

hjqb
汉

iuvc
语

vmpa
双

pkpb
拼

hjqb
汉

ziqb
字

bgpb
编

mavb
码

wnta
陈

jqb
建

unta
文

njta jzpc daqa xota tuta vupa gbvb
南 京 大 学 图 书 馆

1979 ng 6 io 1 ri

汉语双拼汉字编码初探

南京大学图书馆 陈建文

(一)

语言是人类思想交流的工具，文字是记录语言的符号。随着人类社会不断进步，科学技术飞速发展，对文字实现机械化处理的要求也愈加迫切。

在一些技术比较先进的国家中，文字的印刷、电讯和情报检索工程，已普遍地采用电子计算机。可是，在我国国家中，文字工程还基本上停留在落后的手工操作阶段。

给文字机械化添麻烦的根本原因，在于方块汉字本身。所以鲁迅先生曾尖锐地指出：“我认为方块汉字本身就是一个死症，吃点人参，或者想一点什么方法，固然也许可以拖延一下，然而到底是无可挽救的。”“汉字不灭，中国必亡。”毛主席也早就指出：“文字必须改革，要走世界文字共同的拼音方向。”

然而，文字改革也不是一件轻而易举的事情，要化相当的气力和时间。因此，方块汉字在走向拼音化之前，仍然要继续为我国社会主义建设服务。这样，研究如何使方块汉字比较方便地进入电子计算机，从而加速实现四个现代化的进程，就是摆在我们面前急待解决的问题。这是我们汉字编码工作者所面临的光荣任务。

(二)

当然，世界上也有人研究不用代码而直接输入汉字的方法，即大键盘整字输入。在这方面获得明显成效的要首推日本。我们国内也有少数同志在进行探索。其优点是直观性强，适合于照相排版。但缺点是设备笨重，效率低。无论是按键式，还是笔触式，都难于普及。因此，许多学者和科学家不得不寻找新的途径，在汉字编码上下苦功。不少人获得了一定的科研成果。

就目前来看，从输入方式上说有研究中键盘和小键盘两大学派。

中键盘（约在一百个键以上而有的多到几百个键）在使用上，无论是采用主辅键盘式，还是人机对话选择式，部首字根拼合式，一般要按两次键以上才能输入一个汉字。优点是设备较轻便，效率比大键盘稍有提高。但缺点是需要记忆的单元太多，全靠死记硬背，因此实用和推广仍然比较困难。

由于小键盘（约50~100个键）的设备简单，操作方便，所以赞成者越来越多。研究的方法虽然有很多很多，但大体上又可分为形码和音码两大学派。

形码的工作原理是拆字法。各人主张不尽相同。大体是从偏旁部首、三角四角、笔顺笔形、字根字元等等方向上进行研究，也有用拉丁字母或阿拉伯字母按各拆字法规则进行组合的办法。优点是设备简单，能见字识码。但缺点是缺乏可读性，不能见码识字。一旦字形简化或变更时，编码也随着变更，整个编码系统也会受到影响。尤其用在电报通讯方面，还有一定的困难。

音码是从文字的读音上入手进行编码。每个方块汉字的字音，是相对稳定的。它的语音平面，占整个信息量的四分之三。将它们弃而不用，未免太可惜了。因此，去年在青岛召开的全国汉字编码学术交流会上，音码方案占有优势的地位。由于音码能见字识码，又能见码识字，又不会受汉字简化的影响，能够符合文字改革的方向，因此其优点是明显的。

(三)

用音码比较好，但直接就用汉语拼音行不行呢？当然可以，但问题在于汉字的同音字太多，会出现重码。这在计算机的实现上和通讯工作上都是不允许的。因此，还需要改进。最简单的改进方法就是利用现行的汉语拼音方案，在每个拼音字的后面加上形符或义符，来区别同音字。但由于这样做不但多数音节的拼式过长，会给时间、人力和物力造成浪费，而且也同样会出现人造码与音节混淆的情况，抵消了一部分拼音的优点。是否还可以进行改造呢？可以。我国古代注音用的反切法就给了我们一种启示：每个字的读音，都由另外两个字的读音反切而成。用前面一个字的声，用后面一个字的韵。现在我们用两个拉丁字母来表示每个汉字的声和韵，这也就是最基本的一种汉语双拼方案。

