

现代汉语名词槽关系系统初步研究

陈群秀

智能技术与系统国家重点实验室

清华大学计算机科学与技术系, 北京 100084

E-mail: cqx@s1000e.cs.tsinghua.edu.cn

摘 要: 现代汉语名词槽关系系统不仅是现代汉语语义系统的一个重要组成部分, 而且是自然语言处理中汉语语义资源的重要工程之一。本文首先论述了现代汉语名词槽关系系统的研究路线和设计原则, 然后提出了以语类和槽类型来描述名词和它的修饰成分之间的关系的槽关系表示方法, 最后论文给出了槽关系描述的一个例子。

关键词: 现代汉语名词槽关系系统、现代汉语语义系统、汉语语义资源、槽关系表示方法、语类、槽类型、计算词典学

Elementary Research on The System of Relations of Slots Centering on Noun for Contemporary Chinese

Chen QunXiu

The State Key Laboratory of Intelligent Technology and System

Department of Computer Science and Technology, Tsinghua University, Beijing 100084

Email: cqx@s1000e.cs.tsinghua.edu.cn

Abstract: The system of relations of slots centering on noun for contemporary Chinese not only is an important component part in Chinese semantic system, but also is one of the most important function on Chinese semantic resource in NLP. First, this paper introduces the research way and design principles of the system, then presents representation method of relations of slots describing relations between noun and its modify components with category and type of slot. At last, this paper shows one noun examples describing relations of slots.

Keywords: the system of relations of slots centering on noun for contemporary Chinese, contemporary Chinese semantic system, Chinese semantic resource, representation method of relations of slots, category, type of slot, computational lexicography

1. 前言

近十年来, 随着计算机软硬件技术的飞速发展和网络技术的进步, 人类文明也进入了革

*本文承国家社科“九五”重大项目《信息处理用现代汉语词汇研究》的子课题“现代汉语述语动词机器词典的扩充和槽关系研究”(批准号 97@YY001-7)的资助。

命性的变革时期，既从工业化社会进入信息化社会。世界经济正在发生着一场影响深远的变革，那就是信息经济、知识经济在全球的兴起。美国等发达国家在这方面走在最前面。信息化已成为二十一世纪全球不可抗拒的选择，成为国家经济与社会发展的命脉，成为一种新的控制财富的手段，这是来自用枪炮无法阻挡的一种新的威胁与挑战。我国信息产业面临的形势十分严峻，已经是“八国联军兵临城下”。如果不重视这种挑战不抓住这个机会，则发达国家有可能将信息技术作为新殖民化(信息殖民化)的有力武器。中文信息处理产业是否立得起来？立得好不好？关系到我国政治、经济、社会生活的变革甚至我国在世界上的地位，关系到我国的安全生存问题。因此必须研究汉语语言信息处理的理论和方法，研究和建立汉语信息处理共享的强大、可靠、丰富的知识资源。

汉语是我国乃至世界范围的主要交际工具和信息载体，使用计算机进行汉语信息处理是当代信息社会的需要，也是智能计算机系统的重要特征。汉语信息处理的研究和进展，依赖于计算语言学的词汇学、句法学、语义学、语用学的研究和进展，核心问题是语义学。相比而言，句法分析理论和技术(无论是对外语还是汉语)发展得比较成熟和完善，语义学则是难度最大、起步较晚的一个薄弱环节，空白点更多。目前，自然语言理解、汉语信息处理处于一个关键时期，处在取得重大突破的前夜，最重要最困难的是语义学的研究和突破。

