



解放军外国语学院博士文库项目

ИССЛЕДОВАНИЯ ПО ПРОБЛЕМЕ ПОТЕНЦИАЛЬНЫХ
НЕОДНОЗНАЧНОСТЕЙ РУССКОГО ЯЗЫКА ПОД УГЛОМ
ЗРЕНИЯ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ

计算语言学视野下的 俄语潜在歧义研究

◎张禄彭 著



中国出版集团



世界图书出版公司



解放军外国语学院博士文库项目

ИССЛЕДОВАНИЯ ПО ПРОБЛЕМЕ ПОТЕНЦИАЛЬНЫХ
НЕОДНОЗНАЧНОСТЕЙ РУССКОГО ЯЗЫКА ПОД УГЛОМ
ЗРЕНИЯ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ

计算语言学视野下的 俄语潜在歧义研究

◎张禄彭 著



中国出版集团



世界图书出版公司

图书在版编目(CIP)数据

计算语言学视野下的俄语潜在歧义研究/张禄彭
著. —广州:世界图书出版广东有限公司, 2012.11
ISBN 978-7-5100-5360-3

I. ①计… II. ①张… III. ①计算语言学—应
用—俄语—语义学—研究 IV. ①H353

中国版本图书馆CIP数据核字(2012)第241614号

计算语言学视野下的俄语潜在歧义研究

策划编辑:刘正武

责任编辑:程 静

出版发行:世界图书出版广东有限公司

(广州市新港西路大江冲25号 邮编:510300)

电 话:020-84451969 84459539

http://www.gdst.com.cn E-mail: pub@gdst.com.cn

经 销:各地新华书店

印 刷:广州天鑫源印刷有限责任公司

版 次:2012年11月第1版 2012年11月第1次印刷

开 本:880mm×1230mm 1/32

字 数:250千

印 张:9.75

ISBN 978-7-5100-5360-3/TP·0016

定 价:35.00元

版权所有 侵权必究

咨询、投稿:020-84460251 gzlzw@126.com

序 言

“潜在歧义论”(Potential Ambiguity Theory, 简称PA论)是我在1986年研究科技术语自动剖析时提出的一种理论。张禄彭博士在他的专著《计算语言学视野下的俄语潜在歧义研究》中运用这种理论来分析俄语的潜在歧义,进一步发展了这种理论,我感到很高兴。现在,这本著作就要出版了,张禄彭博士要我写个序言,我欣然同意了。

1985年,中国科学院软件研究所所长许孔时教授邀请马希文(北京大学计算机系教授)和我担任该所的兼职研究员,以加强该所语言信息处理的研究。1986年,中国科学院与德国签订协定,要派一个学者到德国合作进行科技术语的自动处理研究,由于这项研究涉及多种语言,需要派出的中方学者不仅要懂得英语,还要懂得德语和法语。我是法国留学归来的,在法国留学期间,我曾经研制过汉-法/英/日/俄/德多语言机器翻译系统,懂得德语和法语,并且做过德语和法语的自动生成,我的条件正好符合协定的要求。于是,许孔时所长征得中国科学院的同意,决定派我到德国去进行这项合作研究。

当时我是中国社会科学院语言文字应用研究所的正式研究人员,在软件所只是兼职,因此,许孔时所长又亲自到语言文字应用研究所给领导做工作,经过多方解释,终于取得了语言文字应用研究所领导的应允。这样,我就可以理直气壮地到德国进行合作研究了。

于是,我在1986年至1988年在德国夫琅禾费研究院(Fraunhofer Gesellschaft, 简称FhG)新信息技术与通讯系统研究所担任客座研

究员，专门研究科技术语数据库和科技术语的自动分析问题。

我在夫琅禾费研究院的VAX 11/750计算机上，使用INGRES关系数据库，建立了汉语术语数据库GLOT-C，并在GLOT-C与夫琅禾费研究院的多语言术语数据库GLOT之间，建立了多语言术语之间的映射关系。

在我建立的汉语术语数据库GLOT-C中，汉语科技术语只有一小部分是单词型术语，如“程序，算法，流程”等，而大部分都是词组型术语，词组型术语可以由两个词构成，如“程序/设计”，或者由三个词构成，如“数字/字符/子集”，或者由四个词构成，如“条件/控制/转移/指令”，或者由五个词构成，如“平均/无/故障/工作/时间”，或者由六个词构成，如“四/分/之/一/平方/乘法器”。为了解释这些词组型术语的内在结构规律，我决定对这些词组性术语进行结构自动分析，从而为汉语科技术语的规范化和新术语的命名，在语言学上提供理论根据，使汉语科技术语的研究工作与汉语语法和语义的研究工作更加紧密地结合起来。

