



陈振宇◎著

逻辑、概率与地图分析

——汉语语法学中的计算研究

 复旦大学出版社

图书在版编目(CIP)数据

逻辑、概率与地图分析:汉语语法学中的计算研究/陈振宇著. —上海:复旦大学出版社, 2020.5

ISBN 978-7-309-14882-4

I. ①逻… II. ①陈… III. ①汉语-语法-计算语言学 IV. ①H14

中国版本图书馆CIP数据核字(2020)第027694号

逻辑、概率与地图分析:汉语语法学中的计算研究

陈振宇 著

责任编辑/王汝娟

复旦大学出版社有限公司出版发行

上海市国权路579号 邮编:200433

网址: fupnet@fudanpress.com <http://www.fudanpress.com>

门市零售:86-21-65642857 团体订购:86-21-65118853

外埠邮购:86-21-65109143

当纳利(上海)信息技术有限公司

开本 787×960 1/16 印张 34.5 字数 520 千

2020年5月第1版第1次印刷

ISBN 978-7-309-14882-4/H·2966

定价:158.00元

如有印装质量问题,请向复旦大学出版社有限公司发行部调换。

版权所有 侵权必究

序

这是陈振宇教授在最近三年中出版的第三本著作,值得庆贺。应该说,他这几年耕耘不辍,收获颇丰,成果较为突出,也逐渐走向了学术研究的快速上升期。这本书内容很多,给人的突出印象是:具有强烈的理论意识,覆盖了从逻辑到语用的广大区域。振宇对当代汉语研究,尤其是语义学的研究,有着清醒的认识和准确的理解。并且,他试图把平时点点滴滴的独到见解,化入自己独有的系统,开拓新的研究道路。

这本书最初的出发点是试图通过引入认知模型与概率论,来重新分析汉语意义问题。后来,他认识到:所有意义分析的基础实际上是“事实性”分析,这就是逻辑上的“真值”,而事实只能是特定可能世界中的事实,所以对事实的定义必然涉及对可能世界的认知定义。那么可能世界是如何造成的,不同的可能世界会有什么特点(更具体地说,是直陈世界与虚拟世界的本质差异),在认知上从一个世界向另一个世界的投射的可能性,等等,就成了十分重要的问题。振宇说,他的这一研究是受到我和其他研究者这些年来有关文章的启发与影响。不过,看得出来,他更多地从哲学上去探讨,并且很重视规律性的总结。

在引入概率论后,振宇试图将一般的命题逻辑、谓词逻辑、广义量化算子研究、语用逻辑等统一在“概率蕴涵”理论中,并把我们通常的概率为1的逻辑推理研究(书中称为“无缝推理”),进一步推广到概率不为1的语用推理中去(书中称为“有缝推理”)。这一想法是他在北大做博士后工作时产生的,早年也曾在某次与逻辑学界的跨学科讨论会上报告过。但是,当时并没有得到与会者的认同。后来,他一直没有放弃这一研究。经过十年的思考,

终于形成了一个较为完整的体系；并且，用概率的方法证明了许多逻辑公式，其中还对有的流行公式做了修正。简单蕴涵关系的提出，以及对传递公式的修正，是其中两个最大的改变。除此之外，他提出的“全句量化”问题（包括“数的一致性”原理和“虚指”原理），也试图彻底改变我们对量化理论的基本认识。这些观点，将来是否能为学界所承认，还有待观察。但是，不管怎样，他的确是从汉语语言现象出发，为汉语语言研究寻找新的路径。这一点是值得肯定的。

书中关于有缝推理的研究是一个重要的亮点，他不但提出了大概率的有缝推理，还提到了小概率的问题；提出了“小预期”，以解释梯级系统；还有特殊的“逆守”关系。跟此前大部分基于描写的语用研究不同，这是一个完全由逻辑公式组成的体系，可以用有关的格式证明。通过证明，不但可以看到语言中的确存在着很多有缝推理，而且第一次有了一个经过语言格式检验的有缝推理类型划分。这是振宇近几年的研究工作的一个成果，是他最近的一个理论创新。