面向机器处理的语义学研究，国内外起步时间均不长。二十世纪 60-70 年代开始，语义学在世界上成为语言学界讨论和研究的热点，各种语义学流派如雨后春笋涌现出来，学说纷呈，提出了一些处理语义问题的模式、理论和方法，但还远没有一个公认的、照顾周全的体系，至于汉语语义系统和汉语语义学模式的建立更是一片未开垦好的科学园地。因此动词、形容词的论旨网格的建立和名词为中心的槽关系系统的建立以及句子情态的研究，不仅理论上具有极大的意义，而且将对我国的中文信息处理作出很大贡献(即有很大的实用意义)。国外形成的一些语义理论和流派(例如解释语义学、生成语义学、蒙太古语义理论、格语法、CD 模型、情景语义学等)，均是以英语等西方语言为基础的，例如影响较大、应用较广的两种语义分析方法(格语法和 CD 模型)，都是建立在“以动词为中心”的前提下。由于汉语不仅是以动词为中心也是以名词为中心的(许多汉语句子可以不包括动词，这是汉语重要特点之一)，所以仅研究动词句不研究名词句难以满足汉语机器理解的要求。另外，光研究动词、形容词中论旨角色的语义组合关系也不够，还要研究动词框架中的名词为中心的槽关系即论旨角色内部的语义组合关系。由于汉语缺乏屈折变化，是语义型语言，句法分析对句子的贡献比英语要小，因此语义分析对汉语机器理解尤为重要。

国际上，动词的格框架理论提得比较早的，菲尔墨 1966 年和 1968 年就提出了格语法、格框架理论。但是直到 30 年之后(即 1996 年)，菲尔墨才向美国基金会申请一项语言工程项目“框架网：基于框架语义学的英语语义词库”，计划研究 5000 个英语动词的语义词库(没有研究情态)，但至今未见这个工程性项目的研究成果公布于世。普林斯顿大学 Miller 等人研制的英语词网 WordNet 旨在从心理学角度建立英语词汇基本语义关系的实际模型。WordNet 目前包含大约 95600 个词条(51500 个单词和 44100 个搭配)，描写了上下位、同义、反义、部分一整体等词汇语义关系。有一些国家将 WordNet 进行了本地化，例如 Euro WordNet(德语、法语、意大利语)、Korean WordNet(朝鲜语)。

国内对汉语词汇语义研究有“八五”期间北京语言学院、清华大学、北京大学联合研究和开发的《现代汉语语义分析词典》。该词典建立了一个较为复杂的事物类语义分类体系(描述上下位关系)和一个语义特征描写体系(描写属性特征等关系)，分析了四万多个词。该词

典存在的主要问题是经语义特征描写过的词数量较少，对其他语汇语义关系的研究尚未真正展开和深入。另外，清华大学和中国人民大学在研究自然科学基金项目《现代汉语述语动词机器词典》时也研制了一个信息处理用现代汉语语义分类体系并应用到该机器词典中和日汉机器翻译系统中。但该分类体系的语义词典尚未完成正在研制中。董振东建立了一个描述概念、概念间关系以及概念所具有的属性的知网(含汉语词语 50000)，已于近日在因特网上发布。东北大学正在探索将基于英文的 WordNet 进行本地化来构造中文名词 WordNet。北京大学在研究了一个实用的含几万词汇的汉语语法词典基础上，也对词汇的语义进行了研究并正在设计语义词典。

国内涉及句义的研究主要有：黄昌宁、陈祖舜以情景语义学为理论依据提出的“关于语义辞典构造的初步设想”。黄曾阳等以 HNC(概念层次网络)理论来建立网络式概念体系和语句的语义表述模式。HNC 以五元组和语义网络表示抽象概念，用“类别符号+挂靠”近似表达具体概念，并设立语言深层网络的局部联想和全局联想的功能。HNC 在语义表达上有自己的特色，在语义处理上不失为一条新路。另外一个有特色有新意的研究是清华大学陈群秀、黄昌宁和中国人民大学林杏光等人研制的现代汉语述语动词机器词典，该词典的理论依据是：原则参数语法、格语法、配价语法和研制者自己的一些想法。该词典是一项大规模的知识工程，包含 3002 个常用的汉语动词义项的有关词形、拼音、词性、动词分类、释义、论元类、义项数、义项序号、论旨模式的基本式和变换式、句例、论旨角色的语类、句法功能、语义限制、论旨实例、否定形式、时态、语义指向动词的后状及论旨模式的扩展式等丰富的词法、句法和语义信息。而且由于是采用传统词典学和计算词典学相结合方法在大量客观的语料(2000 万语料)的基础上研制的，故而词条描写内容丰富详尽、角度新颖。