我的老师朱德熙教授生前在讨论汉语的特点的时候指出：“如果我们把各类词组的结构都足够详细地描述清楚了，那么句子的结构实际上也就描述清楚了。因为句子不过是独立的词组而已。”可见，要解决汉语句子的自动句法分析这个大问题，可以首先从汉语句子的自动句法分析入手，而要解决汉语句子的自动句法分析问题，首先要从汉语词组的自动分析入手。这是我研究汉语词组自动分析的初衷。

汉语的科技术语绝大部分是词组型术语，这些词组型的科技术语，其结构一般比较严谨，其含义一般比较单纯，它们在一定程度上反映了汉语词组结构的规律。根据朱德熙教授的理论，我相信，如果我把汉语词组型科技术语的结构描述清楚了，也就有可能把汉

语的词组结构描述清楚了，并进一步把汉语句子的结构也描述清楚了。正是基于这样的信念，我试图从汉语词组型科技术语的自动分析研究中，找到解决汉语句子自动分析问题的钥匙。

这样，我便在夫琅禾费研究院的 VAX 11/750 计算机上，开始了汉语词组型科技术语的自动分析研究。

科技术语一般都符合单一性的要求，因此，研究开始时，我幼稚地以为，在对汉语词组性科技术语进行自动分析时，不存在歧义 (ambiguity) 方面的困难。哪知，实际情况与我的估计大相径庭。在自动分析术语结构时，我才发现，汉语词组型科技术语中，也存在着大量的歧义问题。

例如，“分割/字符”这个词组型术语，其中的“字符”这个名词，在词汇意义上，可以被分割，也可以具有“分割”这种特性，因此，“分割/字符”可以解释为“分割”某一个“字符”，在句法功能上是“述语+宾语”的“述宾式”，也可以解释为某一个被“分割”的“字符”，在句法功能上是“定语+中心语”的“定中式”。这样，从计算机处理自然语言的角度看来，“分割/字符”就是一个兼具“述宾式”和“定中式”两种结构的歧义术语了。把这两个词组型术语进一步抽象为 V+N 的结构，我们就可以说，V+N 这个结构是一个歧义结构。

进一步分析 V+N 这个结构，我还发现更多有趣的现象。在词组型术语“响应/时间”中，“响应”是动词，标注为 V，“时间”是名词，标注为 N，这也是一个形式为 V+N 的词组类型结构；在词组型术语“取/比例尺”中，“取”是动词，标注为 V，“比例尺”是名词，标注为 N，这也是一个形式为 V+N 的词组类型结构，这两个词组型术语的词组类型结构相同；可是，在句法功能上，“响应/时间”是“定语+中心语”，属于“定中式”，“取/比例尺”是“述语+宾语”，

属于“述宾式”；“响应/时间”和“取/比例尺”的词组类型结构相同，而它们句法功能结构却大不相同。就是 V+N 的句法功能结构被判断为“述宾式”之后，这个句法功能结构的逻辑语义结构还可能不同。述宾式的 V+N 可以解释为“谓词 + 受事者”（“取/比例尺”），又可以解释为“谓词 + 施事者”（“跑/带”），又可以解释为“谓词 + 结果”（“印/字”），又可以解释为“谓词 + 目的”（“归/零”），又可以解释为“谓词 + 方向”（“面向/问题”）。可见，术语的句法功能结构与逻辑语义结构之间也不存在一一对应关系。

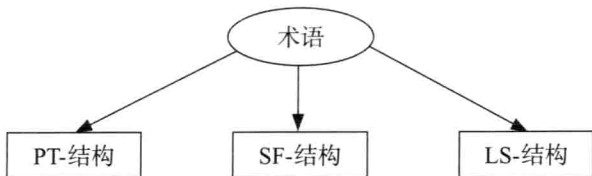
这些现象说明，词组型术语并不像我原来估计的那样单纯，词组型术语仍然有着日常的自然语言中那样复杂的歧义问题。因此，我决定从词组类型结构、句法功能结构和逻辑语义结构三个方面来研究词组型术语的歧义问题。

