



高校社科文库
University Social Science Series

教育部高等学校
社会科学发展研究中心

汇集高校哲学社会科学优秀原创学术成果
搭建高校哲学社会科学学术著作出版平台
探索高校哲学社会科学专著出版的新模式
扩大高校哲学社会科学科研成果的影响力



言说动词概念语义 网络系统构建研究

肖 珊/著

Research on Constructing a Modern Chinese
Speech-act Verbs System Based on
Conceptual Semantics

光明日报出版社



高校社科文库
University Social Science Series

013069127

教育部高等学校
社会发展研究中心

H146.2
25



汇集高校哲学社会科学优秀原创学术成果
搭建高校哲学社会科学学术著作出版平台
探索高校哲学社会科学专著出版的新模式
扩大高校哲学社会科学科研成果的影响力



言说动词概念语义 网络系统构建研究

肖 珊/著

Research on Constructing a Modern Chinese
Speech-act Verbs System Based on
Conceptual Semantics



北航

C1676439

光明日报出版社

H146.2
25

图书在版编目(CIP)数据

言说动词概念语义网络系统构建研究 / 肖珊著. —北京:
光明日报出版社, 2013. 7

(高校社科文库)

ISBN 978-7-5112-5000-1

I. ①言… II. ①肖… III. ①汉语—动词—语义学—
研究 IV. ①H146.2

中国版本图书馆 CIP 数据核字 (2013) 第 157458 号

言说动词概念语义网络系统构建研究

著 者: 肖 珊

责任编辑: 曹美娜

责任校对: 娜仁·格日乐

封面设计: 小宝工作室

责任印制: 曹 净

出版发行: 光明日报出版社

地 址: 北京市东城区珠市口东大街 5 号, 100062

电 话: 010-67078251 (咨询), 67078870 (发行), 67078235 (邮购)

传 真: 010-67078227, 67078255

网 址: <http://book.gmw.cn>

E-mail: gmcbs@gmw.cn caomeina@gmw.cn

法律顾问: 北京天驰洪范律师事务所徐波律师

印 刷: 北京楠萍印刷有限公司

装 订: 北京楠萍印刷有限公司

本书如有破损、缺页、装订错误, 请与本社联系调换

开 本: 690×975 1/16

字 数: 240 千字

印 张: 13.5

版 次: 2013 年 11 月第 1 版

印 次: 2013 年 11 月第 1 次印刷

书 号: ISBN 978-7-5112-5000-1

定 价: 35.00 元

版权所有 翻印必究



序

计算机应用的发展、互联网的广泛使用及智能网时代的到来,给语言学、计算机科学和相关学科,都提出了新的需求、挑战和研究方向。本书是适应这种需求和方向,结合相关学科的最新成果,为智能语义网建设和发展而做的开拓性的语义研究。

网络的智能化发展,一个重要的前提,就是用最经济的模式告诉机器足够多的语言知识,特别是语义知识,让机器理解词语的意义,进而进行句子理解和逻辑推理。并且,只有让机器理解的语义知识和资源越多、越成熟,语言信息处理和网络的智能化程度就越高。

动词是句子和篇章自然语言理解的最重要的核心内容之一。一种语言的动词数量相当多,但是其类型是有限的。怎么从语义的角度,结合人类的认知实际,用新的理论和方法,对其进行一群词、一片词或一族词的语义构成和类别的系统分析、描写和刻画,成为词汇语义研究新的方法和趋势。本书以人体五官中与“口”有关的言说表达动词为对象,进行了面向计算机或语言信息处理,推进网络智能化发展的有益探索和类群的系统研究,其成果是十分难能可贵的,其理论价值和现实意义也是不言而喻的。

在现代汉语学习和使用中,小孩最先学习或习得的言说动词,一般是“说”,并且是以“说”为认知基础,逐步建立了汉语言说动词的系统,完成了这类动词所表事物和行为的系统认知。本书以动词“说”为研究起点,以“词群-词位变体”理论和“概念基元结构”理论为工具,全面系统地进行了汉语言说动词的语义网络建构和概念语义的研究,同时在研究中不囿于既有的研究结论和理论构成,还从一个方面推进了“词群-词位变体”理论的发展。