根据框架语义学，格关系、槽关系和情态是句义的三大语言知识工程。论旨网络描写的是论旨角色间(或与动词间)的语义关系，即事件的各种角色间的语义关系(格关系，框架的语义关系)。在此基础上，再来研究论旨角色内部的语义关系是很有必要的，即研究动词框架的槽关系。槽关系是事件内的偏词和正词之间的语义关系，即研究以名词为中心的词组的定语与中心词之间的关系。所以取名槽关系是因为采用人工智能中框架知识表示中的术语“框架”、“槽”。现代汉语的槽关系是现代汉语语义系统的一个重要组成。研究名词为中心的槽关系，其重点是研究名词。印欧语着眼于谓语动词和时间，汉语着眼于名物和空间，因此名词的研究在汉语研究中具有特别重要的意义。整个句子语义框架中格关系(论旨角色)已进行了较多的研究，而槽关系则还是一片未曾开垦过的处女地。格关系和槽关系共同组成语义知识，可作为中文信息处理和语言学研究的重要知识源而共享，对汉字输入的自动化和智能化、办公室自动化、手写体汉字和印刷体汉字识别的后处理、汉语语音识别的后处理、汉语全文检索和自动文摘、机器翻译、汉语自动分词和自动标注、汉语句法分析和语义分析、汉语自动生成、汉语人机接口等自然语言处理系统的研究和开发都是必不可少的，对汉语语言学研究、汉语教学和对外汉语教学也大有裨益。

清华大学计算机科学与技术系、中国人民大学对外语言文化学院和北京语言文化大学语言信息处理研究所三个单位的计算机界和语言学界工作者联合承担国家社科“九五”重大项目《信息处理用现代汉语词汇研究》(批准号 97@YY001)的子课题“现代汉语述语动词机器词典的扩充和槽关系研究”(批准号 97@YY001-7)，其中一个主要研究目标是研究以现代汉语名词为中心的槽关系，建立一个槽关系系统。

2. 现代汉语名词槽关系系统的研究路线和设计原则

名词槽关系研究在国内外都比较少见,汉语名词槽关系研究尤其是作为机器处理使用的槽关系研究尚未见到。汉语的句子中,名词短语(以名词为中心的名词词组)的分析和理解对句子的整体结构和语义理解有很大的影响,但汉语名词短语常常有着复杂的内部结构,即名词词组中修饰中心词名词的定语可能有多项,而且定语的语类以及定语与中心词名词的语义关系也非常复杂,因为汉语名词词组的定语的多样化以及定语与名词的语义关系的复杂性正是汉语的一个特点和机器处理的难点。因此研究汉语名词槽关系和建立汉语名词槽关系系统的难度很大,何况还没有研究先例和可以参照的外语的相应工作。

汉语语言界研究名词性词组的定语多从定语出现的位置、定语的语类(例:名词、形容词、动词、名词词组、动词词组等)入手,对定语的划分也各不相同,有的把定语划分为修饰性定语和限制性定语(黄伯荣),有的划分为描写性的和限制性的(朱德熙;刘月华),有的划分为称谓性定语、外延性定语和内涵性定语(陆丙甫)。有的划分为限定性定语、区别性定语、描写性定语(张卫国)。存在的问题是有的界限不易把握,划分起来很不一致,而且不能完全反映出各种定语的特点和相互的区别,例如出现的位置、形式上的特点,特别是无法揭示定语与中心词的语义关系是什么。而且一般没有做到系统化,更无法形式化。

