

# 理解模糊概念的动态特征<sup>\*1)</sup>

马谋超

李文馥

孟庆茂

中国科学院心理研究所

中国科学院心理研究所

北京师范大学心理学系

## 摘 要

本文探讨了不同场合下人脑理解模糊概念(量词、修饰词)的动态特点,并且考察了不同赋值区间(论域)、比较级的基数、上下关联(Context)等变量对理解给定的模糊概念的影响,特别考察了上述问题上的年龄差异。

## 前 言

人们普遍承认,凭着“矮胖个、勾鼻子、大鬍子”的模糊信息,比告诉精确的身高体重、勾鼻的弧度以及鬍子的根数更容易从人群中识别出所说的对象。这有力地提示人们,语词充满含混性(模糊性),但不失信息。人脑就是这样把它们巧妙而和谐地统一在一起,从而有效而经济地进行着各种活动。应当说,研究人脑如何把握语义中模糊信息的特点,可以认为是探索人类智能的一个重要方面。

随着知识工程中自然语义的模糊表达的兴起,促进了对语义模糊性的研究。这类研究表现为两个方面:一是旨在如何用模糊数学表达<sup>[1,2]</sup>,一是通过赋值的统计,以确立它的经验基础<sup>[3-8]</sup>。本工作沿着后一方向,进一步探讨人脑理解辞义中模糊信息的动态与发展特点。

所说的动态与发展特点,指的是由辞义引出的模糊性的变化。而模糊性的量化基础则是模糊集的隶属函数。本工作的基本思路就是通过对若干量词或修饰词的赋值作模糊统计,以获得它们的隶属函数,然后,以此作为基础,进而推论作出结论。

## 方 法 与 程 序

受试者:男女大学生102名。幼儿60名,年龄在五岁半至六岁半,是两个幼儿园大班的儿童。

试验材料:事先印好的试验表格。该表格包括三部分:修饰词赋值,如:较差[ ]、差[ ]、较好[ ]、好[ ]等;在不同条件下,对量词“几个”、“很多”赋值,如表1所

\* 本工作为国家自然科学基金资助课题

1) 本文于1989年5月4日收到。

示；在系列变化的论域中，要求对量词“几个”赋值。该论域的系列为[0,10]、[0,20]、[0,30]、[0,40]、[0,50]、[0,100]。

表1 量词赋值设计

论 域		“几个”	“很多”
无比较的基数	[0,10]	[ ]	[ ]
	[0,20]	[ ]	[ ]
比2多	[0,10]	[ ]	[ ]
	[0,20]	[ ]	[ ]
比5多	[0,23]	[ ]	[ ]

赋值方法：要求被试依照指定词的辞义，选取一个适当的数字范围或一个数。赋值规则为：辞义所含的属性越多，选取的数字应越大；辞义越含混，选取的数字范围相应地更宽。如果自认定毫无含混，则可用一个数表示。

对幼儿来说，赋值区间的确定与成人被试所作不同，即不作直接抽象地估量，而是从一定数量的事物中，依照指定词检出合适数量的事物来。对每一个词都务必作两次检出的操作。这两次操作的时距为两小时以上。实验时，物品被平摊在桌面上的特定部位。两次检出操作所得数，被视为区间的两端（方法学上的根据，见本文讨论部分）。由此构成的区间可同成人被试估量的区间比较。

试验程序：按照所述的方法，将102名成人被试，分期分批进行集体试验。幼儿的相应试验在幼儿园里进行。进行试验的方式是在老师的指导语下单个独立完成。

数据处理：把每个被试对特定词所作的赋值区间进行统计。步骤是，首先，找出数轴上每个实数（元素）的频数，然后，除以总数求得隶属频率。根据随机集落影理论<sup>[1]</sup>，这种隶属频率可视为模糊集的隶属度。一个元素的隶属度用以表征它对模糊概念的符合程度。相应地，隶属频率分布被当作隶属函数的描述。

