

Title	中国四角号码检字与电脑汉字输入法现状探讨
Sub Title	中国の「四角コード」漢字検索法とコンピュータ入力法の現状について Current issues in sijiao code hanzi search and computer input method for Chinese language
Author	山下, 輝彦(Yamashita, Teruhiko)
Publisher	慶應義塾大学藝文学会
Publication year	2013
Jtitle	藝文研究 (The geibun-kenkyu : journal of arts and letters). Vol.104, (2013. 6) ,p.177(168)- 195(150)
JaLC DOI	
Abstract	
Notes	
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00072643-01040001-0195

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

中国四角号码检字与电脑汉字输入法 现状探讨

山下輝彦

前言

中国的汉字自其发明以后在长达几千年的发展过程中，不仅是书写汉语的文字，还成为了书写日语和朝鲜语的文字。汉字不像拼音文字那样使用有限的文字来书写无限的概念，而是每一个概念都用一个文字来表达，是一个开放的集合。现在最大的汉字辞典已经收录了6万以上的汉字，理论上汉字可以无限增加。对于这样一个庞大的集合，从几万字中如何检索到一个汉字，历来就是一个很大的课题。汉朝许慎的《说文解字》（公元100年）使用部首检字。两千年来，除个别采用笔画检字法等以外，这种方法一直沿用至今。1925年11月王云五发表了他的四角号码检字法，发明了一种理念完全不同的检字法。

20世纪中期电脑出现了。起初电脑只能使用英文，后来电脑中装进了汉字字型，可以处理汉字，也就产生了把汉字输入到电脑中的大规模行动。英文键盘上有26个英文字母键和数字键，投资最少的办法就是使用电脑上必带的英文键盘输入汉字。于是就出现了电脑汉字输入法这一课题。

汉字检索与汉字电脑输入有很多相似之处，本文根据在中国北京观察到的现况为基础，就这两方面的问题作一些探讨。

1.0 王云五四角号码检字法探讨

王云五出生于1888年，曾经主持商务印书馆长达四十年之久，以创出王云五四角号码检字法闻名于世。当时中国已经开始使用电报，收发电报使用

的是一种四位数的“电报码”。王云五参考了这种数字编码，摆脱偏旁检字的束缚，开创了一条检字法的新途径。众所周知，传统的偏旁检字法存在许多缺点。其中之一就是构成该字的哪一部分是偏旁不容易确定。例如：对于“鷹”字，有的字典把“广”作为偏旁，而有的字典则把“鸟”作为偏旁，查字的时候很难确定。此外，使用偏旁检字法要先找偏旁，找到偏旁后再从同样偏旁的许多字里面选出自己要查的字。这样一来，要经过好几道手续才能查到要查的字。王云五则完全放弃了偏旁，着眼于汉字的四个角，通过观察四个角的笔画，直接变换为四位数来检字。具体地说，把四角的笔画分为10种，每一类给予一个号码，因此称为四角号码。¹

