



教育部人文社会科学重点研究基地重大项目成果丛书

Publication Series: MOE Supported Projects of Key Research Institutes of Humanities and Social Sciences in Universities

语言文学类 Linguistics and Literature

俄罗斯计算语言学 与机器翻译

傅兴尚 许汉成 易绵竹 李向东 主编

语 文 出 版 社

教育部人文社会科学重点研究基地重大项目成果丛书

本书得到教育部人文社会科学重点研究基地基金资助

俄罗斯计算语言学与机器翻译

КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА И МАШИННЫЙ
ПЕРЕВОД В РОССИИ

傅兴尚 许汉成 易绵竹 李向东 主编



语文出版社

图书在版编目(CIP)数据

俄罗斯计算语言学与机器翻译/傅兴尚等主编. —北京:
语文出版社, 2009. 1

ISBN 978-7- 80241- 099- 2

I. 俄… II. 傅… III. ①数理语言学—研究—现状—俄罗斯②机器翻译—研究—现状—俄罗斯
IV. H087-151.2 H085-151.2

中国版本图书馆 CIP 数据核字(2008)第 158127 号

俄罗斯计算语言学与机器翻译

傅兴尚 许汉成 易绵竹 李向东 主编

*

语文出版社出版

100010 北京朝阳门南小街 51 号

E-mail: ywp@ywchs.com

新华书店经销 北京市联华印刷厂印刷

*

787 毫米×1092 毫米 异 16 开本 27.75 印张 467 千字

2009 年 8 月第 1 版 2009 年 8 月第 1 次印刷

印数: 1—1,000 定价: 45.00 元

本书如有缺页、倒页、脱页, 请寄本社发行部调换。

中文序言

近年来,我国计算语言学开始注意引进和介绍国外先进的理论和技术,但是,这些引进和介绍主要是针对欧美等发达国家的,对于俄罗斯的介绍寥寥无几,至于引进那就更是微乎其微了。这是一件非常遗憾的事情。

俄罗斯是我国的友好邻邦,在计算语言学和机器翻译研究方面,俄罗斯的起步比我国早得多。

在计算语言学方面,早在 1913 年,俄罗斯著名数学家马尔可夫(A. A. МАРКОВ, 1856—1922)就注意到语言符号出现概率之间的相互影响,他试图以语言符号的出现概率为实例,来研究随机过程的数学理论,提出了马尔可夫链的思想,他的这个开创性的成果用法文发表在俄罗斯皇家科学院的通报上^①。后来马尔可夫的这一思想发展成为在计算语言学中广为使用的马尔可夫模型(Markov model),是当代计算语言学最重要的理论支柱之一。在用数学思想来研究语言的创新性研究中,我们甚至可以追溯到 19 世纪中叶,早在 1847 年的时候,俄罗斯数学家布良柯夫斯基(Б. Буляковский)就提出了用概率论方法来进行语法、词源和语言历史比较研究的卓越见解了。1958 年,苏联数学家库拉金娜(О. С. Кулагина)采用集合论描述了基本的语法概念,为机器翻译研究奠定了坚实的数学基础。

在机器翻译方面,1933 年,苏联发明家特洛扬斯基(П. П. Троянский)设计了用机械方法把一种语言翻译为另一种语言的机器,并在同年 9 月 5 日登记了他的发明。特洛扬斯基认为翻译可以分为三个阶段,第一个阶段由只懂源语言的编辑,将输入的原文分析成特定的逻辑形式,将带有屈折词尾的变形词还原成原形词,并分析出各个单词的句法功

^① A. A. Markov, Essai d'une recherche statistique sur le texte du roman "Ougene Onegin" illustrant la liaison des épreuve en chaîne, Bulletin de l'Académie Impériale des Sciences de St-Petersbourg, 7, 153 - 162.