## 结 果 与 分 析

表2与表3分别列出了各种条件下对应于量词“几个”与“很多”的隶属频率分布及其加权平均值 $\bar{x}^*$ 。

在表2与表3所列数据基础上作如下分析：

1. 在不同赋值区间中，量词的隶属频率分布 表2与表3列出了[0,10]与[0,20]的隶属频率分布。将它们两两对照（即表2中序号1与3,2与4,5与7,6与8,以及表3中的序号1与3,5与7各列数据）可以清楚看出，不同赋值区间中的量词“几个”、“很多”的隶属频率分布是不同的。其特征表示在双对数座标纸上为一条直线，即幂函数，如图1所示。

$$\bar{x} = \sum_{i=1}^n r_i f_i, \text{ 在这里, } r_i \text{ 代表给定数轴 (论域) 上的实数; } f_i \text{ 代表隶属频率。具体的统计及其数据处理, 参}$$

见马谋超, 词义赋值的模糊统计分析, 心理学报1989, 1.

表 2 量词“几个”的辞义赋值统计结果

	No.	[0,10]	1 2 3 4 5 6 7 8 9 10										$\bar{x}$						
无比较基数	1	f <sub>成</sub>	.52	.71	.75	.44	.23	.05	.03	.01	.01						2.83		
	2	f <sub>儿</sub>	.02	.13	.40	.58	.67	.40	.17	.10	.02						4.71		
	3	[0,20]	1 2 3 4 5 6 7 8 9 10 11 12 13 ... 20										4.35						
		f <sub>成</sub>	.44	.56	.75	.74	.73	.54	.35	.23	.14	.05							
4	f <sub>儿</sub>	.03	.28	.43	.50	.52	.50	.32	.20	.17	.01	.03	.02						5.50
	5	[0,10]	1 2 3 4 5 6 7 8 9 10										4.75						
f <sub>成</sub>		.48	.66	.76	.40	.21	.05												
6	f <sub>儿</sub>	.45	.60	.55	.27	.05	.02						4.45						
	7	[0,20]	1 2 3 4 5 6 7 8 9 10 11 12 ... 20										6.26						
f <sub>成</sub>		.39	.54	.75	.73	.58	.46	.33	.19	.04									
8	f <sub>儿</sub>	.37	.60	.70	.48	.25	.18	.07	.05	.03	.02						4.05		
	9	[0,23]	... 6 7 8 9 10 11 12 13 14 15 16 17 18 ... 23										9.74						
f <sub>成</sub>		.36	.53	.74	.76	.65	.43	.37	.24	.18	.11	.05		.03	.03	.01	.01		
10	f <sub>儿</sub>	.37	.68	.63	.30	.15	.12	.08	.05	.05	.03						8.25		

注: No. 表示序号; [ ] 表示赋值区间; f<sub>成</sub> 与 f<sub>儿</sub> 分别表示成人与儿童的隶属频率;  $\bar{x}$  表示加权平均值。

