



现代汉语维度形容词语义 及其计算机识别研究

伍莹 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

教育部人文社科青年基金项目(编号:12YJC740113)最终成果
湖南科技大学学术著作出版基金资助出版



现代汉语维度形容词语义 及其计算机识别研究

伍莹 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

现代汉语维度形容词语义及其计算机识别研究/伍莹著. —武汉: 武汉大学出版社, 2015. 8

ISBN 978-7-307-16333-1

I. 现… II. 伍… III. ①现代汉语—形容词—语义—研究 ②现代汉语—形容词—自动识别—研究 IV. ①H146.2 ②TP391.43

中国版本图书馆 CIP 数据核字(2015)第 157177 号

责任编辑: 鲍 玲

责任校对: 汪欣怡

版式设计: 马 佳

出版发行: 武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷: 虎彩印艺股份有限公司

开本: 720 × 1000 1/16 印张: 14.75 字数: 209 千字 插页: 1

版次: 2015 年 8 月第 1 版 2015 年 8 月第 1 次印刷

ISBN 978-7-307-16333-1 定价: 36.00 元

序

空间概念是人类最早产生和使用的基本概念之一，是人类思考和理解时间及其他概念的基础。在哲学和自然科学史上，人类对空间的理解涉及不同角度和研究指向。若以物体为参照，人类的空间概念由物体存在空间（所处空间）、物体自身空间（自空间）及空间关系三部分构成。自有语言以来，客观存在的空间及关系经由人类大脑认知处理后，投射到语言中，其基本部分就形成了一个封闭的空间词汇系统。在词汇系统中，表示事物自身空间的词汇通常是描述空间维度的形容词及指称空间概念的处所词（即地名、物体名、单位词等），表示物体空间关系的词汇通常是描述位置关系的方位词和起止空间基元的趋向动词。

空间词语一直是汉语界十分关注的对象，就现代汉语而言，20世纪90年代就有齐沪扬的《现代汉语空间问题研究》、储泽祥的《现代汉语方所系统研究》等。但语言学界的研究一般主要集中在处所词和方位词，而对其他空间性词汇及其空间性的关注则不多。进行系统、全面的空间性探索，以及获得本体研究和应用研究结合起来的成果，一直为语言学和计算语言学界所期待。伍莹的《现代汉语维度形容词语义及其计算机识别研究》一书，是现代汉语空间词语、词语空间要素及其系统研究方面的一个新的进展和突破，并且该书在立论高度、理论深度、知识丰度、解释力度和应用宽度等多个方面，均给人以深刻的印象。

阅读该书我们可以看到，空间维度词包括空间维度形容词、空间维度名词和空间维度量词（米、平方米、立方米），等等。“维”是几何学及空间理论的基本概念，“维”是度量空间的，因而称之为“维度”。几何学一般将维度四分，即零维、一维、二维、三

维。“长”、“宽”、“高”各是一维。零维空间是一个点，无限小的点，在理论上不占任何空间。严格说来，言语作品中所言之“点”一般有大小，占有空间，否则就无法为感官感知和存在。如语言中所言之“线”、“面”是生活、生产和大自然中的事物，因而再细的线都有粗度，再薄的蝉翼都有厚度，即均关系到三维。认知的“图式化”（理想化和抽象化）把语言中所言之物“图示”为零维物体（如星星点点）、一维物体（如米线、毛线、绣花线）、二维物体（如桌面、路面、墙面）。研究不难发现，是认知的加工和中介作用，打通了几何学与语言的“维度”概念和空间观念。维度不仅是物体诉诸人们视角的首要几何特征，而且也是人类将语言所言对象的维度与几何学维度进行“同型同构”认知的客观依据和解释工具。在空间认知中，语言采用了图示“忽略”和“凸显”两种基本策略。如语言对物体零维的表达，就是将其大小忽略不计，图式化为不占空间的点，以凸显和极言其小，如“像针尖、像灰尘”；语言一维物体表达，是凸显其一维而忽略其他维度，如棍子、树、铅笔、藤、绳子、蛇、河、路等通常是唤起对空间长度的心理印象，在汉语里一般搭配使用的量词是“根”、“枝”、“条”等；其二维物体表达，如纸、布、叶子、木板、毯子、门等，是将其两个维度给予了特别凸显，其常用配搭量词为主要显现二维图形“面”的“张”、“块”、“片”，等等；其三维物体表达，如楼、箱子、衣柜、盒子、汽车、球，是将三个维度都予以认知的彰显。^①