1.1 四角号码检字法介绍

首先是笔形和代号

号码0 笔名 头（点和横相结合）

例字 主 病 广 言 的上部

号码1 笔名 横（横及挑、横上钩和斜叉钩）

例字 天（上部） 土（下部）

活（左上角） 兄（右下角）

号码2 笔名 垂（竖、撇和竖左钩）

例字 旧（左边） 山（两边）

千（上部） 力（右边） 则（右边）

号码3 笔名 点（点、捺）

例字 宝（上部） 外（右边）

去（左下角） 迹（下部） 瓜（右边）

号码4 笔名 叉（两笔交叉）

例字 古（上部） 草（上部）

对（左边） 皮（下部）

号码5 笔名 插（一笔纵穿两笔或两笔以上）

例字 青（上部） 本（中部）

打（左边） 中（中部）

号码6 笔名 方（四角整齐的方形）

例字 另（上部） 国（周围）

目（周围） 园（周围）

号码7 笔名 角（一笔向下或向右转折的角形、两笔笔头相接所形成的角）

例字 刀（右上） 亡（左下）

阳（左上） 雪（右下）

号码8 笔名 八（八字形、八字形的变形）

例字 分（上部） 共（下部） 余（上部）

羊（上部） 午（上部）

号码9 笔名 小（小字形、小字形的变形）

例字 尖（上部） 宗（下部） 快（左边）

当（上部） 组（左下部）

检字时还要遵守下列几个原则：

一、取角顺序：

每字按①左上角②右上角③左下角④右下角的次序取四个角的号码。

例：端 0212 烙 9786

二、取角方法

1. 一笔可以分角取号。

例：以 2870 乱 2261 乙 1771 几 7721

2. 一笔的上下两段和别笔构成两种笔形的，分两角取号。

例：半 9050 大 4080 火 9080 米 9090

3. 下角笔形偏在一角的，按实际位置取号，缺角作0。

例：产 0020 户 3020 弓 1702 妒 4340

但“弓”等字用作偏旁时，取2作整个字的左下角号码。

例：张 1223 鄂 6722

4. 凡外围是“口门（門）門”的三类字，左右两下角改取里面的笔形。

例：园 6021 田 6040 開 7744 闕 7721

但上、下、左、右有附加笔形的字，都不在此例。

例：苗 4460 汨 3610 陸 6401 簡 8822

5. 一个笔形，前角已经用过，后角作为0。

例：王1010 全8010 心3300 一1000

三、附号

1. 为了区别四角同码字，再取靠近右下角（第四角）上方一个笔形作“附号”，若这一笔形已被右上角用过（如“决”“速”），则作0。

例：芒 44710 喜 40601 逢 37305 豆 10108

2. 四角和“附号”相同的字，照各字笔画数多少排列。

例：市 00227（五画） 帝 00227（九画）

除了上述的普通情况以外，作为“附则”还对一些笔形作了一些解释。如角形尽量取复笔：“厂”的号码是7120而不是1020等等。

四角号码检字法还有一个口诀为：

横一垂二三点捺 叉四插五方框六

七角八八九是小 点下有横变零头

把规则编成口诀让大家容易记忆是中国文化的一个特色。这个口诀基本上包括了四角号码检字最基本的作法。

1.2 对四角号码输入法的探讨

从信息处理的观点上来看，四角号码因为采用了汉字的四个角来取码，所得到的必然是四位数。而且从所用数字的性质上来看，只能使用0到9的十个数字。但是汉字的笔形何止十种，因此四角号码中所规定的十种笔形中也包括了其他大约50个种类的笔形。这50多种笔形如果各用两位数来转换的话，字码就会变为5位数，因此这种处理是有合理性的。

另外，中国的电码以前也使用四位数的数字。四位数的数字是从0000到9999的，因此理论上只能用来表示1万个字的编码。对于常用汉字这是没有问题的，但是对于字典就不够用了，就会出现重码即两个以上的字使用一个数码的现象。电码是将0000到9999的数字机械地把汉字分配上去的，因此不会有同一个数码表示几个汉字的现象。但是四角号码是根据汉字的四角转换成的，因此四个角形状相同的不同的字当然会转换成同样的号码。根据最近的调查²，四角号码中数字的使用频率并不平均。涂建国的调查统计指出，四角号码所使用的四位数字，在《四角号码检字表》(1926年收字数7334)中，

实际使用号码数为2422个,实际使用号码平均字数为3.03,空号率为75.75。就是说四角号码可使用的1万个数字中只使用了三分之一。

如果使用上面引用的“附号”的话,可以进一步把一部分同码的字分开来。就是说,四角号码的位数是4,但是根据情况可以扩大到5位数。

四角号码完全摆脱了部首的束缚,没有必要考虑汉字的意义即可把汉字转换为号码来查字。如果熟悉了该检字方法,可以提高检字速度。关于这一点,王云五曾经做过实验,并得到了内行的认可。

徐祖友在《王云五与四角号码检字法》³一文中阐述说:

“四角号码检字法只取汉字的四角,完全摆脱了字义的束缚,彻底贯彻以字形为查字的唯一依据的原则,使查字的原则单一化,这可作为四角号码的根本长处。……四角号码检字法以部分笔画代替全部,以笔画种类代替笔画数量,在正确确定所查字的四角号码后,便可直接从字典正文中检得,这无疑极大地提高了检字速度。据王云五当时的试验,运用四角号码法查检一字的平均时间,要比部首法省一分二秒,比笔画法省一分四十五秒;据笔者较长时间的使用比较,四角号码法在学习掌握的初始阶段,固然要相对多花一些时间,但一旦能熟练运用,确要比其他方法便捷。”

