

汉语语义记忆提取的初步研究¹⁾

陈永明 彭瑞祥

中国科学院心理研究所

摘 要

对汉语语义记忆的提取进行了两个实验。实验1是自由联想实验,从十三个句子中选出二十三个词作为刺激词。结果表明,以两词间所保持的常识关系来定义的语义距离,对预测自由联想的作业是有效的。实验2采用反应时技术,对真假命题判断及信息提取的时间进行了考查。实验结果与A. M. Collins等人的结果一致,说明了存储于网络不同水平上的句子其提取时间不同这一事实。结果还表明,在信息提取阶段上,汉字字形的作用不是一个重要的因素。

有关语义记忆的问题,近年来国外的学者进行了不少的研究。这些研究主要是探讨语词的心理学表征方式,以及在理解语句时如何存取和利用这些表征的知识等有关问题。A. M. Collins和M. R. Quillian⁽¹⁾提出了第一个也是最著名的一个表征人们语义记忆的模型,叫做层次网络模型。他们曾指出,提取存储于该模型最低层的信息,所花的时间较短,而提取较高层的信息所花的时间较长。A. Caramazza等人⁽²⁾曾指出,以两词间所保持的常识关系来定义的语义距离,对自由联想的作业和对信息提取时间的作业是个良好的预测器,并断言主观结构的知识与词之间联结的距离能用来预测不同系统语言的作业。但这些研究都是用英文作为实验材料的。本研究的目的,企图探查以汉语作实验材料得到的结果与先前以英语作材料的结果是否一致。虽然英语和汉语均能很好地传递信息、交流思想,但这两种语言在细节上有很大的不同。汉语的语义单位是词。汉语中有不少语词是单字词,但双字词比单字词要多得多。而且,英语使用的是一种拼音文字系统,汉字则是表意文字。大多数汉字是由两部分组成的,即表达字义的部首部分和表达它的读音部分。汉字的这种特点以及中国人的思维习惯是否会对语义记忆及其提取产生某种影响,这是值得探查的。本研究包括两个实验:实验1,探查语义距离与自由联想作业的关系;实验2,探查真句和假句的加工时间特点,以及存储于网络不同水平上的信息与提取这种信息所需时间之间的关系。

实验1. 自由联想实验

实验材料 从小学课本选取十三个句子,并把它们逐一写在卡片上。除第1、3和5句为每句一个刺激词外,其余各句均选出两个词(大多数是双字词)作为刺激词。例如在

1) 本文于1984年7月10日收到。

“花圈献给为人民英勇牺牲的烈士”一句中,选取“花圈”和“人民”作为刺激词。此外,为了便于分析所得的结果,对每个句子都预先画出了表达该句的命题网络。图1是上述语句的命题网络。从该网络中我们可以看出,“花圈”与“烈士”这两个词相距4条链;而“花圈”与“献给”这两个词相距三条链。

被试 18名大学哲学系的学生,男女各半。

步骤 实验是个别进行的。实验时从这些卡片中随机取一张给被试看,并要求他记住卡片上的句子。然后拿走卡片,在被试口述该句子无误后,开始进行自由联想测验。首先给被试呈现一个刺激词,要求被试在看到这个刺激词后,尽快地把由该刺激词首先联想到的刚才默记过的句子中的任何一个词大声报告出来。十三个句子做完一遍以后,再呈现句子中的另一个刺激词,步骤同前述。

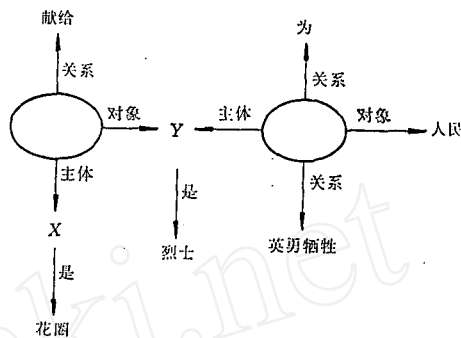


图1 表达句子“花圈献给为人民英勇牺牲的烈士”的命题网络

结果和讨论 实验1的结果见表1,这些结果表明了三点:

(1)命题网络理论者们认为,在自由联想作业中,被试从刺激词首先联想到的句中其他的词,经常是那些在命题网络中彼此靠近的词,即彼此相距的链数较少的词^[5]。但从本实验的结果中可以看到,二十三个刺激词中,被试的反应词符合命题网络假说预测的,仅有九个例子(占39.1%)。看来,企图用这个假说去说明语词提取的机制是颇可置疑的。

(2)与此相反,符合语义距离预测的例子却有十三个(占56.5%)。虽然在自由联想任务中,许多变量能影响被试的作业。但是,词之间在语义上彼此愈接近,或彼此形成了组块,则被提取的可能性就愈大。例如,在第7句中,刺激词“狐狸”,大部分被试的反应是“老虎”(占77.7%)。又如在第6句中,被试对刺激词“花圈”的反应,大部分是“烈士”。显然,这是由于“狐狸”与“老虎”和“花圈”与“烈士”它们之间的常识联系比较密切。看来,用语义距离来预测自由联想任务的作业,对汉语材料来说也是有效的。

(3)此外,在第12句中,被试对刺激词“报纸”或“公园”的反应,大多数是“老头”。看来这是难以用语义距离来说明的。但进一步考虑这个句子的内容,我们看到,“老头”这个词在该句子中占有重要的地位,其他的词都是描述它的。一些被试实验后的主观报告说,在默记了句子后,有些词的印象比较深刻,容易首先想到和说出来。因此,被试对该句的反应是可以理解的。

实验2 真假判断实验

方法 本实验采用对句子所包含的命题作真假判断的反应时技术。实验时,从快门的窗口每次呈现一个句子,要求被试在看到呈现的句子后,立刻作出该句的陈述是真抑或是假的口头报告,同时,测量出从句子呈现到被试作出反应所经过的时间。

表 1 实验 1 的总结果*

句子**	刺激词	反 应									符合网络 距离	符合语义 距离	
		词	次 数	链数	词	次 数	链数	词	次 数	链数			
1	哥哥	弟弟	13/18	4	严厉	3/18	3						√
2	劳动	农民	10/18	3	辛勤	4/18	3				√		√
	庄稼	农民	9/18	4	辛勤	3/18	5	劳动	9/18	5			√
3	水平	比赛	10/18	3	今天	3/18	4				√		√
4	慢慢	面包	7/18	5	小孩	6/18	3	吃	5/18	4			√
	面包	冷	7/18	3	小孩	7/18	4	吃	4/18	4	√		√
5	书	列宁	9/18	5	故事	6/18	3						
6	花圈	烈士	11/18	4	献给	4/18	3						√
	人民	烈士	7/18	3	献给	4/18	4				√		
7	狐狸	老虎	14/18	6									√
	百兽	老虎	8/18	7	吓跑	4/18	3	狐狸	3/18	4			√
8	一群	小鸡	10/18	3	我们	3/18	4				√		√
	食糟	吃	6/18	4	小鸡	5/18	5	米	4/18	4	√		√
9	鸚鵡洲	梅生	6/18	5	信	4/18	3	送	3/18	3			
	立安	梅生	8/18	2	鸚鵡洲	3/18	3	派	3/18	2	√		√
10	草根	鱼	6/18	6	画	5/18	5	马良	4/18	3			
	河边	鱼	7/18	6	马良	5/18	3						√
11	手榴弹	小金花	7/18	5	拉	4/18	3	噩耗	3/18	3			
	噩耗	母亲	6/18	4	小金花	3/18	2	手榴弹	3/18	3			
12	公匠	老头	9/18	6	椅子	4/18	4	报纸	3/18	8			
	报纸	老头	11/18	4	公园	4/18	8				√		
13	玫瑰花	校长	6/18	4	李明	3/18	3	美丽	3/18	3			
	小学	校长	11/18	3	玫瑰花	4/18	7				√		√

* 反应次数少于 3 次的不予列出 **参见附录

实验材料 实验使用了三十八个句子, 其中一些句子的陈述为真, 另一些则为假(真句 20 句, 假句 18 句)。每句的长度从四个字到六个字不等。句子分为三类: 动物、农作物和食品类。这些句子所陈述的内容, 都是一些常识性的东西。表 2 是实验所用的句子种类的例子。

为了便于表示, 我们对陈述属性关系的句子以 P 来表示, 对子集关系的句子则以 S 来表示。英文字母后面附加的数字表示被陈述的信息存储于层次网络模型中不同的水平上。在这里, 零表示网络的最低水平, 2 表示最高的水平。