鲁川主张以谓词类别和受谓词联系的名词类别来划分槽类。他将所定义的格角色(中框角色 24 种、外围角色 32 种)组合成“当事一存在”、“当事一自变”等 28 种“主谓”格式。然后设想将“主谓”的顺序一一颠倒即成为偏正,这偏就是槽类。这种设想尚未作系统化和工程化实验,效果如何不大清楚。林杏光、张庆旭主张根据偏正语块中的偏词和正词的语义关系来划分槽类,并设想有领属、性状、范围、指量、特征 5 个槽类。陈群秀根据在研究“信息处理用现代汉语语义分类系统”中的经验和体会,主张汉语名词槽关系的描写应以语义关系描述为主、句法描述为辅,语义关系的类型设计以语义分类体系为借鉴且宜细不宜粗。清华大学和中国人民大学联合研究小组多次召开小型工作会,讨论本组几个成员(陈群秀、林杏光、张卫国、冀小军)对名词为中心的槽关系研究原有成果和设想的优缺点,决定采用的新设想,研究现代汉语名词槽关系的描述表达问题、研究路线、工作单设计原则和工作单具体设计问题,终于取得共识。

(1) 我们的研究方法、技术路线是基于大规模语料的计算词典学编辑方法与基于语言工作者的语感的传统词典学编辑方法相结合,使动词的论旨网络描述和名词为中心的槽关系研究真正建立在丰富和客观的语言事实基础上。即先从 5000 万的语料库中抽取大量的动词的句例以及名词为中心的词组实例,再用人工进行分析、研究和填写工作单,然后再录入计算机进行统计、分类以及组成系统。

(2) 《现代汉语名词槽关系系统》的工作单设计遵循以下几个原则:

① 工作单设计以名词为轴心,即以每个名词为中心组织一张工作单的信息;

② 研究方法是“语料库方法+联想”,即把主要从机贮语料库中获取的大量例证为依据的计算词典学方法同主要以语言工作者的经验、语感为依据的传统词典学方法结合起来,以使名词的槽关系研究真正建立在丰富的客观的语言事实基础上,同时发挥语言工作者的经验、语感进行联想和归纳。

具体做法:首先从大规模语料库中取出有关名词词组的大量实例,然后由语言工作者人工分析这些含各种定语的名词词组,从中选出三个最有代表性的名词词组分别填入该名

词工作单的词组实例 1、词组实例 2、词组实例 3 的各对应栏目内。在填写时充分发挥语言工作者的经验、语感对词组实例进行联想补齐。槽关系描述项“槽类”以语义描述为主(槽类型)、句法信息为辅(语类)。每个词组实例的“定语 1”、“定语 2”、...、“定语 11”，系指该名词词组中名词前各个定语中排在第一个位置、第二个位置乃至第十一个位置的定语的词形部分。“语类”系指该定语 i 的句法范畴，用“{ }”将其括起，同时有几种语类时用“|”表示分隔，表示“或”的关系。“槽类型”指该定语 i 的槽关系语义类型，即“槽类”。“槽类”(亦即槽关系名)用“[]”括起来表示。“槽类型顺序”系指该词组实例的多个槽关系名的排列顺序。“可能顺序”系指本词组实例代表的那一类词组实例可能有的槽序(槽类型顺序)的联想表达式。语言工作者综合 3 个词组实例的“可能顺序”的联想表达式，再联想归纳出该名词的总的可能的槽关系表达式，形式如[A₁][A₂]*[[A₃][A₄]|[A₅]][A₆]*，其中“*”表示此类槽关系可有 1 到 n 个。“[]”中的 A_i 为槽关系名。

③ “语类”和“槽类型”的设计原则是将“必要性”、“够用原则”、“可操作性”和“计算机处理用”几个原则综合考虑；

④ 欲研究和描写的 3000 个名词的义项选择综合考虑名词在《信息处理用现代汉语语义分类系统》中分布性、名词的常用性和典型性。

3. 现代汉语名词槽关系系统工作单的设计

现代汉语名词槽关系系统的工作单信息分为三部分：①制作信息；②词条信息(包括：词形、拼音、语义分类(填写该名词在《信息处理用现代汉语语义分类系统》中的上位或上上位义类)，义项数目，义项序号，释义)；③名词槽关系信息(包括词组实例 1，词组实例 2，词组实例 3，槽关系表达式联想)。每个词组实例中又提供了“实例框”、十一个定语(定语 i)框、槽类型顺序框和可能顺序框。每个定语框中包含“定语”、“语类”、“槽类型”三个小框。因篇幅原故，“语类”的规定和定义在此不作介绍。

