

汉字识别中部件组合的频率效应^{1)*}

韩布新

(中国科学院心理研究所, 北京 100012)

摘 要

用整字识别、整合识别等任务研究部件组合频率在汉字识别中的作用。发现部件组合频率可促进高频两部件合体字的识别(实验一);在先后呈现两组成部件时这种作用仍然存在(实验二);但在识别高频三部件左右型合体字时(实验三),位于左边的部件组合表现出易化作用(田型字);位于右边的部件组合表现出干扰作用(田型字)。部件组合有与部件类似的频率效应。

关键词 汉字识别, 部件组合, 整合识别, 频率效应。

1 前言

汉字结构一般分为笔画、部件、整字三个层次。笔画影响汉字识别反应时^[1]、记忆广度^[2],即所谓的“笔画数效应”^[3,4];部件对整字识别有易化(位于字的左边或上边时)或干扰(位于字的右边或下边时)作用^[5];而识别部件的反应时与整字结构类型有关,识别上下字中的部件较慢^[6]。实际上字形结构还有一潜在层次——部件组合^[7],即一个汉字中同时存在的两个部件,比如“部”字中的“立一口、口一阝”。部件组合是介于部件和整字的一个中间层次。不同部件组合组成汉字的个数不同。这种次数的差别,对整字识别有无影响?部件组合同部件的作用是否相似?本文采用整字识别和整合识别等实验范式对上述问题进行了研究,比较部件组合作为整体属性和局部属性对整字识别的不同影响。

2 方法

2.1 被试 北京某高校三年级本科生 64 名,年龄 19—24 岁。其中实验一 18(男 9、女 9)人;实验二 16(男 12、女 4)人;实验三 30(男 20、女 10)人。

2.2 仪器 HH AT286 微机、640×350 分辨率彩显和声控开关各一台,话筒一支。

2.3 程序 被试坐位,眼睛平视距显示器 60 厘米(cm)完成整字识别或整合识别任务。1.8cm×1.2cm 大小的白色汉字呈现于黑色屏幕背景中央。识别反应时由计算机自动记录,反应正确与否由主试判别并经键盘输入计算机。从刺激呈现到被试读出字音为识别反

1) 本文初稿于 1995 年 3 月 22 日收到,修改稿于 1996 年 2 月 10 日收到。

* 部分内容曾在首届亚洲语言信息认知加工研讨会(1994 年 12 月,悉尼)作英文专题报告。

应时。刺激呈现时间为 25 毫秒(ms)。

3 实验一:识别两部件合体字

3.1 材料 72 个两部件合体字,高频、低频各半,上下、左右结构各半;笔画数均在 10 画以下。组合频率均值分 3.33、1.03、0.26、0.10%四个等级^[8]。部件频率相近,均可命名。

3.2 结果与讨论

整字识别反应时及错误率列于表 1。重复测量多因素方差分析(MANOVA)表明整字频率主效应显著,识别高、低频字的反应时分别为 555 和 602ms, $F(1,17)=43.91$, $P<0.001$; 错误率分别为 3.6 和 12.2%, $F(1,17)=44.69$, $P<0.001$ 。两项指标一致验证了字频效应。

部件组合频率的主效应显著, $F(3,68)=5.39$, $P<0.01$;在其由低到高的四个水平下,识别反应时分别为 609、586、548 和 571ms,除 609 与 586 外,其它三组差异显著。部件组合频率的这种作用反映了两个组成部件的紧密程度,同时表现出同字频类似的整体特征。组合频率高,一方面意味含有此类部件组合的字较常见,易于识别,故表 1 中反应时随组合频率增加而减少;另一方面在识别目标字时需要搜索的范围比较大,此时作为一组字共同特征的组合便成了尽快识别目标的阻碍因素,使识别变慢,故表 1 中反应时在部件组合频率最高时有所回升。所以说,识别反应时是两种过程综合作用的结果。

