色标注时,动词所支配论元就有“对象”和“与事”。袁毓林<sup>[7]</sup>还作了相应的解释,他认为“与事”是“动作、行为的非主动参与者”,而“对象”是“感知性动作、行为的对象和目标”。很显然这种做法是缩小了“对象”的外延。本文认为不能为了分离而缩小一个概念。如果依托本文对“对象”语义范畴的认定的话,“对象”可以与“与事”进行融合。

首先,关于“对象”和“与事”的争论,它们只是概念的不同,而所指有很多交叉。因为“对象”是汉语语义角色系统的自有产物,它和格语法中的“Dative”有着千丝万缕的关系。格语法体系的创始人 Fillmore,也把像“like、meet”之类的动词的后接论元标为“Dative”。甚至有人就把 Fillmore 的“Dative”等同于“对象”。例如丛琳<sup>[8]</sup>指出:“对象”就是表示由动词所确定的动作或状态所影响的有生物。这一定义和 Fillmore 对“与格(Dative)”基本一样。而汉语中的“与事”取名“参与之意”,分化了一部分“Dative”的特征,一般只是用于带三个论元的动词结构中,其目的是为了凸显和“受事”(Patient)的对立。

其次至于“对象”和“与事”内部语义差别,是双胞胎间的差别。双胞胎肯定是有差别,但是两者的相似之处正好可以区别于其他语义角色。比如“对象”的“关涉性”特征和“与事”的“参与性”特征所指其实一样,只不过“与事”的“参与”在句法表层上表现出一种不可割舍性,省去后句子不能自足;而“对象”的“参与”是自由的,可由透视域自由选择。例如:

- (10) A 我向[小李]授予了“学习标兵”的称号。(与事)  
B 我授予了“学习”标兵的称号。\*

C 我授予小李。\*

(11) A 我向[小李]学习技术。(对象)

B 我学习技术。

C 我学习小李。

(12) A 我对[小李]很尊敬。(对象)

B 我很尊敬。\*

C 我很尊敬小李。

以上例句中,(10)句的“小李”属于“与事”,具有一定的受动性,其核心动词“授予”属于“交接类”动词,或者说是“给”类动词,意义上决定了动作的发生的施事者、给予的事物、还有被给予的“对象”即接受者“小李”都不可或缺。后两句的“小李”属于“对象”,但是“小李”语义成分的去留完全由具体的语境决定或者说由动词决定。(11)中的“小李”显然是“学习”这个动作的“参与者”,它的存在对“学习”这个动作的发生具有导向性,而且“技术”也是一个必有成分,是“学习”的内容,虽然省去后都说得通,但是语义侧重不同。而(12)的“小李”可以说和“尊敬”这个动作无关,但是却不能省略。因此从必有和可有这一特性上,“对象”可以说比“与事”涵盖更广。既不能把所有不能省略的都归入“与事”,也不能把所有能省略的归入到“对象”,但是可以把那些不能省略的“与事”归入到“对象”的范畴。例如(10)中的“小李”不能省略,归入到“对象格”中也基本符合“对象格”的特征,可以认为是“授予”的“对象”,而不是“与事”。

再次,这两种语义角色在同一句中不能共现。理论上,“与事”和“对象”共现,起码得需要带三个论元的动词。那么这三个论元的分配关系,最起码应该有施动论元、直接受动论元和间接受动论元。不管语义关系如何变化,只要一个动作发生,作为一个动作的直接受动者肯定应该标为“受事”,至于另外一个,不管是“对象”还是“与事”,也只能出现一者,根据语言的经济性原则,也只需保留一个。

另外,对于中文信息处理来说,虽然细化语义角色的划分可以更加准确地反映动词和名词间的语义关系,但是要考究这种细化在面对这样两个相似的语义角色时有没有实施的必要。与其通过人为的划分来区分这两种语义角色的不同,还不如扩大“对象”的范畴,这样既实现了对广义“受事”的进一步划分,而且凸显了和典型“受事”的区别性特征,在保证语义角色理想的精细度的同时,又注重了局部语义关系的一致性。