此书的成果不仅是面向机器的,也是面向对外汉语教学的。在谈到面向机器和对外汉语教学研究的关系时,著名语言学家邢福义先生曾说过这样一段话:机器是没有语感的外国人,外国留学生是有语感的机器。意思是说,机器



和外国人学习语言具有内在的一致性，好的语言研究，是人机共享的。肖珊的《言说动词概念语义网络系统构建研究》，不仅是面向计算机和智能语义网的研究和探索，更是面向对外汉语教学的新的尝试，并且该书从另一个方面展示了第二语言之教学和研究新的方法和思路。

本书是作者博士毕业论文，在其毕业后不久就能付诸出版，实乃可喜可贺。作为她的博士论文的指导教师，更是由衷感到高兴。在此书出版之际，写了上面一些与其著作内容相关的话，献给读者和作者，是为序。

萧国政

2012年6月18日于武汉大学珞珈山



前言

当今社会是一个网络信息的时代，自互联网之父蒂姆伯纳斯·李（Tim Berners-Lee）80年代提出万维网（WWW）构想以来，互联网进入飞速发展阶段，各种网络技术日新月异、应用层出不穷，互联网已经融入人们的日常生活、工作和学习，成为人们相互交流、沟通和互动不可或缺的平台，网络为我们开启了一个全新的世界，而也正是这个世界信息的极大丰富和爆炸式发展，在为人们生活、生产、学习等各方面带来便利的同时，人们也对其提出了更高、更完善的需求。

智能网络的建设和实现需要自然语言处理技术的支持，对于我国而言，就是要利用计算机对各种汉语信息特别是语义信息进行一种自动化处理。汉语是一种特殊的语言，尤其表现在词汇方面（自动分词比英语等拼音文字难度大得多），成为制约中文信息处理发展的瓶颈，但也正是这种特殊性的存在使得在语义信息处理上更要关注词语语义的本身，细化语义描写的颗粒度，尤其是语义内部成分、要素的联系和区别，而汉语词汇的内部规则与句子的语法规又存在着相似的一致性，因此我们可以说一旦提高了词汇语义的处理技术，势必会为句子的处理带来突破性的意义。

在汉语词汇研究学界，一直以来都是以动词的研究为主，这不仅仅是因为动词在数量上比较庞大，而且也因为动词是句子中衔接其他语义成分的纽带，其语义表现特征最为典型、研究价值比较大，因而建设词汇语义资源知识库就应该从动词开始。当然动词的类别及语义特征各不相同，要选取一个好的入手点和带普遍意义的对象范畴也不是一件容易的事情。

言说动词是动词类中有代表性的一类，其中“说”最为典型、最有概括性。对它感兴趣是从很简单的语言用法开始的，如“他说了一句”、“他说了我一句”词汇形式都是一样的，但是两个“说”的语义却完全不同，前一个“说”是“说话”，后一个“说”是“批评、责备”，因此我们可以说后者是



“说”在基本语义上的延伸；而我们再进一步扩大研究范围，发现无论何种句子语境下的言说动词（非“说”这种词语形式）都可以用“说”附加一定的限制条件这样的规律来进行替换，这一规律也几乎可以囊括所有的言说动词。也就是说我们可以构建一个以认知起点词“说”为语义基点的、较为完整的现代汉语言说动词概念语义网络系统。

随着研究的深入，我们发现言说动词的语义还可以按照包含的语义结构的对立分为三大类：一是只凸显说话动作行为本身的言说动词；二是体现说话动作行为的各种性质（诸如传息、回馈、情感等）的言说动词；三是凸显采取某种方式方法（如交互、介入等）来进行说话行为的言说动词，这三大类将整个网络系统分为三大词群，即：一般言说动词同位词群、性质言说动词同位词群和方式言说动词同位词群；但我们说词也不乏“例外”，总有一些调皮的言说词汇要“跳出”这些词群之外，自成一体，这些词之间有语义上的交叉，交叉的语义又组成新的言说动词，新的言说动词则同时具有两个或以上的语义，如：“商讨”是并列式的“商量”与“讨论”的语义交叉；“控诉”是连动式的“陈述”后“请求”的语义交叉等等，但是不论语义如何变化和跳脱，所有的言说动词都还是建立在以“说”为概念语义的立足点和生发点之上的。