根据 Chomsky 的上下文无关语法 (Context-Free Grammar, 简称 CFG) 中的 Chomsky 范式 (Chomsky Normal Form, 简称 CNF)，我采用二叉单标记树形图来表示汉语词组型术语的结构。这种由许多层二叉树枝构成的树形图，是以二叉树枝作为其结构的基本单元。二叉树枝上的两个相邻结点的词类或词组类型组成的结构，叫做术语的词组类型结构 (Phrase Type Structure, 简称 PT-结构)。树形图中某一层级的两个相邻树枝结点上的句法功能信息，叫做术语的句法功能结构 (Syntactical Functional Structure, 简称 SF-结构)。树形图中某一层级的子树中两个相邻树枝结点的逻辑语义信息，叫做术语的逻辑语义结构 (Logic-Semantic Structure, 简称 LS-结构)。

任何术语都包括 PT-结构、SF-结构和 LS-结构这三种层次各异的结构，它们之间的相互作用，决定了术语的字面含义的基本内容。我们常常可以对术语的含义做出“望文生义”或者“顾名思义”的解释，正是这三种结构在我们头脑中相互作用的结果。因此，我们用

严格的科学方法来分析这三种不同的结构，就有可能揭示这种“望文生义”或“顾名思义”现象的某些实质，从而对术语的字面含义做出科学的解释。

术语的 PT-结构、SF-结构以及 LS-结构之间的关系可表示如下：



如果我们能够根据术语的 PT-结构，通过有穷步骤，自动地推算出术语的 SF-结构，并进而推算出术语的 LS-结构，那么，就可以做到词组型术语的自动分析（Parsing）。

然而，对于汉语来说，这是一个颇为复杂和相当困难的研究课题。

汉语术语的特点是，这三个结构之间，在绝大多数情况下，不存在一一对应关系。同样的 PT-结构，可以解释为不同的若干个 SF-结构；同样的 SF-结构，又可以解释为不同的若干个 LS-结构。

朱德熙教授在《汉语句法中的歧义现象》一文中，提出了“歧义格式”这个概念来从理论上概括汉语中歧义结构的类型。他认为，句子的歧义“是代表了这些句子的抽象的‘句式’所固有的”^①，因此，他主张用“歧义格式”来概括汉语中的同形歧义结构。

朱德熙教授的这种见解是很有价值的，因为语言中的任何一个有结构歧义的形式，都不是孤零零地存在的，它往往代表具有某种格式的许许多多形式。抓住歧义格式是研究歧义的必要途径。

^① 朱德熙，《汉语句法中的歧义现象》，载《现代汉语语法研究》，1980年，171页，商务印书馆。

但是，朱德熙教授的关于“歧义格式”的见解，还有不完全之处。我们在词组型术语的自动分析中发现，“歧义格式”所反映的类别的歧义，在具体的术语中有时存在，有时并不存在。当我们把具体的单词代真到歧义格式中的范畴符号（也就是类别符号）中，而使歧义格式变为具体的词组型术语的时候，有的词组型术语中仍然可以保持歧义格式原有的歧义，而有的词组型术语中，歧义格式原有的歧义却消失得无影无踪了。

例如，根据我们前面的分析，在汉语词组型术语中，V+N这个结构是一个“歧义格式”，当我们把其中的范畴符号代真为“分割/字符”的时候，其句法功能可以解释为“述宾式”，也可以解释为“定中式”，具有“述宾一定中”歧义，确实是有歧义的。可是，当我们把其中的范畴符号代真为“取/比例尺”的时候，其句法功能只能解释为“述宾式”，不能解释为“定中式”；当我们把其中的范畴符号代真为“响应/时间”的时候，其句法功能只能解释为“定中式”，不能解释为“述宾式”。在这两种情况下，“歧义格式”V+N中原有的歧义完全消失了。

“歧义格式”竟然不再存在“歧义”，岂非咄咄怪事！

这是我在汉语词组型术语自动分析中一个重要的发现。这个发现意味着，朱德熙教授的关于“歧义格式”的见解并不是无可非议的“不刊之论”，这种见解难以解释我在汉语词组型术语自动分析中发现的“歧义格式”不再存在“歧义”的这种普遍存在的现象。事实上，朱德熙教授的“歧义格式”只具有歧义的可能性，并不一定具有歧义的现实性，因此，“歧义格式”这个名称也是不尽恰当的。

根据我在汉语词组型术语自动分析中的这种发现，我又进一步来研究英语中最常见的、公认的“歧义格式”VP + NP1 + Prep + NP2。

在这个“歧义格式”中，当我们把VP代真为saw，把NP2代真为a boy，把Prep代真为with，把NP1代真为a telescope时，得到的“saw a boy with a telescope”是有歧义的，其意思可以是“看见一个戴着望远镜的男孩”，也可以是“用望远镜看一个男孩”。