语言表达的空间维度类型，除零维外的维度还有三类，但本书从语言的真实语料及前贤和同世的研究中，概括出汉语维度表达基本形容词有7对：“大、小”、“高、低”、“宽、窄”、“深、浅”、“长、短”、“粗、细”和“厚、薄”，并运用认知语言学理论和若干相关理论，较为全面和详细地描写了这7对维度形容词运用于各类对象而形成的种种语义及其特征，其描述和论证给人很多启发，

^① 并且在这个意义上，应反过来说，几何学的物体思维是对语言物体思维的再抽象。

有的甚至是惊喜。而更难能可贵的是，本书在充分剖析语义的基础上，构建了7对维度形容词的计算机语义识别流程和模型，进行了由语言本体、语言理论向语言应用的延伸。我们知道，在自然语言信息处理中，最大的难题之一是对句子的语义识别，要解决这一问题，词义是句子语义的基础，对词义进行区别于传统工具书的描写和模型建构，对计算机识别自然语言是不可或缺的。

本书不仅介绍了维度形容词本身的语义，而且还阐述了与维度形容词搭配的名词的语义特征。这类研究，能给计算机维度形容词及相关组合的语义自动识别带来极大便利。此前，计算语言学关于语义自动识别已经触及动词、名词、形容词等的大类，而其发展需要并期待学界对形容词等大类词中的次类进行研究。本书是对形容词一个语义次类的研究，并且该次类及其他词类之次类的相关研究，是学界和技术界目前和今后一个时期内仍需进一步着力的重要课题。本书所进行的研究具有开拓性、创新性和系统性，其研究对于计算机自然语言识别，对于词汇语义学、词典编撰，对于对外汉语教学，都具有不言而喻的学术价值和应用意义。

此书即将付梓之际，笔者再次详读了全书，写了上面一些话，以求教于读者、作者和方家，是为序。

萧国政

2015年4月于武汉大学

前 言

空间由物体存在空间和物体空间关系构成。客观存在的空间经由人类大脑认知处理后形成了一个封闭的空间词汇系统。现代汉语空间词汇包括两类，一类是空间指称词，表示指称空间概念的词汇，主要有地名、物体名词、机构名词等处所词；另一类是空间描写词，表示描写空间面貌或空间关系的词汇，描写物体空间面貌的参数是空间大小和空间形状，反映在语言中表现为空间维度词和空间形状词；描写物体空间关系的参数是空间方位和空间距离，反映在语言中主要表现为空间方位词和表示距离的“远、近”。空间维度词包括空间维度名词、空间维度形容词和空间维度量词。空间维度形容词是描述物体各个维度特征的形容词，通常认为现代汉语共有8对维度形容词，分别为“大、小”、“高、低/矮”、“长、短”、“宽、窄”、“深、浅”、“厚、薄”、“粗、细”和“远、近”。本书只研究前7对维度形容词，不包括“远、近”。严格地说，“远、近”主要用来描述两地之间的空间距离或空间位置（即事物的空间关系），而不是描写事物本身的空间维度，这些维度形容词聚合起来形成了一个封闭的词汇系统，在这个系统中，维度形容词本身具有空间义和隐喻义，两者相互关联，各维度形容词的语义也相互关联，从而形成了一个小型的词汇语义系统。

