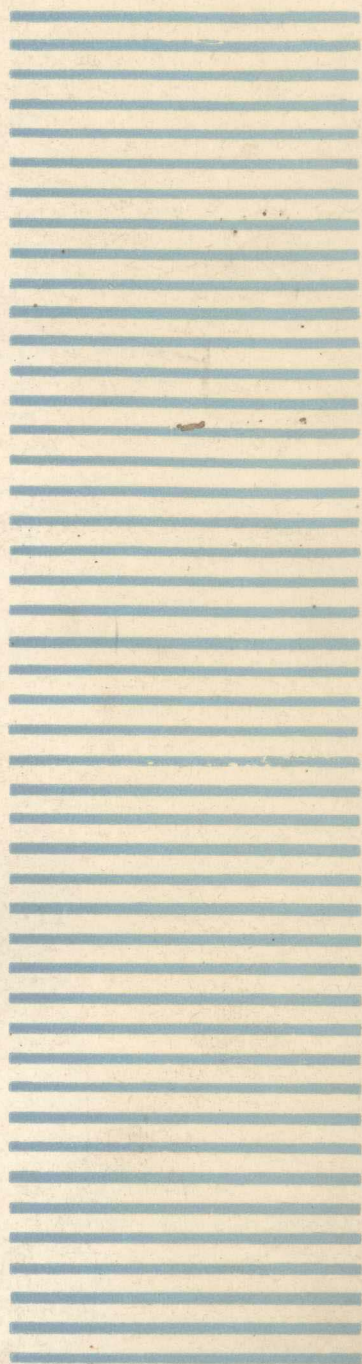


语文论丛

3



H10-53

~~7/3~~

11/3

语文论丛

(第3辑)

上海市语文学会编

语文论丛

(第 3 辑)

上海市语文学会编

上海教育出版社出版

(上海永福路 123 号)

新华书店上海发行所发行 祝桥新华印刷厂印刷

开本 787×1092 1/16 印张 11.5 字数 244,000

1986 年 12 月第 1 版 1986 年 12 月第 1 次印刷

印数 1—1,700 本

统一书号: 9150·67 定价: 1.95 元

目 录

语言的社会本质和语言学的性质	高天如(1)
试论语言的离散性和模糊性	钱 锋(7)
音位学略说	王德春(16)
意义和信息	程雨民(26)
从训诂学上来看古汉语的基本词	张世禄(32)
天干地支之起源与作用	罗君惕(41)
古汉语常用词问题的探讨	陆心国(51)
试论古今字和通假字的界线	傅承德(58)
评高本汉为《广韵》拟音的基础——四等洪细说	潘文国(66)
先秦修辞理论试探	李金苓(76)
重评《马氏文通》.....	申小龙(85)
关于稳定旧体诗歌押韵标准的刍议.....	金云峰(94)
入声消失年代试探	顾汉松(102)
论教师的语言修养	朱 川(105)
上海市区少年所使用的上海话语音	许宝华 汤珍珠 汤志祥(112)
关于汉语句子格局分析的若干设想.....	陆致极(121)
说插语.....	范 晓(129)
鲁迅论修辞	宗廷虎(138)
惯用语浅说	董天琦(146)
缩略语词形成基础简论	李熙宗(155)
评国语罗马字“词类连书”——兼谈汉语拼音的连写问题	颜逸明 徐子亮(159)
再论声韵字母	李公宜(167)
聋人手势语初探	蔡浩中(175)

语言的社会本质和语言学的性质

——马克思主义经典作家语言理论学习札记

高天如

马克思主义经典作家关于语言的理论，虽说专著不多，而马克思、恩格斯、列宁、斯大林和毛泽东同志的关于哲学、政治经济学、科学社会主义的大量著作，对于语言的本质、语言的起源和发展，以及语文学学习和语文政策性的问题等等，均有所述，给我们留下了丰富的理论遗产。本文试就语言的社会本质和语言学的性质问题，谈谈学习马克思主义经典作家语言理论的一些体会。

肯定语言的社会本质，是马克思主义经典作家论语言的一个基本观点。马克思在《资本论》中讲到商品的价值时，就认定语言同使用品的价值规律一样，是“社会的产物”。列宁在讲到“语言是人类最重要的交际工具”时，亦着重说明语言是保证经济的发展，以及社会各阶级、阶层的划分和联系起来的重要条件（见《论民族自决权》）。斯大林的《论语言学中的马克思主义》等文，不仅解释了语言既依存于社会而又服务于社会的一般特征，而且集中地表述了作为特殊社会现象的语言的专门特点，即“语言是作为人们交际的工具、作为社会中交流思想的工具为社会服务的，这个工具使人们能够相互了解并调整他们在人类活动的一切范围中（包括生产的领域，也包括经济关系的领域，包括政治的领域，也包括文化的领域，包括社会生活，也包括日常生活）的共同工作。”从而，论定语言既不是基础也不是上层建筑，可又是“属于社会现象之列的”，“社会以外，无所谓语言”。

马克思主义经典作家在把语言列为社会现象、肯定语言为社会服务的工具作用的前提下，为揭示语言社会本质的专门特点，其论述是多方面的。要言之，有语言的全民性问题、语言的民族特征问题，以及语言与思维的关系问题等。这些方面的科学理论，对于我国现代语言学的发展曾经发生过并日益显示出它的积极作用。现分述如下：