由于逻辑问题会涉及我们对万事万物内在规律的认识，因而又自然地引出了如何从调查数据去获得规律的问题。这是本书最后走向地图分析的原因。振宇不满足于仅仅讨论语义地图，而是希望更为根本地建立一套基于地图分析的隐性规律挖掘理论与操作方法。为此，他还专门编写了程序，进行计算的实践和验证工作。这种研究方式，在汉语学界是相当少见的。在语义地图的研究中，人们一直说这一分析方法具有很强的生命力，可以适用于语言研究的各个方面，甚至可以适用于人类社会研究的各个方面；但是，真正去进行变革性探索的并不多。振宇的这个研究，可以说是填补了汉语学界的一个理论空白。本书中主要的研究内容虽然还是围绕着具体的语法案例展开的，但是叙述和论证的方式已经完全超越了语言学。特别是其中的“俱乐部”隐喻，应该可以适用于任何一个社会系统。希望这一研究范式今后能推广到其他研究领域，以便取得更大的研究成果，走出一条全新的研究道路。

当然，振宇的研究方法和论证方式也有一些需要改进的地方。特别在术语使用和表述方面，需要更加适应语言学既有的传统和人文学者的阅读习惯。

振宇博士毕业于复旦大学，师从戴耀晶教授。后来于 2016 至 2018 年来北京大学中文系做博士后，跟我进行了广泛的学术合作，一起发表了好几篇计算语言学方面的论文。他对汉语语言学研究有着极大的热情，十余年来矢志不渝，付出了艰辛的努力，终于取得了一批优秀的研究成果，并形成了自己独特的研究风格。可以说，在一些研究方向上，他站在了当今汉语研究的前列。我希望他能够继续努力，今后取得更大的成就。

袁毓林

2018 年 9 月 18 日

于京北橡树湾

目 录

序	001
引 言	001
第一部分 真值	002
第二部分 概率蕴涵	002
第三部分 量化与单调	003
第四部分 地图分析	004
第一章 概念与真值	006
1. 概念类型	007
1.1 认知路径——索引与摹状	009
1.2 认知类型——索引词、概括词、先验词、实证词、 操作词、漂流词、超验词	013
1.3 命名类型	020
2. 对概念的言说	026
2.1 定义	026
2.2 解说	029
2.3 概念的逻辑表达	031
3. 概念之间	034
3.1 概念之间的关系	034
3.2 等值替换	036
4. “相对真值”观	038

4.1	“可能世界分层”理论	039
4.2	直陈世界与事实性	042
4.3	个人信念、公共信念与宣告	044
4.4	实指与虚指	047
4.5	无指与虚指的区别	050
4.6	实指论元与虚指论元	052
4.7	为什么不使用“模型论”	055
5.	事实性问题	056
5.1	存在预设及其有效性	056
5.2	叙实性	057
5.3	内涵逻辑	060
6.	意向活动的表达	060
6.1	意向谓词	061
6.2	意向名词	062
6.3	(广义)情态	064
6.4	世界信息与世界设定	068
6.5	摹状性词语	071
6.6	存在——质与量	074
6.7	回指原则	078
第二章	透明性	081
1.	索引性等级	081
2.	透明算子	089
2.1	意向行为及其达成	091
2.2	感知、记忆及其证据性	098
2.3	知识的获得、拥有与失去——缺省推理	105
2.4	关于可控性行为的知识	111
2.5	过去的认识状态	119
2.6	以主体的知识为背景	122
2.7	以主体的事实性判断为背景	133
2.8	言者的认识	147