因此本文把“与事”融入到“对象”是基于汉语理论和信息处理的实际需要出发的。

<sup>①</sup>主受动关系有时候会受到时态的制约,比如“我要跟小李离婚”和“我跟小李离婚了”。第一个句子中“我”可以看作是主动提出离婚者,而“小李”可能是被动接受。但是第二句变成完成时,就已经看不出是谁主动了,可以是“我”,也可以是“小李”。这种时态的制约,只对“与事格”有用,而“对象格”和动词的主受动关系没有时态的限制,比如:“我们要请教老师”和“我们已经请教了老师”,这两句不管是什么时态,“老师”都是“被请教”的对象,具有受动性,而“我们”是动作“请教”的发出者,具有主动性。



### 3 对象名词自动识别

#### 3.1 对象名词所涉动词的提取

因为“对象格”语义角色的划分是伴随动词的类别的,所以要想在动名搭配中提取并识别“对象名词”必须依托动词的分类。换句话说,只有心理动词和交际动词才有接“对象名词”的可能,如果不受此类动词支配,基本够不上“对象名词”的“资格”。

事实上语义角色的划分大致也涉及到动词的划分。“对象格”主要涉及到两类动词,一种是心理动词(如,喜欢、热爱、相信等),一种是交际动词(如,汇报、教育、抗议等),当然还有一些其他类别的词,比如“违反”后面也可以接“对象”。

对于心理动词的提取,主要是选取了形式上能加“很”或者“打心眼里”的二元动词,比如“很尊敬老师”、“打心眼里认为这件事不对”。对于交际动词的提取,根据汪梦翔<sup>[6]</sup>的判定,提取了所有三元动词和一些不能加“很”,而前面可以加“向、对、跟”的动词,如“向他递了一包烟”。结合《动词大词典》和《动词用法大词典》对1 540个动词进行了考察,提取出能接“对象”的动词387个,外加自己再查阅《现代汉语词典》(第六版)以及根据中文Propbank的语料,又额外增加了96个,一共483个对象动词。

需要指出的是并不是所有这些动词所接的名词都属于“对象”的语义范畴,同一个动词有可能接“对象”,也有可能接“受事”。比如“学习雷锋”就是“对象”,“学习英语”就是“受事”,因为“雷锋”只是“关涉者”,而“英语”是“学习”的“内容”,具有“处置性”。

不过经过观察发现,只要是心理动词,它们所接的常规宾语基本都可以判定为对象宾语。而有些交际动词所接的常规宾语一般也都可判为对象宾语,如“帮助”后面的名词就应该是“对象名词”。

交际动词虽然具备接“受事”和“对象”的可能,但是一般情况下,“受事”多半是交际的内容,而“对象”多半是交际行为的共同参与者,且一般由人来充当,所以只要是交际动词后带有[+人]或者具有[+生命性]的名词一般可以判定为对象宾语。比如:“教育孩子”,“学习雷锋”中的“孩子”和“雷锋”具有[+人]的特征,因此这些可以看作纳入“交际对象”的语义范畴。而“了解情况”、“解释原因”中虽然“了解”和“解释”可以划入交际动词范畴,但是“情况”和“原因”不具有[+人]的特征,因此都不属于“对象格”的语义范畴。不过有些即使不是人,但是带有人的某些社会属性,或具有拟人性的也可以纳入“对象”。比如“对抗政府”中的“政府”虽然不是具体的人,但是人类社会属性的产物,从某种程度上也认为具有[+人]这一特征。

#### 3.2 对象名词的识别

基于此,在对对象类名词进行识别时,应该从两个

角度来进行:

第一是从动词来着手,需要建立一个对象动词库B,并且要把这类动词分为两类,一类是常规宾语必为“对象”的动词(B1),一类是常规宾语可能为“对象”的动词(B2)。之所以这样分,而不采取心理动词和交际动词的分法,是因为除心理动词外,有些其他类动词所接的宾语也只能是对象宾语,如“教育、违反、帮助、对抗”等。据此,把之前收录的493个动词中的280个动词归入到B1,把另外213个归入到B2。

