

副词语义指向自动识别的路径探讨和个案分析

赫琳 吴迪

[摘要] 研究现代汉语副词语义指向的计算机识别,是从语义入手探索句处理的一个尝试。我们提出的基本思路是:充分地利用海量语料,详尽考察现代汉语副词语义指向的不同情况,探寻不同语义指向的约束条件和机制,并对其进行形式化的描述,然后构建计算机自动处理策略和流程模型,编写程序,使之在计算机上得以实现。据此,我们以否定副词“不”为样本,全面探讨了其语义指向的各种情况,构建了计算机自动识别的流程图。

[关键词] 副词;语义指向;计算机识别;不

[中图分类号] H08 [文献标识码] A [文章编号] 1671-881X(2009)04-0459-05

一、要研究的问题

在自然语言信息处理中,当前遇到的瓶颈是句处理的问题。所谓“句处理”,就是怎么让计算机处理、理解自然语言中一个句子的意义,怎么让计算机生成一个符合自然语言规则的句子。制约句处理的重大难题就是要解决语义问题,这也是自然语言处理的关键。句子的语义问题涉及两个层面:一是句子各成分本身的意义,另一个是句子各成分组合所产生的意义。句处理必须对句子的语义问题进行分层处理。既要弄清各成分本身的意义,又要弄清成分与成分组合所产生的种种意义。一般来说,成分本身的意义相对简单、明确,成分的组合意义十分复杂,而且成分组合意义的复杂性更导致了句处理的复杂性和艰难性。本文拟探讨的语义指向问题就是句子成分组合中的复杂问题,是句处理不可回避的重要问题。

语义指向是指句中某个成分在语义上跟哪个成分直接相关。某成分语义指向的不同,直接影响句子的意义。例如:

饭我们都吃了。

例子中的“都”在意义上既可能指向“饭”,也可能指向“我们”。如果指向“饭”,句义是:饭被我们吃光了;如果指向“我们”,句义是:我们都吃了饭了。显然,同一成分的语义指向不同,句义完全不同。如果不能正确判断语义指向,就会误解句义。而语义指向在句法上往往没有明显的形式标记,这就给语言理解带来一定的困难,尤其是给计算机的自动识别带来障碍。要解决自然语言的句处理问题,必须突破语义指向的自动识别问题。

自从20世纪80年代语义指向分析进入语言研究领域以来,我国语法学者从不同角度对语义指向进行了研究。总的来说,主要包括两个方面:其一,运用语义指向分析法对语言事实做具体的分析,探寻其句法和语义上的对应规律。有的着眼于句子成分,考察了补语、状语、定语等句法成分的语义指向问题,有的从词类出发,考察了副词、形容词、代词、名词、动词以及介词结构等的语义指向问题。其二,从

理论上对语义指向本身进行探讨,说明其产生的背景、性质、内涵、范围、表现形式以及在语法分析中的作用等。总的来看,语义指向研究虽然涉及的内容较为广泛,但较多的是现象分析,较少归纳规律;较多的是个别分析,较少系统考察;面向应用的更少,几乎还没有见到针对自然语言计算机处理的语义指向研究。总之,我们对语义指向问题的认识还十分有限,尤其是对语义指向的规律缺乏把握,对语义指向的形式标记捕捉不多,加上已有的研究立足点大都不在应用,没有为计算机处理语义指向储备好充足的知识,所取得的成果与语言信息处理的要求之间存在着很大的距离。因此,要想实现计算机对语义指向的自动识别,还有大量工作要做。

我们拟在语义指向研究既有成果的基础上,紧紧围绕“为语言信息处理所用”这一目标,进一步深化、细化语义指向研究,以现代汉语副词语义指向为样本来解剖麻雀,具体探讨现代汉语副词不同语义指向出现的条件和可以捕捉的形式标记,描绘现代汉语副词语义指向网,建构副词语义指向的流程图,从而为副词语义指向计算机自动识别策略的制定和程序编写,提供依据甚至蓝本。这无疑对于实现语义指向的计算机自动处理乃至整个基于语义的自然语言处理都具有重要意义。