双拼方案所用字母少，形式美观，拼式短。26个拉丁字母全部能够条件变读。其中多数字母声韵合用，前声后韵。所用的声韵母经过周密考虑，进行了合理安排。所有的音节都由两个字母组成，界线清楚而语调，连写不易出错，而且还便于记忆。在印刷、电讯等方面使用，不改动西文打字键盘的现成设备，效率

可以大大提高。但是，这种简单形式的双拼方案，仍然存在看大量同音字一即会出现许多同码的问题，这个问题不解决，还是不能使用的。

如果在双拼方案音节前面或后面，加上区别同音字的一两个符号，就成了一个完整的汉语双拼汉字编码。

各种双拼编码方案用来区别同音字的符号也都是不同的。有的按词类、词义、有的标偏旁部首，形体结构，起末笔笔形，出现频率等等，也都各有特点。其中一机部机械工业自动化研究所扶良文同志的专著《双拼方案及其在汉字工程中的应用》，说理深刻、透彻，受到普遍的赞扬。《英华大字典》的编者郑易里同志也给予很高评价，说他的报告“无异是第一次在汉语拼音问题上爆炸出来的一颗原子弹”。更有说服力的是扶良文和其他一些同志的双拼编码，均已在电子计算机上做过实验，获得了成功。今年二月，中国汉语双拼工程研究会筹各组，在北京主办一次“汉语双拼工程汇报表演”，内容包括工程简介、幼儿园教学、小说朗读、打字、手语、灯语、旗语、音响语、速记、电报、带调双拼盲文等。与会者对精彩的表演报以热烈的掌声。

因此，不难看出，汉语双拼方案有着比较明显的优点。

(四)

我所编写的汉语双拼字编码，是以我自己设计的汉语双拼文字方案(II)为基础的(见附表一)，其特点是没有重码，并能准确读音。结构是每个编码都由四个字母组成：前面两个字母是双拼音节，第三个字母是声调符号，第四个字母是区别同音字的顺序号。

已经学会汉语拼音方案的同志，要掌握我的汉语双拼文字方案(II)并不难。声母zh、ch、sh换成了y、w、v，其余的都不变。韵母改动较大，原来两个以上的字母，全由一个字母代替。说明第一和第二两条与汉语拼音方案相似，只需要记住三、四两条就行了。

汉语普通话总共四百多个音节，全部可以用汉语双拼文字书写。在一般情况下，无需标调。在播音员正音手册，教科书和工具书中，声调符号不可省略。声调符号用字母表示：阴平p，阳平t，上声v，去声q，轻声x。

同音同调的双拼音节后，再加上顺序号，就成了汉字编码。例如：把，音ba，上声v，顺序号a，汉字编码就成了bava。在新华字典中，同音同调的方块汉字，一般不超过26个。超过此数的，用下列字母标调：阴平k，阳平r，上声w，去声s。超过52个的，再用下列字母标调：阴平l，阳平m，上声z，去声y。余此类推。按用户需要，自己还可以续码。将整个康熙字典的用字全编进去，也不会有重码。

当然，每个编码的顺序号都不能随便加。它主要是参考汉字频度表，对号入座的。众所周知，每个方块汉字在大量的出版物中，出现的次数并不相等。就拿北京新华印刷厂编印的“汉字频度表”来说，它的“综合频度表”是由政治理论、新闻通讯、科学技术、文学艺术等四类频度表综合而成的。所查内容包括86本书，104本期刊，7075篇文章。总字数：21,629,372个字。单字数：6374个字。最多的出现830,302次，最少的只出现1次。按综合频度表编号顺序，掌握921个字，就可以看懂各类出版物百分之九十以上。掌握2346个字，

就可以看懂各类出版物百分之九十九。

一个汉字字音读得准的译码员，要在较短时间内容掌握九百多个汉字编码，并不是一件十分困难的事情。如果结合词、词组、句子，甚至一小段文章，来记汉字编码的顺序号，那就更能收到较好的效果。