能,为此,他创造了一套逻辑分析符号。第二阶段是利用他的翻译机,把源语言的原形词和逻辑符号转换成目标语言的原形词和符号。第三阶段由只懂目标语言的编辑,把目标语言的原形词和符号转换成目标语言。特洛扬斯基认为,他的翻译机只能在第二阶段作为自动词典来使用。不过他相信,只要能够建造出一部专门处理逻辑分析过程的机器,总有一天,上述的整个翻译程序都能够用机器来实现。1939年,特洛扬斯基在他的翻译机上增加了一个用“光元素”操作的存储装置;1941年5月,这部实验性的翻译机已经可以运作;1948年,他计划在此基础上研制一部“电子机械机”(electro-mechanical machine)。但是,由于当时苏联的科学家和语言学家对此反映十分冷淡,特洛扬斯基的翻译机没有得到支持,最后以失败告终了。

1954年美国乔治敦大学进行世界上第一次机器翻译试验,成功地把俄语自动地翻译成英语之后不久,在同一年出版的第10期苏联《数学》杂志上就马上刊登了苏联科学院科技信息所所长潘诺夫(Д. Панов)的文章,介绍了美国乔治敦大学俄英机器翻译试验的成就,这篇文章的发表有力地推动了苏联组建自己的科研队伍、开展机器翻译研究的进程。1955年在别尔斯卡娅(И. Бельская)领导下的苏联科学院精密机械与计算机技术所机器翻译研究组进行了苏联的第一次机器翻译试验,将应用数学文本从英语译成俄语,这是世界上继美国之后进行的第二次机器翻译试验。我们可以说,在计算语言学和机器翻译研究方面,俄罗斯和前苏联的学者做了很多开创性的研究,取得了很大的成绩。

中华人民共和国成立后,俄罗斯和前苏联在计算语言学和机器翻译方面的研究成果传入我国,我国也开展了计算语言学和机器翻译的研究。1956年,国家便把机器翻译和计算语言学的研究列入了我国科学工作的发展规划,成为其中的一个课题,课题的名称是:“机器翻译、自然语言翻译规则的建立和自然语言的数学理论”。1957年,中国科学院语言研究所与计算技术研究所合作,开展俄汉机器翻译的研究。1959年,他们在我国制造的104大型通用电子计算机上,进行了俄汉机器翻译试验,翻译了9个不同类型的、较为复杂的句子,建立了我国第一个机器翻译系统。我国第一个机器翻译系统是俄汉机器翻译系统,直接受到了俄罗斯和前苏联的计算语言学和机器翻译研究的影响,这个机器翻译系统的主要设计人刘涌泉教授就曾经到当时的苏联专门学习机器翻译的方法,掌握了机器翻译的关键技术,可以说,我国的计算语言学和机器翻译研究首先是在俄罗斯和前苏联计算语言学和机器翻译研究的影响下开始的。

我于1957年高中毕业后,考入北京大学地球化学专业本科就读,一心想研究化学元素在地球上的分布规律。当时我才十九岁,求知的愿望非常强烈,对于新事物极为敏感,北京大学图书馆丰富的藏书吸引了我,我成为了图书馆的常客,整天泡在图书馆的书海之中。一个偶然的机,我在北京大学图书馆馆藏的1956年出版的美国《信息论》(IRE Transaction, Information Theory)杂志上,读到了美国语言学家乔姆斯基(N. Chomsky)的论文《语言描写的三个模型》(Three models for the description of language),被乔姆斯基在语言研究中的新思想深深地吸引了。乔姆斯基在他的文章中,提出了形式语言和形式文法的新概念,他把自然语言和计算机程序设计语言置于相同的平面上,用统一的数学方法进行解释和定义,提出了语言描写的三个模型。用数学方法描写的这三个模型是这样地抽象,它们既可以用于描写自然语言,又可以描写计算机程序设计语言。我预感到这种语言的数学描写方法,将会把自然语言和程序设计语言紧密地结合起来,在信息的处理和研究中发挥出巨大的威力。与此同时,我还在北京大学图书馆中,看到了俄罗斯数学家库拉金娜(О. С. Кулагина)在1958年的《控制论问题》上发表的《在集合论基础上确定语法概念的一种方法》(Об одном способе определения грамматических понятий на базе теории множеств)^①,库拉金娜采用集合论描述了基本的语法概念,提出了“族”(семейство)、“域”(окрестность)、“构形”(конфигурация)等语法概念的数学模型,为机器翻译研究在计算语言学的基本理论方面奠定了坚实的数学基础,使我的眼界大开。库拉金娜的成功坚定了我使用数学方法来研究自然语言问题的决心,于是我在1959年毅然从理科转到中文系语言学专业从事语言学的学习,从此走上了研究计算语言学和机器翻译的道路。