据说,这个检字法发表后的1928年,经全国教育委员会通过,当时国民政府的大学院(后来的中央研究院)通令全国推广采用四角号码检字法。中华人民共和国成立后的1964年,文化部、教育部、中国文字改革委员会和中国科学院语言研究所联合组成的汉字查字法整理工作组提出“拼音字母查字法、部首查字法、四角号码查字法(对王云五的四角号码检字法作了修改)和笔形查字法(即笔画查字法)”4种草案,推荐给文化、教育和出版界试用。⁴

在上述四种草案中,拼音检字法与其他三种是完全不同的检字法。人们查辞典的目的有所不同。就汉语来说,人们在写文章的时候,如果已经知道发音和意思,但是不知道或忘记如何书写的时候,可通过辞典来查字。在这种情况下,因为不知道字形,就需要通过发音来查。由于中国推行汉语拼音的历史还不长,这样的字典以前是不存在的。民国时代的《国语辞典》⁵第一次从发音上检字,当时使用的是注音字母。新中国成立后使用拼音字母查字,

其典型的字典就是《新华字典》。然而,如果在读书读报过程中遇到不能读并且不知道意思的汉字,就不能使用拼音检字法了。用发音来检索的字典大致相当与日语的“国语辞典”,而以汉字为检索的字典相当于日语的“汉和辞典”。汉语中使用汉字的数目远远多于日语,需要查的字数也大大超过了日语,因此汉语更需要像四角号码这样的速度快的检字法。

1.3 四角号码输入法逐渐消失

对于四角号码输入法的现状,笔者在北京作了一项调查(2013年2月)。受笔者提问的60人中,正在使用四角号码的人为0。而且,一直出版四角号码检字法辞典的商务印书馆现在也只出版一种四角号码检字辞典。这样一个有效的检字法为什么没有被利用呢?对此,韩黎新在《四角号码检字法亟待重视推广》一文中总结为以下4点:

“1、掌握四角号码有一定难度。四角号码需要记忆的内容多,但又难以记忆,而且全部记忆后还需要融会贯通,再经过较长时间的练习才能会使用。(中略)每天需要记忆很多内容的学生,对于这些难以记忆的内容,只好‘敬而远之’。

四角号码重码字多,也使人‘心生畏惧’。据对13247字的四角号码统计,字数最少的9字头号码共有651个,有7个号码的重码字在9字及以上,9702号重码最多,共有16字(中略)

最主要的困难在于,四角号码还有新旧之分。(中略)单就现有的工具书没有统一使用一种四角号码而言,新旧之分就为学习四角号码、给汉字编码查字带来很大的困难。(中略)

2、四角号码不受重视,采用四角号码检字法或附有四角号码检字表的工具书少,学了四角号码难有用武之地;大学教材中的四角号码教学内容也少,小学也只教拼音、部首、笔画三种检字法,不教四角号码。

3、(中略)传统的部首检字法、笔画检字法尽管有部首不易确定、汉字笔画偏多等困难,检字非常慢,但由于《新华字典》《现代汉语词典》等采用‘多开门’的方式设置部首,国务院1956年正式发布了《汉字简化方案》,中国文字改革委员会1964年公布了《简化字总表》,部首不易确定、汉字笔画偏多等困难有一定程度的减轻。

4、在新华社1948年公布的以蒋介石为首的43名头等国民党战犯的名单中,四角号码的创始人王云五被列为第15号战犯。(中略)这样一个国民党政府的大员、要员、头等战犯,尽管有学术成就,但在1949年新中国成立后也一定要被‘封杀’,这也就影响了对四角号码的研究、评价和推广。1964年,虽然查字法整理工作组也推荐了四角号码,但要用四角号码却不能提王云五,四角号码遭受冷落也就在所难免。”