请看下面一句话的编码练习：

oipa-dzqa 定 ivqa 要 jite 极 daqa 大 dexb 地 tita
提 gvpa 高 ynva 整 geqa 个 yipa 中 ndta 华 mhta 民 zuta
族 dexta 的 kepa 科 xota 学 unta 文 hdaqa 化 vlva 水 pzta
平。

在这 21 个字的编码中，我们只需要记住 jite (极) 和 dexta (地) 两个编码的顺序号，其余 19 个汉字编码的顺序号，都是 a，用不着特别记忆。这样的例子不是巧合，也不是少数。yipa 中 gota 国 gfqa 共 wjva 产 axva 党 yipa 中 ixpa 央 uiva 委 ibta 员 hlqa 会这十个汉字编码的顺序号，都是 a，记忆就不花什么时间。

为什么顺序号是 a 的编码出现频率如此高呢？说来也并不奇怪。

汉语音调总共有四百多个音节有字。加上四声，总共 1280 个音调有字。少部分音调只有一个字。如：甬、北、白、蹭、忖、草、藏、猜、打、大、端、短、过、们、能、我、走、揍等等，它们的顺序号肯定是 a。

确定同音字中出现频率最高的办法有两种。一种是作工程心理学上的实验；另一种是数理统计。先作工程心理学的实验：找一部分不同年龄、不同性别、不同职业、不同文化程度的人，

请他们将若干组同音字中最先想到的那个字写出来，然后求平均数，其结果与数理统计基本吻合。它们的顺序号也都是a，道理很简单：出现频率越高的字，在人们头脑中印象越深。

经过长时间反复实验，人们将会发现，各种资料译成编码后有70%以上的顺序号是a。

当然，对那些出现频率不高的字码，还是要靠死记硬背。初学者可查编码本。

那么，这样的汉字编码究竟有哪些用处呢？

一、可代替目前使用的电码收发电报。因具有“可读性”，翻译电文时，一般用不着看译码本。

二、报道新闻和记录新闻使用电传机打码，效率可以大大提高。

三、因为没有重码，凡电子计算机编辑、排版、图书情报检索，可以做到准确无误。又因为是等长码，可以安排自动跳格，用不着按空格。

四、因为字、码的读音统一，可以参照它编写同音字典，有利于推广普通话，为汉字走向拼音化找到一条捷径。

五、根据它编写报刊、杂志、图书、资料、档案、文献等索引。在没有汉字显示设备的情况下，可以利用西文电子计算机进行检索。还可以用于机器翻译。

(五)

我的汉语双拼汉字编码，是按音序排列的。对一些见形不知音的冷僻字，也用不着望“文”兴叹，可以按照“礼”字笔顺查

码法，迅速找到各个字的编码。

笔顺查码法，实际上就是笔顺查字法。按多数人笔顺习惯排检汉字的笔顺查字法，其笔形选定和取笔方法，各有不同。但许多同志认为，基本笔形号码化，能使汉字排检定位简单易行。基本笔形分几种，如何定位，尚无统一认识。例如，主张基本笔形分五种的，号码排列就有 $P_5 = 120$ 种。其中，正好由五个基本笔形构成的字就有礼、央、史、丙、弁等。

主张笔向号码查字法的李金铠、李毅民同志，将基本笔形分为五种，用五个号码来代表：

名称	横	竖	撇	点	折
笔形	一		丿	丶	乚
号码	1	2	3	4	5

虽然李金铠、李毅民同志的笔向号码法尚未得到推广，但他们选定的笔形号码却被普遍采用。

又因为按这种基本笔形顺序书写，正好构成一个“礼”字，为了记忆方便，厉为民同志建议，将这类查字法取名为“礼”字笔形（或笔顺）查字法。

经过对多种查字法进行仔细地研究比较，吸收各家的长处，我试编“礼”字笔顺查码法，希望得到同志们批评指正。

“礼”字笔顺查码法规则有四条，简述如下：

第一、横1竖2撇是3，点捺作4 5折弯。

每个单体字与合体字的偏旁部首，都要按笔顺排序。从左齐位，由小到大。查字时，先看第一位，再看第二位。例如：

一	1	干	112	井	1132	十	12
二	11	王	1121	开	1132	土	121
三	111	弄	·132	开	1132	圩	·115
幸	1112			天	1134	坛	·1154