在上世纪50年代和60年代,由于东西方的隔绝,在学术领域中,很难找到英文的文献,不过,当时要找到俄文文献并不十分困难,为了跟踪国外计算语言学和机器翻译的发展情况,俄文的文献帮了我的大忙,我以俄文的文献作为中介,间接地了解到不少的情况。在当时极为封闭的学术气氛下,俄文成为了我了解国外学术动态的最主要的语言工具,它就像一扇敞亮的窗子,使我有可能会艰难地跟踪着国外学术的发展步伐。我从《语言学问题》(вопросы языкознания)、《控制论问题》(проблемы

^① О. С. Кулагина, Об одном способе определения грамматических понятий на базе теории множеств, 《проблемы кибернетики》, вып1, стр 201—214, 1958.

кибернетики)、《语言学中的新事物》(новое в лингвистике)等俄文文献中,了解到国外学术的最新进展,大大地拓宽了我的学术视野。

在这个时期,我还阅读了列夫辛(Ревзин)的《语言模型》(модель языка)以及阿赫玛诺娃(О. С. Ахманова)、梅尔楚克(И. А. Мельчук)等的《语言研究中的精密方法(О точных методах исследования языка)》等专著,对于自然语言处理中使用的集合论、数理逻辑、概率论和数理统计等方法有了深入而系统的认识。俄罗斯和前苏联在计算语言学和机器翻译方面的研究以及他们通过俄语对于西方学术的介绍,成为了中国学者间接地了解西方计算语言学和机器翻译进展的一个重要的途径。

改革开放以来,我国打开了国门,很多青年学者掌握了英语,英语文献成为了我们了解国际计算语言学和机器翻译进展的重要来源。但是,由于青年学者中懂得俄语的人很少,不能阅读俄文的文献,我们对于俄罗斯和前苏联的计算语言学和机器翻译的介绍和了解非常不够,这是一个很大的缺憾!

《俄罗斯计算语言学和机器翻译》弥补了这个缺憾。本书全面地介绍了俄罗斯以及前苏联学者在计算语言学和以机器翻译为代表的自然语言信息处理系统领域取得的成就,包括基础理论探讨、语言词汇知识库的建造、形态自动分析、句法自动分析、语义自动分析、句法—语义一体化分析、文本自动处理、语料库语言学、机器翻译、信息检索系统、自动文摘和计算机辅助语言教学软件等实用系统的开发等内容,涉及到了计算语言学的各个方面,使我们对于俄罗斯计算语言学和机器翻译的研究以及他们取得的成就,获得一个鸟瞰式的认识。

俄语是一种典型的屈折语,形态变化非常丰富,名词有6个格以及单数与复数的变化,还有阳性、阴性和中性的差别,动词有人称(person)、时态(tense)和体(aspect)的变化,形容词不仅有性、数、格的变化,还有一般级、比较级、最高级的变化。与英语比起来,俄语的形态变化丰富得多,俄语如此丰富的形态变化,为充分揭示形态分析、句法分析、语义分析、篇章分析提供了极为充分的语言数据,可以用更加多样的语言数据,来检验基本上以英语的语言数据为基础而建立的自然语言处理的理论和方法,发现其不足,从而促进自然语言处理研究的发展。