拙作是在博士毕业论文的基础上修改、完善而成的，毕业不久就能得到部分资助出版的机会，个人甚为高兴。所以首先要感谢教育部高等学校社会发展研究中心以及光明日报出版社各位工作人员的支持，同时也要感谢我的博士生导师萧国政先生从开题直到毕业后论文的进一步修改上的指导，同门的师兄胡憺、师姐郭婷婷、李圃给我的文章修改也提供了帮助和指点，师妹胡莉也为我在毕业后能在母校翻阅最新资料给予了极大的协助。

概念语义网络系统的构建并非一朝一夕，言说动词的研究可以说只是“万里长征的第一步”，希望能起到“抛砖引玉”的作用，能使更多对语言信息处理感兴趣的有识之士及同仁来共同关注，帮助他们能有更多新的发现和发展，那就足以欣慰了。

作者

2012. 6. 19

于中国地质大学（武汉）南望山



CONTENTS 目录

导 言 / 1

第一节 选题缘由及背景 / 1

第二节 国内外研究现状 / 3

一、面向信息处理的词汇语义知识库研究 / 3

二、言说动词研究 / 9

第三节 研究目标及对象 / 13

一、研究目标 / 13

二、研究对象 / 14

三、语料及术语符号说明 / 16

第一章 言说动词同位词群的建构 / 17

第一节 同位词群建构的理论基础 / 17

一、“词群-词位变体”理论 / 17

二、“语义基元结构”理念 / 19

第二节 言说动词同位词群的构成 / 21

一、言说动词概念语义及“说” / 21

二、三大同位词群的构成及群位关系 / 25

三、语义交叉变体的提出 / 27



第二章 一般言说动词“说1”的基元结构及同位词群 / 31

第一节 一般言说动词“说1”的意义 / 31

一、“说1”的同位变体 / 31

二、“说1”的下位变体 / 34

第二节 “说1”的语体变体——叙述类词群及基元结构 / 35

一、叙述类言说动词的意义 / 35

二、“叙述”的同位变体 / 36

三、“叙述”的下位变体 / 37

第三节 “说1”的过程变体——开口类词群及基元结构 / 39

一、开口类言说动词的意义 / 39

二、“开口1”的同位变体 / 40

二、“开口1”的下位变体 / 41

第四节 “说1”的频率变体——唠叨类词群及基元结构 / 42

一、唠叨类言说动词的意义 / 42

二、“唠叨”的同位变体 / 43

三、“唠叨”的下位变体 / 45

第五节 “说1”的音强变体——叫喊·低语类词群及基元结构 / 45

一、强音叫喊类词群及基元结构 / 45

二、弱音低语类词群及基元结构 / 52

第六节 “说1”类语义基元结构及词群概览 / 55

第三章 性质言语动词“说2”的基元结构及同位词群 / 56

第一节 “说2”的传信息变体——告知类词群及基元结构 / 56

一、告诉类词群及基元结构 / 57

二、透露类词群及基元结构 / 60

第二节 “说2”的内容变体——誓谎类词群及基元结构 / 62

一、誓言类词群及基元结构 / 63

二、谎言类词群及基元结构 / 66

第三节 “说2”的祈行变体——使令类词群及基元结构 / 68

一、请求类词群及基元结构 / 69

二、劝告类词群及基元结构 / 72

三、召唤类词群及基元结构 / 76



四、命令类词群及基元结构	/ 80
五、警告类词群及基元结构	/ 82
第四节 “说2”的祈言变体——询疑类词群及基元结构	/ 84
一、一般问类词群及基元结构	/ 85
二、慰问类词群及基元结构	/ 88
三、探问类词群及基元结构	/ 90
四、盘问类词群及基元结构	/ 92
第五节 “说2”的回馈变体——应答类词群及基元结构	/ 95
一、一般回应类词群及基元结构	/ 95
二、肯定回应类词群及基元结构	/ 98
三、否定回应类词群及基元结构	/ 104
第六节 “说2”的观点变体——评价类词群及基元结构	/ 109