关于语言的全民性问题，是马克思主义经典作家论述语言社会本质的一个最具现实意义的论题。马克思早就说过“语言本身是某一集体的产物”（《马克思未发表的手稿》），并指出：“文法规则就不会因为讲解它的是一个信教的托利党人还是一个自由思想者而有所改变。”（《卡·马克思关于现代社会中的普及教育的发言记录》）这一关于语言的全民性的论点，在斯大林批判马尔的“阶级语言论”时，更得到充分的令人信服的论证。因为，事实

上只有“全民语言的存在”，而不存在什么“阶级语言”（见《马克思主义和语言学问题》）。可是，在三十年代，马尔的理论也曾经被当作“用唯物史观研究语言”的新学说介绍到我国来。这对当时我国正在蓬勃兴起的“大众语”问题的讨论和“拉丁化新文字”的提倡造成了某些理论上的混乱。所谓“统治阶级的言语”、“新兴阶级的普通话”等“阶级语言论”的观点，只是在斯大林的《马克思主义和语言学问题》发表以后，才逐步得到纠正。正如马叙伦所说：“斯大林的马克思主义语言学说给中国语言学、文字学研究指引了新的光明的道路”，“帮助我们批判错误的旧语言学说，首先我们应当清除马尔语言有阶级性的谬论的影响”^①。正是由于有了正确的语言理论准备，建国后的汉语规范化工作和汉字改革工作，才得以顺利而迅速地发展。

关于语言的民族特征问题，是同语言的全民性相联系的又一重要论题。马克思主义经典作家历来视语言为民族的特征。斯大林就这么说过：“马克思主义说语言的共同性是民族的最重要标志之一。”（《马克思主义和语言学问题》）这说明每个民族都有自己的共同的语言。当然，这不等于是说不同的民族就必然运用不同的语言。而是从它的历史渊源看，语言总是作为某一民族或其他人类共同体的标志而存在的。不同的语言，终究反映着不同民族的语言习惯。

因此，我们在考察语言的社会本质时，应该看到语言虽具有为社会全体成员服务的性质，但又是民族的界限的。语言与社会的关系，不仅表现为同社会经济和社会制度的联系，实际上语言同民族本身的发生和发展更有直接关系。所谓社会的分化或统一影响着语言的分化或统一，最终还是通过民族的分裂或融合实现的。即使同一民族的分隔或不同民族的聚居，其语言上的分化或统一，由于语言习惯的缘故，也不是立即就会发生的。所以说，“民族和民族语言的特点是具有非常的稳固性以及同化政策的巨大抗拒力”。（斯大林《民族问题和列宁主义》）列宁在《论大俄罗斯人的民族自豪感》一文中宣称：“我们酷爱自己的语言和自己的祖国”。语言和民族的关系就是如此不可分离。

可是，在我国“五四运动”时期，却曾经出现过以西文替代汉文的主张，说什么“lamp、pen、ink之类，也以采用原字为是，免得再用什么‘洋灯’、‘洋笔’、‘洋墨水’这些鄙俗的词儿”，“什么‘音译’、什么‘意译’，都是吃饱饭，没事干，闲扯淡”^②。更有甚者，在倡导改革汉字时，竟有人主张用“最佳最易之万国新语”（即世界语）替代汉文的使用。这些主张，都理所当然地不能为现实所接受。为什么？一个重要的原因就是由于这些主张没有重视语言的民族特征，人为地使语言的发生或发展脱离了使用该语言的民族的语言习惯。文字的形体可以是国际性的，文字所记载的语言则不能不具有民族性。当然，一种民族语言的发展变化，总会包括对外族语言成分的吸收和利用。事实上，利用外语成分，已成为丰富和发展语言的途径之一。毛泽东同志在《反对党八股》一文中，提倡“向人民群众学习语言”，“学习古人语言中有生命的东西”，同时也提出“要从外国语言中吸收我们所需要的成分”。但是，那“不是硬搬或滥用外国语言，是要吸收外国语言中的

① 见《学习斯大林的语言学说》，载《中国语文》1953年第6期。

② 见《汉字革命》，载《国语月刊》汉字改革号。

好东西，于我们适用的东西”。可见，吸收外语成分，是以我为主的，是以本族语的需要为前提的。而要达到“适用”，被吸收的外语成分，还必须在本族语的体系上造成。以通常的借词为例，即便是音译的，其音系也要符合吸收语的习惯。无疑，这都是由于语言具有民族特征的缘故。