3. 主题与信息顺逆结构	149
第三章 概率蕴涵	152
1. 概率	152
1.1 简单概率	153
1.2 汉语主观概率的表达	160
1.3 条件概率	164
1.4 一般的合取公式、一般的析取公式、贝叶斯公式	170
1.5 条件概率的类型——全称量化、优势条件、随机 条件、存在量化、部分量化、隶属度量量化	170
2. 基于模型的概率计算	173
2.1 不相容模型	173
2.2 映射模型	174
2.3 两两配对模型	176
2.4 超级配对模型	177
2.5 独立性	180
2.6 关联度	183
3. 全集的影响	185
3.1 交集与并集	185
3.2 概率蕴涵的类型——蕴涵、衍推、预设及相互 蕴涵、相互衍推	187
3.3 为什么不采用真值表和实质蕴涵?	193
4. 因果与蕴涵	196
4.1 使因与理由	197
4.2 使因导致的蕴涵关系	201
第四章 推理	205
1. 推理的结构与语言学表现	205
1.1 推理的结构	205
1.2 语篇中推理各部分的表现	206
1.3 推理的有效性类型	209
2. 如何保障无缝推理	210

2.1	同一世界律	211
2.2	对象一致律	211
2.3	概率与蕴涵传递律	213
3.	无缝推理的模型	220
3.1	存在式	221
3.2	概率之和公式	221
3.3	选择式	222
3.4	添加式	223
3.5	传递式	227
3.6	贝叶斯公式	232
3.7	衍推公式	232
	【附】 计算推导举隅	234
4.	对无缝推理和有缝推理的验证	242
4.1	预期语篇、反预期语篇和小预期语篇	242
4.2	有效推理的类型	245
5.	无缝推理	246
5.1	无缝推理与例外句	246
5.2	无缝推理举隅	249
6.	大概率推理	252
6.1	概率性大前提的来源——概括、语用改造和语义和谐	253
6.2	类比推理	258
6.3	溯因推理	262
6.4	时间和谐	264
6.5	结果和谐	265
6.6	其他和谐	272
7.	逆守关系	276
8.	小概率推理和小预期	285
8.1	梯级的两个方向——无缝与小概率	285
8.2	“极端”句——“(连)X都/也Y”	288

第五章 量化	291
1. 量化及其语义类型	291
1.1 理论背景	291
1.2 简单量化	293
2. 量化的语言实现	296
2.1 直接量化与间接量化	297
2.2 全称量化的语义结构	298
2.3 直接量化算子	301
2.4 数的一致性	304
2.5 预先存在与全部量	312
2.6 虚指的通指性质	315
3. 任何、每、所有	324
3.1 强无限性——“任何”与“从来”	326
3.2 单数性与融合性	332
3.3 功能差异	339
3.4 内部量化和外部量化	342
3.5 “都”——共现还是不共现	357
4. 间接量化的性质	362
4.1 语义性和语用性	362
4.2 唯一性与非唯一性	365
4.3 量化中主题与焦点的性质	368
5. 进一步的发展	375
5.1 更多的语义要求	375
5.2 “各”的可分别性	378
5.3 “总”	381
5.4 “任何”的选择性限制	383
第六章 单调性	387
1. 双联说与单联说	388
2. 存在与全量	391
2.1 存在量化上升规则	391

2.2	全称量化下降规则	394
2.3	全量否定下降规则	396
2.4	判断上升规则	398
3.	量性焦点	399
3.1	数值/比例不提升规则	399
3.2	数值/比例的语用提升	403
3.3	偏量提升规则	406
3.4	主观量提升规则	408
3.5	部分量上升规则	410
4.	全句量化——数的一致性和扩展的一致性	412
5.	单句蕴涵提升所涉及的其他问题	422
6.	充分、必要、充要及极限条件复句	428
第七章	交际控制地图	434
1.	“交际控制”理论	435
1.1	隐性控制	435
1.2	一个例子——“俱乐部”隐喻	437
1.3	交际控制图各部分的名称与性质	440
2.	“交际控制图”的计算原理	442
2.1	理想状态下的经典算法	442
2.2	完全加权算法	445
2.3	“赢家通吃”算法	448
2.4	“赢多输少”算法	451
2.5	各种算法的比较	452
3.	扩展的交际控制图	456
3.1	“正负双向”数据计算	456
3.2	“基元对”数据计算	457
3.3	“广义基元对”数据计算	460
4.	对交际控制图的进一步分析	463
4.1	蕴涵概率	463
4.2	简图与中心点	466