如韩黎新所述,四角号码检字法是通过学习才能使用的一种方法,这种学习就需要一个熟练的过程。而且这种检字法会出现重码,即一个数码表示许多字的现象。如果要想减少重码,检字法的规则就要定得非常细,这就增加了学习检字法的麻烦。王云五在创造四角号码的时候还没有电脑出现,还没有完善的信息处理理论作后盾,因此四角号码存在着一些缺陷。关于这一点有人进行了考察研究,并提议说应对四角号码的编制作一些改良。

“从统计可以看到,0区间(即0000—0999)的实用号码最少,2、4区间的最多。不仅如此,而且笔者在分析、统计时,还发现各区间中数码0的出现率较低,数码2、4的出现率较大。这是因为0只代表一个笔形,且汉字中这种角形出现较少,数码2、4代表的笔形则不仅较多,而且在汉字中出现也较频繁。笔者设想能否将2、4所代表的部分笔形的变体调整给其他数码,使号码的分配合理一些;至于复杂,我想只要熟练,也就简单了,歧义问题,可以通过修改‘四角法’的查字规则来解决;这样一来,四角法作为优秀的字典码就当之无愧了。”⁶

笔者以为,四角号码对于速查繁体字是最有效果的,因为繁体字的四个角比较明显。简体字使许多汉字的四角变得模糊了。例如“亿”的繁体字“億”就比简体字的四个角容易找出。因此简体字的使用很可能对四角号码查字法造成了混乱。但是如果中国官方重视四角号码的话,完全可以制定一套经过改良的统一的四角号码,使这种检字法发挥其作用。

对于因王云五的个人经历致使四角号码在中国国内难以推行的说法,尚需继续考察,难以立即下结论。不过,中国国内对四角号码的批判,有的曾经相当强烈。杨继本在一篇题为《汉字结构数码检索法探究》的文章中列举了“规则嫌繁多,难熟悉掌握”、“强行规定代号、不利记忆运用”、“一笔

分角取号，造成感知困难”等四角号码检字法的缺点后这样写道：

“总之，这种检索方法基本上可说是脱离劳动人民在三大革命实践中创造、认识和使用汉字过程中所总结出来和确定下来的字形结构实际，而主观地确定字形代号和割裂地、拼凑地确定角形并给予角号及呆板地排定角序，使人消极地反映字形，在实际上抹煞了人同字之间的辩证关系，抹煞了人的活生生的能动作用，违反马列主义认识论的原理。在一定意义上可说还起着培养机械唯物主义、唯心主义形而上学的不良作用。（中略）汉字检索方法的改革，对于批判剥削阶级利用方块字垄断文化实行愚民政策的罪行，帮助劳动人民实现文化翻身，更好地掌握马列主义和科学技术，巩固无产阶级专政，具有积极意义。”⁷

对于这段评论，笔者认为其观点值得商榷。与其他检字法比较，四角号码的规则不能算是太多的，不能认为是“规则嫌繁多，难熟悉掌握”。而且汉字是一种非常复杂的文字，任何一种字形检索法都必须把一些笔形归纳在一起，要想让每一种笔形变换为数字时都具有理论依据是很困难的。四角号码的可贵之处正是摆脱了偏旁的束缚，彻底地以视觉形状来编码的，因此它是对“劳动人民实现文化翻身”具有积极意义的。杨继本的这种批判应该说是不妥当的。

对于重码的问题，通过上面引用的四角号码“附号”的规则，可以解决一部分问题。查字典用的系统没有必要要求不出现重码，因为这在检索上不会发生太大的问题。

2.0 汉字电脑输入法探讨

与汉字的检索法有密切关系的是汉字的电脑输入法。汉字检索在第一部汉字字典出版的时候就已经开始了。但是，汉字的电脑输入是上世纪中叶以后电脑开始处理汉字后才开始的。当然，中国的电码就是一种汉字输入系统。拍发电报时，把汉字改变为电报码用莫尔斯信号发出，收到信号的对方再把电报号码转换为汉字阅读。这就是使用电码打电报的原理。电报号码如果用来作为输入电脑的方法来使用是完全可以的。但是，电报号码是机械性的通过给汉字赋予4位数号码来创制的，没有重码，易于使用，但是电码不利于