表 2 实验 2 所用的句子种类例子

真 句	假 句
P0 鸭梨是黄的	P 杏子是方的
P1 白菜有纤维	P 芹菜有皮
P2 燕子有皮肤	P 企鹅有四足
S0 冬瓜就是冬瓜	S 枣子是核桃
S1 广柑是水果	S 大葱是麦子
S2 乌鸦是动物	S 野鸡是家禽

为了考察汉字字形对信息提取时间是否有助长作用,实验还使用了下述三个句子:麻雀是鸟;大雁是鸟;白鹤是鸟。根据汉字频度表(内部发行),雀、雁、鹤这三个字同属于不常用级的字。但是,其中的鹤字由两部分组成,它的部首恰好是鸟字,而其他两个字都不含有这一部首。

被试 18名被试,同实验1。

步骤 被试面对快门坐着,告诉他当快门打开的时候将呈现一个句子,要求他对这个句子陈述的事情尽快作出判断。如果认为句子陈述的事情是真,则大声说“真”;如果认为句子陈述的事情是假,则说“假”。句子呈现的次序是随机的。

结果和讨论

(1)被试对不同类型和网络不同水平上的句子的反应时是不同的,其各自的平均值见表3和图2。

表3 不同水平上的属性句和子集句的反应时平均值(秒)

句子	平均值	标准差
P0	1.497	0.278
P1	1.505	0.332
P2	1.801	0.514
S0	1.236	0.192
S1	1.411	0.299
S2	1.630	0.491

从表3中我们可以看到,陈述属性关系的句子,其反应时平均值比较长,而陈述子集关系的句子,其反应时平均值较短。实验结果还表明,处于网络中水平较高的句子,其反应时的平均值也较大。我们把同一类型句子间的平均值两两进行t考验,结果除了P0和P1两组句子之间的反应时差异不显著外,其余均达到显著的水平。总的说来,我们实验所得的结果,与Collins等人实验的结果颇相一致。

当然,我们也可以对存储于同一水平上的各个句子的反应时平均值进行变异数分析,看看它们之间的差异情况。如果同一水平上的三个句子的反应时平均值的差异不显著,那么,就可以说,这组句子中的每一句对被试来说是等价的,从而更好地说明各组之间反应时的差异确是起因于句子的不同水平。变异数分析的结果见表4。

在这六组句子中,P2和S1这两组句子其平均值的差异达到了显著的水平。这是由于它们

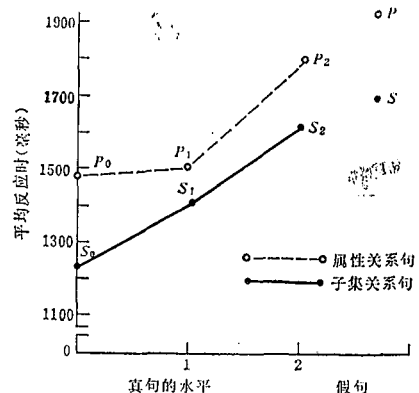


图2 不同类型句子的平均反应时

表 4 在同一水平上每三个句子的反应时平均值(秒)

句 子	平 均 值	标 准 差	F	P
P0: 公鸡会啼 菠菜是绿的 鸭梨是黄的	1.481 1.421 1.598	0.355 0.266 0.453	F(2.32)=1.603	>0.05
P1: 老鹰有翅膀 白菜有纤维 香蕉有皮	1.414 1.621 1.476	0.229 0.668 0.300	F(2.34)=1.559	> 0.05
P2: 燕子有皮肤 萝卜有根 石榴有维生素	1.969 1.422 1.980	0.774 0.323 0.722	F(2.34)=6.616	<0.01
S0: 麻雀是麻雀 冬瓜是冬瓜	1.231 1.237	0.230 0.209	F(1.17)=0.010	>0.05
S1: 大雁是鸟 黄瓜是蔬菜 广柑是水果	1.280 1.413 1.514	0.358 0.375 0.304	F(2.34)=4.983	<0.01
S2: 乌鸦是动物 扁豆是农作物 桔子食品	1.652 1.592 1.668	0.725 0.498 0.510	F(2.34)=0.177	>0.05