4.3 控制性	468
4.4 歧解	469
5. 语法学研究案例——世界语言“人称标记”语义	
地图	471
5.1 以往的研究	471
5.2 CCG 地图研究	477
第八章 性质类型地图	484
1. “性质类型”理论	485
1.1 隐性类型	485
1.2 一个例子——“分组”隐喻	486
1.3 性质类型图各部分的名称与性质	490
2. 对性质类型图的进一步分析	491
2.1 均匀性与熵	491
2.2 正态分布	492
2.3 家族相似性和飞地	494
2.4 地图的分割原则	495
2.5 性质选择的好坏	497
3. 扩展的性质类型图	497
3.1 性质的权重	497
3.2 单向地图与单向组	500
4. 多次聚类与类型层次——成都话“哇”字句	502
4.1 数据调查以及“哇”字句基本格局	503
4.2 对“哇”字句的自动聚类	514
参考文献	524
后记	535

引 言

本书是教育部人文社会科学研究规划基金项目“现代汉语句法与语义计算研究”(批准号:13YJA740005)的最终成果,也是拙著《汉语的小句与句子》(2016,复旦大学出版社)和《汉语的指称与命题》(2017,上海人民出版社)的姊妹篇,是在这两本书的基础上展开的进一步讨论。我们默认读者已经详细读过这两本书,因此相同的内容如非必要本书不再重复。本书是最近几年作者《汉语语法学原理》系列研究中的第三本,关于这一系列的有关情况,已在前两本书中作了介绍,不再重复。

全书围绕着两条主线进行安排:一是从汉语的语句出发,去进行逻辑意义分析,尤其是增加认知与语用分析,从而适应于复杂的语言现象;二是将概率论引入逻辑体系,从根本上消除逻辑基本定义中的“隐患”,并将无缝推理扩展到有缝推理,最终扩展到超级匹配模型,从而为复杂现象中隐性规律的发掘创立一条新的研究道路。这两条主线交织在一起,形成“概念、真值、概率蕴涵、无缝推理、有缝推理、语用逻辑、地图分析”的有序系列,先认知后概率,先无缝后有缝,先语义后语用,先四格表后超级匹配,将全书内容贯通为一个整体。

我们认为,不存在孤立的、自足的逻辑意义体系。逻辑的基础在于“真值”与“蕴涵”两大范畴。真值不是某个抽象的、无差别的可能世界的存在判断,而是具体的两大可能世界——直陈世界和虚拟世界——中的存在。这两大世界的对立,也是人的基本认知模型的反映:说话者(认知者)赋予了语句所代表的各种概念以实指和虚指的地位,分别从外延与属性内涵来把握它们。蕴涵也不能由抽象的、无差别的真值表来定义,而是代表着概念(它们各是一个集合)之间的投射关系,应该用条件概率来刻画,这一概率不需

要等于 1 或 0,甚至不需要接近于 1 或 0,就可以做出有效的推理,称为有缝推理,它在社会语言活动中是最为常见的推理。

全书十章 35 万余字,分四部分。

第一部分 真 值

从认知与语用的角度重新定义了意义理论的基础——真值系统,更为彻底地奉行相对真值观。

第一章提出概念的认知模型,以及相对真值观。从两大认知路径“索引”与“摹状”出发,把概念分出索引词、概括词、先验词、实证词、操作词、漂流词和超验词等各种类型。这是对弗雷格、罗素以来分析逻辑理论的继承与发扬。