关于语言与思维的关系问题，同样是一个涉及语言社会本质的论题。马克思这样说过：“思维的社会性质是同思想一起存在于语言中，就象价格是同商品一起存在着一样。”（《1857—1858年的经济手稿》）对于这个问题，马克思主义经典作家虽然没有从心理学的角度进行具体的或实验的研究，但他们已从哲学的高度为考察语言与思维的关系提出了具有理论意义的指导原则，即语言和思维是相区别的两种社会现象，而彼此又是紧密地联系着的。这在马克思恩格斯合著的《德意志意识形态》一书里，早有明确的表述：“‘精神’从一开始就很倒霉，注定要受物质的‘纠缠’，物质在这里表现为震动着的空气层、声音，简言之，即语言。”语言是实践的“现实的意识”，“意识一开始就是社会的产物”。这说明，思维是“精神”范畴的现象，而语言是隶属于“物质”范畴的，两者不能混为一谈，但它们又总是“纠缠”在一起的。语言与思维的这种辩证关系，对于认识语言的社会本质至关重要。现代国外某些语言学流派，虽然取得许多有成效的研究成果，但由于处理不好语言与思维的关系，往往造成方法论上的一些缺陷：或是将语言与思维分离开来，或是把语言与思维混同在一起。心理社会学语言学派的创始人索绪尔，举起“心理”和“社会”两面旗帜，肯定了作为符号系统的语言的社会性质。但他把社会的“语言”（langue）和个人的“言语”（parole）人为地对立起来，则不能不说明其理论既排除了语言的物质性，又把语言与思维的联系分割开了。因为，他认为语言“纯粹是心理的”，那包括发音的言语本是“心理物理的”，语言成了没有音响的“两面心理实体”。尽管他亦声称“语言是一个表示观念的符号系统”，而他所谓的“符号系统”仅仅是作为同思维不相联系的形式化的概念而存在的，作为记载着人类思维成果的语义被排斥于“语言”之外了。语言学成了专门研究符号系统的形式和规律的学科^①。因此，索绪尔的语言理论，在语言与思维关系的问题上，并没有达到全面的认识。

至于把语言与思维混同在一起的理论表现，观念主义语言学的“模式论”就是一例。这种理论首先把语言归结为思维的模式。萨丕尔说：“语言，作为一种结构来看，它的内面是思维的模式。我们所要研究的，与其说是言语的物理事实，不如说是这抽象的言语。”^②而后，又以这种“思维的模式”推论它必然制约人们对于问题的思考。这一点，“模式论”的另一代表人物沃尔夫表述得最为清楚。他在《美洲印地安人的世界模式》一文里说：“河比人的‘真实世界’为他们的语言模式所制约，在性质上与欧洲人的世界完全不同。”“近十年来物理化学的发展都是由使用欧洲语言模式的人推动的，或可以说，所有这些科学都是由使用欧洲语言的人创造出来的。”对此，除了种族主义的偏见外，其理论上的错误则是把语言与思维混为一谈了。在我国，也曾有人一度认为，根据汉语介词的

① 见索绪尔《普通语言学教程》，商务印书馆1980年出版。

② 见《言论》，商务印书馆1964年出版。

研究可以让我们了解汉族人民的思想方式具有原子主义和表象主义两个特点。这显然是受了“模式论”的影响，而又缺乏科学的语言理论指导的结果。

近年来，《北方论丛》等刊物开展语言与思维问题讨论时，又有人对语言思维之间的相辅相成关系提出质疑，认为语言固然离不开思维，而思维却不一定依赖或凭借语言。他们以聋哑人的思维、未掌握语言的幼儿的思维，以及事物概念先于命名、直观思维先于语言等现象，说明存在“没有语言的思维”，并认定思维可以脱离语言而单独存在和发展。那么，能否由此而动摇马克思主义经典作家关于语言与思维关系的理论原则呢？不能。这是因为马克思主义经典作家关于语言与思维的论述，首先是把它们置于人类社会的形成和发展的广度上来考察的。那不是从“个体”方面而主要是从“群体”方面立论的。恩格斯在《劳动在从猿到人转变过程中的作用》的著名论文中，有这么一段话：“首先是劳动，然后是语言和劳动一起，成了两个最主要的推动力，在它们的影响下，猿的脑髓就逐渐地变成人的脑髓”。这里说的是劳动和语言对于人脑形成的作用，当然也就说明了语言对由人脑所进行的思维的作用。而取得这种作用的结果，单凭生命有限的“个体”显然是不能奏效的。因此，即使存在着聋哑人或幼儿的思维，也不足以证明人类思维是在不依赖或不凭借语言的条件下形成并发展的。至于某些事物及其概念先于命名的现象，如果不囿于该事物及其概念同所命名的语词的直接对应性，也不难作出合理的解释。因为先于命名的新概念，也还是在语言的条件下发生的，只不过它对于语言的凭借所采用的是“以其知喻其所未知”的方式，即借助词组或句子的形式，运用所掌握的其他语词来造成或表述新的概念。其次，马克思主义经典作家所谓同语言相联系的思维，是指在直观思维基础上形成的一种人类的高级精神活动，即所谓抽象思维或逻辑思维。列宁在《黑格尔〈哲学史讲演录〉一书摘要》中所说，“感觉表明实在，思想和词表明一般的东西”，就把作为直观思维的“感觉”和联系着词的作为抽象思维结果的“思想”，区分得清清楚楚。而且，只有人类思维才具备的第二信号系统，恰恰是由语言促成的。这已为巴甫洛夫的科学实验所证明。因此，我们就不应以存在不借助语言的直观思维来否定语言与思维相辅相成关系的科学论断，或以非人的猿的“思维”来说明思维的起源先于语言的起源。在研究语言和思维问题的过程中，尽管可以从事多方面的探索，作出这样或那样的具体解说，但就人类社会的思维和语言及其相互关系言，马克思主义经典作家理论的真谛，无疑是客观存在的反映，是我们应该遵循的准则。