表 3 量词“很多”词义的赋值统计结果

	No.	[0,10]	1 2 3 4 5 6 7 8 9 10										$\bar{x}$							
无比较基数	1	f <sub>成</sub>	.01	.05	.16	.36	.60	.71	.57	.29						7.67				
	2	f <sub>儿</sub>	.20	.07	.20	.35	.43	.53	.28						8.03					
	3	[0,20]	... 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20										14.99							
		f <sub>成</sub>	.01	.01	.03	.08	.09	.25	.29	.38	.42	.50		.66	.55	.50	.49	.40	.35	
4	f <sub>儿</sub>	.02	.02	.08	.13	.18	.27	.25	.28	.40	.40	.40	.35	.33	.27	.18	.12	?		
	5	[0,10]	1 2 3 4 5 6 7 8 9 10										7.67							
f <sub>成</sub>		.01	.05	.13	.35	.59	.69	.54	.30											
6	f <sub>儿</sub>	.18	.25	.47	.53	.43	.25	.13						6.92						
	7	[0,20]	... 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20										14.05							
f <sub>成</sub>		.01	.02	.05	.12	.17	.20	.38	.39	.52	.52	.60		.68	.53	.46	.40	.33	.28	
8	f <sub>儿</sub>	.08	.22	.28	.25	.38	.37	.35	.43	.38	.23	.23	.25	.25	.20	.15	.08	.07	?	
	9	[0,23]	... 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23										17.28							
f <sub>成</sub>		.01	.01	.03	.03	.13	.16	.22	.27	.28	.51	.55		.57	.61	.55	.58	.41	.36	.33
10	f <sub>儿</sub>	.03	.13	.18	.20	.35	.42	.37	.32	.27	.22	.30	.32	.35	.27	.23	.17	.12	.07	?

注: “?”表示作加权平均值没有意义。

2. 不同比较基数中的量词隶属函数 表 2 与表 3 中分别列出了量词“几个”与“很多”在比较基数 2 与 5 两组相应的隶属函数(见表 2 与表 3 中序号 7 与 9)。将它们的加权平均值  $\bar{x}$  加以比较, 可以明显看出有 3 个单位的相差。可是, 把比较基数 5 的比较