可是，如果我们把VP，NP1，Prep，NP2等范畴符号代真为别的单词或词组的时候，这个“歧义格式”中的歧义却消失了。请看如下的例子：

She sent the ticket to New York (1)

(她把票寄到纽约)

She lost the ticket to New York (2)

(她把到纽约的票丢失了)

He cooks dinner for the children (3)

(他为孩子们做饭)

The company sells toys for children (4)

(这家公司出售儿童玩具)

在(1)中，动词sent表示传送，具有趋向性，介词词组to New York作它的状语，不作名词词组the ticket的定语，“歧义格式”中的歧义消失了；在(2)中，动词lost表示丧失，不具有趋向性，介词词组to New York作名词词组the ticket的定语，不作动词lost的状语，“歧义格式”中的歧义也消失了；同样地，在(3)中，介词词组for the children作动词cooks的状语，表示目的，而不作名词dinner的定语，“歧义格式”中的歧义也消失了；在(4)中，介词词组for the children作名词toys的定语，而不作动词sells的状语，“歧义格式”中的歧义也消失了。

这说明，在研究歧义问题时，我们归纳概括出来的“歧义格式”中所反映的歧义，并不是“现实的歧义”，而是一种“潜在的歧义”

(potential ambiguity); 这种潜在歧义只有歧义的可能性, 没有歧义的现实性。当用具体的单词去代真“歧义格式”中的范畴符号时, 在所形成的具体的句子或词组中, 这种潜在歧义有可能继续保持, 也有可能不再继续保持而消失得无影无踪了。在歧义格式的研究中, “潜在歧义”是一个值得特别注意的、带有普遍性的语言现象。

在汉语的“歧义格式”中, 也同样存在着潜在歧义的问题。例如, “VP + 的 + 是 + NP”是汉语中的一个歧义格式, 其中的VP是一个双向动词, “VP + 的”做主语, “是 + NP”作谓语, 整个格式是一个主谓结构, 由于主语部分的“VP + 的”可以是施事, 又可以是受事, 因而产生了歧义。

例如, 我们把VP代真为“反对”, 把NP代真为“少数人”, 得到“反对的是少数人”这一句子, 可以理解为“提反对意见的是少数人”, 这时, 主语“反对的”是施事, 表示反对者, 也可以理解为“所反对的是少数人”, 这时, 主语“反对的”是受事, 表示被反对者。

当歧义格式“VP + 的 + 是 + NP”代真为如下的句子时, 这种歧义都一直保持着:

“看的是病人”可以理解为“正在观看某种情况的是病人”(“看的”是施事), 也可以理解为“被看的是病人”(“看的”是受事);

“关心的是她母亲”可以理解为“她母亲关心某人某事”(“关心”是施事), 也可以理解为“被关心的人是她母亲”(“关心”是受事);

“扮演的是一个演员”可以理解为“一个演员扮演了剧中某个非演员的角色”(“扮演的”是施事), 也可以理解为“被扮演成一个演员”(“扮演的”是受事);

“援助的是中国”可以理解为“中国援助了别国”(“援助的”是

施事),也可以理解为“别国援助了中国”(“援助的”是受事);

“相信的是傻瓜”可以理解为“相信某种情况的人是傻瓜”(“相信的”是施事),也可以理解为“所相信的人是傻瓜”(“相信的”是受事)。

但是,如果我们把歧义格式“VP + 的 + 是 + NP”代真为“关心的是分数”时,只可以理解为“所关心的事是分数”,“关心的”只能是受事,而不可能是施事,因为“分数”不可能去关心什么东西,这样,歧义格式中的潜在歧义也消失了。

如果把歧义格式“VP + 的 + 是 + NP”代真为“反对的是战争”时,只可以理解为“被反对的东西是战争”,“反对的”只能是受事,而不可能是施事,因为“战争”作为无生命的事物,不会去反对什么东西,这样,歧义格式中的潜在歧义也消失了。

上述语言现象说明,在自然语言的歧义研究中,当我们把具体的歧义词组或歧义句子概括为某种抽象的歧义格式的时候,这种抽象的歧义格式中所包含的歧义只是一种潜在的歧义。这种潜在的歧义在该歧义格式被代真为其他的词组或句子时,有可能继续保持,也有可能消失。这是自然语言歧义格式研究区别于自然语言的一般句法研究的一个重要特点,我们在自然语言的歧义格式的研究中,不可不注意这一个重要特点。在这些研究的基础上,我提出了“潜在歧义论”。