马克思主义经典作家关于语言本质的理论，无论关于什么方面的论述，皆以肯定语言的社会性为前提。那么，语言的社会性本质是否亦规定语言学的社会科学性质呢？

这本来是一个无庸置疑的问题。语言学为社会科学，因为语言是社会的产物，语言学所研究的对象是社会现象。在现代语言学史上，虽曾出现过自然主义语言学派的理论主张，不久就销声匿迹了。可是，由于现代科学技术的发展，利用自然科学手段研究语言的新学科不断产生，语言学是否仍具有社会科学性质的问题，又重新被提出来。早在实验语音学兴起后不久，我国就有人认为“语音学乃自然科学之一种”，因为“语言之发

音,完全取决于生理的及物理的条件,其方法与生理学及物理学所用者初无二致”。^①现在有人除认为“研究语音的生理和物理特性的生理语音学和声学语音学”为自然科学外,还提出“结构语言学的研究对象是语言的结构系统,所以很难说这样的语言学还是一门社会科学”^②的问题。这说明,重新提出语言学性质问题的依据,有研究方法的因素,也有研究对象的因素。因此,这同样是一个涉及语言的社会本质的问题。

科学的性质一般是由其研究对象的性质所规定的,语言学也是这样。当然,研究中所采用的自然科学手段,作为手段本身及其应用技术是由自然界规律所使然的。因而关于这种手段的学问,无异是自然科学。可是,对于现代语音学来说,其所采用的自然科学的实验技术,仅仅是一种研究手段,语音的社会性质并未因此而有所改变。现代音位学是建立在对于音素的精确分析和描写的基础之上的,它离不开自然科学技术或知识的应用。然而,音位系统的确立又是以语音的辨义功能为条件的。所谓语音的“区别特征”,实际上还是以语音的社会性差异为基础的。如,辅音中塞音的清、浊特征,在英语为区别性特征,在汉语普通话中则为非区别性特征,而塞音的送气、不送气的特征,在汉语为区别性特征,在英语又为非区别性特征。为什么同样的语音自然特征,却有“区别特征”和“非区别特征”(或称“多余特征”)的分别?显然所依据的是语音的辨义功能。语音的辨义功能,那是在语言发展的历史长河中约定俗成的。它同语音的自然属性虽有联系,却是非本质的联系,因而语言不同其音系也各异。这说明社会性始终是语音的基本属性,那生理的、物理的属性只是造成语音的物质基础。吴宗济先生在《试论普通话语音的“区别特征”及其相互关系》里说到,某一音位变到另一音位的质变,其“程度范围是随着语言社会、语言环境以及个人特点而大有伸缩的”,定出音位“主要依据这个特定语言的社会习惯”。^③至于“人工语言”,自当别论。不过,“人工语言的存在,并不妨碍自然语言的社会功能,它可以辅助自然语言更好地完成社会功能。”^④因此,无论是实验语音学,还是应用自然科学手段研究语言的其他新学科,如统计语言学、代数语言学、工程语言学等,如果它所研究的对象涉及人类的自然语言,就不可能不具有社会科学的性质,因为语言终究是社会的产物。

同理,以语言的结构系统为研究对象的结构语言学也还是一门社会科学。尽管,它的研究方法亦要求有如同自然科学方法一样的精确性,但语言的结构,毕竟不是社会之外的自然现象。它如同语音的系统一样,其约定俗成的社会性质依然存在。正是基于这样的事实,现代普通语言学和其他语言学家,关于语言的定义,才会产生诸如“社会制度”、“社会规约”、“社会传统”之类的表述。马克思主义经典作家对于语言结构的性质尽管未作专门论述,但他们关于语言社会本质的理论,是不排斥语言结构的社会性的。斯大林就说过这样的话:“不能把语言及其结构看作是某一个时代的产物。语言的结构及其

① 见岑麒祥《语音学概论》,中华书局1941年版。

② 见《现代语言学的发展趋势》,载《语言研究》创刊号。

③ 见《中国语文》1980年第5期。

④ 见《语言的社会功能,语言的全民性》,载《中学语文教学》1981年第5期。

语法构造和基本词汇，是许多时代的产物。”（《马克思主义和语言学问题》）因此，如果确立起“社会以外，无所谓语言”的观念，那就不得不承认，社会以外亦无所谓语言的结构，因为语言的结构是包含在语言之中的。

而今，把语言作为一种社会现象来研究已日益为人们所注重了。当前兴起的社会语言学就是生动的说明。就社会语言学已经取得的研究成果看，其内容亦很广泛。有语言的统一和分化问题，有词语的消长和流变问题，也有语言的各种社会变体问题，等等。集中到一点，都是联系社会条件、社会环境或社会背景来研究语言的。当然，这不等于说唯有社会语言学才能把语言作为一种社会现象来研究，只有它才具备社会科学的性质。社会语言学只是语言学的一个分支，语言学的分支是多种多样的。作为语言学的分支，它们的区别在于研究语言的方面的不同、方法或手段的不同、宗旨的不同，决非其研究对象的本质属性的不同。因此，无论冠之以什么名称的语言学，只要它所研究的是语言现象，就具有社会科学的性质，都应该把语言作为一种社会现象来研究。吕叔湘先生指出：“把语言作为一种社会现象来研究，是语言学的第三次解放。”^①这不是没有普遍意义的。