记忆，所以只能供专业人员使用。电脑输入法与汉字检字法一样，也要求容易学习。但是两者的要求并不相同。例如，检字法一次一般只检索一个字，不要求速度有多快。而输入法则要求速度越快越好，因为输入的字数一般是以篇章为单位的。上世纪80年代为了输入汉字，中国曾出现了电脑输入编码热。很多人都参加了输入编码的研究工作，创出各种各样的输入法，呈现出了“万马（码）奔腾”的气势。到现在为止，仅在中国国内登记在案的输入法就有500多种。很多优秀的汉字输入方法通过了国家汉字信息处理委员会的鉴定，并且获得了国家专利，并进入了商品化阶段。⁸

至今为止，常用的输入编码有“智能ABC”、“刘氏双音输入法”、“自然码输入法”、“太极码（两笔字型）汉字输入法”、“五十字元输入法”、“长春双凸输入法”、“四笔声形输入法”、“六笔声形输入法”、“表形码”、“五笔字型汉字输入法”、“大易输入法”、“郑码输入法”、“仓颉输入法”、“层次四角输入法”等。⁹

初期的编码方案，大都受到了当时电脑硬件方面的制约：电脑的内存非常小，处理器速度慢，不能启动词典软件。因此，大部分方案都是采用单字直接键入的方式。因此在编码方面字形码比较多。这些输入法中，值得一提的是仓颉输入法。¹⁰

2.1 繁体字输入法——仓颉输入法

仓颉输入法以传说中的汉字创始人仓颉命名，是以台湾所使用的繁体字大五码（Big5）为基础编制的。该输入法自1980年发表以来在使用繁体字的地区得到了相当的普及。这是一个标准的形码输入法，把24个字根称作“仓颉字母”安排到键盘的24个英文字母上。这24个仓颉字母都是独立的汉字，其本身也可以单独使用。

24个仓颉字母的分配如下：

1、哲理类

仓颉字母	日	月	金	木	水	火	土
英文字母	A	B	C	D	E	F	G

2、笔画类

仓颉字母	竹	戈	十	大	中	一	弓
英文字母	H	I	J	K	L	M	N

3、人身类

仓颉字母	人	心	手	口
英文字母	O	P	Q	R

4、字形类

仓颉字母	尸	廿	山	女	田	卜
英文字母	S	T	U	V	W	Y

当然24个仓颉字母不可能概括汉字所有类型的部件。因此，这24个仓颉字母下面又挂着78个派生出来的或者形状类似的笔画，被称作“辅助字母”。例如，仓颉字母“戈”里面包含着字形“广”，如果要输入“床”字，就须键入“戈+木(I+D)”。

仓颉输入法的取形方法与汉字的笔顺基本一致：从外到内，从上到下，从左到右。对于由几个字根组成的可拆分的一个汉字，先将其分为2部分。第1部分称为“字首”，剩余部分称为“字身”。整个一个汉字最多变换为5个码，即5个英文字母。其中，字首最多用2个码，字身最多用3个码。如果字首太复杂，可取字首中开始和最后的码共2个。字身复杂时，取1、2和最后一个码。连体字则以第1个笔形为字首，其余的为字身。字首取1码，字身取第2、3、4码。下面对于可拆分的字与连体字，各举3个例子：

可拆分汉字

汉字	字首	输入字首字母	字身	输入字身字母	打键
明	日	日	月	月	AB
櫻	木	木	嬰	月、金、女	DBCV
騎	馬	尸、火	奇	大、一、口	SFKMR

连体字

汉字	输入字母	打键
正	一卜一中	MYLM
兩	一中月人	MLBO
面	一田卜中	MWYL

仓颉输入法开发较早，很早就被应用于电脑上，其最大优点是重码率低。在尚未使用电脑内软件词典变换功能的情况下，这种输入法得到了广泛采用。人们经过学习以后可以使输入速度达到相当高的水平。