第二、对于多数合体字，斩头去尾留中间。

合体字占汉字总数90%以上。同偏旁同部首的字可以归并。码头码尾皆省略，分别用圆点“·”和短横“-”表示，码段究竟取几位数字合适，要与其它同码头的字进行比较，以区分开来为限。例如：

自	325111	ziqa	又	4134	unta
泉	·12-	ntql	斌	·11-	bkpc
臭	·13-	⊖wqqa	斋	·13-	yypb
		⊖xsqh	刘	·25	lstb
鼻	·2-	bita	吝	·251	lkqb
新	·11-	hjpd	芥	·32	qite
艱	·12-	nxqa	剂	·25	jiqg
蹉	·21-	yapm	齋	·321-	jipx
剔	·25	oiql	嫻	·4-	ljtm
鞠	·2-	hqpa	素	·5-	unvc
龜	·3-	uhqc			
忌	·4-	xipe			

第三，少数笔顺有分歧，上下左右分正偏。

不赞成笔顺查字法的同志，往往会找一些笔顺有分歧的字来进行非难。其实，如能严格遵守自上而下、从左至右、先正后偏

的原则，这个问题也是不难解决的。例如：

上 211 小 534 水 5534 刀 53
九 35

第四，笔形笔顺全相同，按照音序排后先。

这类字，为数不多。按照汉语双拼音节排先后次序。例如：

士 121 央 25134 入 34 刀 53
士 121 史 25134 入 34 力 53

附表一：

汉语双拼文字方案（Ⅱ）

一、声母表：

b	玻	p	坡	m	摸	f	佛
d	得	t	特	n	讷	l	勒
g	哥	k	科	h	喝		
j	基	q	欺	x	希		
y	知	w	蚩	v	诗	r	日
z	资	c	雌	s	思		

二、韵母表：

a	啊	o	窝	e	鹅		
i	衣	u	乌	û	迂	r	儿
y	哀	w	欸	v	熬	q	欧
â	呀	ê	耶	c	腰	s	忧
d	娃	t	歪	l	威	ô	约
j	安	n	恩	x	昂	h	翰
g	烟	k	因	m	央	z	英
b	弯	p	温	m	汪	f	翁
b	冤	p	晕			f	雍

三、说明：

1. 声母 y w v r z c s 自成音节时，后面加韵母 i；
2. 声母 j q x 和韵母 û â ê ô 组成音节时，韵母上的附加符号可以省去；
3. 韵母自成音节时，分别写成：

oa	啊	ua	窝	oe	鹅	oi	衣	eu	乌	iu	迂	er	儿
ai	哀	ei	欸	ao	熬	ou	欧	ia	呀	ie	耶	iv	腰
ua	娃	uy	歪	ui	威	io	约	an	安	en	恩	ah	昂
ij	烟	in	因	ix	央	ih	英	uj	弯	un	温	ux	汪
ib	冤	ip	晕			if	雍						

4. 普通话四百多个音节中，有附加符号的音节只有十二个：

nû lû lâ bê pê mê âê tê nê lê nô lô.

为便于文字机械化，可以将它们改写成：

nr lr ld bt pt mt dt tt nt lt nl ll.

四、拼写练习：

un zi bi xu gy ge, iv zq vi je un zi gf tf

文 字 必 须 改 革， 要 走 世 界 文 字 共 同

de pk in fx xm.