一、评价类词群及基元结构	/ 109
二、称赞类词群及基元结构	/ 111
三、批责类词群及基元结构	/ 114
第七节 “说2”的情感变体——致意类词群及基元结构	/ 120
一、道谢类词群及基元结构	/ 120
二、道歉类词群及基元结构	/ 122
三、道喜类词群及基元结构	/ 124
四、骂类词群及基元结构	/ 125
第八节 “说2”的论述变体——解·说类词群及基元结构	/ 128
第九节 “说2”的量度变体——繁简类词群及基元结构	/ 132
一、详说类词群及基元结构	/ 132
二、简述类词群及基元结构	/ 133
三、赘述词群及其基元结构	/ 133
第十节 “说2”的受损变体——嘲诬类词群及基元结构	/ 134
一、嘲笑类词群及基元结构	/ 134
二、诬蔑类词群及基元结构	/ 138
第十一节 “说2”类语义基元结构及词群概览	/ 141
第四章 方式言语动词“说3”的基元结构及同位词群	/ 145
第一节 “说3”的交互变体——论辩类词群及基元结构	/ 145



一、交谈类词群及基元结构	/ 146
二、讨论类词群及基元结构	/ 150
三、商量类词群及基元结构	/ 154
四、辩论类词群及基元结构	/ 157
五、争吵类词群及基元结构	/ 161
第二节 “说3”的介入变体——插接类词群及基元结构	/ 165
一、插话类词群及基元结构	/ 166
二、接话类词群及基元结构	/ 168
第三节 “说3”的策略变体——“宣引”类词群及基元结构	/ 170
一、宣布类词群及基元结构	/ 171
二、引用类词群及基元结构	/ 174
三、转述类词群及基元结构	/ 174
第四节 “说3”类语义基元结构及词群概览	/ 176
第五章 语义交叉变体	/ 177
第一节 语义交叉变体的类型	/ 177
一、言语内交叉	/ 178
二、言语外交叉	/ 179
第二节 语义交叉方式	/ 180
一、并列式交叉	/ 180
二、偏正式交叉	/ 181
三、连动式交叉	/ 182
四、选择式交叉	/ 183
五、叠式交叉	/ 184
第三节 语义交叉变体归类	/ 185
第六章 结 语	/ 189
参考文献	/ 192
后 记	/ 201



导 言

第一节 选题缘由及背景

21 世纪的到来是信息化社会全面发展的时代,随着计算机和互联网应用的普及和深入,人们越来越觉得计算机科学与人工智能研究带来的一系列成果正逐步改变着社会的生产和生活,计算机科学成为 20 世纪以来发展最快而且对人类影响最大的学科。

然而在如今的世界范围内,信息量之大、信息传递速度之快是以往任何时候都无法比拟、无法企及的,我们每一个人都好像是一叶扁舟飘荡在无边的信息海洋之中,单凭一己之力断不可能全面地掌握到自己所需要的全部信息,当然就更谈不上对这些信息按照不同的个人需求做一步处理,因此采用何种手段来高效地获取、处理和运用信息显得至关重要。

网络信息作为当今社会极其重要的资源之一已渗透到人类生活中的方方面面。语义网是第三代网络的表现形式,2001 年由万维网(简称 WWW 或 Web)的创始人蒂姆伯纳斯·李(Tim Berners-Lee)正式提出的,设想构造一种能够辨认、识别和自动处理人类语言的智能网络,具有人一样的判断和推导能力。例如:它可以让计算机分清“doctor”这个单词意思何时是“医生”何时是“博士”并扩展到“医院”“护士”“打针”“学位”等一系列密切相关词群上;在读者看网络新闻时,它能够轻松、迅速、准确地从成千上万的网页中过滤出读者感兴趣的标题、导语等。也就是说它是万维网的延伸和扩展,其目的就是为了解决目前计算机不能理解网页内容的语义和网上有用信息查准率较低的问题。对于我国而言,就是要利用计算机对各种汉语信息进行一种自动化处理。许嘉璐说:“中文信息自动化处理每提高一步,给我国的科学技术(包括国家的信息化)、文化教育、经济建设、国家安全所带来的效益,将是无法用



