



手写汉字识别中的语境作用浅析

周昌乐

【内容提要】 研究语言信息处理中语境的作用,是目前语言学的一个重要课题,已越来越引起语言学界的重视。本文针对手写汉字识别问题,较为详细地分析了不同尺度上语境所起的作用。主要包括作为字形结构整体的字境作用,作为语言上下文的词境和句境作用;并强调了不同层次语境跨越尺度的相互作用。一般而言,本文中所给出的主要观点,也是适用于其它语言现象中类似语境作用分析的。

一、引言

日本学者西椋光正在《语境研究论文集》的前言中指出:“人机对话是人类美好的科学未来,可语境却成为了这一尖端信息工程中的一个老大难,目前还没有一个科学且系统的语境理论,因而还做不出足以应付千变万化的语境来实现人与机器之间的应用自如的对话程序”^①。实际上,不仅对于人机对话是这样,即使对书面汉语的单字识别,也同样如此。也即,语境在手写汉字识别过程中起着相当重要的作用。不同的是,手写汉字中的语境,不但指一般的语言上下文,而且也指字形整体结构中的上下文(我们称为字境)。这样对于手写汉字的识别而言,根据作用范围的不同尺度,语境可以分为字境、词境和句境三个不同层次;而对于手写汉字的正确识别,也就成为一种大语境套小语境,多层次跨越尺度相互作用的结果。

二、整体字境作用

有关研究业已表明,汉字虽然可以看作是笔画、部件构成的,但当一组笔画以一定关系构成汉字后,随着作为一个整体反复认读的经验逐渐积累,在识别时会表现出一种整体关联性。

比如罗春荣的阅读材料实验^②证实了在汉字识别中存在着“整体优先”原则:汉字识别是一个从整体到局部的过程。郑昭明^③和陈焯之^④的两项研究也为汉字的整体性作用提供了直接证明:一方面,自上而下信息对个别部件识别具有促进作用;另一方面,结构整体性产生的一种“抗分解”倾向,对单个部件识别起抑制作用。这一结论也为喻柏林等人的实验^⑤所证实。另外,虽然尚无一致的结论,但字形结构方式(也称为组合图式)对单字识别结果产生作用的可能还是存在的^{⑥-⑩}。

实际上,很早的研究就发现^⑪,学生在认字过程中,对字形并不作专门分析,而是把多个因素作为一个整体予以感知。例如,对视觉呈现的一个熟字,学生一般不作细致的横竖撇捺笔画和上下左右结构的分析,而是在瞬间便能把握这一字形。对此问题的实验

表明：从整体辨认过程来说，学生首先辨认出字形的大致轮廓，然后才是各组成部分及其关系，而字的细节部分常常被忽略。比如对于书写潦草的“楼”字，即使右下方的“女”部写得很模糊，由于字境中整体结构关系的作用。同样能正确识别其为“楼”字。

汉字识别研究中字境作用的另一个结论是，在认字过程中，部件重要性程度的空间分布是对称破缺的。一般说来，一个字的不同笔画或部位所提供的信息量的多少与重要性程度是有区别的。据早期周先庚^⑧的研究发现，字的上半部对识别更为重要。赵振飞^⑨进一步指出，字的上下端之间的信息量分布差异比左右之间更是悬殊。彭瑞祥、喻柏林^⑩把一个字划分为四个相等面积的部位，发现对合体字来说，左上角包含的信息量最多，右下角最少；而对于独体字则无差异。刘英茂^⑪也证实了字的各部位的这种线索能力极大地依赖于字形结构形式，部位的重要性与笔顺次序相一致。也即字形辨认顺序通常是，先整体后局部，先轮廓后内含，先上部后下部，先左部后右部，先熟悉后生疏。另有研究发现^⑫，被试对汉字的识别首先获得清晰印象的是字形的边角突出笔画（头尾，偏旁，外框等）。在阅读时，人们并不是一开始就把每个汉字的字形都看得很清晰，而是根据获得的整体轮廓的模糊印象去迅速而有选择地加以认识的。

综上所述，就形状知觉而言，人类识字是经由整体属性及分层结构的选择等字境，以多种途径综合而成的^⑬。汉字结构的整体性和有选择的可分解性是字境作用法则中两个相互关联的方面。很明显，对于一个复杂整体的识别，必须通过对这一整体的各组成部分进行分析把握之后才能完成；然而，如果在分析时没有一个整体字境作为指导，对这个复杂整体的识别就连一步也不能深入下去。