接着作者提出了“可能世界分层”理论,分出三个主要的可能世界层次:“现实世界”(说话者所在的世界)、“直陈世界”(说话者作为(语言)事实直言说或思考的世界)和“虚拟世界”(说话者假设的世界,或者是他完全虚构的反事实世界,或者是他不能确定事实性的非事实世界),并重点讨论了直陈世界与虚拟世界的关系。书中还根据具体语料,尽可能详尽地归纳总结了各种造成虚拟世界的意向活动表达方式。

第二章重点讨论虚拟世界中存在的事物如何向直陈世界投射的问题,其中既有认知上的倾向,也有语用上的倾向,甚至涉及说话者的信念,如索引性等级、意向行为的达成、证据性(传信)、过去的认识状态、主题信息、缺省推理、背景化等因素,书中都给予了充分的关注。

第二部分 概 率 蕴 涵

引入概率论重新定义各种(逻辑)意义关系,将命题逻辑、谓词逻辑、广义量词理论及形式语用学研究,整合在统一的“概率蕴涵”理论之中。

第三章用条件概率分别定义了简单蕴涵、衍推、预设等意义关系,以及全称量化、优势条件、随机条件、存在量化和部分量化等,并且指出“全集”(当前言语活动涉及的论域)在意义关系中具有决定的地位,当 X 蕴涵 Y 时,如果全集大于 Y,则有 $\sim Y$ 蕴涵 $\sim X$,得到衍推关系;如果全集等于 Y,则

有 $\sim X$ 蕴涵 Y ,得到预设关系;衍推与预设是平行互补的,但是在语用中我们往往默认全集是无限的,因此衍推解读是无标记的优势选择,而预设仅在特殊的情况下存在。

第四章重新省视各种逻辑推理公式。大部分公式在概率蕴涵中也成立,只是用概率计算再证明一次。但有一些公式却需要根据语言中的情况进行修正,如传递公式1,需要做两个修正:传递要无缝,必须大前提的条件概率为1,当 $P(X_3|X_2)=1$ 时, $P(X_3|X_1)$ 大于等于 $P(X_2|X_1)$,而一般所说的传递率是这一公式的特例;另外,传递有可能陷入“自残困境”,这时因为主观与客观原因, X_3 会落入 $\sim X_1$ 的区域之中,从而得到荒谬的结果。

本章还专门研究了无缝推理,并给出了三组验证格式(预期语篇、反预期语篇和小预期语篇),它们可以将有效推理进一步分化:无缝推理只能用于预期语篇,不能用于反预期和小预期语篇;大概率的无缝推理可以用于预期语篇和反预期语篇,不能用于小预期语篇;小概率的无缝推理可以用于小预期语篇和反预期语篇,不能用于预期语篇;逆守关系则只可以用于反预期语篇,不能用于预期和小预期语篇。本章还给出了不少汉语中的无缝推理实例。

第三部分 量化与单调

第五章研究量化系统。其中“全句量化”(也称为“间接量化”)是本书的一大发明。首次提出,可以通过“数的一致性”原则从语义上进行“全句量化”,以此解释否定句、复数事件句的特殊语言现象。首次提出,虚指认知中存在“角色——演员”的矛盾关系,可以触发无限的解读,以此解释通指句、条件句中全称量化的来源:它们是虚指中的单数性存在,是唯一角色,但却可以由其他世界中的任一成员来担任,所谓“任意性”的实质是无限性,即不加以限制,只要符合这一角色的属性内涵,任何演员都可以担任它。在上述理论的基础上,从历史与当代、语义与语用等方面分别研究了汉语中“每、所有、任何、都”的使用条件、功能性质与演化方向。

第六章专门讨论单调性问题。认为如果 X 蕴涵 Y ,则在 F 表示存在量化时, $F(X)$ 蕴涵 $F(Y)$; F 表示全称量化时, $F(Y)$ 蕴涵 $F(X)$ 。另外从传递律可以得到偏量的提升规则。与以往研究完全不同的是,由于引入“数的一致