表 1 整字识别反应时(RT±SD,单位 ms)和错误率(单位%,括号中)

部件 组合(%) 频率	低频字		高频字	
	上下字	左右字	上下字	左右字
0.10	603±67(20)	633±72(7)	604±52(9)	596±94(4)
0.26	609±47(13)	604±76(18)	596±88(2)	537±27(4)
1.03	577±35(2)	593±53(0)	512±47(2)	510±26(4)
3.33	626±72(22)	572±61(16)	541±51(4)	544±32(0)

整字频率×部件组合频率有交互作用, $F(3,51)=4.70$, $P<0.01$ 。识别高频字时,部件组合频率四水平下反应时分别为 600、566、511 和 542ms,除 542 和 566 外,其它各组差异显著;识别低频字时各水平的反应时无显著差异。因为高频字的组合频率包含了部分字频信息(两部件合体字尤其如此),两者协同作用对反应时影响较大;而低频字的组合频率不一定低。

部件组合频率与整字频率的作用相关,因为二者都涉及对特定目标的熟悉性。若部件组合组字次数为一,组合频率实际上就是该字的整字频率。比如“血”和“半”形成的组合只存在于“衅”字中。若其组字次数大于一,则对该组合的熟悉性包括因整字频率产生的熟悉性和因部件组合本身的组字频率而产生的熟悉性,比如“日”和“一”形成的组合,除了组成“旦”字以外,还含于“查、量、坦”等 40 余个字中。

4 实验二:整合识别两部件合体字

部件组合频率对识别高频两部件合体字有易化作用。假若破坏其完整性,比如拆开部件组合,这种易化作用将有何变化?

4.1 程序 刺激字同实验一。两部件间隔 120ms 呈现,分别持续 25ms。第一个部件呈现至被试读出目标字为识别反应时。一半字先呈现第一部件,另一半先呈现第二部件。两部件按其原字中的相对位置呈现,在屏幕上不重叠,合成完整的字形。

4.2 结果与讨论

两部件合体字整合识别反应时和错误率列于表 2。重复测量 MANOVA 表明整字频率的主效应显著,整合识别高、低频字的反应时分别为 572 和 673ms, $F(1, 15) = 13.61, P < 0.01$ 。结构类型主效应显著,上下字和左右字的整合识别反应时分别为 670 和 574ms, $F(1, 15) = 50.32, P < 0.001$; 错误率分别为 39.4 和 29.3%, $F(1, 15) = 16.16, P < 0.001$ 。说明将上下呈现的两个部件整合成一个完整的汉字比较难。因为上下字结构紧密^[6],将其两部件拆开先后呈现,对其结构完整性的破坏影响较大,因此识别反应时长。

表 2 整合识别反应时(RT±SD,单位 ms)和错误率(单位%,括号中)

部件组合(%) 频率	低频字		高频字	
	上下字	左右字	上下字	左右字
0.10	909±284(43)	627±222(28)	627±140(38)	640±112(29)
0.26	664±96(33)	601±136(38)	608±100(43)	512±160(21)
1.03	663±200(43)	566±177(28)	537±140(40)	508±168(42)
3.33	753±282(40)	598±260(38)	604±180(35)	540±194(13)

部件组合频率由低到高的四水平下,整合识别反应时分别为 701、576、569 和 624ms,其中 701ms 与其它三组的反应时有差别, $F(3, 45) = 5.46, P < 0.01$ 。说明在两组合部件先后呈现时,部件组合频率所表现的组合部件之间的连接紧密程度确实在识别整字时起了重要作用,但对这种紧密性的一定程度的破坏影响了组合频率的作用。两部件先后呈现破坏了部件组合的完整性,但部件组合频率、整字频率、结构类型三个属性对命名反应时仍然有显著作用,说明此时自上而下的有关整体属性知识仍然起着很重要的作用。

部件组合频率在两部件合体字中具有整体特征是其作用的关键。那么,若作为一种局部属性(如在多部件合体字中),部件组合频率是否会有另一种作用呢?