3.2 简体字输入法——五笔字型

在简体字的输入法中，在大陆最流行的被认为是“五笔字型汉字输入法”，简称“五笔字型”。这也是根据汉字的字形输入的输入法。该输入法把汉字的笔画分为5种：¹¹

代号	笔画名称	笔画走向	笔画及变形
1、	横	左→右	一 “才”的最后一画 ¹²
2、	竖	上→下	丨 “到”字最后一画
3、	撇	右上→左下	丿
4、	捺	左上→右下	㇏、
5、	折	带转折	乙 フ ㄣ “乃”右边的一画“刀” 右上的一画

这里的笔画是象征性的。不管笔画的长短，只要方向相同就归纳为一类。基于这一原则“㇏”与“、”属于一类。

这5种起笔笔形在键盘上分为5个区：

1区（横起笔）GFDSA 2区（竖起笔）HJKLM 3区（撇起笔）TREWQ
4区（捺起笔）YUIOP 5区（折起笔）NBVCX

这就是各个键的区号，按照这个顺序对每一个区的键上再加一个“键位”号码。这样一来从1区起，各个区键与键位（号码）的对照关系如下：

G=11 F=12 D=13 S=14 A=15 H=21 J=22 K=23

L=24 M=25

T=31 R=32 E=33 W=34 Q=35 Y=41 U=42 I=43

O=44 P=45

N=51 B=52 V=53 C=54 X=55

与其他字形输入法相同，各个键位上也被赋予许多有关联的汉字部件即字根，以G11为例，这个键位上包含了“王旁青头戈五一”7个字根。五笔字型86版共设计了130多个基本字根，平均起来一个键位要负责输入差不多5到6个字根。学习五笔字型的第一步就是记忆这些字根的位置，然后按照拆分汉字的规则决定打字键位。

为了减少重码率，五笔字型有时候需要添加“末笔交叉识别码”。例如：“叭”与“只”的字根及字根的顺序完全相同，形成了重码。一个为左右结构，一个为上下结构。分别变换为K WY K WU，最后的Y是左右结构的识别码，U是上下结构的识别码。这样就解决了这两个字重码的问题。

2.3 对字形输入法的探讨

以上对中国汉字检索与电脑汉字输入法中比较常用的两种方法简单地作了介绍。汉字是几千年来积累起来的书写系统。在造字之前并没有规定应该遵照什么样的规则来造字，因此很多字都是约定俗成的。汉朝的许慎在《说文解字》中提出的“六书”也只是对已经造出的汉字的造字法作的总结。从许慎的时代到现在，又出现了新的汉字造字法，因此汉字的结构越来越复杂。本来的文字学是专门研究汉字起源及结构的学术领域，现在的文字学开拓了汉字属性和应用的研究领域。汉字的应用研究中包括了汉字人机界面、应用及计算机处理等新的领域。汉字输入法也应用了这一方面的理论。特别是在拼音输入方面，新的输入法不断出现。根据笔者对50人的电脑汉字输入法的调查发现，所有的人都在使用拼音输入法。就是说使用五笔字型等字形输入法的人已经非常少。80年代北京街道上的电脑训练班广告上几乎都写的是五笔字型培训。当时，要想使用电脑，必须要学好五笔字型。如上所述，电脑开发初期，一般是单字输入，如果使用单字拼音输入，即使加上四声，同一个单音节的拼音会调出几十个汉字来。拼音输入法得不到普及的一个社会原因是当时的电脑相当昂贵，不可能每人一台。一般每台电脑都配上了专业

的打字员，也就是电脑录入人员。录入人员是一种职业，再复杂的字形输入法他们都可以学会并熟练地操作。他们一般是看着别人手写的稿件来录入电脑，因此字形输入法是非常合适的。字形输入法难学，反而保证了他们职业的稳定。当电脑普及而且人们普遍使用电子邮件等以后，人们需要自己键入自己想写的文章，此时，字形输入法就会给电脑使用者带来困难。如果采用字形输入，即使已经学会，有一段时间不使用电脑，也会忘记。而且，因为大部分电脑用户是一边想一边写的，不是先有稿件而看着录入，用拼音是非常自然的。