本书选择现代汉语维度形容词作为研究对象，主要有如下几个方面的原因：

- ①汉语学界对描述空间关系的空间方位词研究较多；
- ②现代汉语维度形容词的语义相当复杂；
- ③计算机识别维度形容词语义的困难。

本书在前人研究的基础上，通过利用最充分的海量语料，详细

考察维度形容词和名词搭配的实际情况，探寻维度形容词和名词搭配的组合机制和语义限制，分析维度形容词的空间义及其语义特征，探索维度形容词隐喻义演变的机制和过程，构建维度形容词语义自动识别的计算机模型。具体的研究内容主要有三个方面：7 对维度形容词空间义的研究；7 对维度形容词隐喻义的研究；7 对维度形容词语义自动识别的计算机模型构建。其中，维度形容词空间义的研究是本书研究的重点。维度形容词的空间义是其隐喻义产生的基础，也是计算机语义自动识别模型建构的起点，因此，只有对维度形容词的空间义进行全面而深入的研究，才能准确而清晰地阐释维度形容词的隐喻义及其演变的过程和方式，才能使建构的计算机语义识别模型具有更有效的应用价值。

本书虽已完成并即将出版，但是还有一些特别有研究价值的课题因时间和精力有限没能深入考察和探究，如汉语维度形容词与其他语言中维度形容词的对比研究；维度形容词与其他量度形容词之间的语义关联分析等，希望今后与对这一课题感兴趣的同仁一起继续探讨和研究。

本书的研究成果得到了“2012 年教育部人文社会科学青年基金项目（编号：12YJC740113）”的资助和“湖南科技大学学术著作出版基金”的资助，在此一并感谢。同时还要感谢我的博士生导师萧国政教授在本书创作过程中给予的大力支持和悉心指导。

伍 莹

2015 年 5 月于湖南科技大学

目 录

第一章 绪论	1
第一节 空间、空间维度和空间维度词	1
一、空间与空间描述	1
二、几何学、语言学和现实世界中的空间维度	3
三、空间维度词的词性界定与意义分类	5
四、维度形容词的界定和分类	7
五、维度形容词在现代汉语形容词语义系统中的地位	9
第二节 理论基础	10
一、范畴理论	10
二、隐喻理论	13
三、语义范畴理论	16
四、体验哲学和人类中心论	17
第三节 维度形容词国内外研究综述	19
一、理论语言学界研究现状	19
二、计算语言学界研究现状	31
三、国内外维度形容词研究述评	31
第二章 “高、低/矮”语义及其计算机识别分析	35
第一节 “高、低/矮”的空间义	35
一、“高、矮”维度义的语义特征	36
二、与“高、矮”搭配的名词的语义特征	50
三、“高、低”的位置义	54
四、“低、矮”空间义的异同	55
第二节 “高、低”的隐喻义	57

一、数量域	58
二、等级域	59
三、评价域	61
四、心理域	62
五、时间域	62
第三节 “高、低/矮” 计算机语义识别模型构建	64
一、“高、低/矮” 计算机识别流程	64
二、“高、低/矮” 计算机语义识别模型构建	68
 第三章 “宽、窄” 语义及其计算机识别分析	69
第一节 “宽、窄” 的空间义	69
一、“宽、窄” 维度义的语义特征	73
二、与维度义“宽、窄” 搭配的名词的语义特征	86
三、“宽、窄” 面积义的语义特征	90
四、与面积义“宽、窄” 搭配的名词的语义特征	94
第二节 “宽、窄” 的隐喻义	96
一、范围域	98
二、心理域	98
三、政策域	100
四、经济域	100
第三节 “宽、窄” 计算机语义识别模型构建	101
一、“宽、窄” 计算机语义识别流程	101
二、“宽、窄” 计算机语义识别模型构建	105
 第四章 “深、浅” 语义及其计算机识别分析	106
第一节 “深、浅” 的空间义	106
一、“深、浅” 维度义的语义特征	107
二、与维度义“深、浅” 搭配的名词的语义特征	115
三、“深、浅” 的位置义	125
第二节 “深、浅” 的隐喻义	128
一、时间域	129