性”原则和“扩展的一致性”原则，书中将语用因素引入了单调性研究，从而发现了语言中微妙的调节机制，证明同一句话在不同的上下文场景下其单调性可以不同。

第四部分 地图分析

前面各章实际上主要是研究简单的两项(最多再加上少数几项)之间的关系，但是在同一个语言现象中，有可能涉及十几项、几十项的要素，它们同时起着作用，但并不是平均着力，会因为语言使用而分成各种倾向性，代表着各种隐性规律。常规的逻辑无法帮助我们，所以本书提出了一个从实际调查数据中挖掘概率蕴涵逻辑关系的操作方案——地图分析。

第七章针对语法研究中存在的、大量的模糊、弥散、隐性的分布与规律，以“交际——控制”理论作为切入点，将多项分布的数据看成多人参与的交际活动，将它们的共现模式与频次，看成交际的权重，通过赢家通吃、赢多输少等算法，计算它们之间的关联情况，并最终绘制在 CGG 地图上，直观地展示整体系统的群簇与分布，计算它们的条件概率关系，从而挖掘出隐藏的倾向性规律。

第八章介绍基于地图的自动聚类方法。以成都话句尾语气词“哇”为样本，全景式地展示语料分析、数据集成、软件应用、地图分析的全过程，包括各个语用因素之间的关联图，以及所有语料的多层聚类，最后探讨语气词功能的弥散与收敛。

需要特别提到的是，地图分析部分是与妹妹陈振宁合作的成果。

综上所述，本书的目的是试图建立基于具体语言现象的、与语言研究紧密结合的语法学计算体系。书中主要是原理性的论述，具体研究已在我们发表的各种论文中展示，书中所选案例主要是为了说明理论。

本书的研究方法总地来说是演绎性的，每一部分首先给出基本理论框架，然后基于框架和语义公理的设定，对各种现象进行形式化、运算、证明、解释和预测；与此同时，还紧扣汉语的历史与现实，在语料中深入挖掘汉语现象的历史因素与语用因素，以此分离出偶然性的部分，以尽力避免“过度解释”的错误。

语言学与纯粹的逻辑、哲学和数学不同，不仅需要建立理论系统，做出

运算与解释,而且必须考虑我们的运算与解释是否与语言的事实相符合。即使理论再高妙,如果与事实不符,那就说明它不适合语言,或者说是它忽略了什么重要的东西,导致了错误的结论。作为经验科学的语言学中的语义学,不存在什么“不证自明”的先验体系。

怎样验证我们的理论的有效性?需要进行语言调查。一是针对语料库进行统计与分析,共时语料的分布要符合相关项目的倾向,历时语料的语法化先后要符合逻辑机制的先后。二是注重语言比较,一种语义学解读首先必须符合汉语的事实,不能无条件地将其他语言(英语)的研究成果移植到汉语研究中来。三是要通过语言中的常见格式,造出一些语句,然后请母语者来判断其对错,即采纳检验法,从而得出母语者真实的心理倾向性。

如果调查的结论与我们的理论相符,则说明我们较为成功地描写了语言现象,反之则需要反思。当然,反思不一定是因为我们理论错了,也有可能是调查方法上的失误。所以需要双向反思。

最后,按传统的观念,语法学中的计算研究主要有两个方面:1) 语句本身的意义解读,就是对其中单位的真值、指称、主观概率、量性、量化、预期,以及对语篇中无缝推理和有缝推理的研究,它们都是由基本的逻辑运算系统控制的,是语义学和语用学的基础。2) 对语言现象进行统计,并从统计概率中挖掘出隐性的规律,即分布、蕴涵、关联、独立和聚类等关系。因此,语言学中关涉的数学,主要是“数理逻辑”与“概率统计”。不过我们认为,上述内容全都可以统一在集合论和概率论的体系之中,它们不是两个相互独立的数学系统,而是一个统一的数学系统;只有通过概率论改造的数理逻辑,才是符合语言实际的逻辑系统。