第二是从名词着手。虽然对于那些必接对象宾语的动词来说不需要考虑名词特征,但是对于有些既可以接“受事”又可以接“对象”的动词来说,需要借助所接名词的特征来进行语义角色的识别。这就需要建立一个带有“[+人]”特征的名词库A,这里的“[+人]”名词必须具有[+生命性]和[+社会性]。比如“学习雷锋”中的“雷锋”具有[+人],标为“对象”;而“学习英语”中的“英语”是[-人],因此不能标为“对象”。其实北大的CCD已经在名词分类上做了相关[+人]或[-人]的属性划分,把其中11 422个带有[+人]特征的词项纳入到了名词库A中。当然这些还不包括人名,很显然人名一般都带有“[+人]”的特征,而且一般是未登录词,为此采取了倪吉、孔芳、朱巧明<sup>[9]</sup>基于可信度模型的中文人名识别方法,来识别人名,程序流程简称“RM”。

不过有时候对于一些特殊的带有领属关系转移的交际动词来说,有时候“受事”也可能是“人”,比如“我过继他一个儿子”。依靠原先思路,计算机可能把“他”和“儿子”都标为“对象”。而一般情况只要不是并列结构一般不可能同时出现两个一样的语义角色,因此采取默认取值的方法,将远离动词的论元默认为“受事”,也即是:把间接宾语(近宾语)标为“对象”,而直接宾语(远宾语)不管是否具有[+人]的特征都不能纳入到“对象”范畴。

另外对于处于非宾语位置的对象名词,可以采取“格标记”和动词特征相结合的策略。一般情况下,对象名词的常规位置是动后,如果位于动词前,则一般会由“对、向、跟、与……相”(K<sub>对象</sub>)等引导,如“我向他请教”。当然并不是所有这些标记引导的名词都可以视作“对象”,如“我向东边走去”中的“东边”就是“方位”而不是“对象”。还必须结合动词来判定。一般情况下,一个名词受到“对、向、跟、与……相”等介引,又受到“对象动词”所支配,那么基本可以认定这个名词的语义角色应该是“对象”。对于那些处于动前又没有格标记引导的,只能根据之前动词和名词的属性来进行识别。比如“我的青春我做主”这样的句子,这里可以先确定动词“做主”是“常规宾语必为对象的动词”,然后利用语块分析(chunking paring)把“做主”所支配的名词为“青春”确定为“对象”。

针对这些特征可以采用如下的方案来对对象宾语

进行自动标注。首先建立带有人性特征描述的名词库A和对象动词库B(该库中动词都是能带对象名词的),下设两个分库B1(该库动词所支配的常规论元必为对象)和B2(该库动词所支配常规论元可能为对象)。然后再确定动词V所支配的NP是否有“K<sub>对象</sub>”(对象格标记)引导,如果有则继续考察动词和词库B的匹配情况。动词如果能够和词库B匹配,则NP直接标为“对象”。如果不能和词库匹配,则进入E(exit),识别结束。如果没有“K<sub>对象</sub>”(对象格标记)引导,则将动词V与B库进行匹配,如果匹配失败则进入E(exit),识别结束。如果与B1匹配成功,则NP直接标为“对象”。如果与B2匹配成功,需让NP先进入RM进行人名识别,如果是人,则NP标为对象,如果不能识别,则进入名词库A匹配,如果匹配成功,则输出“对象”,如果匹配不成功,则进入E(exit),识别结束。对于输出的结果,如果遇到两个NP都标为“对象”,则自动默认就近NP。如图1。

接下来以三个短语“向领导请教问题”(P1)、“尊敬老师”(P2)、“告诉你一件事”(P3)为例,进行具体识别程序的说明:首先观察组块中有没有对象格标记引导的名词,P1有“向”,P2和P3没有,所以对于P1来说,直接进行动词匹配就可以完成“对象”的标注,“请教”与动词库

B匹配成功,则“向”后的名词“领导”标为“对象”。而P2和P3需继续匹配动词,“尊敬”属于B1,则所支配名词“老师”直接标为“对象”。“告诉”属于B2,此时还需要考察所支配的NP的特征,这时“告诉”后接两个NP,其中“你”是“人”,而“事”不是“人”,所以只有“你”标注为对象。如图2所示。