“潜在歧义论”从实质上改进了朱德熙教授关于“歧义格式”的理论,把“歧义格式”的理论更加深化了,也更加泛化了。我们可以把“歧义格式”看成是“潜在歧义”转化为“现实歧义”的一种特殊情况。

“潜在歧义论”并不是我灵机一动、脱口而出的一种“看法”(point of view),而是我在研究汉语词组型术语时,通过计算机分

析“歧义格式”而得到的一个重要“发现”(discovery)。这个发现,对于揭示自然语言中的歧义,是行之有效的。

张禄彭博士在他的这部著作中,运用“潜在歧义”的理论和方法来研究俄语,他进一步发现,潜在歧义不仅存在于词组型术语中,而且也存在于语音、形态、词汇、句法、语义、语用等多个语言学层面以及词素、词汇、句子等各个语言单位层次之中。他明确指出,潜在歧义是自然语言中的一个普遍现象。自然语言处理分为语音、形态、词汇、句法、语义、语用、篇章等多个步骤,其本身正可以看作是在以上多个步骤中的潜在歧义消解,每一层的歧义消解都是在为进行下一步处理做准备。

在语音学层面,他考察了词素的语音潜在歧义、单词的语音潜在歧义和句子的语音潜在歧义;在形态学层面,他考察了词素的形态潜在歧义和词形的形态潜在歧义;在词汇学层面,他考察了兼类词中的潜在歧义;在句法学层面,他考察了词组的句法结构潜在歧义和句子的句法结构潜在歧义;在语义学层面,他考察了词素的语义潜在歧义、单词的语义潜在歧义、词组的语义潜在歧义和句子的语义潜在歧义;此外,他还考察了语用学层面的俄语潜在歧义。

张禄彭博士的这些研究,向我们展示了一幅俄语潜在歧义现象的栩栩如生的图景,说明了潜在歧义确实分布在俄语的各个方面。正如D.Jurafsky和J.H.Martin在《自然语言处理综论》一书中指出的,在自然语言处理中需要的复杂的语言知识可以分为语音学与音系学、形态学、句法学、语义学、语用学、话语学六个方面,而这六个方面的语言知识存在着一个令人吃惊的事实,那就是:“语音和语言计算机处理的绝大多数或者是全部研究都可以看成是在其中某个层面上的消解歧义。”^①

① D.Jurafsky & J.H.Martin.《自然语言处理综论》(冯志伟、孙乐译),电子工业出版社,第3页,2005年。

张禄彭博士的这部著作还讨论了有关“潜在歧义”的一些理论问题。他认为,从广义上讲,歧义指的是一个语言形式对应多个意义或者结构,但是,在计算语言学中研究计算机自动处理潜在歧义与理论语言学中讨论人处理现实歧义有所不同。“潜在歧义”指的是计算机处理自然语言过程中语言形式的一个输入对应多个意义或者结构输出的情况。他依据潜在歧义与现实歧义的对应关系,把潜在歧义分为“Ⅰ型潜在歧义”和“Ⅱ型潜在歧义”;并讨论了潜在歧义与多义、模糊等概念的联系和区别。他认为,从计算语言学角度来看,自然语言处理过程中的潜在歧义具有多角度性、动态性、两面性、可消解性、多层性、复杂性等特性。潜在歧义的形成在语言符号属性、计算机处理过程、交际表达需要等三方面具有复杂的原因。

张禄彭博士的这部著作还讨论了俄语潜在歧义的消解方法和策略。他认为,自然语言处理最重要的任务之一就是尽量消解自然语言中的各种潜在歧义,包括消极的Ⅰ型潜在歧义和与现实歧义不符的Ⅱ型潜在歧义。大多数潜在歧义都可以依据语境进行消解,语境分为语言内语境和语言外语境。潜在歧义的消解需要分类进行:一般地说,歧义的消解依据语言学层面由低到高、语言单位层次由小到大的顺序;先消解Ⅱ型潜在歧义,再消解Ⅰ型潜在歧义;Ⅰ型潜在歧义中,消极歧义需要消解,而积极潜在歧义是说话人故意制造的一语双关的表达效果,不仅不会造成交际失误,反而可以使语言变得风趣幽默或者寓意深远,因此积极的Ⅰ型潜在歧义是不需要消解的。他指出,当前最重要的歧义消解策略有基于规则的策略、基于概率统计的策略、基于语义的策略和人机交互的消歧策略四种,并分别举例说明了这四种歧义消解策略在俄语潜在歧义消解中的运用。