二、研究思路

要实现现代汉语副词语义指向的计算机自动识别,必须摸清现代汉语副词不同语义指向出现的条件和可以捕捉的形式标记,并把它们形式化。因此,我们首先要做的工作,就是最充分地利用海量语料,详尽考察现代汉语副词语义指向的不同情况,探寻现代汉语副词不同语义指向的约束条件和机制,并对其形式化的描述,然后构建计算机自动处理策略和流程模型。

我们知道,现代汉语中副词主要做状语,在语义上可能前指,也可能后指;既可以指向相邻成分,也可以指向间隔成分,在不同的情况下具体指向不同。本研究就是要找出不同指向出现的条件,并在此基础上编制流程图,使计算机根据有关的条件执行相应的命令,做出相应的判断,从而使整个系统成为一个可以自动运作的过程。

现代汉语副词可分为程度副词、时间副词、范围副词、否定副词、语气副词、频率副词等。其中,时间副词语义指向单一,流程简单;范围副词、否定副词、程度副词、频率副词情况较复杂,编制流程图需要具体详尽的条件限制。即使是同类副词,不同的词具体运行情况也不相同,所以我们拟对副词进行分类讨论,对同类中的特殊副词进行个别讨论。具体包括程度副词和时间副词语义指向的计算机识别,范围副词“都”、“只”语义指向的计算机识别,否定副词“不”、“没(有)”、“别”语义指向的计算机识别和频率副词“又”、“也”语义指向的计算机识别,等等。

确定了副词的类别之后,我们将从具体的副词入手,以具体副词为关键词,利用大型语料库和巨量生语料查找全部用例,考察其上下文,确定其语义指向,探索其语义指向的条件和标记,然后将这些条件和标记系统化,抽象其规律,找出其相互制约的因素,根据这些因素起作用的程度和先后顺序编制流程图。

三、个案分析

下面我们以否定副词“不”^①为样本,全面探讨其语义指向的各种情况,构建计算机自动识别的流程图。

现代汉语研究中,关于否定的成果颇丰。有的是从否定辖域的角度进行,有的是从否定焦点的方面着手。但不管是否定辖域还是否定焦点,研究的都是否定词到底否定的是什么成分,其实也就是否定副词的语义指向问题。

在否定副词“不”指向的方向性这一问题上,学者们的研究结论基本一致,即指后。但是,“不”否定的究竟是其后哪一个或哪几个成分,各家看法不一,归纳起来有三种观点。

第一种,指向“不”后的全部成分。持此种观点的是吕叔湘先生。

第二种,指向焦点。持此种观点的有徐杰、李英哲、沈家煊、刘顺等。

第三种,指向毗邻成分,即紧邻的成分。持此种观点的有钱敏汝、李宝伦、潘海华等。

由上述可知,关于副词“不”的语义指向还有一些问题需要进一步研究。为了弄清副词“不”的语义指向的全面情况,我们利用北京大学现代汉语语料库,详尽考察了“不”的语义指向状况。根据我们的考察和研究,含有否定副词“不”的句子分为两种情况:一种是“不”后出现带有“的”或“地”的修饰语,另一种是“不”后没有出现带“的”或“地”的修饰语。我们发现,“不”的语义指向与“的”、“地”的有不具有明显的相关性。这条规律为我们解决副词“不”语义指向的计算机自动识别问题提供了重要线索。为此,我们对“不”的语义指向与“的”、“地”的具体相关性进行了详细的考察,现分述于下。

(一)“不”后有带“的”或“地”的修饰语

据我们考察,“不”后有带“的”或“地”的修饰语时,“不”就否定紧随其后的修饰语。究其原因,主要是因为定语带“的”、状语带“地”就表明定语和状语与被修饰语之间结合不甚紧密,“的”、“地”起到了间隔修饰语与被修饰语的作用,所以“不”只能否定紧随其后的成分,而不能否定被“的”、“地”隔开的成分。