4 实验论证及分析

为检验该标注方法的有效性,选取了1998年的人民日报作为实验数据。因为目前对同形异义的动词区分不是很成熟,为方便匹配,将动词库中具有不同义项但是具有同一形态的动词进行去重,B库得到466个动词,B1库为279个动词,B2为187个动词。与1998年的人民日报匹配后发现,有421个动词涵盖,总的出现次数为37 850次,其中B1库动词为14 278次,B2库中的动词为23 572次。

随机提取了含有B1库和B2库动词的句子,各500句。然后利用上文提到的标注方法,对受这些句子动词作用的NP进行语义角色标注,然后利用人工对自动标注的结果进行了检验,具体结果如表1所示。

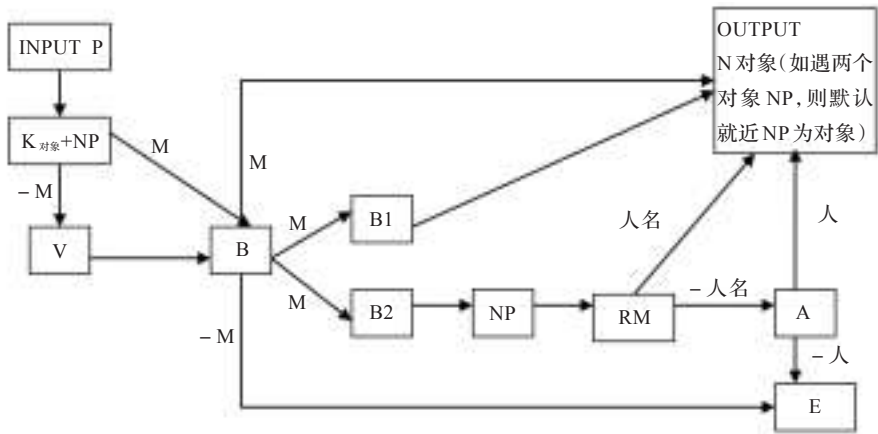


图1 “对象格”识别方案

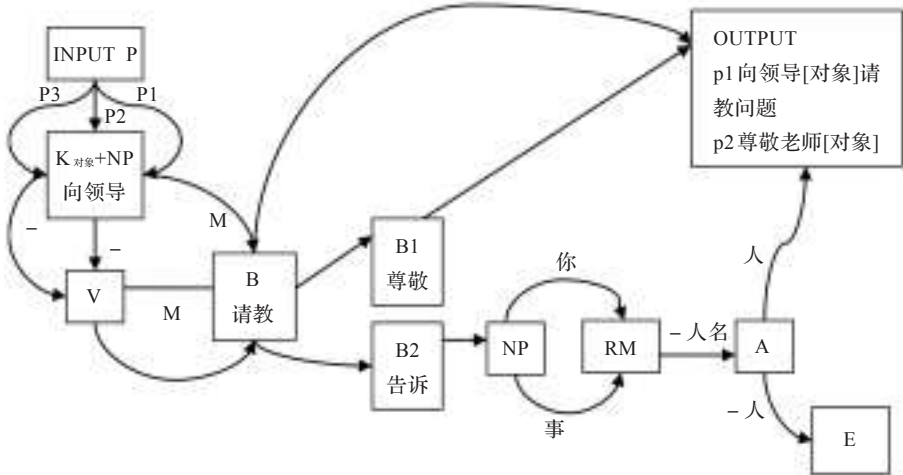


图2 “对象格”识别实验流程图

表1 基于生语料的标注统计

动词库	句数	标注数	标注率/(%)	准确标注数	正确率/(%)
B1	500	334	66.8	308	92.2
B2	500	292	58.4	247	84.6