比如，对于手写的“地”字，要将该字当作为“地”识别出来，就必须把构成该字的“土”和“也”识别出来；而要把“土”识别为提土旁的“土”而不是“士”，反过来又离不开“地”这个字的整体字境。就像词句的理解依赖于组成词句的字义理解，而字义的确反过来同样也依赖于词句上下文语境一样；汉字识别是离不开作为整体结构信息的字境作用的。正如格式塔心理学家所强调的，形状知觉产生于该形状部分之间的关系中，而部分特性就人们所确定的内容来说，依赖于它们所处整体的全部关系，部分只在整体之中起作用，离开了整体的部分是没有意义的。可见，那些反映整体字形结构和全部组元之间关系的字境在汉字识别中至关重要。

三、词句语境的作用

除了字境所起的作用外，在手写汉字的识别中，语言上下文中的词句语境也起着相当重要的作用^⑭。特别是对于形似汉字的区分和疑难汉字的确认，更是起着决定性的作用。这一点同确定任何语言片断的结构和意义时所起的作用是相同的^{⑮⑯}。

通常，单字识别是在一定的词句语境中进行的。识别过程往往是自上而下和自下而上两种加工共同作用的结果。这不仅表现在单字的整体结构中，也同样表现在词句语境中。不同的是此时单字只是作为词句语境的组成部分而出现。

有许多研究^{⑰⑱}证实，双字词境、句子语境和课文语境等各种尺度类型的语境都对单字识别有显著的影响。当然就影响强度而言，不同类型的作用是不同的^⑲。另外，朱晓平的研究^⑲结果也表明，受到语境自身条件和单字特点的影响，语境效应在字形特征分析，

语词到达及后期整合的各阶段都有可能发生。虽然,语境作用范围和利用语境信息的策略是依赖于条件的,但不同尺度的语境单位都对单字识别起着影响作用是肯定的。

双加工理论认为^②,语境对单字识别的影响是通过两种不同的加工方式进行的,一种是自动的激活扩散加工,另一种是有意的注意加工。当目标刺激正好是预期要出现的刺激时,语境会起促进作用;而当目标刺激与所预期的刺激不一致时,语境则会抑制目标的认同。这也为 Neely 所作的工作所肯定^③。例如在识别形似汉字“士”或“土”时,如果有上下文的语境“战士”或“土地”,那么因为“战士”和“土地”的构词预期可能性很小,于是当在“土”和“士”之间选择时,这样的词境上下文就成为一种关键的指导信息。更一般地,如果遇到“X州”这样的搭配时,对疑难字“X”的识别必然可以限在州名中寻找,并从这些构成州名中的字集里产生一个指导信息去指导单字的识别。不过,当考虑到字词搭配的多重性问题,比如像“土兵”和“士兵”,前者可以指少数民族的兵士,后者则用于一般正规军的兵士称谓。那么此时词境信息不足以起到有效的作用,因此就必须进一步借助于句境上下文信息,甚至文本上下文信息。同样的道理,句境甚至文本语境,也将对单字的识别产生影响作用。

归纳起来,词句语境可以对汉字识别过程产生重要的影响作用,特别是对形似汉字的区分,疑难汉字的确认以及加快一般汉字的识别速度等,尤为重要,这也已为我们手写汉字机器识别的实验所证实。^{④~⑥}

四、结语

通过上述分析,字境、词境、句境等不同层次的语境信息都在不同程度上对手写汉字的识别起到重要作用。特别是按照双加工观点,对于每一层次的语境作用,都可以看作是自上而下和自下而上二个方面相互影响的结果;或者按乔姆斯基的后期语言理论^⑦,看作是支配和约束的结果。于是如果采用吴长安自足语境的说法(理解一个语言单位所涉及该语言单位上下文的范围,称为该语言单位的自足语境^⑧),那么不同层次的语境反映的是不同尺度的影响作用。这样,汉字识别过程中的语境作用就可以看作是一种多层次跨越尺度的多重语境作用,是一种大语境套小语境的自相似性体系。这正是我们所要阐述的观点。

注:

- ①西棋光正:《语境研究论文集》,北京语言学院出版社,1992,11,北京。
- ②罗春荣、纪桂萍、房路江:《从汉字识别看视觉信息检测的整体优先性》,《心理学报》,第二期,第184—189页。
- ③高尚仁、郑昭明:《中国语文的心理学研究》,文鹤出版有限公司,1982。
- ④Henry S. R. Kao & Rumjahn Hoosain, *Linguistics, Psychology, and the Chinese Language*, Centre of Asian Studies University of Hong Kong, 1986。
- ⑤喻柏林、冯玲等:《汉字的视觉—知觉任务效应和汉字属性效应》,《心理学报》,1990. 2,第141—148页。
- ⑥郭德俊、彭聃龄、张素兰:《汉字字形特征及使用频率对汉字识别的影响》,《普通心理学与实验心理学研究进展》,成都科学出版社。
- ⑦彭聃龄:《阅读的认知心理学研究》,《北京师范大学学报》,1989. 5。