基点位移 3 个单位数与比较基数 2 的基点重合时,它们之间则显出良好的一致性,如图 2 与图 3 所示。

这一事实可以作如下解释:两者相差的距离正好是被比较基数的差数本身。实际上,这是实验条件给定的。而曲线的形状才是真正表征人脑对量词辞义的主观反映。由

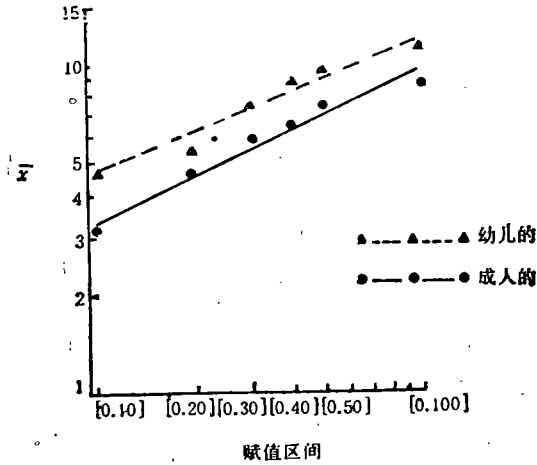


图 1 赋值区间与“几个”之间的关系

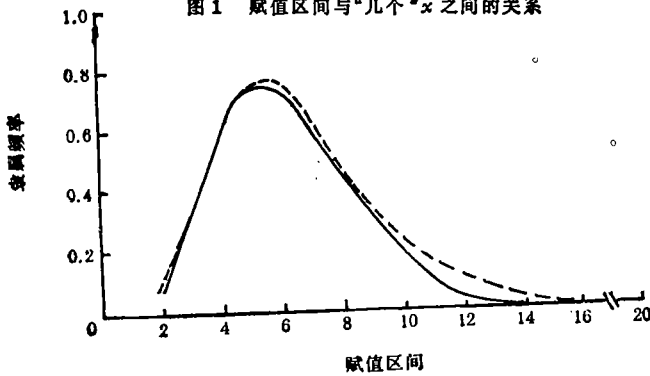


图 2 位移后的“比 5 多几个”与“比 2 多几个”隶属函数的比较

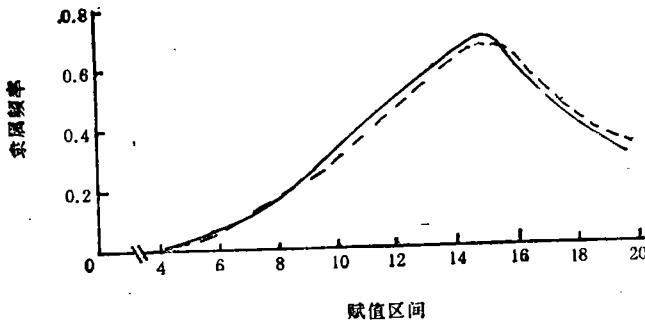


图 3 位移后的“比 5 多很多”与“比 2 多很多”隶属函数的比较

此有理由认为,判别比较基数是否成为制约人脑理解模糊概念的变量,其判断应当是曲线的形状,而不是它们之间在数轴上的差数。图 2 与图 3 表征出的一致性说明比较基数不是真正心理语言反应的真正变量。

另一点值得提及的是,以 2 与 5 作比较基数的论域,虽然在形式上不同,前者是 $[0, 20]$ ,后者是 $[0, 23]$ ,但是,实际上,区间的差数均等于 18。

3. 不同年龄被试的相应隶属函数特征 在无比较基数条件下,由图 1 表征的直线截距,幼儿的明显高于成人。在表 2 中同样条件下的  $\bar{x}$  值,也表征出上述的特征。

在不同的比较基数条件下,相应的函数如图 4、图 5 和图 6 所示。

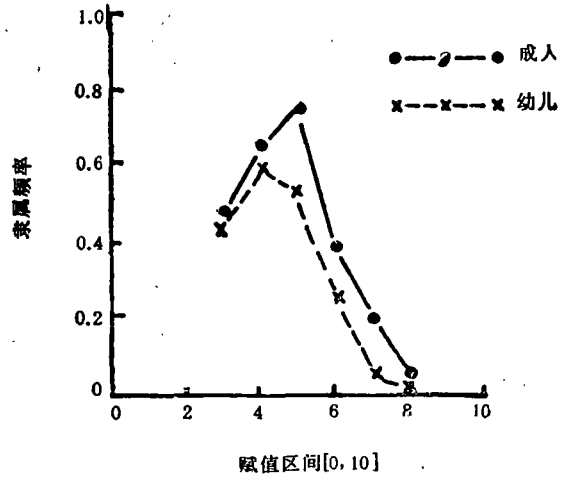


图 4 [0, 10]论域上,“比 2 多几个”的隶属函数

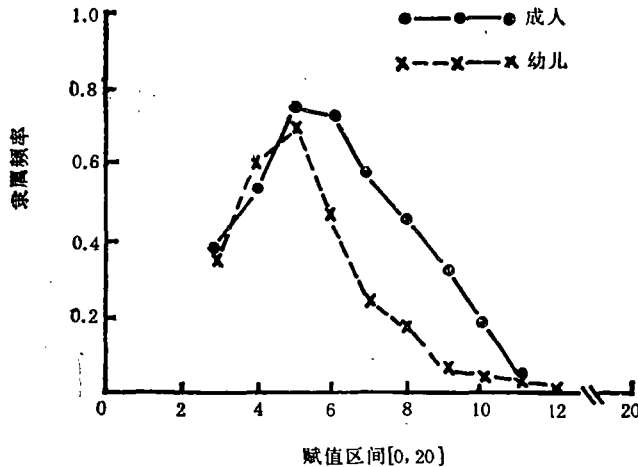


图 5 [0, 20]论域上,“比 5 多几个”的隶属函数

图 4、图 5 和图 6 一致显示出幼儿的曲线高于成人的相应曲线。这个趋向还可从表 3 中 $[0, 10]$ 论域上,“比 2 多很多”的  $\bar{x}$  得到印证。

值得注意的是,在 $[0, 20]$ 的论域和 $[0, 23]$ 的论域上,无论无比较基数条件,还是比较基数为 2 或 5 的条件下,对于量词“很多”,幼儿的曲线均没有象在 $[0, 10]$ 论域上显示的那样有规律,并且它们的隶属度均没有能超出 0.5 的。如此低的函数值和无规律的曲线,有可能推断,在这些论域上,量词“很多”在幼儿的心理语言词汇表中尚未真正形成。