二、颜色域.....	131
三、知识域.....	132
四、情感域.....	132
五、数量域.....	133
六、程度域.....	134
第三节 “深、浅” 计算机语义识别模型构建	135
一、“深、浅” 计算机语义识别流程	135
二、“深、浅” 计算机语义识别模型构建	138
 第五章 “长、短” 语义及其计算机识别分析	139
第一节 “长、短” 的空间义	139
一、“长、短” 空间义的语义特征	140
二、与“长、短” 搭配的名词的语义特征	147
第二节 “长、短” 的隐喻义	153
一、时间域.....	153
二、篇幅域.....	153
第三节 “长、短” 计算机语义识别模型构建	154
一、“长、短” 计算机语义识别流程	154
二、“长、短” 计算机语义识别模型构建	156
 第六章 “厚、薄” 语义及其计算机识别分析	157
第一节 “厚、薄” 的空间义	157
一、“厚、薄” 空间义的语义特征	158
二、与“厚、薄” 搭配的名词的语义特征	159
第二节 “厚、薄” 的隐喻义	164
一、数量域.....	165
二、情感域.....	165
三、味觉域.....	166
四、质量域.....	166
第三节 “厚、薄” 计算机语义识别模型构建	167
一、“厚、薄” 计算机语义识别流程	167

二、“厚、薄”计算机语义识别模型构建	169
第七章 “粗、细”语义及其计算机识别分析	170
第一节 “粗、细”的空间义	170
第二节 “粗、细”的隐喻义和转喻义	180
一、原因转喻结果	181
二、结果转喻原因	182
三、声音域	182
四、性格域	184
五、品质域	184
第三节 “粗、细”计算机语义识别模型构建	187
一、“粗、细”计算机语义识别流程	187
二、“粗、细”计算机语义识别模型构建	189
第八章 “大、小”语义及其计算机识别分析	191
第一节 “大、小”的空间义	191
第二节 “大、小”的隐喻义	195
一、时间域	196
二、抽象量域	200
第三节 “大、小”计算机语义识别模型构建	202
一、“大、小”计算机语义识别流程	202
二、“大、小”计算机语义识别模型构建	203
第九章 结论与展望	205
一、研究结论	205
二、研究展望	209
参考文献	211

第一章 绪 论

第一节 空间、空间维度和空间维度词

一、空间与空间描述

哲学和自然科学史上对空间的理解多种多样，其中影响最大的观点有三种。古希腊哲学家亚里士多德把空间定义为“某物的包含者，就是此物的场所，而没有此物的内容”，也就是说，空间是一种包含物体的东西（傅勇林、张愚，1998：53）；牛顿的绝对空间论认为“绝对的空间，就其本性而言，是与外界任何事物无关，而且永远都是相同的和不动的，处所是物体所占空间的部分”，在牛顿看来，空间是绝对的，是初始的自足范畴。空间是无限的广延，并将所有物体纳入其中，但与物体相独立，即空间是“空”的容器（赵亮，2008：69）；莱布尼茨提出的相对空间哲学认为“空间与物体密切相关，抽掉物体的空间是不存在的，是某种纯粹相对的东西，是一种并列的秩序”，也就是说，空间是相对的，取决于物体，空间不先于物体存在，空间是多个物体相互位置的秩序，考察单个物体独立于其他物体的“广延”是没有意义的，空间即关系或秩序，只存在于多个物体间（杨明天，2004：78）。现代哲学的空间观认为，空间与物体不可分割，空间依赖于物体存在，是物体存在的基本形式。空间指物体存在的广延性，具有连贯性，相对的间断性和三维性。