之所以出现自动标注率不高的情况,主要是在判断NP是否和“对象动词”有关联上出现了问题。一方面,有些“对象动词”无NP可接或所接NP可以省略,如“不管未来如何,我们始终坚持”,这里即使找出了对象动词“坚持”,也没法找到它所接的“对象名词”。另一方面,在有些句子中,虽然动词能够判断是否属于对象动词,但是很多时候动词后面常接VP结构或并列NP结构,难以确定哪个和动词V发生真正的关系,如“坚持解放思想”,这里的“思想”计算机可能就理解为受“解放”支配,而不受“坚持”支配。再比如,“我们要坚持经济建设”,这里自动标注就默认把“经济”标为“对象”了。基于此,应该借助依存树库的标注技术来对NP和对象动词的关系进行进一步分析。选取了哈工大的依存树库作为平台,随机提取了含有B1库和B2库动词的句子,各500句,发现在依存书库中,实际能找到的被“对象动词”所支配且进行标注成功的“对象名词”有75%左右,这比在实际语料中提取的结果有所提高。理论上B1或B2库中的动词为及物动词,应该都有名词支配,但在实际语料中,及物动词不一定都支配有NP,或者NP省略。不过从人工对这些标注成功的句子来看,正确率有相应有所提高,如表2所示。

表2 基于依存树库的标注统计

动词库	句数	标注数	标注率/(%)	准确标注数	正确率/(%)
B1	500	375	75.0	348	92.8
B2	500	372	74.4	324	87.1

另外,B2动词库中“对象”的标注正确率相对于B1动词库来说偏低,这主要还是在“NP”的属性判定上出现了问题,一方面来源于名词库A还是不够大,很多“属人”名词不能和人库A匹配,另一方面就是很多名词虽然不带有“人性”,但是在使用可能隐喻“人”的含义,这些都给成功标注造成了影响。

5 余论

本文根据中文信息处理学界对语义角色精细化的要求,从广义的“受事”中剥离出“对象”这一语义角色,而且又和“与事”进行了整合,其目的是为了避免标注的混乱、保持语义特征的一致性,同时保证了和典型的“受

事”角色的区别性特征,并探讨了机器识别“对象”名词的具体方法。这一研究对将来计算机进一步深入理解和区分动名搭配结构打下了一定基础,但这一研究仅为一个初步的探索,还有很多值得深入研究的地方。比如本文所提取动词主要是基于《动词用法词典》,虽然占到了词典总量的近3成,但是对于整个动词的总量来说还是很少,因此肯定还有一些未收录的对象动词,以后还必须对对象动词库进行相关规模的扩大。

此外,在名词库中对“人”的属性的判断,是基于CCD中名词的概念分类,在实际的语料中,还有一些既非人名又具有隐喻特征的名词还无法识别。比如,“你可以学习斑竹”,这里“斑竹”(论坛的版主),带有“人”的特征,但这种情况利用此方法计算机目前还不能识别。因此,今后还需要在名词属性自动识别和对象动词的外围扩展上继续深入展开。

参考文献:

[1] 贾彦德.对汉语语义格的认知和划分[J].语文研究,1997(3).  
[2] Fillmore C J.“格”辨.[M].胡明扬,译.北京:商务印书馆,2002.  
[3] 鲁川,林杏光.现代汉语的格关系[J].汉语学习,1989(5).  
[4] 鲁川.动词大词典[M].北京:中国物资出版社,1994.  
[5] 孟琮,郑怀德,孟庆海,等.汉语动词用法词典[M].北京:商务印书馆,1987.  
[6] 汪梦翔.对象格语义范畴及其相关语法理论研究[D].武汉:华中师范大学,2012.  
[7] 袁毓林.语义角色的精细等级及其在信息处理中的应用[J].中文信息学报,2007(4).  
[8] 丛琳.“给+NP”中NP的语义范畴[J].北京教育学院学报,2001(9).  
[9] 倪吉,孔芳,朱巧明,等.基于可信度模型的中文人名识别研究[J].中文信息学报,2011(3).  
[10] 陈昌来.论语义结构中的与事[J].语文研究,1998(2).  
[11] 丁伟伟,常宝宝.基于语义组块分析的汉语语义角色标注[J].中文信息学报,2009(5).  
[12] 林杏光.进一步深入研究现代汉语格关系[J].汉语学习,1993(5).  
[13] 林杏光,王玲玲,孙德金.现代汉语动词大词典[M].北京:北京语言学院出版社,1994.  
[14] 袁毓林.论元角色的层级关系和语义特征[J].世界汉语教学,2002(3).  
[15] 袁毓林.一套汉语动词论元角色的语法指标[J].世界汉语教学,2003(3).