4. 关联 (Context) 因素对模糊集隶属函数的作用 所谓关联因素,在这里,指的是一个维度上,原词并存在着另一个量词修饰词,比如,在“累”的维度上,与原词“累”一起的还有“很累”,而在“好”的维度上,却仅有原词“好”,而缺“很好”。表 4 列出两个维度上不

同修饰词的 $\bar{x}$ 。

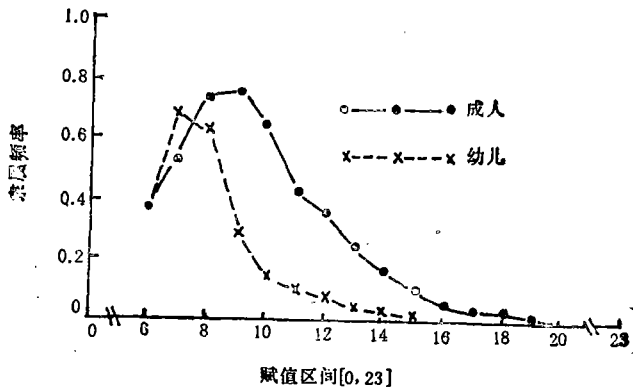


图6 [0, 23]论域上, “比5多几个”隶属函数

表4 修饰词词义的赋值 $\bar{x}$

差	较差	较好	好
1.70	3.40	6.95	8.80
不累	稍累	累	很累
1.53	3.49	6.22	8.94

在表4中,“好”与“累”同是无量词修饰的原词,但是,它们的 $\bar{x}$ 却相差甚远,或者说,“好”的 $\bar{x}$ 更接近于“很累”的。图7示出了“好”与“累”、“很累”的隶属函数曲线。

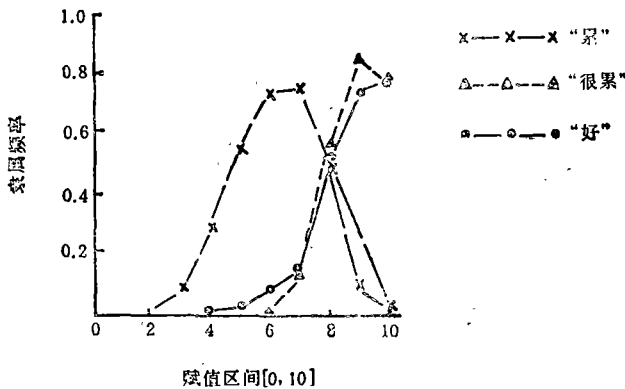


图7 修饰词“好”、“累”与“很累”的隶属函数

该曲线簇清楚表明,“好”的隶属函数曲线的形状,若说与“累”的类同,不如说,它同“很累”的更为相似或接近。

## 讨 论

在本工作中,关于检测幼儿辞义模糊性的方法问题是值得探讨的。其要点是,给以一个特定词的词义,幼儿必须分开进行两次操作。两次操作所得到的数据被分别作为区

估量的两个端点,而接弃将它们视为两次独立的随机试验结果。这种做法在方法论上的根据是,词义本身不是分明的。因此,在评价量表上不该是一个点,应是一个区间。实际上,幼儿从一堆扣子中,依照给定的量词,比如“几个”,进行捡取的操作时,无论如何一次只能是某个数量。但是,这决不意味着此时此刻,论域中的其它数量就不能成为“几个”(集合)中的一部分元素。从逻辑上说,在“几个”的命题下,既然一次操作捡取了5个;另一次捡取了2个,那么,虽然没有检出3个与4个,但是可以推断,它们也必然包含在其中。因此,把2与5作为他对“几个”估量区间的两个端点是合理的。