马克思主义经典作家语言理论的科学价值是显而易见的。尽管在他们的著作里，也会存在某些未必全然符合语言实际的具体论述，如说“语法的特点在于，它得出词的变化规则”（《马克思主义和语言学问题》），对此，如果作为普遍性的结论看，就没有能概括汉语的实际。但是，他们关于语言的一些基本理论无疑是正确的，是客观规律的反映，并且成为理论语言学的一个重要组成部分。因此，不能认为“他们从来没有研究过语言学的研究对象、语言研究的理论原则和具体方法、以及语言发展的内部规律等问题”^②。无庸讳言，马克思主义经典作家并非专门的语言学家，他们亦未专门从事作为语言学的主要任务“语言发展的内部规律”的研究和语言学“具体方法”的研究。可是，能否由此推论出马克思主义经典作家就“从来没有研究过语言”，在他们的著作里也没有写到过“语言研究的理论原则”呢？当然也不能。如前所述，马克思主义经典作家关于语言的理论是丰富的。不仅如此，即便是关于语言内部规律的研究，也不是绝对没有。例如，恩格斯的《法兰克时代·法兰克方言》，就是研究德国古代的历史和语言的著作。它汇集丰富的语言资料，通过比较研究，论述了法兰克语的流变及其内部规律。

综上所述，马克思主义经典作家关于语言的基本理论，不仅对于语言研究有指导意义，而且在理论语言学的发展史上，对于建立科学的语言观亦具有重大的历史意义。如果以马克思和恩格斯所写的《德意志意识形态》这一部论及若干语言问题的巨著为标识的话，它作于1845—1846年，是时西方的现代普通语言学诸流派尚未兴起，从理论上系统阐述“语言”社会性的《普通语言学教程》的问世，几乎比之晚了大半个世纪。而马克思主义的早期著作，无疑已经完成了关于语言研究的若干基本原理的奠基工作。

① 见《语言作为一种社会现象》，载《读书》1980年第4期。

② 见《语言研究必须解放思想》，载《江汉论坛》1980年第1期。

试论语言的离散性和模糊性

钱 锋

引言

自十九世纪以来,科学界有一种有趣的现象。一方面,是数学这个精密化的工具在完全置语言学不顾的情况下蓬勃发展;另一方面,则是语言学界的有志之士紧紧跟随着数学的脚步,他们探索、尝试着每一种新的数学分支,要想用之于语言研究精密化的宏伟大业。但是,却始终没有尽如人意。

语言的一个特点是离散性。在众多的语言现象中,或多或少具有连续性的唯一现象是语流。但任何一种语言中,语流也只有在人脑中切分为单个的元素之后才能为人们所理解。再者,用精密的物理仪器分析之下,语流也呈现着离散的分隔界限。

语言的另一个特点是模糊性。模糊性与离散性之间有着辩证的关系。离散性说的是有界限,模糊性说的是界限却又不分明的。这就象现代物理学的“粒子性”和“波动性”一样,是一种根本的性质。

世界上有离散的事物,因此而产生了离散数学;世界上有模糊的事物,因此模糊数学(国际上更多是称作模糊子集合论——Fuzzy subsets)应运而生。模糊数学也早已有人用于语言研究,在我国,也开始有人谈论这个问题。但是,似乎定性的议论多,定量的分析少。就是说,没有指出究竟该在那些地方以及怎么样应用模糊数学。

因此,本文的论说方法便是:

(1) 论证语言具有上述两种辩证的性质,阐述离散——模糊数学乃是语言精密化研究的出发点;

(2) 指出语言学各部门应用离散——模糊数学进行刻划的例子,试各说明一、二个問題,以引起大家进一步研究的兴趣。

1. 语言的离散性

1.1, 什么是离散性?因为它是与连续性相对立的,而且人们对前者的认识较后者要晚,所以我们可以从后者开始讨论。

连续性在数学上有严格的定义。例如,假如有一个实函数 $f(x)$,则在实数轴上可以选到任一个数列 $\{x_n\}$,对于点 a ,只要

$$x_n \rightarrow a$$

就有

$$f(x_n) \rightarrow f(a)$$

则我们说, $f(x)$ 在 a 点是连续的。

在上面连续性的定义中, 事物存在的空间是一个“连续统”。所谓连续统, 就是一堆不可数或叫不可列的东西。

离散性可以在上述连续性的基础上加以考察。离散性的事物不是分布在一个连续统上, 它们或者是单个而有穷的, 或者是单个、无穷而可数的。总之, 它们能与自然数(1, 2, 3, ...)建立一一对应的关系。这样的事物, 不管它具有什么样的性质(函数关系), 都不可能获得连续性。

1.2, 在时间上连续分布的语流, 只有在大脑中切成离散的一段一段并使每一段有意义, 才能使整个语流成为一句话。这是语言在时间上的离散性。

即使有一些短的、常见的语流, 我们已经感不到切分过程, 仍不能作为语言连续性的佐证。这是因为在大脑的长时记忆中已经有了一个代表整个语流意义的映象(我们称之为言语心象——verbalimage), 才使我们不假思索地听懂。因而, 这个语流在时间轴上包含在更大的语流中时, 也还是作为单个的实体出现的, 这恰恰证明了离散性。