由此可见,张禄彭博士的这部著作大大地拓展了我原来提出的

“潜在歧义论”的研究范围，对于自然语言处理的各个层面，都是很有用处的。

这部著作的不足之处是实验设计比较简单，算法分析比较薄弱。张禄彭原来是学习俄语出身的，他在博士期间才开始学习计算语言学，并且决心献身于计算语言学事业，因此，在实验和算法方面，我们对于他不应当过于苛求。

计算语言学研究是用计算机研究和处理自然语言的一门新兴的边缘学科。我过去说过，计算机对自然语言的研究和处理，一般应经过如下四个方面的过程^①：

第一，把需要研究的问题在语言学上加以形式化，建立语言的形式化模型，使之能以一定的数学形式，严密而规整地表示出来；这个过程可以叫做“形式化”。

第二，把这种严密而规整的数学形式表示为算法，这个过程可以叫做“算法化”；

第三，根据算法编写计算机程序，使之在计算机上加以实现，建立各种实用的自然语言处理系统；这个过程可以叫做“程序化”。

第四，对于所建立的自然语言处理系统进行评测，使之不断地提升质量和改进性能，以满足用户的要求；这个过程可以叫做“实用化”。

因此，为了研究计算语言学，不仅要有语言学方面的知识，而且，还要有数学和计算机科学方面的知识，这样计算语言学就成为了一门界乎于语言学、数学和计算机科学之间的边缘性的交叉学科，它同时涉及文科、理科和工科三大领域。

我希望从事计算语言学研究的青年学者同时具备语言学、数学

^① 冯志伟，《计算语言学基础》，商务印书馆，第2页，2001年，北京。

和计算机科学的知识，成为文理兼通、博学多识的人才。对于那些没有同时具备语言学、数学和计算机科学知识的青年学者，至少在自己原来所学的专业应当是精研通达的内行，在另外两个专业不是似懂非懂的外行，这样，才有可能有效地进行计算语言学的研究。因此，每一个决心研究计算语言学的人，都有必要进行更新知识的再学习。对于计算语言学者来说，“活到老，学到老”绝不是一句装扮门面的空话，而应当成为身体力行的座右铭。

目前，张禄彭博士已经认识到更新知识的必要性和重要性，他除了继续学习俄语和语言学知识之外，还开始学习数学和计算机科学的有关知识，这使我感到非常的高兴。

希望张禄彭博士继续努力，不断学习新知，为我国计算语言学的发展做出更大的贡献。

冯志伟

2012年5月19日 于北京

摘 要

自然语言处理是人类进入信息时代所面临的重要课题。20世纪60年代,由于计算机科学和语言科学的相互融合产生了计算语言学,其研究目标就是使计算机自动处理人类的自然语言,其中首要问题乃是识别和理解自然语言。潜在歧义是自然语言中普遍存在的现象,自然语言处理的绝大多数或者是全部研究均可以看成是在某个语言学层面或者语言单位层次上的消歧。在自然语言处理过程中,计算机可能遇到的歧义是自动处理的中间产物,对人类来说不一定具有现实意义,因此是潜在的歧义。

本书通过系统梳理计算语言学界关于歧义自动消解的研究,尝试提出并论证自然语言处理潜在歧义消解论的工作假设,并在俄语自动处理的工程实践中加以验证。面向自然语言处理的潜在歧义消解论是对冯志伟先生所提出潜在歧义论的继承和发展,是对计算语言学框架内潜在歧义消解问题研究的系统化阐释。

本书着重研究以下四方面内容:

第一,综述与本课题相关的研究成果。通过对计算语言学产生和发展轮廓的简要回顾,考察了自然语言处理研究的两大方法论流派——符号派与随机派。符号派主要由具有语言学背景的学者组成,他们为基于规则的消歧策略提供了方法论依据;随机派主要来自统计学专业和电子学专业的人员,他们提出了基于概率统计的消歧策略。自然语言处理是计算语言学研究的核心内容,贯穿于自然语言处理过程各个步骤的潜在歧义及其消解问题成为计算语言学难以解决的瓶颈。歧义问题研究自古有之,乔姆斯基关于句法歧义的理论在整个语言学界引起很大反响,激起大批学者从事句法歧义自