然而,两次捡取操作所得到的两个数据视为估量区间的两个端点,并不意味着真正的端点不可能在这两点之外。从理论上说,确实,操作多次更能逼近于真实的界线。不过,这已不是问题的本质了。

现在,让我们再讨论一下本文中图1所表征的下面一点,即在各个论域上,成人的 $\bar{x}$ (对应于量词“几个”),尽管随论域的扩大而增高,但是始终未能超出10。对此,有两个解释:1.在成人看来,量词“几个”与“十几个”是有严格界线的。其临界点正是10;2.在心理语言的范畴里,它们之间的界线并非如此严格,正如图1中幼儿的 $\bar{x}$ 所表征的那样(见[0, 100]上的 $\bar{x}$ )。由此,设想将论域进一步扩大,比方[0, 150],在此场合下,成人的 $\bar{x}$ (“几个”)也可能突破10的界限。究竟如何尚待证实。

## 结 论

1.本工作经验地表征了人脑对量词“几个”、“很多”以及“差”、“较差”、“较好”、“好”,“不累”、“稍累”、“累”、“很累”等辞义的理解具有典型的模糊性质。这些模糊性质受论域、关联因素的影响表现出动态的特性。与此联系,有如下2、3两点。

2.给定量词“几个”、“很多”的范例研究指明,论域的变化导致了相应的隶属函数的改变。两者之间的函数关系在给定的范围内遵循幂定律。

3.在心理语言的范畴内,词义的赋值存在关联(Context)效应。

4.年龄差异明显地表现在给定的量词“几个”、“很多”的隶属函数上。成人被试在给定的所有条件下,对于“几个”、“很多”词义的隶属函数均显示出规律性;而幼儿被试只能对“几个”的词义作出有效的赋值,至于对量词“很多”,除了在小论域[0, 10]外,均无有效数据。由此表征模糊性的年龄差异。

5.在本工作中,比较级的基数没有显示出对给定量词“几个”、“很多”引出的模糊集函数形状以实质性的影响。

## 参 考 文 献

- [1] Zadeh, L. A. , "A computational approach to fuzzy quantifiers in natural languages". *Computers and Mathematics with Applications*, 1983,9,149—184.
- [2] 陈国权编著, "知识工程中自然语义的模糊表达", 1989, 科学出版社。
- [3] Hersh, H.M. and Caramazza, a., "A fuzzy set approach to modifiers and vagueness in natural language". *J. Exper. Psychology, General*, 1976, V. 105 N. 3, 216—228.
- [4] Pipino, L. L, Van Gigch, U. P and Tom, G.  
"Experiments in the representation and manipulation of labels of fuzzy sets." *Behavioral science*, 1981, 26:216—228.
- [5] 张南纶, "随机现象的从属特性及概率特性", *武汉建材学院学报*1981, 1, 2, 3期。
- [6] 马谋超, "心理模糊性的测量", *心理学报*, 1981, N. 1, 64—75。
- [7] 马谋超, 邵正强, "模糊集合论的某些经验研究", *模糊数学*, 1985, N. 3, 53—62。
- [8] 马谋超, "修饰词词义赋值的模糊统计分析", *心理学报*, 1990, 1, 53—62。
- [9] 汪培庄, "模糊集与随机集落影", 北京师范大学出版社, 1985。

## EXPLORING CHARACTERIZATION OF DYNAMICS IN HUMAN COMPREHENSIVE FUZZY CONCEPTS

Ma Mouchao Li Wenfu Meng Qingmao

*Institute Psychology, Department of Psychology,  
Academic Sinica, China Beijing Normal University*

### Abstract

This paper explores characterization of dynamics in human comprehensive fuzzy concepts as qualifiers or modifiers under given different conditions. It is considered that variables such as assigning numerical interval (discourse universe), cardinal numeral of comparison and context factor exert an influence on comprehension of the fuzzy concepts. Emphasis is placed on the examination of nature of age shown in between adult and child the point in question.