1.“不”后出现带“的”的修饰语,也就是“不”出现在定语中,这时“不”指向紧邻其后的成分。例如:
他喝了不新鲜的牛奶。

“不”否定的是“新鲜”。诚如钱敏汝所说,定中结构中否定载体的否定范围不会超过定语。

2.“不”后出现带“地”的修饰语,这时“不”指向紧邻其后的成分。例如:

她不高兴地^②瞪了他一眼。

这一类修饰语往往是对施动者的情态进行描写。其后的动词一般不是单纯形式,而是“动+了”、动结式、动趋式、动词重叠式(“VV”和“V了V”)、“动+着”等。例如:

鲍里斯·别尔曼在与她谈话时,不高兴地提到了格利戈里·阿尔库斯。(动+了+宾)

我很不高兴地走开了。(动结式)

不高兴地走了出去。(动趋式)

不高兴地故意清清喉咙。(VV)

骑兵们不高兴地抿了抿嘴。(V了V)

他不高兴地嘀咕着。(动+着)

这些动词代表的动作往往都是有终结点的,也就是说,意义上是有界的。在“不”和“没”的否定分工上,“不”否定的是无界成分,“没”否定的是有界成分。因此,上述各动词是无法用“不”来否定的,所以“不”在这里否定的只能是修饰语。

(二)“不”后没有带“的”或“地”的修饰语

“不”可以出现在状语、补语、主谓谓语句的谓语中,“不”后面可以是一个成分,也可以有多个成分。根据我们的考察,只要“不”后未出现带“的”或“地”的修饰语,“不”就能否定其后所有的成分。究其原因,主要是因为在这类格式中,“不”后的所有成分结合紧密,构成一个整体,所以“不”只能否定其整体,而不能只否定其部分成分。

1.“不”后只有一个成分。

这咖啡不香。 小明跑得不快。

2.“不”后有两个成分。

(1)“不”+状语+动/形
这句话不太懂。 小明跑得不很快。

(2)“不”+动+宾

爷爷不吃面食。 张老师不是我们的数学老师。

3.“不”后有三个成分。

(1)“不”+状语+动+宾
我不在学校读小说 她不天天吃食堂

(2)“不”+动+宾 1+宾 2

张老师不教我们数学。 我就不告诉你那件事。

(3)“不”+动 1+兼语+动 2

大家不选老王当组长。

4.“不”后有四个成分。

“不”+状语+动+宾 1+宾 2

张老师不只教我们数学。

上面讨论的例子的谓语一般只有一个动词或形容词。连动句中能用“不”否定的有三种情况:

(1)后一个动词/短语表示的动作行为是前一动词/短语表示的动作的目的。例如:

她不去开门。

(2)前一个动词/短语表示后一个动词/短语所表示的动作的方式(或手段、工具)。例如:

他不骑车上班。

(3)前一个部分表示肯定的意思,后一个部分表示否定的意思,但前后两个部分表示的意思一样,其中后一个部分用“不”来构成否定。

孩子紧紧拉着她的手不放。

显然,当“不”后没有带“的”或“地”的修饰语时,“不”指向其后所有的成分。

综上所述,当“不”后有带“的”或“地”的修饰语时,“不”指向修饰语,也就是其后毗邻成分;“不”后没有带“的”或“地”的修饰语时,“不”指向其后全部成分。“的”和“地”成为区分“不”语义指向的标志词。计算机识别就可以以“的”和“地”为切入点,首先看“不”后是否有助词“的”,如果有,“不”指向其后毗邻成分;如果没有,再看“不”后是否有助词“地”,如果有,“不”同样指向毗邻成分;如果没有,“不”指向其后所有成分。依据这个顺序,我们可以画出否定副词“不”的语义指向流程图(如图 1)。