的 拼 音 方 向。

附表二：

汉语双拼文字音位表

字母表	字母名称	相当于汉语拼音的					
		声母音位	韵母	母音位		机械规定	
				在声母jqx后			
a	Y		a	啊	ia	呀	
b	ㄅ	b	uan	弯	üan	冤	
c	ㄘ	c	iao	腰			
d	ㄉ	d	ua	蛙			在d, l后读ia呀
e	ㄜ		e	鹅	ie	耶	
f	ㄈ	f	ong	翁	iong	雍	
g	ㄍ	g	ian	烟			
h	ㄏ	h	eng	鞞			
i	ㄣ		i	衣			
j	ㄐ	j	an	安			
k	ㄑ	k	in	因			
l	ㄌ	l	ui	威			在n, l后读üe约
m	ㄇ	m	uang	汪	iang	央	
n	ㄋ	n	en	恩			
o	ㄛ		o, uo	窝	üe	约	
p	ㄆ	p	un	温	ün	晕	
q	ㄑ	q	ou	欧			
r	ㄖ	r	er	儿			在n, l后读ü迂
s	ㄙ	s	iou	忧			
t	ㄊ	t	uai	歪			在b, p, m, d, t, n, l后读ie耶
u	ㄨ		u	乌	ü	迂	
v	ㄨ	sh	ao	熬			
w	ㄨ	ch	ei	欸			
x	ㄒ	x	ang	昂			
y	ㄩ	zh	ai	哀			
z	ㄗ	z	ing	英			

aiud	霭	anpj	盒	aotf	逃	ba	吧
aigb	肱	arpk		aotg	聾	bapa	八
aigb		anva	铵	aoth	葵	bapb	巴
aita	昂	anvb	俺	aoti	螫	bapc	吧
antb		anvc	按	aotj	鳌	bapd	八
ahqa	盘	anvd	掩	aotk	翱	bape	扒
ahqb		anve	揩	aotl	磨	bapf	疤
ai		anvf		aotm		bapg	芭
aipa	埃	anqa	按	aova	拗 ⁽⁻⁾	baph	捌
aipb	挨 ⁽⁻⁾	anqb	棠	aovb	袄	bapi	芭
aipc	唉 ⁽⁻⁾	anqc	暗	aovc	媪	bapj	芭
aipd	哎	anqd	岸	aovd		bapk	耙
aipe	底	ange	胺	aoqa	奥	bapl	
aipf	鎰	anqf	黯	aoqb	澳	bata	拔
aipg	埃	anqg	犴 ⁽⁻⁾	aoqc	傲	batb	跋
aiph		anqh		aoqd	懊	bate	颀
aita	呆 ⁽⁻⁾	anpa	耍	aoqe	拗 ⁽⁻⁾	batd	鼐
aitb	癌	anpb	气	aoqf	拗	bate	
aitc	皓	anpc	鞍	aoqg	鳌	bava	把 ⁽⁻⁾
aitd	捱	anpd	按	aoqh	鸫	bavb	靶
aite	联	anpe	庵	aoqi	圪	bavc	钹
aitf		anpf	请	aoqj	蚕	bavd	
aiva	矮	anpg	鹤	aoqk	鼻	baqa	把 ⁽⁻⁾
aivb	喂 ⁽⁻⁾	anrh	厂 ⁽⁻⁾	aoql		baqb	霸
aivc	藹	anpi	广 ⁽⁻⁾	aote	敷	baqc	爸

