

论汉字部件与英文字母的认知对应关系

韩布新 林仲贤

(中国科学院心理研究所)

摘要 首先讨论了汉字部件和英文字母在文字学和形体特征上的六方面共同点及其在语言认知方面和文字发展史上的相似作用,其次指出英文字母与汉字笔画的非对应关系;最后通过分析两者的一些不同点,说明尽管从形体结构和语言认知心理学角度来看,汉字部件与英文字母存在着一种认知对应关系,但并不能将两者硬性地等同类比。

关键词: 部件 字母 认知对应

1 前言

汉语和英语的书面文字均可划分出一定的结构层次,从知觉过程中的整体和局部加工之关系出发,比较文字结构层次在认知意义上的相对应关系,可以为中、英文各自研究(比如文字信息的阅读、自动化处理等)领域内有关研究结论的类推提供逻辑上的依据。

分析部件作用,必将成为汉字(尤其是手写体)机器识别研究的突破口^[1]。因为在该领域的研究一直以 Marr 所倡导的视觉计算理论为指导思想,目前遇到不少困难^[2]。人类对于视觉图形的识别,到底是采用自下而上的分析计算—表征处理方法,还是采用自上而下的整体表征方式,即字形的整体和局部信息在识别中的相对作用,是争议的关键所在。部件作为字形结构的中间层次,在整体和局部之间处于承上启下的重要位置;同时它作为一种可能为许多汉字所共有的结构元素,是联结个别汉字与整体汉字集合的桥梁,在无字库汉字系统的研制中也具有重要的意义。故此,本文从以下几个方面讨论了汉、英文字结构体系的基本单位——汉字部件与英文字母在语言认知中的关系。

2 汉字部件与英文字母的认知对应

2.1 在两种文字结构系统中有一些相似特征

汉语和英语书写文字的形成历史很不相同,但在结构层次上有许多相似之处,如表 1 所示。表 1 中“笔画”是指下笔后按一定方向连续写成的笔迹,以提笔处为终点。比如汉字中的“丶、一、丿”和英语中的“c、e”。“部件”是汉字中反复出现的、并能分割出来的有固定形体的笔画组合块,比如“讠、匚”等。“笔画(或部件)组合”是笔画(或部件)以相离、相接或相交等连接方式形成组合关系,如“八、口、十(相、丕、未)”等^[2]。

表 1 中英文书写文字的结构层次

汉字	英文
汉字或词(character or word)	单词(word)
部件组合(combination of components)	字母串(letter cluster)
部件(component)	字母(letter)
笔画组合(combination of strokes)	笔画组合(combination of strokes)
笔画(stroke)	笔画(stroke)

根据表 1 所列情况,不难看出在文字学上部件和字母有以下六个方面的共同点。

(1)均为小于整字(单词)、大于笔画的一个结构层次。

(2)是单字的最小不可切分单位,但均可以一定数目形成具有独立意义的语言单位或可构成其它字的、有一定出现次数的组块。在字母串与部件组合之间有一种对应关系,因为它们有四个共同点:①在字符集合中的存在受约定俗成的规则制约;②可以有一定的意义,比如英文单词中的“UN”前缀表示否定意义,而汉字中含有“氵”部件的多表示与水有关的事物;③与字的界限不明确——可以是成字,也可以是成字的一部分,如“相”和“an,he”等;④在成字中可处于不同的位置。

但是部件组合和字母组合同时又有不同点。字母串是一维线性排列,字母间相对位置只有前、中、后的离散型连接关系;而部件组合则是二维的,部件间有相离和相接甚至相交型相接关系,其位置关系可以是上下或左右型,也可以是混合型的。

(3)部件或字母总数远远小于单字(词)的数目。英语有26个字母,而《新英汉词典》(1985年2月增补本)收词5万多;《Webster's Third New International Dictionary》^[3]收词则达45万以上,英文一般交流用词也有一万左右。汉字部件数目随汉字集合大小而变化,但部件总数约是汉字总数的1/20—1/10。汉字总数在5万以上^[4],现代汉语书面用字总数不到一万。《信息交换用汉字编码字符集——基本集》(GB2312-80)仅收入了6763个汉字。这些汉字包含了567个部件^[1]。由于汉字中绝大多数为形声字,还有一部分会意字,所以部件在音、义方面的作用要比字母突出。