1.3, 那么, 能否证明语言在空间上的连续性呢? 例如取地理上的线段作为坐标轴, 研究定义在这一线段上方言的变化^①。地理线段当然是一个连续统, 方言变化于是也就获得严格的连续性了。这样, 就可以用微积分或甚至是拓扑学来研究方言的差异分布了^②。

但是, 这实际上也似是而非, 因为:

一、地理线段固然是连续统, 但居民点却是离散的, 而离开了居民点便无所谓方言;

二、即使把居民点硬看作是不可列的排列在地理线段上, 其上的方言仍不可能有类似在微积分中

$$x_n \rightarrow a$$

$$f(x_n) \rightarrow f(a)$$

这样的性质。因为这里 $f(x)$ 也是一个实数。方言则完全不是这样。方言不可能笼统地用一个实数来表示。我们认为, 也许它可以用一个向量来表示:

$$f(Y, C, F)$$

这里, Y 代表语音, C 代表词汇, F 代表语法。而它们又都是离散的。

1.4, 我们的结论: 语言实体, 不论是语音、词汇、语法和修辞中的单位, 都是离散的。把形式与意义结合起来考察, 语言中没有数学上所说的那种严格连续的东西^③。

研究语言的离散性, 当然不能硬用连续统数学, 有现成的离散数学可以应用。离散数学, 就是探讨离散性的事物之数量关系和空间结构的数学。离散数学中的逻辑, 集论, 函数, 关系, 图论等等, 都可以恰当地用之于语言研究。^④

① 陈汉清, 朱建颂: 用数学方法描述方言的差别, 《方言》, 1979年第一期。

② 陈汉清, 朱建颂: 论拓扑学在方言研究中的应用, 《语言研究》, 1981年创刊号。

③ 钱锋, 冯志伟: 试论模糊数学在方言研究中的应用, 《华东师大学报》, 4(1983)。

④ Stanat, D. F., D. F. Mcallister: Discrete Mathematics in Computer Science, Prentice-Hall, 1977.

2. 语言的模糊性

2.1, 模糊性是一种较之离散性更为深刻的性质, 模糊性伴随着人而产生。可以说, 凡有人参加的活动, 就必然或多或少带有模糊性。语言的宿主是人, 语言于是有模糊性。

2.2, 模糊性的含义我们可以从明确性推得。明确性是指事物之间有确定不移的界限, 模糊性即是指事物没有确定不移的界限。

模糊性与含混性不同。含混是指信息的缺乏。在补充了信息之后, 含混性即得到消除。模糊性并非信息量不够, 而是由事物本身的含义或人们对它们的解释所决定的。

这通过一个例子即可说明。语言中的歧义是一种含混性, 在补充了信息之后, 歧义就消除了。但是, 造成歧义的, 是语义上的模糊性。在歧义消除后, 模糊性依然存在。

模糊性与概率性也不同。概率与随机性相联系, 而模糊与可能性相联系。可能性则又是比随机性要深刻的一种性质。例如, 名词“人”所表示的就是现实的一种范围、一种可能性, 而不是一种概率性, 因此, “人”的含义也是模糊的。

2.3, 语言之有模糊性的根本原因之一是语言单位是与概念相联系的, 而概念几乎全是模糊的。举例: 我们面前有一本东西, 我们不能断言说它是“书”还是“小册子”。显然, 并不是缺乏关于这两者的信息, 诸如颜色、厚薄之类, 而是由于在概念上这两者就是模糊的。这种模糊性与其说是客观事物本身固有的, 还不如说是人们在使用词语时带入的, 是人赋予语言的。

2.4, 语言模糊性的起源, 从原理上说要追溯到语言本身起源的那些岁月。

犹豫以无知为前提。概念的不精确性起源于没有能力作出判决, 后者则又是由于无知。这里说的无知是广义的, 指的是碰到了一堆相互关联的矛盾。我们认为, 人类在创造、发展语言的漫长过程中曾面临着这样四种矛盾。

(1) 具体与概括

人看到的只是一本一的书, 概念也是从具体的书提炼出来的, 但是“书”这个概念却又要概括所有的书。

(2) 个别与全貌

概念是由客观经过感觉、映象在大脑中形成的, 这些个别的東西通过感觉和映象就混成了一个复合的东西, 不再是孤立的了。书就是这样一种界限模糊的复合映象。

(3) 原则与灵活

语言与其他有人参与的事物一样, 会在使用的过程中不断得到改造或“磨损”。在形成概念时可能有一个原则或依据, 但在以后经年累月的使用中, 不同的人会发挥灵活性而对它加以增补, 而这种增补又会进入全民语言。

(4) 无知与有知

人在相当无知时所“创造”的语词, 也要容纳以后他们已经相当有知时的认识。

2.5, 由此可见, 语言, 正如其他有人参与的事物一样, 都或多或少带有约定俗成或主观的性质, 尽管这种主观也许以一个民族为单位体现出来。有的学者认为, 凡带有

主观性的东西，就依赖于人的判断或评价，就很难应用严格的数学方法分析。我们不同意这种观点。问题不在于能否应用数学，而在于应该用什么数学。也就是说，我们如果要对一个对象进行严格的科学描述而又碰到困难，那么，我们不应该企图改变对象来适应我们的工具，相反，应该改造工具以适应这种对象。把微积分用于方言研究可说是前者，应用模糊数学则属于后者。