- ⑧彭瑞祥、喻柏林：《不同结构的汉字再认的研究》，《中国心理学会普通和实验心理专业委员会》，1980年学术年会论文。
- ⑨彭瑞祥、张武田：《速示条件下再认汉字的某些特征》，《心理学报》，1984. 16，第49—53页。
- ⑩喻柏林、曹河圻等：《汉字形码和音码的整体对部件识别的影响》，《心理学报》，1990. 3，第232—239页。
- ⑪Just, M. A. & Carpenter, P. A., *The Psychology and Language Comprehension*, Allyn and Bacon, Inc. 1987.
- ⑫曹传泳、沈晔：《小学儿童分析概括和辨认汉字字形能力的发展研究》，《心理学报》，1965年第1、2期。
- ⑬Chou, S. K. (周先庚), *Reading and legibility of Experimental Psychology*, 1930. 13, pp. 332—351.
- ⑭赵振飞：《汉字的不同部位对再认的影响》，《华东师范大学物理系学士论文》，1986年。
- ⑮Liu, In-mao (刘英茂), *Recognizing of fragment-deleted characters and words*, *Computer Processing of Chinese & Oriental Languages*, 1984, pp. 276—287.
- ⑯杨继本：《汉字心理在汉字信息输入编码上的应用》，《心理科学通讯》，1987年第3期。
- ⑰路浩如、杨源远：《手写体汉字识别问题综论》，《计算机应用与软件》，1994年第2期。
- ⑱谭力海：《现代心理学关于单词再认的研究—影响单词再认的因素》，《心理学探新》，1988年第2期。
- ⑲张志公：《现代汉语》（第四章“语义与语言环境”），人民教育出版社，1982. 1，北京。
- ⑳张普：《汉字信息处理与语境研究》，西模光正编《语境研究论文集》，北京语言学院出版社，1992年，北京，第516—540页。
- ㉑朱晓平：《课文语境条件下的单词识别》，中国心理学会工业心理学会专业委员会1990年学术年会论文。
- ㉒McClell, J. and Rumelhart, D. An interactive—activation model of context effects in letter perception, Part I, An account of basic findings, *Psychological Review*, 1981. 88, pp. 375—407.
- ㉓郑绍明和 Shui-I Shih, The nature of lexical access in Chinese, Evidence from Experimental on Visual and phonological priming in lexical judgment. In-mao Liu et al. (Eds.), *Cognitive Aspects of Chinese Language*. Vol. 1. Asian Research Service, 1987.
- ㉔谭力海、彭聃龄：《中文语词识别过程：对语境和词频的作用机制的探讨》，《心理科学通讯》，1989年第2期。
- ㉕谭力海、彭聃龄：《快速呈现条件下语境和词频对中文语词识别的影响》，《心理科学通讯》，1989年第2期，第1—6页。
- ㉖朱晓平：《汉语名字语境对单词识别的效应》，《心理学报》，1991年第1期。
- ㉗朱晓平：《语境条件下的单词识别》，《华东师范大学心理系博士学位论文》，1989。
- ㉘Posner, M. T. & Snyder. C. R. R. Facilitation and inhibition in the processing of signals, In Rabbitt, P. M. A. & Dornic. S. (Eds.) *Attention and Performance V*, New York Academic Press, 1975 (b).
- ㉙Neely, J. H., Semantic Priming and retrieval from lexical memory: Theories of inhibitionless spreading activation and limited-capacity attention, *Journal of Experiment Psychology: General*. 1997, pp. 106, 226—254.
- ㉚苗兰芳、张森、周昌乐：《基于N联字的汉字识别后处理研究》，《中文信息学报》，第8卷第2期，1994年。
- ㉛金雄伟、苗兰芳、周昌乐：《一种汉字识别疑难语段的确认方法及其实现》，《中国青年计算机研究新进展》，西北工业大学出版社，1994年9月。
- ㉜张雄伟、周昌乐：《一种形似汉字区分识别原理》，《模式识别与人工智能》，第8卷第4期，1995年。
- ㉝诺姆·乔姆斯基：《支配和约束论集》，中国社会科学出版社，1993年，北京。
- ㉞吴长安：《语境的范围及其在语法分析中的作用》，《语言学通讯》，1989年第3—4期。

(责任编辑 陈晓菲)