(4)按约定俗成的发音、词义和连缀等规则组成整字。

(5)2—3个部件的汉字最多;4—5个字母的英文单词最多^[5]。

(6)日文假名、朝鲜文、越南文等的字母均由汉字部件形变而来,说明汉字部件可蜕变为拼音文字的字母。这是汉字部件和英文字母相对应的例证。如表2所示^[6]。

表2 亚洲几种语言文字与汉字的渊源关系

语言	字母名称	创建年代	字母举例及发音	构成字(词)的方法
日文	平、片假名各73个	5世纪后	あ[a:]、カ[ka]	“安”的草体、“加”的左偏旁;拼写时从左至右书写成字
朝鲜文	颜文字母40个	1444年创制, 1954年重订	ハ[s]、ㅍ[Φ]	与“人”字、“丩”部件形近; 拼写时由字母组成方块字
古越南文	字喃	10世纪后	𠄎 [Mát/tròì]	借用汉字造字方法构成方块字

2.2 在文字阅读理解时存在认知对应关系

认知心理学关于字词识别研究的结果表明英文字母与汉字部件有相似的认知作用——局部特征对整体知觉存在一定的影响,这种影响同时受到整体特征的制约。Gough等曾在词汇决策和再认任务中使用含有不同数目字母的单词为刺激,发现在一定范围内,增加一个字母则平均要增加17.5ms时间来读出它。因此字母是词的知觉单位,词的认知要以字母的认知为中介。Adams提出书面形式词的认知加工包括字母和词两种水平的分析:其一是字母图式之间的相互激活;其二是竞争的词汇图式需要寻找其内部的字母信息以使它的组成部分具体化^[7]。Solso等研究发现字母和字母串的频度是单词识别的单调递增函数。Gibson在研究字母视觉特征时,曾提出分析字母特征的四项标准:①用与区别比较的关键特征只出现在一些字母中,另一些字母则没有;②对特征的辨认不受明度、大小、透视条件变化的影响;③这些特征使得每个字母都具有独一无二的形体外观;④特征数量是有限的,相对较少^[8]。这些标准可类推于分析部件。

关于汉字识别的研究发现,部件频率和部件组合频率作为一种局部属性,对于整字识别有干扰和易化两类趋势不同的作用。这种作用同刺激字的呈现方式有关。整字呈现时,位于左边或上部的局部熟悉性有助于整字识别,表现出对于整字识别的易化作用;位于右边或下部的局部熟悉性对于

整字的知觉加工表现出干扰作用。部件频率和部件组合频率的作用受整字属性,特别是整字频率的影响^[9]。

2.3 在文字发展上,字母和部件的作用相似

各种古文字的前身是图画文字,原始的象形文字都来自这种复制客观事物的图画文字。在拉丁语系中,象形字发展成了一个字母;而在汉藏语系中,一部分象形字则形成了独体字。以描摹词所反映事物为基础,以图象的各种灵活运用为手段构成了“表意文字”体系^[10];GB2312-80的6763个汉字中共有独体字318个,构成了567个部件的主要部分^[5]。

2.4 英文字母与汉字笔画不对应

字母和笔划的视觉特征复杂性不同,在对两者进行加工时,知觉深度不应处在一个水平上。笔画书写时只有一个起点和落点,一笔到底;而字母大多数有多个起点、落点(O、C、L、S、V、W、Z等除外),大写尤其如此,由点、直线、曲线相交、相连或组合而成,这与部件的组成特征更相近。再者,笔画数与汉字总数的比率要比字母数与英文单词数的比率小。

3 不能等同类比

部件和字母虽然存在着认知对应关系,却并非完全等同,因为它们在以下几方面不同:①字母组成单词是一维线性排列,而部件组成汉字则排列在二维方块空间;②视觉复杂性不同(部件可拆分成数个笔画;字母则线条简单,数目有限);③部件与字(或笔画)的分界不清,比如成字部件;④字母都有固定读音;少数部件不能音读;⑤字母有大小写之分。

4 结语

总之,汉字部件和英文字母在字词识别中的作用和地位有许多共同点,有认知对应关系。将两者作这样的匹配比其它匹配更合理些。数百个部件或26个字母作为基本结构单元,构成了数以万计的汉字或单词。在现代语言文字信息处理中,两种语言出现了差距——汉字信息的计算机处理已经实现,但在输入、识别等方面还存在相当的障碍。这种差距原因何在?要解决这个问题,是于着眼机器系统原理的改进(比如搞中文计算机),还是在比较人类阅读、学习和利用各种不同语言系统的规律中寻找答案?诸如此类的问题,还需要进一步探讨。