本书是教育部人文社会科学重点研究基地第二批重大项目成果,是作者们多年研究心血的结晶。计算语言学和机器翻译是语言学、数学和计算机科学的交叉学科,而本书的作者都是俄语专家,他们的学术背景是语言学,几年来,他们努力进行知识更新的再学习,逐渐地熟悉了数学和

计算机科学,现在,他们在数学和计算机科学方面,虽然还不是精研通达的内行,却也不是似懂非懂的外行,他们已经掌握了数学和计算机科学的基本知识,在知识更新的基础之上,他们把语言学与数学和计算机科学巧妙地结合起来,写出了这样的专著,这是难能可贵的,相信本书的作者们在探索计算语言学和机器翻译的道路上一定会取得更加出色的成绩。

冯志伟

于北京后拐棒胡同寓所

2008年8月8日

Предисловие

В современном мире новых информационных технологий большое значение приобретает компьютерная лингвистика. Она составляет то, что сейчас получило название *lingware*, то-есть лингвистическое обеспечение компьютерных технологий. Этот термин введен по аналогии с *software* (программное обеспечение) и *hardware* (техническое обеспечение). Лингвистическое обеспечение необходимо для построения лингвистических процессоров - программ, обрабатывающих устные и письменные тексты на естественных языках для ввода их в компьютер или для вывода из компьютера результатов обработки текстов. Лингвистические процессоры облегчают работу с компьютером для человека. Создание человеко-машинных интерфейсов, лингвистических сред, благоприятных для человека-пользователя, и составляет предмет современной компьютерной лингвистики. Это является как бы внешней стороной существования компьютерной и прикладной лингвистики - решение проблем, существующих вне собственно языкознания как науки.

С другой стороны, практика прикладной и компьютерной лингвистики подтверждает (или не подтверждает) теоретические положения и теоремы, создаваемые в рамках лингвистики теоретической, общего языкознания и других его разделов (исторического, сопоставительного, математической лингвистики и т. п.). Теория, таким образом, нуждается в проверке практикой.

В настоящей монографии рассматриваются наиболее важные вопросы компьютерной лингвистики, а именно: основные этапы становления и развития этой науки, вопросы моделирования лингвистических явлений, автоматический морфологический анализ естественных языков, автоматический синтаксический анализ, авто-

матический семантический анализ, проблемы морфологического, синтаксического и семантического синтеза. Освещаются также наиболее важные аспекты разработки и использования современных действующих систем автоматической обработки естественно-языковых текстов. Сюда включаются системы распознавания устной речи, автоматизированного информационного поиска, искусственного интеллекта, экспертные системы, системы машинного перевода и системы, соединяющие машинный перевод с человеческим (переводческая память), системы и структура дистанционного обучения. Рассмотрены методы оценки эффективности действующих систем автоматической обработки текстов и целый ряд теорий, лежащих в основе этих методов.

Все вышеуказанные вопросы освещены на основе новой и новейшей литературы по предмету. Безусловно, всякая наука базируется не только на разработках сегодняшнего дня, но и учитывает уроки истории этой науки. В настоящей монографии нет специального рассмотрения исторических путей развития той или иной концепции. Однако составители монографии исходили из имеющихся современных оценок пройденного компьютерной лингвистикой пути. Такие оценки усиливают доказательность выводов и определение перспектив развития тех или иных информационных и лингвистических технологий.

Компьютерная лингвистика развивается очень быстро. Во многих странах разрабатываются интеллектуальные системы, работающие с естественным языком. Обилие естественных языков и появление на мировой арене все новых естественных языков в рамках растущего делового, научного и политического общения, с одной стороны, а с другой - утверждение и расширение единой всемирной информационной сети Интернет создают некоторые координаты и стимул для ускоренного развития информационных технологий, ориентированных на естественные языки. Поэтому то, что сегодня кажется новым словом в науке, завтра уже устаревает. Однако мы надеемся, что нынешняя монография внесет свой вклад в дело совершенствования информационных технологий, в образование и воспитание новых

кадров, в производство новых научных и технических идей и их приложений.