bgjc 標	bgqf 辦	bipa 遍	biqk 痺	biqj 筭	bjvd 板
bgql	bgqs 卞	bipb	biql 庇	bick 啤	bjve 敝
bg	bgqh 汴	bita 鼻	biqm 菟	bicl 踣	bjvf
bgpa 邈	bgqi 汴	bitb 孛	biqn 碧	bicm 啤	bjqa 办
bgpb 編	bgqj 忤	bitc	biqo 晁	bicn 僻	bjqb 半
bgpc 鞭	bgqk 弁	biva 比	biqp 鉍	bico 浣	bjqc 拌
bgpd 鯽	bgql	bivb 笔	biqq 弊	bisp 霰	bjqd 伴
bgpe 砭	bh	bive 彼	biqr 敝	bisq 薜	bjqe 辦
bgpf 煊	bhpa 崩	bivd 鄙	biqs 毖	bisr 泌	bjqf 扮
bgpg 蛄	bhpb 緇	bive 批	biqt 篦	biss 淅	bjqg 緋
bgph 蓮	bhpc 喘	bivf 俾	biqu 婢	bist	bjqh 棼
bgpi	bhpd 紡	bivg 妣	biqv 壁	bj	bjqj
bgva 扁	bhpe	bivh 訖	biqw 壁	bjpa 般	bk
bgvb 敗	bhta 甬	bivi 乚	biqx 陞	bjpb 班	bkpa 吳
bgvc 匾	bhtb	bivj	biqy 懷	bjpc 斑	bkpb 吳
bgvd 編	bhva 緇	biqa 必	biqz 裨	bjpd 搬	bkpc 嫻
bgve 確	bhvb	biqb 辟	bisa 弼	bjpe 扳	bkpd 彬
bgvf 薺	bhqa 聚	biqc 壁	bisb 肅	bjpf 頒	bkpe 頻
bgvg 窆	bhqb 蛭	biqd 避	bisc 控	bjpg 癩	bkpf 撲
bgvh	bhqc 蹠	biqe 閉	bisd 膏	bjph 癩	bkpg 縵
bgqa 變	bhqd 迸	biqf 臂	bise 賁	bjpi 放	bkph 候
bgqb 便	bhqe 鋪	biqg 半	bisf 必	bjpi 放	bkpi 鏤
bgqc 通	bhqf 覓	biqh 秘	bisg 嬖	bjva 板	bkpj 邪
bgqd 辯	bhqg	biqi 巾	bish 昇	bjvb 版	bkpk 幽
bgqe 辯	bi	biqj 蔽	bisi 算	bjvc 阪	bkpl

bkqa 揆	bopc 剥	botr 鵠	btta 剔(-)	buqs 埠	bvuf 蕨
bkqb 冀	bopd 瓊	bots 鉞	bttb 釐	bugh 鉅	bvvg 鴉
bkqc 娛	bope 拔	bott 傅	bttc	buqi 埔	bvvh
bkqd 睽	bopf 鉢	botu 鮑	btva 瘰(-)	buqj 舘	bvqa 扱
bkqe 骸	bopg 疲	botv 毫	btvb	buqk 哧	bvqb 暴
bkqf	boph 俘	botw 踣	btqa 剔(=)	buql	bvqc 爆
bn	bopi 幡	botx 百(=)	btqb	bv	bvqd 刨
bnpa 奔(-)	bopj	boty 𦵏	bu	bvpa 匏	bvqe 抱
bnpb 鑄	bota 薄(=)	botz	bupa 逋	bvpb 胞	bvqf 豹
bnpc 拼	botb 伯(=)	bova 𪔐(-)	bupb 晡	bvpc 剝(-)	bvqg 鮑
bnpd 賁	bote 勃	bovb 跛	bupc	bvpd 炮(=)	bvqh 瀑(-)
bnpe	botd 博	bovc	buta 𪔐	bvpe 抱	bvqi 鈞
bnva 本	bote 駁	boqa 薄(=)	butb	bvpf 苞	bvqj
bnvb 本	botf 柏(=)	boqb 𪔐(-)	buva 补	bvpg 裱	bw
bnvc 眷	botg 泊(-)	boqc 𪔐	buvb 捕	bvph 匏	bwpa 卑
bnvd	both 𪔐	boqd 𪔐(-)	buvc 堡(=)	bvpi 煲	bwpb 杯
bnqa 奔(=)	boti 鉅	boqe	buvd 哺	bvpj	bwpc 悲
bnqb 𪔐	botj 𪔐	boxa 卜(=)	buve 卜(-)	bvta 薄(-)	bwpd 碑
bnqc 𪔐	botk 搏	boxb 𪔐	buvf	bvtb 匏	bwpe 𪔐(-)
bnqd 𪔐	botl 𪔐	boxc	buqa 不	bvte	bwpf 庠
bnqe 𪔐	botm 𪔐	bt	buqb 𪔐	bvva 保	bwpg 𪔐
bnqf	botn 𪔐	btpa 𪔐	buqc 步	bvvb 宝	bwph 𪔐(-)
bo	boto 𪔐	btpb 𪔐	buqd 布	bvvc 𪔐	bwpi
bopa 波	botp 𪔐	btoc 𪔐(-)	buqe 𪔐	bvvd 堡(-)	bwva 北
bopb 播	botq 𪔐	btpd	buqf 𪔐	bvve 𪔐	bwvb