中有个别句子的难度太高或太低所致。例如，“萝卜有根”作为 P2 组的句子，相对于其它句子来说，难度太低了一些，致使该组句子的平均值差异达到显著的水平。但其他四组句子其平均值的差异都不显著。因此，除了极个别的句子以外，处于网络同一水平上的句子，对被试来说是等价的。

(2) 被试确证不同类型的真句和假句时，其句子加工所需的时间是不同的。表 5 列出了被试对真句和假句的反应时的平均值。

表 5 不同类型的真句、假句的反应时平均值(秒)

句 子	平 均 值	标 准 差	t(17)	P
P真句	1.600	0.335	3.782	<0.01
S真句	1.445	0.302		
P假句	1.923	0.463	3.180	<0.01
S假句	1.692	0.337		

从表 5 中可以看出，属性句真句的反应时平均值显著地大于子集句真句的平均值。同样，属性句假句的反应时平均值也显著地大于子集句假句的平均值。而且，属性句真、

假句的反应时平均值之间的差异和子集句真、假句的反应时平均值的差异,均各自达到显著的水平(前者: $t(17)=4.546, P<0.001$; 后者: $t(17)=4.584, P<0.001$)。总起来说,不管属性句是真抑或是假,其反应时的平均值都相应地大于子集句的反应时的平均值。同样,不管是哪种类型的句子,假句的反应时的平均值都大于真句的反应时的平均值。图 3 形象地给出了上述情况。

(3)我们在实验 2 中使用了三句各自指不同鸟名的句子,以便考察汉字字形对提取时间的影响。根据层次网络模型的假设,它们在网络中处于同一水平。结果见表 6。

表 6 不同鸟名的三个句子的反应时平均值(秒)

句 子	平 均 值	标 准 差
麻雀是鸟	1.226	0.272
大雁是鸟	1.279	0.358
白鹤是鸟	1.613	0.420

从表 6 中我们看到,最后一句的反应时平均值比前两句大些。变异数分析的结果为: $F(2,34)=13.958, P<0.001$ 。这结果表明,虽然“鹤”具有“鸟”字的部首,但它的字形对提取时间并没有明显的促进作用。这个结果与喻柏林等人^[4]的结果颇为一致。她们曾用 12 个彼此意义上无关联的,但每个字内部都含有相同的子模式(或相同的部首)的汉字,让被试进行短时记忆实验。结果表明,对短时记忆的容量并没有改进。看来,在信息的提取过程中,汉字字形的作用不是一个重要的因素。而事实与其概念之间联结的强度,对信息提取的时间具有重要的影响。我们知道,在人们的日常生活中,现实的麻雀是经常遇到的,而现实的白鹤却是很少碰见的。所以,“鹤”字虽有“鸟”的部首,但其平均反应时仍然比前者长。

根据层次网络模型的代表方式,检验一个句子时所涉及到的主要过程,就是对网络进行搜索,以便寻找一条把句子前部的主项和后部的宾项联接起来的合适的通道。这是层次网络模信息加工的特点。在我们的实验结果中,不同类型的句子或同一类型而在网络不同水平上的句子,其平均反应时所以有明显差别,也很可能主要是由加工句子时的搜索过程造成的。一个事物,往往具有许多属性,因此,从它发出许多链来联结它所具有的各种属性。所以,在确证它是否具有某种属性时,将会有搜索较多数量的链的可能性。相反,属于标明概念之间子集关系的链,数量较少,在加工语句时,可能需要加以搜索的方向较少。人的信息加工的方式,基本上是串行的,而不是平行的。所以,加工前一类句子的时间就会长一些,而后一类句子的加工时间就会短一些。同样,对那些处在网络不同水平上的句子的加工时间,也可以用搜索过程来说明。当然,语句加工的时间,也并不是唯一地由搜索过程所决定的。在某些情况下,其他的因素也会有重要的影响。有的被试在实

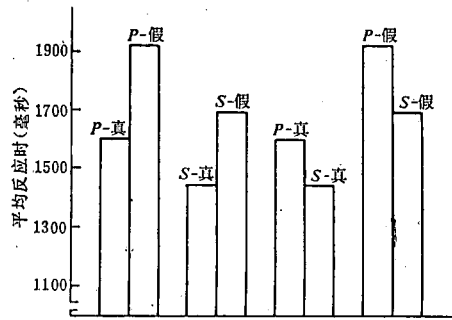


图 3 P 真句、P 假句和 S 真句、S 假句的平均反应时

验后的主观报告中说,事物的特征如果是经验中熟识的,就容易判断,事物的特征若是隐藏的,判断句子就难一些。例如,“燕子有皮肤”或“石榴有维生素”,对这样的句子必需稍加考虑以后才能作出判断。这也说明了为什么网络同一水平上的各种句子之间的反应时会有差别。