5 实验三:识别多部件合体字

本实验用识别三部件合体字任务,探讨部件组合频率作为局部属性时的作用。

5.1 材料 90 个三部件左右型合体汉字。各部件均可音读。一次次切分出偏旁部首后剩下的两个相邻部件为目标部件组合。例见表 3。高频(1—3 级)、低频(4、5 级)^[9]各半;目标

表 3 实验三刺激字举例

部件组合(%) 频率	高频字部件组合方式			低频字部件组合方式		
	田	田	田	田	田	田
高	况	部	侧	拐	郡	靴
中	湿	创	谢	肮	刚	渺
低	挖	叔	辩	谍	颖	狱

组合频率分 3.33、0.58 和 0.08% 三水平^[8]; 部件有三种组合方式: 田型、田型和皿型。

5.2 结果与分析

三部件左右型合体字识别反应时与错误率列于表 4。重复测量 MANOVA 表明整字频率的主效应显著, 识别高、低频字反应时分别为 632 和 743ms, $F(1, 29) = 365.26, P < 0.0001$; 错误率分别为 4.3 和 18.4%, $F(1, 29) = 75.43, P < 0.0001$ 。部件组合频率的主效应显著, 低、中、高三水平下识别反应时分别为 683、703 和 676ms, $F(2, 58) = 8.77, P < 0.001$; 703 与 683、676 差异显著。部件组合作为局部特征, 低频时干扰识别, 高频时易化识别。

田型、田型和皿型三种部件组合方式下识别反应时分别为 661、705 和 696ms, $F(2, 58) = 38.40, P < 0.0001$; 错误率分别为 7.3、14.3 和 12.3%, $F(2, 58) = 18.10, P < 0.001$, 首次切分后单部件在笔顺序列前端(左边)的字识别快, 错误率低; 而部件组合在前端的则识别慢, 错误率高。

表 4 整字识别反应时(RT±SD, 单位 ms)和错误率(单位%, 括号中)

部件组合频率(%)	低频字部件组合方式			高频字部件组合方式		
	田	田	皿	田	田	皿
低	694±91(7)	770±105(36)	746±73(13)	574±57(2)	654±56(3)	662±75(9)
中	754±74(16)	794±90(21)	714±68(11)	629±75(2)	667±80(7)	660±68(3)
高	722±80(17)	757±86(17)	741±93(26)	596±53(0)	590±73(1)	653±81(12)

整字频率×部件组合频率有交互作用, $F(2, 58) = 3.64, P < 0.05$ 。识别高频字时, 部件组合频率低、中、高三水平下的识别反应时差异显著; 而识别低频字时没有显著差别。整字频率×部件组合方式的交互作用显著, $F(2, 58) = 15.80, P < 0.001$ 。识别高频字时, 田型、田型和皿型字的反应时分别为 600、637、658ms, 识别皿型字最慢; 错误率分别为 1.3、3.8、7.8%, 以皿型字为最高。而识别低频字时, 三类型反应时分别为 723、774 和 734ms, 以田型最慢; 错误率分别为 13.3、24.9 和 16.9%, 以田型字最高。部件组合频率×部件组合方式有交互作用, $F(4, 116) = 10.38, P < 0.001$ 。识别田型字时, 部件组合频率低、中、高三水平下整字识别反应时分别为 712、730 和 673ms, 后两者差异显著; 而识别田型字时, 部件组合频率低、中、高三水平下的识别反应时分别为 634、692 和 659ms, 前两者差异显著。即识别田型字和田型字时, 部件组合频率分别表现出易化和干扰作用, 中频的反应时最长。识别皿型字时, 部件组合频率无显著作用。

本实验中识别低频字时, 部件组合频率无显著作用, 同实验一类似; 而实验二识别高频字和低频字中, 部件组合频率的效果无差别, 说明整字频率的作用因分开呈现部件受到破坏。实验一、三的结果说明, 不论部件组合频率是整体还是局部属性, 其作用均受整字频率的影响。