以上四种哲学空间观，除牛顿的空间观以外，其他的哲学空间

观都认为空间和物体是密不可分的。莱布尼茨强调空间是物体之间的关系，而现代哲学空间观强调空间是物体存在的广延性。实际上，空间应该包括物体之间的关系和物质存在的广延性两个部分。“空间不仅是物质存在的广延性……同时又构成了各种不同的物质存在状态的一种外部联系”（王鹏令，1985：79）。因此，我们认为空间由物体的存在空间和物体的空间关系构成，物体存在空间是空间的核心，是第一位的，而空间关系是在物体存在空间的基础上形成的，是第二位的。物体存在空间具有广延性，描写物体存在空间的参数为空间大小和空间形状；空间关系具有拓扑性（结构性），描写空间关系的参数为空间方位和空间距离。现实中的存在空间和空间关系反映到语言中表现为空间词汇，语言中表达物体的存在空间的词汇是处所词，主要是地名、物名和机构名；表达物体的空间关系的词汇主要是方位词，如“上、下”、“左、右”、“前、后”等。

空间不能独立存在，也不能被人们直接接触和感知，人们所能接触和感知的是空间中的物体，因此，对空间的描写实际上就是对空间中的物体进行描写，要描写空间的面貌，其实就是描写空间中物体的外形的完形及其结合。由于形状和大小是物体外形的基本构成部分，所以对空间的描写，其实就是描写物体及其组成部分形状信息的表达方式（卡波耶娃，2000）。物体的形状和大小是描述空间的两个参数，反映到语言中就表现为空间形状词汇和空间维度词汇。

综上所述，我们认为现代汉语中空间词汇包括两类，一类是空间指称词，表示指称空间概念的词汇，主要有地名、物体名词、机构名词等处所词；另一类是空间描写词，表示描写空间关系或空间面貌的词汇，主要有空间方位词（“前、后”，“左、右”，“上、下”等）、空间形状词（圆、扁、尖、直、弯等）和空间维度词（“高、矮”，“长、短”，“宽、窄”等），空间维度词包括空间维度形容词、空间维度名词和空间维度量词（米、平方米）。本书研究的维度形容词属于空间描写词的空间维度词，如图 1-1 所示。

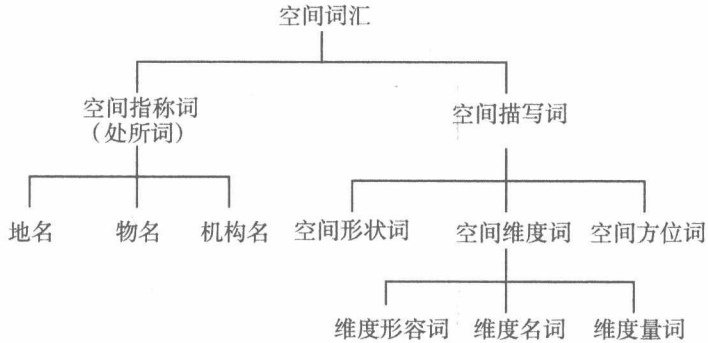


图 1-1 空间词汇分类

二、几何学、语言学和现实世界中的空间维度

“维”是几何学及空间理论的基本概念。构成空间的每一个因素（如长、宽、高）叫做一维。“维”是度量空间的概念，因此，人们通常也称之为维度。空间维度是用来度量空间的，是描述空间面貌的参数之一。在几何学中，维度可以分为零维、一维、二维、三维。零维空间是一个点，无限小的点，不占任何空间，点就是零维空间。无数点集合排列之后，形成了线，直线就是一维空间；无数的线构成了一个平面，平面就是二维空间；无数的平面构成了一个立体空间，立体空间就是三维空间。

语言中的物体没有严格意义上的点（点没有大小，而语言中物体一定有大小，否则就不存在），也没有严格意义上的一维线形事物和二维平面事物（线只有一个维度，面只有两个维度，而语言中再细的线都有粗度，再薄的蝉翼都有厚度）。但语言中的物体可以被图式化（理想化和抽象化）为零维物体（如星星点点）、一维物体（如米线、毛线）、二维物体（如桌面、路面）。因此，几何学中的概念或术语“维度”，可以借用到语言学中用来表示事物的空间概念。维度是物体的首要形状特征，我们把语言中的物体按