2.4 拼音输入法

使用拼音的拼音输入法最大的缺点是重码，如果是单音，查阅字典可以知道，一个拼音“yi”可以有100多个汉字。这就是成为拼音输入的瓶颈，成为人们认为拼音难以输入电脑的原因之一。我们所能看到的古代汉语书面语，基本上是一个汉字一个词的。但是现代汉语的词，大都是两个汉字以上。把汉字组成的词为单位输入进电脑，作为电脑处理的单位，应该是最合理的。于是，就出现了汉语词联想输入法。这种做法有些像日语的游戏「尻取りゲーム（日语单词联珠游戏）」一样，输入一个汉字后，电脑屏幕上会出现由该汉字起头的词汇，让用户选择。现在许多智能手机的手写输入法中都添加了这一功能，提高了输入效率。但是这种输入法还不是真正的词语输入，要真正解决单字输入的瓶颈，一定要把单词作为处理单位。如果使用双音节处理，从减少同音也就是减少重码这一点上看是非常有利的。作为参考，试举日语的例子来考虑同音词问题。

日语中的汉字如果用“音读（来自古代汉语发音的日语读法）”，同音字会非常多，即使是两个音节，用音读来读，也会有很多的同音词。

例如，发音是「こうせい kosei」的日语词汇在《广辞苑》中共有30个词¹³，即

公正 孔聖 広西 江西 江青 向性 好晴 行星 坑井 攻勢
更正 更生 厚生 後世 後生 恒性 恒星 苟生 荒政 校正
高声 控制 甦生 硬性 較正 構成 興盛 鴻声 薨逝 曠世

与其相比，汉语的语音系统比日语复杂得多，因此即使使用不带声调的

汉语来读这些词（这些词不一定都与汉语词汇一致）同音也很少，只有三组：厚生・後生、校正・較正、恒性・恒星，是发音相同的。如果加上声调，则只有厚生・後生、校正・較正两组发音是相同的。可以说，汉语如果使用双音节以上的单位来输入的话，很大程度上可以克服重码的瓶颈。

山东大学盛玉麒教授在谈及汉字以词语的方式来处理的时候这样写道：

“联想输入法是汉字编码从单字编码向词语编码跨越期的一个创举，它突破了人们‘单字本位’的思维局限，启发了后来的词语输入法，在汉字编码发展史上具有重要的学术地位。

（中略）汉字只是一种书面符号而已。由于汉字在记录汉语时不分词连写，就使得汉字与汉语的词语不成对应。而人们的思维和交际，都是以词语为单位的。所以，中文信息处理的发展方向必须以词语为基础。

实际上，汉字只是记录汉语的书面符号系统，是一种可视的信息载体。把汉字放到汉语的系统中，就会发现，汉字在应用中并不是一个个孤立的个体，而是相互之间具有某种相关性的联系。这些相关性密切的汉字串实际上就是一个一个的词语。孤立的汉字只是一个假象，语言中可以自由运用的最小单位应当是词。把着眼点从字移向词，许多问题就豁然开朗了。

汉字编码从单字编码向词语编码的跨越，不但找到了解决单字重码率难题的一个重要途径，也标志着中文信息处理在理论上从字处理到词语处理的一个发展。”¹⁴

盛玉麒教授的这段话意味深长。作为书写符号的汉字是一个字一个音节，与汉语的语音形式相一致。但是，作为有意义的语音串，现代汉语一般需要多音节。长期以来对汉语的理解，受汉字这种文字的影响，总以为每个音节都是有意义的词，我们必须认识到“孤立的汉字只是一个假象，语言中可以自由运用的最小单位应当是词”。

虽说汉语的词在说汉语的人的脑中都有一个基本的概念，这种概念对于说汉语的人来说基本上是一致的。但是要想让电脑来分词却有很大的困难。汉语正书法中规定，在书写时词与词之间不空格，连续书写。按这种方式连

续键入的汉字就需要由电脑来分词。以往的词输入，先要将大量的词汇录入电脑，建立起词语库，当电脑使用者键入汉字时，使用电脑软件查找词语库，由拼音转换为词，确定为汉字。作为电脑操作，一般是使用者每键入一个单词按一下空格键，表示这是词的分界线。但是，这样一来，双音和多音节词的变换比较容易，夹在双音节或多音节之间的单音节词输入时就会出现变换选择项目太多的问题。要解决这个问题，实现超越单词的词组或整句输入，就需要进行电脑自动分词、词语使用频度、语法处理等软件的智能处理来解决。通过处理，可以决定该拼音所表示的词是用什么样的汉字来表示的。现在，中国在这一方面已经有了长足的进步，达到了实用的水平。