2.6, 实际上，模糊性不仅是语言所特有的性质。二十世纪科学技术的一个重要发现是：我们要对一个事物、一个科学研究对象进行精确描述时，往往还会碰到这样的情景：随着事物的复杂性增加，人们对它们进行精密描述的能力便减小。其实，“不精密”的描述并不是人类的“马虎”，恰恰相反，它倒应看作是一种进步。模糊描述便有一些显著的优点，我们从下面的讨论将会看到。

3. 离散——模糊数学对于语言研究的指导意义

3.1, 离散——模糊数学对语言研究的指导意义，目前仍未引起语言学家，甚至是那些赞成把现代数学用于语言研究的语言学家的普遍重视。在国内，有志于语言精密研究的学生往往被鼓励去选修微积分，而不是离散数学或模糊数学。在国外也是如此。1978年，美国在这方面的著名学者、麻省大学的B. H. Partee教授出了一本书：Fundamentals of Mathematics for Linguistics，只字不提模糊数学。因此，有必要举出经典数学在某些方面行不通，而模糊数学却行得通，乃至离散数学也行不通，只有模糊数学才能解决问题的例子。

3.2, 经典数学(连续统数学、微积分等)行不通的地方，前面结合方言已经讨论过了。下面是一个离散数学也行不通，只能用模糊数学解释的例子。

假定我们用离散数学中集合论来研究语言学中的语义问题。我们要研究一种语义相近、但有差别，而差别却又是模糊的这一类词。以“书”和“小册子”作为代表。

运用经典集合论的方法，我们可以从数学上证明下列定理(证明从略)：

[定理1] 任何页数有限的书都是小册子。

这个结论当然是不能接受的。但是，我们从严格的经典数学出发，为什么会得出违背常识的结论？不是说集合论错了，而是它用的不是地方。如果硬要套到它所不适用的语言学对象，就会得出毫无意义的结论。

3.3, 现在，让我们换一个观点，即模糊数学的观点，仍然来研究语义学。

用模糊数学的观点来看“书”和“小册子”的语义差别，就要抛弃下面这种推理方法：对任何一本书S而言，要么S是小册子，要么S不是小册子，无“中间道路”可言。我们宁愿这么说：对任何一本书S，我们感兴趣的只是，S在多大程度上是“小册子”？如果页数很小，则S是小册子的可能性程度增大，如果页数相当大，则S是小册子的可能性程度就减小。于是，我们有一个新的定理：

[定理2] 任何一本书都在可能程度t上是一本小册子。其中 $0 \leq t \leq 1$ 。

显然，这样的结论更符合语言的实际。在语义学上，涉及那些语义相近而差别模糊的情况时，我们不是说：在什么确定的情况下一定要用甲，而在另一种确定的情况下则一

定要用乙。我们是说：通常在这种情况下用甲，而通常在那种情况下用乙，等等。

4. 语言学应用离散——模糊数学的方法和途径

4.1, 本节中,我们将要简单讨论离散——模糊数学可以用于哪些语言学部门以及如何应用。

本节的基本出发点是关于模糊子集的定义,陈述如下:

设有一个讨论问题的范围,讨论涉及的个体都在其中选取,这一范围称之为论域,记为 U 。

U 上的一个模糊子集是由一个隶属函数 μ 决定的:

$$U \xrightarrow{\mu} [0,1]$$

也即

$$0 \leq \mu(x) \leq 1 \quad x \in U$$

或者说,模糊子集 u 也就是由其隶属函数 μ_u 来下定义的。

4.2, 语音学(音位学)

在语音学和音位学上,如果引入模糊子集的概念,则问题可能讲得更清楚。

4.2.1, 模糊音位学 任一个音位 y 都可以看成全部音素论域 U 上的一个模糊子集,它的隶属函数是 μ_y 。为了方便,我们仍用传统的国际音标符号表示它。

于是,对任何的 $u \in U$,都可求出 $\mu_y(u)$,但各个 $\mu_y(u)$ 的数值不一样。举美国英语音位 $/\phi/$ 为例。从物理学上,我们可以精确地测定(或更精确地说是规定), $[\phi]$ 的频率、幅度、音色、音长应是多少。显然,没有谁能够绝对准确无误地发出这个 $[\phi]$ 来,所有现实存在的,只是一些与 ϕ 相近的音。它们与元音 ϕ 接近的程度,可以用 μ_ϕ 来表示。

μ_ϕ 是如何求得的?它实际上是对很多人的发音进行调查的结果。如果把问题简化,则可以用下面的方法推导 μ_ϕ 。

对 1000 个人(实际调查情况当然要复杂得多)进行调查发现:

其中把 ϕ /发成 (ϕ) 的有 819 人

把 ϕ /发成 (a) 的有 74 人

把 ϕ /发成 (o) 的有 50 人

把 ϕ /发成 (α) 的有 40 人

把 ϕ /发成 $(\phi\perp)$ 的有 10 人

把 ϕ /发成 $(\phi\uparrow)$ 的有 3 人

把 ϕ /发成其他的有 4 人

则

$$\mu_\phi(\phi) = 0.819$$

$$\mu_\phi(a) = 0.074$$

$$\mu_\phi(o) = 0.05$$

$$\mu_\phi(\alpha) = 0.04$$

$$\mu_\phi(\phi\perp) = 0.01$$

$$\mu_\phi(\phi\uparrow) = 0.003$$

如果我们分别用 $y_1, y_2, \dots, y_6$ 来代表这6个音, 则 $/\phi/$ 这个模糊子集又可表示为

$$/\phi/ = 0.819/y_1 + 0.074/y_2 + 0.05/y_3 + 0.04/y_4 + 0.01/y_5 + 0.003/y_6 + \beta$$