Марчук Ю.Н

Москва

2005.08

前 言

当今信息化社会主要有以下几个显著特点:信息存储方式的数字化;信息处理方式的自动化;信息传输方式的网络化;信息应用方式的产业化。语言是信息的主要载体,语言信息的可计算处理成为信息技术研究领域的主流,同时也是举世关注的焦点和热点。语言活动(包括语言交际、语言研究、语言教学、语言翻译等)要现代化,计算机性能要智能化,所以语言学和计算机科学两个学科进一步发展的瓶颈问题在很大程度上取决于计算语言学(Компьютерная лингвистика),应该说,自然语言处理(Обработка естественного языка)的实现是语言现代化和计算机智能化的根本保证,也是当今信息时代具有代表性的前沿学科。在该领域,俄罗斯(包括苏联)在理论构架、资源建设、数学建模、工程实现等各个方面都有悠久的传统和鲜明的特色,取得了公认不斐的辉煌成就,在国际上有广泛的影响。我们觉得有责任和义务深入学习和研究这些成果,并系统地展现给我国的专家和同行。

一般认为,计算语言学是随着 1946 年第一台电子计算机的问世应运而生的。其实早在 20 世纪初期,俄罗斯学者就提出了机器翻译的思想。1933 年 9 月 5 日,彼得·斯米尔诺夫—特洛扬斯基(Петр П. Смирнов—Троянский)在莫斯科登记一项专利,其内容是要求保护“制造一台在从一种语言翻译成另一种语言或多种语言时能选择和打印词汇的机器”的专利。但是,将机器翻译作为自然语言处理领域的研究课题,并取得初步应用性成果的,则首推美国乔治敦大学(Georgetown university)于 1954 年进行的世界上第一次俄英机器翻译实验。20 世纪 70 年代以来,随着科学技术的迅猛发展和人类社会信息化进程的明显加快,人们对于作为信息重要载体的语言研究更加关注,使得属于人文科学传统学科的语言学在整个现代科学体系中的地位发生了显著变化,进而成为一门先导科学。

语言学具有自身的文化本原性,但与当前其他学科的交叉性联系也日益明显,计算语言学正是在这种学术背景下得以快速发展。这种迅猛势头一直延续至今。在俄罗斯,学者们把计算语言学也称之为工程语言

学(Инженерная лингвистика, ИЛ)或数理语言学(Математическая лингвистика)。这从一个侧面说明,计算语言学是文、理、工交叉的年轻学科。

我们知道,自然语言(естественный язык)是人类独有的交际工具,和人类独有的思维能力密不可分,语言是人类最重要的思维工具。计算语言学研究,归根结底,就是要实现计算机对自然语言的理解,建造模拟人类语言能力、模拟人类从事言语活动时思维过程的装置,通过这些装置实现言语作品的生成、言语作品的理解和言语作品的转换。

人之所以具备语言能力是因为具有两种系统:一是静态的语言知识和语言规则系统。根据认知科学的观点,这个系统以网络的形式存储在人的大脑中;二是动态操作系统。即在言语场景、表达或理解等交际欲望的激发下快速激活静态系统中相关节点,通过已有知识经过一系列认知操作推知未知信息,从而实现预期的目标。这个过程虽然是瞬间即逝的,但是完全可以分解为有序的步骤,由这些步骤组成的过程体现为动态操作系统。

自然语言处理,不论处理哪一平面、哪一层级的语言信息都要从两个方面入手。一是把处理对象的有关知识和规则进行模式化分析和形式化表达,储存在计算机中,作为处理过程中的信息源系统。当然,这个系统一定是概括的,而不是具体的;是精确的,而不是模糊的;是元语言形式化表达的,而不是自然语言表达的。二是把由处理对象到处理目标的操作过程分解为单元性步骤,体现为由程序指令构成的算法。当然这种算法与信息源系统是相互制约的,而不是彼此孤立的;程序指令之间是相互依存的,而不是互不影响的;整个算法可以是多方式的,而不是唯一不变的。