金钱的数额来计算的。”^①

网络的建设和实现需要自然语言处理技术的支持。20 年以来,我国的计算机专家和语言文字学家在中文“汉字处理”上曾付出了不少的心血,也取得了不俗的成绩。然而随着时间的发展,计算机专家与汉语学家发现“字处理”并不能解决汉语语言处理的一切问题,中文信息处理要达到相当程度的自动化必须关注机器翻译这一块,要想机器理解自然语言传递的信息,一个重要而根本的前提就是我们应该怎样用最经济的模式告诉机器足够多的语言知识,在许多专家都感到基于概率统计、基于规则等分析技术已经不可能达到这一目标的时候,以词义为基础的语义资源建设就应运而生,并且也越来越受到计算机学界和语言学界的重视。之所以以词义为基础,不仅仅由于汉语词汇本身的特殊性(自动分词比英语等拼音文字难度大得多),而且也恰恰由于这种特殊性致使信息处理上更要关注词语语义的本身,从本质上而言,汉语的词汇内部规则与句子的语法规则存在着相似的一致性,因此词汇本身处理技术提高了,自然会为句处理带来突破性的意义,同时深入分析词汇本身的语义,尤其能帮助细化语义描写的颗粒度,也必然为计算机的自动化处理与“习得”起到巨大的促进作用。

现代汉语的词类分实词和虚词两大类,实词当中又以名、动、形三类占绝大多数,而这三类中又以动词的数量和语义表现特征最为典型、研究价值最大,所以词义资源知识库的建设首要就是研究动词。当然,动词的研究范围相对宽泛,限于篇幅应有所选取。我们知道,人由“眼”“耳”“鼻”“口”“身”来感知世界,分别对应人体的五大机能“视觉”“听觉”“嗅觉”“味觉”和“触觉”,相应也就构成了人体活动的基本词汇(如“看”“听”“闻”“吃”“打”等)及与它们相关的同义词词群。这其中,“口”的功能十分特殊,它除了有直接感知事物(主要是消化食物)的基本功能之外,还有更高级的功能——言语表达功能,是最能代表人类区别于一般动物特征的、体现人类社会性的沟通方式,因而既具有普遍性又带有特殊性,映射到词汇系统上就是以“说”为代表的一系列错综复杂的言说动词。作为人类言语系统研究的最重要组成部分,言说动词语义研究也自然成为语言信息处理的焦点。它的研究有助于从一个侧面探讨自然语言是如何生成和被理解的,既可以帮助语言学

^① 参见许嘉璐:《现状和设想——试论中文信息处理与现代汉语研究》,《中国语文》2000 年第 6 期,第 490 页。



家更加准确地进行语义结构的形式化描写,也可以帮助计算机学者对动词系统乃至整个自然语言系统的计算机模拟起到启发和推导作用,与此同时也是一项对更好地建设词汇语义知识库的探索性工作。

第二节 国内外研究现状

一、面向信息处理的词汇语义知识库研究

词汇语义资源的建立源于西方,迄今已有不少学者基于不同的目标研制了多种类型、满足不同需求的词汇语义知识库,它们各自发掘出概念间的不同语义关系,并以此构建出有不同应用价值的语义聚合网络。有必要先对在世界上影响比较大的、有代表性的国外词汇语义知识库做扼要的介绍和分析,然后再谈它们对中文信息处理技术特别是汉语语义资源建设工程的启发和影响,以及国内学者在这一方面所取得的相关成就。

(一) 国外研究现状

1. WordNet

截至目前,语义知识库研究方面影响最大、成果最显著的当属普林斯顿(Princeton)大学乔治·米勒(George A Miller)等人研制的英语词网数据库WordNet。WordNet体系简洁、内容翔实,已经成为自然语言处理领域里最重要的公用语义资源和语义知识库方面的国际标准。WordNet是一个基于心理语言学原则的机器词典,但它又不是一部词典,它最大的区别就是用大家熟悉的拼法来表示词形,用同义词集Synsets(在一定上下文中可以互换的同义词形的列表)来表示词义,用词义代替词形组织词汇信息,包含同义、反义、上下位义、整体部分义等多种语义关系的描写,其目的是想为广大读者提供依概念而不是依字母顺序查找词典词汇语义知识方面的帮助。