至于真、假句判断时间上的差别,我们认为,这是由于真句所陈述的东西对被试来说是现有的、随时可以利用的信息。而假句所陈述的东西则不是现存的,被试在验证句子时,假句陈述的东西,出乎于他的预期之外。因此,被试对假句作出判断就相应地需要较长的时间。我们的分析也为宋均*(1983)的实验所支持。他曾让小学生在速示条件下再认一些短句。例如“老牛吃青草”,“老牛吃白菜”。结果表明,被试对前一类句子的再认成绩优于后一类句子的成绩。

结 论

(1)在实验1中,我们的结果表明,在自由联想任务中,从语义记忆中提取信息与语义距离之间有某种程度的相应性。两词间语义上愈接近,联想关系就愈密切,则被提取的次数也就越多。语义距离作为自由联想作业的预测器,同样也适用于汉语材料的作业。

(2)在实验2中,我们的结果表明,汉字字形的特点,不是影响提取时间的重要因素。而事实与概念之间联结的强度,却是影响信息提取时间的有效因素。

(3)以汉语作实验材料,对Collins等人提出的语义记忆的层次网络模型进行了验证。实验结果表明,不同类型的句子和网络不同水平上的句子,其提取时间是存在着差异的。

附 录

1. 严厉的哥哥打顽皮的小弟弟。
2. 田里的庄稼是辛勤劳动的农民种出来的。
3. 今天的比赛我保证打出水平来。
4. 小孩慢慢吃冷的面包。
5. 列宁常常把书里的故事讲给别人听。
6. 花圈献给为人民英勇牺牲的烈士。
7. 狡猾的狐狸借着老虎的威风吓跑了百兽。
8. 我们看见一群小鸡正在食槽里吃米。
9. 梅生派立安把两封紧要的信送到鹦鹉洲去。
10. 马良到河边用草根在地上画水中的鱼。
11. 小金花等来的是她妈妈拉响了手榴弹跟敌人同归于尽的噩耗。
12. 公园里椅子上坐着一位穿蓝衣服的老头正在读报纸。
13. 李明送一束美丽的玫瑰花给北京市第一中心小学的校长。

* 私人通信,宋均,儿童对短句的识读。

参 考 文 献

- (1) Collins, A. M. & Quillian, M. R. Experiments on Semantic Memory and Language Comprehension. In L. W. Gregg (Ed.) *Cognition in Learning and Memory*. New York, Wiley, 1972.
- (2) Caramazza, A., Hersh, H. & Torgerson, W. S. Subjective Structures and Operations in Semantic Memory. *Journal of Verbal Learning and Verbal Behavior*, 1976, 15, 103—117.
- (3) Anderson, J. R. (Ed.) *Cognitive Psychology and Its Implications*. W. H. Freeman and Company, San Francisco, 1980, 113—116.
- (4) Yu Bo-lin, Jing Chi-cheng, and Sima He. STM Capacity for Chinese Words and Phrases with Simultaneous Presentation. In *Issues in Cognition Proceedings of a Joint Conference in Psychology*, 1984.

PRELIMINARY STUDY OF SEMANTIC RETRIEVAL
FOR CHINESE LANGUAGE

Chen Yongming and Peng Ruixiang
(*Institute of Psychology, Academia Sinica*)

Abstract

Two experiments were conducted to investigate the semantic retrieval for Chinese language. In experiment 1, twenty-three two-character words selected from thirteen sentences were used as stimuli in a free association test. The results indicated that semantic distance in terms of common-sense relations between words is useful to predict the performance of the free association test. In experiment 2, the true/false judgement reaction time technique was used. Thirty-eight (20 true and 18 false) sentences were used as stimuli in a true/false judgement test. The findings coincided with the results obtained by Collins and Quillian. The difference of retrieval time between levels in which the sentences were stored was verified. The results indicated that in the stage of information retrieval, the morphological effect of Chinese characters is not an important factor.