5 参考文献

- 1 韩布新. 汉字部件信息数据库的建立. 心理学报, 1994; 26(2): 147-152
- 2 张忻中. 汉字识别技术. 北京: 清华大学出版社; 南宁: 广西科技出版社, 1992: 7, 16-25
- 3 Gove PB. Webster's Third New International Dictionary. Cambridge, Mass: The Riverside Press, 1961: The introduction
- 4 陈一凡, 胡宣华. 汉字键盘输入技术与理论基础. 北京: 清华大学出版社, 1994: 2
- 5 韩布新. 汉字属性信息数据库在汉字识别研究中的应用. 心理学动态. 1993; 1(4): 29-35
- 6 季羨林主编. 中国大百科全书·语言文字卷. 北京: 中国大百科全书出版社, 1988: 33, 329, 499
- 7 张必隐. 阅读心理学. 北京: 北京师范大学出版社, 1991: 97, 100, 261
- 8 Han Buxin. Frequency effect of constituent in Chinese character recognition. In: Jing QC, Zhang HC, Peng DL eds. Information Processing of Chinese Language. Beijing: Beijing Normal University Publishing Co, 1994: 87-98
- 9 王凤阳. 汉字学. 长春: 吉林文史出版社, 1989: 70-75, 215, 268

(1995-04-27 收稿; 1995-07-11 修回)

ENGLISH ABSTRACTS

SYNTACTIC PROCESSES IN CHINESE READING; SOME PRELIMINARY STUDIES

Hsuan-chih Chen, Wai Wah Hung
(Department of Psychology, The Chinese University of Hong Kong)

Syntactic processes in reading Chinese texts were investigated in two experiments in which the subjects read short stories for comprehension. By means of a mouse-controlled sequential display technique, the reading time of each individual character was measured as an on-line indicator of the comprehension processes. The basic manipulation in Experiment 1 was to change a critical word in a story, so that various combinations of lexical, syntactic, and semantic information were violated. Various types of linguistic information associated with the critical word were distinctively processed on-line during reading for comprehension. In Experiment 2, the possible influences of structural and contextual factors in pronoun assignment were examined. A target pronoun was preferentially assigned to the first possible antecedent, regardless of the contextual constraints provided. The results are discussed in terms of syntactic processing at and above lexical level during the comprehension of Chinese texts.

Key words: reading processes, reading Chinese, syntactic analysis.

COMMENTS ON R. J. STERNBERG'S COMPONENTIAL THEORY OF INTELLIGENCE

Li Qiwei, Jin Yu
(Department of Psychology, East China Normal University)

The component is the basic unit which R. J. Sternberg put forward for the analysis of intelligent operations. The componential analysis describes the mental mechanisms responsible for intelligent behavior. The componential theory of intelligence is a part of R. J.

Sternberg's triarchic theory of human intelligence, as it alone cannot account for intelligence. The paper not only introduces features, categories and interactions of components, but also shows individual and aging differences in the functions of different components. In short, the componential theory is a valuable exploration in the working principles of the black box of the brain in a software sense. However, it is limited because of the difficulty of falsification.

Key words: intelligence, triarchic theory of intelligence, components of intelligence, mental mechanism, falsification.

ON THE COGNITIVE CORRESPONDENCE BETWEEN CHINESE CHARACTER COMPONENTS AND ENGLISH LETTERS

Han Buxin, Lin Zhongxian
(Institute of Psychology, Academia Sinica)

The cognitive correspondence between Chinese character components and English letters is discussed as follows: First, they share 6 orthographical similarities; Second, they have similar effect in the cognitive process of language information; Third, they have similar position in the development of the written language system; Fourth, Chinese character strokes do not correspond with English letters. Furthermore there are still some different aspects between components and letters.

Key words: component, letter, cognitive correspondence.

THE NATURE OF RISK--AN ASSOCIATION TEST

Xie Xiaofei
(Psychology Department, Beijing University)

Xu Lianchang
(Institute of Psychology, Academia Sinica)

In order to probe into the nature of risk in the mind of people, we work out an association test. With the word "risk" as the stimulating word, the subjects go through three free associations. Many response words as well as some valuable results are also obtained.

Key words: risk, association test.