在WordNet当中,收录了约11000个动词,动词在语义准则的基础上被分成15个文件,很多文件的名字都是来自于这些语义上一致的词汇群中为首的最高层的动词,这些最高层的动词功能作用类似于“核心动词”,再通过语义关系从它们推导出构成该语义领域的其他动词,动词与动词之间按特定的语义关系(转精、反义、推演和致使)连接起来,这种语义关系被称为一种推演(蕴涵)关系。但由于WordNet并不把词语分解成更小的有意义的单位,它认为这是义素分析法(Componential Analyses)应该关注的,而且WordNet也不



是在文本和话语篇章水平上来描述词和概念的语义,因此 WordNet 中没有包含词语在特定的篇章话题领域的相关概念关系,即就词语来分析词语。由于只有词语知识而没有世界百科知识,许多 WordNet 的用户都对 WordNet 中缺乏跟语义处理的细节相匹配的句法信息而感到遗憾,而我们知道句法对动词而言最为重要,目前, WordNet 的每个动词同义词集中包含了及物性和论元类型的基本信息,但有关这些论元的性质的细节就很少提到。这反映了 WordNet 结构不甚严谨之处,并且语义颗粒度较大,应用范围也有限。

2. FrameNet

20 世纪末美国语言学家菲尔墨 (C. J. Fillmore) 在框架语义学的基础上提出了框架网络 (FrameNet), 框架语义学的中心思想是词的意义描述必须与语义框架相联系, 那么 FrameNet 的建立就是通过词元、框架、框架内的元素、语义类型等各项语义信息揭示词汇的本质属性并抽象为概念, 通过逻辑关系使这些概念关联起来, 使得研究者将具有共同认知结构、支配相同类型的语义角色的一类词语集中用一个框架描述, 同时构建框架与框架之间的静态、动态的复杂关系, 最终形成一个形式化可共享的概念网络体系。在所给定的涵义下, 词汇单元的框架语义要描述框架元素的结合方式和在框架中的分布情况, 这种“题元角色关系”^① 包括继承关系、透视关系、总分关系、先后关系、起始关系、致使关系、使用关系、参照关系。

与 WordNet 相比, FrameNet 的先进之处在于框架与词汇之间、框架与框架之间、框架元素与框架元素之间的对应语义关系更为细致丰富, 通过这种语义联系, 使得计算机在明确词汇含义、确定词汇所属的框架类型, 进而在词义消歧、机器翻译、信息抽取或问答系统方面能发挥不小的作用, 无论对于语言理论研究自还是然语言处理都有着十分重要的意义。

在 FrameNet 当中, 在动词分类上, FrameNet 分为 13 个类别的领域, 60 个框架, 共 1905 个词项, 这些词既有类别的说明, 也列举了其他同类的词项, 指出了词汇与框架的具体细致关系, 每个动词的框架语义表达都包含了所有语义和语法组合可能性的规则, 与此同时标出语料库中的例句及说明、指出句型。虽然 Framenet 中关于词汇与框架关系描述较充分, 但是具体到每一个动词本身, 也没有进行更细致的意义上的分解, 也没有对一类同义或近义的词元

^① 冯志伟:《从格语法到框架网络》,《解放军外国语学院学报》2006 年第 29 卷第 3 期,第 1 页。



之间的关系进行刻画,寻找区别和联系。同时,FrameNet 中的动词类别相对有限,且对如 be、belong、include 等静态动词这块并无说明,例句部分也未指出句型特征,对于不合语法的句子也没有显示出来。除此之外,FrameNet 的动词这部分也有整个 FrameNet 所面临的疑题:能覆盖绝大多数的文本资料的动词语义框架是否有限,框架起止的标准,动词所涉及的情景数量及覆盖面,情景与框架、框架与动词之间的对应关系的限制条件是什么等等,这些都是 FrameNet 中存在、尚未解决而又不可忽视的问题。

3. VerbNet

到了 21 世纪,美国宾夕法尼亚大学的玛莎·帕默(Martha Palmer)等教授开始建构 VerbNet,一个在线的英语动词词典。VerbNet 的创建基础是贝斯·列文(Beth Levin)的动词分类标准,他认为句法框架能直接反映潜在的语义,动词的词义会影响其句法的行为,因此对动词的句法进行更为详细的研究,主要是依据动词是否能在成对的句法框架中出现来划分动词的类,它与 WordNet 中的同义词集相连,将动词的类更加细化,使每一个动词的子类提供更多的语义特征和句法特征。