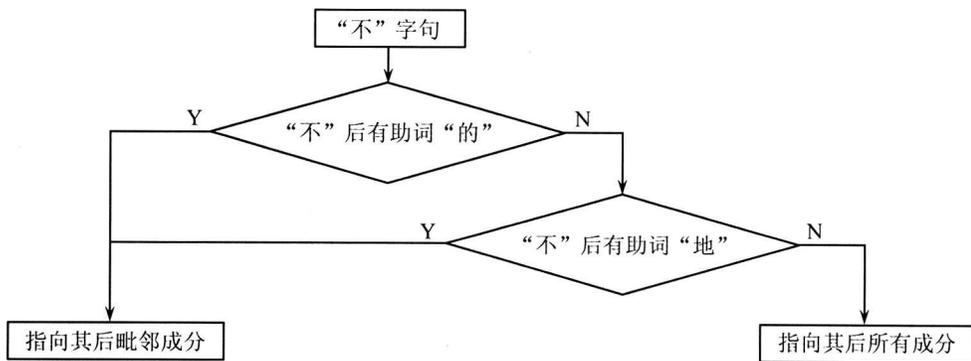


图 1 副词“不”的语义指向流程图

计算机专家们可以根据图 1 编写计算机自动识别副词“不”语义指向的程序。

本文探讨了现代汉语副词语义指向的计算机识别问题,重点讨论了副词“不”的语义指向。语义指向的计算机识别研究,为我们提供了一个新的视角来反观现代汉语中的一些语言现象。所以本文的研究成果可以直接用于语言学领域,以利用副词语义指向的研究思路和方法来研究其他词类,从计算机识别这一新的角度重新探讨各种词类的语义指向,从而对语义指向的性质、内涵、对象和范围等有一个全新的思考和定位。同时,我们的研究成果也可以直接用于语言信息处理领域,计算语言学家们可以将我们已经形式化的条件表示为算法,使之在计算上形式化,根据流程图编写程序,使我们的研究成果最终在计算机上得以实现。

注 释:

- ① 我们在北京大学现代汉语语料库中输入关键词“不”,得到语料约10万条,人工删除不合格语料,获得可用语料约1万条。
- ② 我们在北京大学语料库中搜到的带“不高兴地”和“不满意地”的语料共301条,我们对其后出现的动词及动词词组进行了穷尽性考察,尚未发现例外。

[参 考 文 献]

- [1] 陆俭明:《关于句处理中所要考虑的语义问题》,载《语言研究》2001年第1期。
- [2] 税昌锡:《语义指向分析的发展历程与研究展望》,载《语言教学与研究》2004年第1期。
- [3] 吕叔湘:《疑问 否定 肯定》,载《中国语文》1985年第4期。
- [4] 徐杰、李英哲:《焦点和两个非线性语法范畴:“否定”“疑问”》,载《中国语文》1993年第2期。
- [5] 沈家煊:《不对称和标记论》,南昌:江西教育出版社1999年版。
- [6] 刘 顺:《现代汉语的否定焦点和疑问焦点》,载《齐齐哈尔大学学报(哲社版)》2003年第2期。
- [7] 钱敏汝:《否定载体“不”的语义——语法考察》,载《中国语文》1990年第1期。
- [8] 胡建华:《否定、焦点与辖域》,载《中国语文》2007年第2期。
- [9] 石毓智:《肯定和否定的对称与不对称》,北京:北京语言文化大学出版社2001年版。

(责任编辑 何坤翁)

Computer Identification of the Adverb Semantic Orientation: Path & Case

He Lin¹, Wu Di²

(1. School of Chinese Language & Literature, Wuhan University, Wuhan 430072, Hubei, China;

2. Basic Department, Military Economic College)

Abstract: The recognition of the semantic orientation of the adverb on the computer is a new temptation to discuss sentence processing starting from semantic. We propose basic idea: taking advantage of language materials, checking all variants of semantic orientation of the modern mandarin, exploring and describing formally the conditions and mechanics for different semantic orientation, and then constructing auto-process strategy and procedure diagram, so that the recognition can be realized by computer. According to theory above, we choose *Bu*(不) as an example to examine all cases of semantic orientation and get the corresponding programming.

Key words: adverb; semantic orientation; computer identification; *Bu*(不)