语言模式化(Языковое моделирование)是计算语言学的基本方法,这反映在以下阐述中。“计算语言学是应用语言学的一个分支,指利用计算机科学这一工具—程序、信息组织技术以及信息处理技术对某种条件、场合、领域语言行使的功能进行模式化。也指语言学和其交叉学科中对语言可计算模式的诸多应用。”(Энциклопедия “кругосвет” : компьютерная лингвистика)

不难发现,计算语言学与其他语言学分支学科的不同主要体现在:1)研究目的上。计算语言学要实现各种语言功能的模拟,即语言信息的自动化处理(如,机器翻译、信息查询、语言知识的自动获取和表达等),同时,也要关注如何利用语言信息处理技术促进理论语言学研究,在这方面但不限于这方面表现明显的如,语料库语言学(Корпусная лингвистика);

2)研究手段上。计算语言学研究手段主要是计算机技术、信息技术、控制论、人工智能等,体现为自动化、快速化、准确化和实证化;3)研究方法上,可归结为模式化,包括语言知识的模式化和实现过程的模式化。

计算语言学研究不仅跨越各级语言单位(词素、词、句子、篇章等),而且覆盖语言的各个侧面(语音、语法、语义、语用等),这和只把某类语言单位作为研究对象或只把某一语言侧面作为研究对象的语言学分支学科(前者,如词汇学;后者,如语义学)也是不同的。

一般计算语言学围绕文法和算法两大方面以及二者的契合对接展开研究。文法是语言范畴、语言规则的总称,体现为静态的语言知识库,同时包括近几年逐步得以重视的基于语料库的语言统计模式;算法是以文法为基础实现特定语言处理目标的指令集,体现为动态的操作过程。然而,不论是文法,还是算法,都需要模式化。

本书《俄罗斯计算语言学与机器翻译》,顾名思义,就是以国际计算语言学研究为背景,全面展示和阐释俄罗斯以及前苏联学者在计算语言学和以机器翻译为代表的信息处理系统研发领域取得的科研成就,包括理论阐释、语言知识库建造和应用、形态、句法、语义、篇章信息的自动化处理、语料库语言学、机器翻译、信息查询系统和教学软件等实用系统的开发诸多内容。与此同时,我们合理吸收这些成果,提出自己的语言处理方略,以及某些具体语言处理目标的实现策略和技术。在研究中所涉及的语言模式化对象和信息处理对象主要是俄罗斯民族语言——俄语。俄语是典型的屈折语,形态手段异常丰富,如何充分顾及其语言个性特点,实现语言信息的自动化处理,在研究内容和研究方法上都体现出一系列特点,这也是本书区别于其他计算语言学著作的一个鲜明特色。

本书各个部分的作者分别是:

傅兴尚——第一章的第一、二、三、五节;第二章的第一、二、三、四节;第三章;第四章的第一、二、三节;第五章的第一节;第六章的第一、三节;

许汉成——第七章;第八章;

易绵竹——第一章的第四节;第五章的第二、三、五节;第九章的第七节;第十章的第二、三节;

李向东——第九章的第一、二、三、四、五、六节。

此外,本书增列两篇由李锡胤先生和张家骅先生发表的相关学术论文,分别是第二章的第五节和第六章的第二节,笔者的研究生叶其松和吕红周结合硕士学位论文工作,分别撰写了第四章第四节和第十章第一节。