VerbNet 最大的优势和特色就是更清晰地分析语义成分、谓词论元结构和与单个的动词有关的句法模型,特别是清楚地体现了动词语义与句法行为之间的关系,具体来说包括成员集合、句法框架、句法框架中包含的语义谓词、题元角色对每个句法框架中论元的选择限制等;整个动词词典系统中动词大类与子类的层级关系明确,句法框架则在其基础上列举了动词成员的所有可能的句法形式,并赋以特定的标签,框架中的语义谓词则显示论元及其事件之间的关系,题元角色的论元限制将现有的论元同已公开的层次本体相关联,确定其语义类型,也就是说 VerbNet 并非重新建构新的系统,而是整合了现有的几个构建得比较好的系统如 WordNet、FrameNet 等,结合它们的优势,改进它们的不足。

在 VerbNet 中,它将具有共同句法及题元角色的动词都囊括在同一类中,动词类的分类以及类与子类之间的上下位语义关系描写比较充分、细致,可这样划分出来的动词类就并非同义动词集(Synset),而只是拥有相似句法表现的动词聚类,因此就不可能将动词与动词之间的语义关系表现得如 WordNet、FrameNet 中那么明显,更不用说动词本身更为深入细致的语义特征描写了,由于它的语义主要都体现在句法中,所以它的研究对象不在词语本身而在句法行为。当然由于本身研究对象锁定为动词,所以它词汇的覆盖面也当然不及 WordNet 和 FrameNet 那么广。



总而言之,就动词的语义关系来说,WordNet 提供了动词类别中具有同义、推演等关系方面的知识,FrameNet 提供了共同的情景下的动词关联起来的方式知识,VerbNet 提供具有相同的句法语义特征的动词的关联知识。

(二) 国内研究现状

1. 知网 (HowNet)

1999 年由中国科学院、计算机语言信息工程研究中心的董振东教授等人于在网上发布了多年的研究成果——词汇语义资源知识库“知网 (HowNet)”,他们定义这个知识库的描述对象为汉语和英语的词语概念,揭示了概念与概念之间以及概念所具有的属性之间的关系。

知网的构建不是像国外的语义知识库那样建立在某种具体的已有的理论基础之上的(比如构建 FrameNet 的框架理论学),而是建立在人类普遍的哲学思想、对于世界知识的感知基础上。知网有自己独特的哲学认识,它认为世界上一切事物(物质的和精神的)都在特定的时间和空间内不停地运动和变化,它们通常是从一种状态变化到另一种状态,并通常由其属性值的改变来体现。知网的运算和描述的基本单位是:万物(其中包括物质的和精神的两类)部件、属性、时间、空间、属性值以及事件,同时,它也认为知识是一个包含了概念与概念之间的关系以及概念的属性与属性之间的关系的系统,无论对于个人还是对于全人类,这个系统是无限开放的,意义和知识寓于关系之中,因而只有从关系入手才能理解、表达意义和知识,对人、对计算机都是如此。就是基于这样的思想指导之下,整个知网所体现出来的特点就是:把概念分解为最小的意义单位“义原”,用义原代替义项来描述概念,确定七类顶层概念,从具体的语言文本中获取 90 个细致且明确的语义角色,并加之采用自行设计的知识数据描述语言(KDML)使每一个词语的概念描述都丰富和精确,大大提高词语意义的计算能力。

正是因为知网是自足的、自我完善的,它有一套设计好的分类体系,对动词来说,它有自己的动词框架和相应的概念描述语言与之配套,反映每个动词概念之间的共性和个性,动词与动词之间有相应的属性值差别能够较好地反映动词间的联系和区别。但就如知网研发者自己所说的那样,“关系是知网的灵魂”^①,知网中对于词语之间的关系尽管描述细致且准确,但它对于

^① 参见董振东,董强,郝长伶:《知网的理论发现》,《中文信息学报》2007 年第 21 卷第 4 期,第 5 页。