4. 现代汉语名词槽关系系统的槽类型设计

槽关系研究包括槽类和槽序。我们拟根据偏词和正词(中心词)的语义关系来划分槽类，那么，这样的语义关系有多少类？多个槽类的排序顺序如何？对这两个问题我们反复研究和讨论，决定重点研究槽类型，而对槽序的研究采取在工作单上根据几个词组实例的实际槽序加上语言学工作者的语感和经验进行联想归纳，对每个名词归纳出联想的槽关系表达式。待名词描述达到一定数量规模时再对同语义类名词的槽序和槽关系表达式进行统计，期望从中能发现同语义类名词槽序的规律。

槽类型系指槽类、槽关系。即词组实例 i 的定语 j 与名词之间的语义关系类型。槽类型(即槽关系名)用“[]”括起来表示。我们使用的槽类型(槽关系名)是在本联合研究小组几个成员的研究成果基础上设定的，第一层槽类型为 21 种，第二层为 49 种，两层共 70 个类型，大致分为属性、状态、关系三大块。

属性包括共有的属性(事域、事情、领属、来源、去向、数量属性、顺序属性、频度属性、指量属性、空间属性(位置属性、距离属性、空间范围属性)、时间属性(时点、时段、历时)等)和部分共有属性(又分具体物共有、抽象物共有)。部分共有属性非常丰富，包括物理属性、化

学属性、具体物社会属性(人除外)、生物属性、抽象物属性、内部结构情节属性。物理属性最丰富类型最多,包括材料属性、颜色色调属性、亮度透明度属性、形状属性、大小属性、物态属性、重量属性、温度属性、浓度纯度属性、气味道属性、音质属性、硬度强度密度属性、光洁平整度属性、新旧完整度属性、导电导热耐磨性属性、弹性延展性松紧性属性和型式构件属性 18 种。化学属性包括酸碱性属性、可燃性属性、耐腐蚀性属性、化合性质属性四种。具体物社会属性(人除外,因为人在生物属性中单列了一个“人的社会属性”)包括品牌属性、质量属性、价值属性、价格属性、功能属性 5 种。生物属性包括生理属性、心理属性、人的社会属性三种。抽象物属性包括抽象物状态、抽象物评价、抽象物范围、抽象物类型、抽象物一般属性。内容结构情节属性涵盖的范围比较大。状态部分包括事情状态、具体物状态(人部分除外)、时空状态三种。关系属性则包括社会关系、位置关系、时间关系、逻辑关系、因果条件目的关系、比较关系、部分全体关系、包含关系等 8 种。

下面是部分槽类型的定义和例子(工作单填写规范中的部分内容):

(3) 领属:系指有、所属的关系;通常用人称代词、疑问代词、名词或名词词组来充当。

例如: 我的一生, 谁的山地车, 中国民航, 孩子们的节日, 老师的职责, 国家的主权, 海水的深度

(4) 来源:系指来源、出处:

例如: 农村来的孩子, 商务印书馆出的《现代汉语词典》, 进口彩电, 姥姥为我买的花裙子, 来自首都科技战线的近百位科技工作者, 每年都要接收抢救来自北京市区、郊区及周边的地区的数十例重度中毒患者, 与李素梅配合的是来自山东的新秀, 23 岁的程晓燕

(5) 去向:系指去向、目标:

例如: 到日本去的中国留学生, 走向罪恶深渊的他, 出口产品

(6) 数据属性:区别数量的多少、规模的大小、篇幅的长短等;通常由数量词、表示规模的形容词或词组来充当。

例如: 一个人, 两台计算机, 五尺布, 6 斤米, 七里路, 长达 8 万字的学术论文, 大量农村人口流入城市, 大批上山下乡知青返城, 大规模集成电路, 三分之二的员工, 75% 的人员, 超过三分之二的经验年龄在 35 岁以下, 到 1998 年 6 月“箭”导弹已进行了几次不成功的实验