本书大部分章节作为独立的论文曾在近几年公开发表,所以该成果

应该是一部专题研究性质的著作。虽然不可能面面俱到,但重点问题都涉及到了,同时力求全篇的系统性和统一性。全书由傅兴尚最后修改、定稿,对部分章节进行了增删。

本书是教育部人文社会科学重点研究基地第二批重大项目成果,项目批准号为:01JAZJD740006,得到教育部社政司、黑龙江省教育厅、黑龙江大学的资助,在此深表谢忱。俄罗斯联邦著名机器翻译专家,莫斯科国立大学教授玛尔丘克(Марчук Ю. Н.)百忙中为本书写了俄文序言,我国著名计算语言学家冯志伟教授也欣然地为本书写了中文序言,使此书增色不少,我在此表示诚挚的敬意和谢意。

经过多年的艰苦努力和辛勤探索,《俄罗斯计算语言学与机器翻译》这本书终于呈现在读者的面前,作为尽心尽力的研究者,我们感到非常欣慰,但同时也有些忐忑不安,由于水平和能力所限,一定有许多问题研究的不够深入、甚至有许多纰漏,敬请各位专家和同仁批评指正。

傅兴尚

2005 年 10 月

哈尔滨

目 录

第一章 计算语言学的几个基本问题	(1)
第一节 计算语言学导言	(1)
第二节 语言知识的表征	(2)
第三节 语言模式化	(11)
第四节 语言自动机的建造及工作原理	(21)
第五节 俄罗斯计算语言学研究机构	(27)
第二章 俄语词汇知识库的建造和应用	(32)
第一节 词和词汇知识库概说	(32)
第二节 建造词汇知识库的基础和原则	(40)
第三节 俄语词汇的信息结构	(44)
第四节 俄语词汇知识库的建造技术	(47)
第五节 Е. Падучева 等设计的《词汇信息库》概览	(56)
第三章 俄语形态信息的自动化分析	(70)
第一节 俄语文本的单元化分析	(70)
第二节 俄语形态信息的自动化分析	(72)
第三节 俄语词汇形态信息的统计学习	(83)
第四章 俄语句法信息的自动化分析	(89)
第一节 句法自动化分析的几个基本问题	(89)
第二节 基本构句块及其识别算法	(90)
第三节 俄语句法结构的模式化描述及操作原理	(96)
第四节 形容词扩展模式的形式化	(103)
第五章 俄语语义信息的自动化分析	(119)
第一节 语义中介语及其计算原理	(119)
第二节 词汇语义资源及语义分析	(129)
第三节 意念词典和语义词典	(146)
第四节 MAMM 模型与语句意义表示	(156)

第五节 一种基于 TM 技术的机助翻译系统的语义 处理策略	(167)
第六章 俄语句法语义的一体化分析	(176)
第一节 引言	(176)
第二节 B.A. Ty3oB 的俄语形式化理论	(177)
第三节 基于函数模型的句法——语义一体化计算	(194)
第七章 俄语文本信息的自动化处理	(210)
第一节 文本著作权归属的自动分析	(210)
第二节 文本功能语体的自动分类	(223)
第三节 内容分析: 概念、类型与方法	(240)
第八章 语料库语言学	(265)
第一节 语料库和语料库语言学的概念	(265)
第二节 语料库的类型	(270)
第三节 俄语语料库的建造和发展	(272)
第四节 语料库的标注经验与规范	(305)
第五节 语料库检索工具的基本功能	(325)
第六节 小结	(333)
第九章 俄罗斯机器翻译: 历史与现状	(335)
第一节 机器翻译思想的产生	(335)
第二节 俄罗斯机器翻译发展历程	(336)
第三节 机器翻译方法	(342)
第四节 机器翻译理论流派	(354)
第五节 俄罗斯机器翻译系统	(363)
第六节 总结与展望	(374)
第七节 与 UNL 接口的机器翻译系统 ETAP—3	(377)
第十章 语言信息处理系统	(388)
第一节 信息检索系统	(388)
第二节 基于语义分析的文本自动文摘研究	(393)
第三节 计算机辅助语言教学若干问题研究	(399)
参考文献	(407)