其维度数量可分为四类：零维物体、一维物体、二维物体和三维物体。零维物体是指很小的物体，它的大小几乎可以忽略不计，可以图式化为不占空间的点，如跳蚤、灰尘等。一维物体是指一个维度特别凸显的物体，即在一个维度上的延伸大大超过在其他两个维度上的延伸的物体，这类物体如棍子、树、铅笔、藤、绳子、蛇、河、路等。二维物体是指两个维度特别凸显的物体，即在两个维度上的延伸比在第三个维度上的延伸大得多的物体，这类物体如纸、布、叶子、木板、毯子、门等。三维事物是指三个维度都有相当的广延的物体，通常没有任何一个维度上的广延比其他维度上的广延具有显著的优势，这类物体如楼、箱子、桌子、衣柜、盒子、汽车、球等。

语言中物体维度的划分与几何学意义上的维度划分既有相似性也有差异性。语言中对物体维度的刻画反映了人类与外部世界相互作用形成的身体经验以及事物概念化的方式，这种独特的认知方式融入语言中并体现在语言意义的各个方面。语言再现的空间维度并不是现实世界的复制，而是经过认知思维过程的图解，是图式化（理想化和抽象化）的结果。例如，现实中的所有物体都是三维的，即使一粒肉眼看不见的灰尘都有大小，但人类通过图式化过程，凸显物体的某一个重要的特征，如凸显事物三个维度中的一个维度，如树、藤，忽略其他不重要的非本质特征，即忽略树、藤的粗度，得到一维物体的图式。同样，人类通过图式化过程凸显事物三个维度中的两个维度，如木板、布、地毯，忽略不重要的第三个维度，即忽略木板、布、地毯的厚度，最终形成了二维物体的图式。也可以忽略物体所有三个维度，如颗粒、跳蚤等，而得到零维物体的图式。

现实世界中的空间只有三个维度，然而语言为了描述这三个维度却采用了许多不同的形容词，其形容词的数量已经超过了现实中文度的实际数量，汉语中表达空间维度的形容词通常有“大、小”，“高、低”，“宽、窄”，“深、浅”，“长、短”，“粗、细”，“厚、薄”等。现实中的同一维度在语言中可以根据事物的不同形状、大小、维度的内外位置以及方位和方向等因素采用不同的维度

形容词来表示。

语言不是外部客观世界的简单镜像反映，而是认知主体和外部客观世界互动的结果，是基于人类身体经验对客观外部世界的概念化的符号表达系统，即自然语言是客观外部世界经过认知处理后折射的结果，如图 1-2 所示。

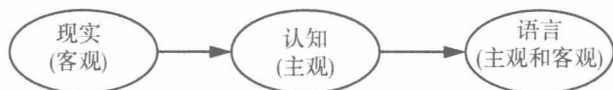


图 1-2 现实、认知、语言的关系

因此，语言中的零维物体、一维物体、二维物体和三维物体，以及采用超过实际维度数量的维度形容词来描写现实中的维度，都是基于人类身体经验对客观世界中物体的认知概念化的结果，而不是客观世界中物体的真实反映。

三、空间维度词的词性界定与意义分类

在《现代汉语词典》（第 5 版）中，无标记维度词如“长”、“高”、“宽”、“厚”、“粗”等都有两种词性，一种是形容词，在句中作定语或谓语，另一种是名词，通常与数量短语连用。例如，《现代汉语词典》对“高”的解释：①从下向上的距离大；离地面远：高楼大厦 | 这里地势很高。②高度：那棵树有五米高 | 书桌高八十厘米。词典中标示义项①是形容词，义项②是名词。有标记维度词如“短、低/矮、窄、薄、细”等都是形容词，而没有名词的词性，因为它们不能跟数量短语连用。

我们认为无论在句中作定语或谓语的维度词还是与数量短语组合的维度词都是形容词，而不是名词，即“高楼、高山”中的“高”是形容词，“那棵树有五米高”中的“高”也是形容词。邢福义（1965）在《谈“数量结构+形容词”》一文中指出形容词也可以受数量结构的修饰，如“一丈多高”、“六七尺宽”、“六公斤重”等，并从句法、语义的角度论证了数量形结构不是名词性