(7) 顺序属性:区别序次;由表示序次的“第、初”等与数词与或数量词组合或单独由表示序次的词例如“首先、最后”来充当。

例如: 第七天, 第三位, 本世纪最后也是最大的一次亚洲体育盛会, 首先进入体育场的是中国运动员, 开始了本届亚运全第一枚金牌的争夺战, 迎着朝阳的第一线光辉, 第一道天险, 中国共产党第十一届三中全会会议, 中国第三次国内革命战争

5. 用语类和槽类型表示槽关系的一个例子

愿望 语义分类: {感觉心境} 释义: 希望将来能达到某种目的的想法。

词组实例 1: 没有实现的中国人民以各种途径各种方式表达的承办两千年奥运会的美好的强烈愿望

定语 1 没有实现(的) 语类: {VP} 槽类型: [抽象物状态]

定语 2 中国人民以各种途径各种方式表达(的) 语类: {S} 槽类型: [来源]

定语 3 承办两千年奥运会(的) 语类: {VP} 槽类型: [内容结构情节属性]
 定语 4 美好的(的) 语类: {A} 槽类型: [抽象物评价]
 定语 5 强烈 语类: {A} 槽类型: [抽象物评价]
 槽类型顺序: [抽象物状态][来源][内容结构情节属性][抽象物评价][抽象物评价]
 可能顺序: [抽象物状态]*[来源][内容结构情节属性][抽象物评价]*

词组实例 2: 中国与南亚各国人民的共同的进一步发展经贸合作关系的这一迫切愿望

定语 1 中国与南亚各国人民(的) 语类: {NP} 槽类型: [领属]
 定语 2 共同的(的) 语类: {A} 槽类型: [抽象物评价]
 定语 3 进一步发展经贸合作关系的(的) 语类: {VP} 槽类型: [内容结构情节属性]
 定语 4 这一 语类: {DL} 槽类型: [指量属性]
 定语 5 迫切 语类: {A} 槽类型: [抽象物评价]
 槽类型顺序: [领属][抽象物状态][内容结构情节属性][指量属性][抽象物评价]
 可能顺序: [领属][抽象物状态]*[内容结构情节属性]*[指量属性][抽象物评价]*

词级例 3: 傅老昨天在会上转达的深受欢迎的……积极愿望

定语 1 傅老昨天在会上转达(的) 语类: {S} 槽类型: [来源]

⋮

槽关系表达式联想: [来源][领属][指量属性][数量属性][内容结构情节属性]*[抽象物评价]

6. 结语

名词槽关系研究在国内外都是罕见的,尤其是象现在这样以语义描述为主、句法描述为辅的描写方法还是第一次,况且还是 3000 个名词的工程性描写,又是计算词典学方法与传统词典学方法相结合,因此技术难度大,工作量也大。例如名词词组实例的大量抽取与人工分析都有技术难度,工作量巨大。目前我们正在进行现代汉语名词槽关系系统的软件设计和填写实验,希望将来把此槽关系系统与我们原来的现代汉语述语动词机器词典(论旨网格描述)集成起来,联合给大家提供丰富可靠的词义分析、句义分析的语义信息和句法信息,成为汉语语义分析的共享资源。

参 考 文 献

- [1] 陈群秀:“信息处理用现代汉语语义分类体系的设计思想”,《计算机时代的汉语和汉字研究》,清华大学出版社,1996年11月。
- [2] 林杏光,张庆旭:“现代汉语槽关系研究”,《语言工程》,陈力为、袁琦主编,清华大学出版社,1997年10月。
- [3] 张卫国:“三种定语、三类意义和三个槽位”,《中国人民大学学报》,1996年第4期。
- [4] 朱德熙:《定语和状语》,上海教育出版社,1984年。
- [5] 陆丙甫:“定语的外延性、内涵性和称谓性及其顺序”,《语法研究和探索》,北京大学出版社,1988年。
- [6] 刘月华等:《实用现代汉语语法》,外语教学与研究出版社,1983年第一版。
- [7] 黄伯荣,廖序东主编:《现代汉语》,甘肃人民出版社,1983年第三版。
- [8] 鲁川:“谓词框架说略”,《汉语学习》,1992年第4期。