这里, β 是一个很小的修正因子。

或者

$$/\phi/ = \sum_{i=1}^6 \mu_i / y_i + \beta$$

这种描述方法的好处是:

- (1) 避免了用一个抽象的音位概括许多实际不同的发音这种不明确的说法;
- (2) 确切地显示了一个音位到底以哪些具体的发音, 即音位变体而存在的, 它们的比例又是如何, 或者以那些方音存在(如上例), 它们的比例又复如何;
- (3) 这样的表达式可以参加进一步的运算, 例如, 求两个音位的“差别”(距离)、“复合”等等。

4.3, 词法与语义问题

4.3.1, 模糊语义学

我们在§3.2—§3.3中已经举了如何应用模糊子集合论的方法描写语义的例子, 可以看出, 这对于细致地刻画那些语义相近的词儿之间的差别有特殊的意义。

4.3.2, 语义空间与语义向量

把模糊数学与离散数学结合起来, 可以比较精确地描绘词义, 特别是能有一种比较同一语言中或不同语言之间近义词的方法, 这就是语义空间和语义向量。

语义空间是各语义坐标所组成的一个多维空间, 下面是一组动词可能的语义坐标:

x_1 : 意念	x_5 : 定着
x_2 : 主动	x_6 : 能力
x_3 : 贯注	x_7 : 结果
x_4 : 好奇	

显然, 这是目前信息辞典中属性的推广。我们可以规定在每一坐标上取值: $+1, 0, -1$ 。这样, 每一个动词都可以有一个语义向量, 用 $v(x)$ 表示,

英语

$$v(\text{see}) = (-1, -1, 0, 0, 0, 0, +1)$$

$$v(\text{look at}) = (+1, +1, +1, +1, +1, 0, 0)$$

$$v(\text{watch}) = (+1, +1, +1, +1, -1, 0, 0)$$

日语

$$v(\text{みる}) = (+1, +1, +1, +1, 0, 0, +1)$$

$$v(\text{みえる}) = (-1, -1, -1, -1, 0, +1, +1)$$

$$v(\text{ながめる}) = (+1, +1, -1, +1, 0, 0, 0)$$

把英语与日语中的近义词两两比较, 它们之间的差异就一清二楚了。

也许, 更好的办法是把“意志”、“贯注”、“好奇”等等都看成是模糊子集, 再研究汉

语、日语动词的语义相对这些模糊子集的隶属函数) 就是模糊

$$v(\text{看见}) = (0, 0, 0, 0.5, 0, 0.5)$$

$$v(\text{みる}) = (0.75, 0.7, 0.7, 0, 0, 0.7)$$

等等。这就是说,“看见”的词义大致上是:“意怠”、“专注和生动”定着都是0,好奇是0.5,而结果则是0.9。由此看出,它与同义词日语みる实际上是很不同的。

4.3.3, 词类划分问题

在《面向人文科学的现代数学引论》中,我们曾讨论离散数学和汉语词类划分的潜在用途。那里指出,词类划分恰似一个代数划分,也应该建立在等价关系之上,即这种关系R应该满足:

$$\text{自反性: } R(x, x)$$

$$\text{传递性: } R(x, y) \wedge R(y, z) \implies R(x, z)$$

$$\text{对称性: } R(x, y) \implies R(y, x)$$

这里,我们进一步指出,实际上在语言学中并不存在离散数学中的这种精确关系,而只存在着模糊关系。模糊关系 $R(x, y)$ 并不永远为1或为0,而是在0与1之间变化。

我们建议可以从一种特殊的模糊关系即模糊替代关系入手来研究这个问题。模糊替代关系并不要求在100%的情况下, x 与 y 都可相互替代,而是象前面的例子一样,只要求一定数量的情况下可以替代就行了。同样,上面的三种性质也是模糊的。我们假设以有一个隶属函数 μ , 这时

$$\text{自反性} \quad \mu \implies R(x, x)$$

$$\text{传递性} \quad R(x, y) \wedge R(y, z) \xrightarrow{\mu} R(x, z)$$

$$\text{对称性} \quad R(x, y) \xrightarrow{\mu} R(y, x)$$

我们在《面向人文科学的现代数学引论》(讲义)中,曾曾经用离散数学完全等价用乎句子的语法、逻辑分析。

因为语言有模糊性,所以可以把模糊子集合吸收到句子的逻辑分析中来。或者说,从这里用的是模糊逻辑的方法。

4.4.1, 量词的作用域问题

在汉语里我们说:

我想不是这样

但是,英语里表达同样的意思,却说

I don't think so.

在《面向人文科学的现代数学引论》(讲义)中,我们曾经讨论用命题逻辑和谓词逻辑来分析汉语句子的问题。如果作这样的分析,上述两种语言的区别就很清楚。

汉语 $P(\sim Q(x))$

英语 $\sim P(Q(x))$