6 结语

部件组合频率在识别高频字时有干扰和易化两方面的作用, 对识别低频字无显著影响。识别高频两部件合体字时, 部件组合作为整体结构, 其频率由低到高表现出易化作用,

在最高段表现出一定的干扰作用;识别高频三部件合体字时,作为局部结构,识别田型字时有易化作用,识别田型字有干扰作用。

致谢:本文据导师荆其诚教授和林仲贤教授指导下所完成博士学位论文的一部分修改而成。在此谨向他们致以衷心的感谢!感谢何海东同志帮助编写实验程序。

参 考 文 献

- 1 Just M A, & Carpenter, P A. The Psychology of Reading and Language Comprehension. Boston: Allyn and Bacon, Inc., 1987: 312.
- 2 张武田, 杨德庄. 汉字词笔画数对短时记忆容量的影响. 心理学报, 1987, 19(1): 79—85.
- 3 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, (2): 1—6.
- 4 朱晓平. 汉语句子语境对单词识别的影响. 心理学报, 1991, 23(2): 145—152.
- 5 Han B X. Frequency effect of constituent in Chinese character recognition. In: Jing QC, et al (Eds.), Information Processing of Chinese Language. Beijing Normal University Publishing Co., 1994: 87—98.
- 6 喻柏林等. 汉字形码和音码整体性对部件识别的影响. 心理学报, 1990, 22(3): 232—239.
- 7 韩布新. 部件组合——潜在的汉字结构. 中文信息学报, 1995, 9(3): 27—32.
- 8 韩布新. 汉字部件信息数据库的建立. 心理学报, 1994, 26(2): 147—152.
9. 上海交大计算机系等单位编, 汉字信息字典, 北京: 科学出版社, 1988, 10.

附表 1 实验一、二因素设计及刺激字

部件 组合(%) 频率	左右字		上下字	
	高频字	低频字	高频字	低频字
3.33	加如叶	叭叮扣吐泪	古另台穴只	呆旦香杲杲
1.03	计利放	柏肚斩付呐	呈皇孟苗虽	胃吊寺旨萎
0.26	池打伏	伐仿灿垃肋	否李旱岩香	帛杰男芍秃
0.10	祖找状	钐炊冯劫砍	泵导亩怎突	宝甬胃劣辛

附表 2 实验三因素设计及刺激字

部件 组合(%) 频率	高频字部件组合方式			低频字部件组合方式		
	田	田	田	田	田	田
高	涂略轻联况	敏劲胡数部	侧例倒雌树	峭舵袒晦拐	魂勋敛毓郡	靴訛咖挪咧
中	旋湿腔拖将	鼓创散剥教	渐概谢脚倾	禅呛肮捏蝎	雏鹁剔鹊刷	鸿坳卿绑渺
低	摆临挖伤播	凯叔款飘荆	柳傲掀辩癣	瑛谏淹聘吮	颧颧甄鸨剃	湃溺狄撇嗽

FREQUENCY EFFECT OF CHARACTER COMPONENTS COMBINATION IN THE RECOGNITION OF CHINESE CHARACTERS

Han Buxin

(Institute of Psychology, Academia Sinica, Beijing 100012)

Abstract

Three experiments with different paradigms were conducted to investigate the frequency effect of character components combination in the recognition of Chinese characters. The result of experiment 1 indicated that the combination frequency had significant facilitating effects upon the RT when the task was to name the whole character with 2 components. While in experiment 2, presenting the 2 components separately, the combination frequency also show significant facilitating effect. When naming the right-left type character combining with 3 character components in experiment 3, the combination frequency had facilitating effect when they were on the left position of the whole character (as in 田 type characters), or inhibitory effect when they were on the right position (as in 田 type characters).

Key words Chinese character recognition, components combination, integrative recognition, frequency effect.