张普在《步入信息社会的汉语和汉字（三）》一文中对这一方面作了详细的阐述：

“经过艰苦的努力，自动分词技术已有了较大的突破。目前，中国的自动分词技术的正确切分率已从79%上升到95%以上。在‘七五’期间，北京师范大学、清华大学、北京航空航天大学、山西大学、杭州电信工程学院等单位的自动分词系统的正确率均达到99%以上。其中北京师范大学的自动分词专家系统，其知识库除采用了词法知识外，还采用了句法、语义等知识，已达到了国内外领先水平。

词处理水平的提高，反过来又推动字处理技术的进一步发展。一些优秀的汉字编码键盘输入系统，都从汉字编码进入了汉语词（语）编码，除建立字库外，又建立了词（语）库。自动分词技术的发展使大规模词频统计工作很快取得成果，各种更加科学的大型汉语通用词库和专用词库在七五期间陆续推出。有了词处理手段，一些字处理阶段无法解决的问题也很快迎刃而解。汉语拼音输入方法所碰到的同音字太多的问题，借助词处理就好办了。”¹⁵

现在电脑的硬件技术和软件技术已经足够使拼音输入法更加实用，更加方便。但是，并不是所有的人都能使用拼音输入法。使用拼音输入法第一个前提是要掌握普通话，并能够使用拼音。但是在年龄大的中老年人中还有一些普通话发音不标准的人，他们使用拼音输入法就会遇到很大的困难。可幸的是，对于普通话发音不标准的用户，输入软件中有加入“暧昧发音输入”

的功能。比如，不能区别声母“n”与“l”、“zh”“ch”“sh”与“z”“c”“s”发音的人或者韵母前鼻音与后鼻音不分的人，可以按照自己的发音习惯键入，由输入软件处理后转换为正确的汉字。而且考虑到有的人如果全部使用方言的发音可能会更容易输入，已经有人研制出了各种方言的拼音输入法，如广东话拼音输入法、上海话拼音输入法等。但是什么是这些方言的规范发音，使用什么样的拼音方式，现在还比较难以确定，因此使用价值究竟有多大，尚待考察。总而言之，使用最先进科技技术在很大程度上已经使拼音输入法在现在的中国成为主流。

对于拼音输入法的普及，中国国内有很多人感到担忧，他们认为这样会让许多人忘记汉字的写法。五笔字型创始人、中国发明协会副会长王永民、新华社《瞭望》周刊编委杨桃源曾在《光明日报》上发表文章说使用拼音会使人“提笔忘字”。¹⁶他说，根据某报对两年来刊出的错误进行分类统计，“拼音输入法”的错误率是“字形输入法”的9倍。文章还说：

“另外一项调查显示：在我国长期使用‘拼音输入法’的用户中，有‘提笔忘字’经历者超过95%。这是因为许多人长期在电脑、手机上‘打拼音’，已经习惯了用拼音‘代替’汉字。越来越多的人‘提笔忘字’，甚至不会写字，这是千百年来中华文化中从未有过的一件怪事，这种现象与‘拼音输入法’有极大关系。（中略）

汉字的‘音’是衣，‘形’才是身。‘弃形留音’，等于‘舍身存衣’。‘拼音输入法’离开了对汉字造字元素的直接思考和运用，可以毫不夸张地说，汉字必将因此而形神俱灭，汉字本身所固有的文化遗传基因，也将因此种输入方式大行其道而丧失殆尽。”

对于本篇文章的说法，笔者不能苟同。有些观点似乎与普通语言学的观点是不一致的。例如，认为录入错误比率高是拼音输入法所致就很难令人信服。普通人抄写文章也会出现错误，这些一般是因为负责这一工作的人员的疏忽造成的。拼音输入可能容易出现同音字的错误，笔形输入也可能容易出现字形类似的错误，如果负责录入工作的人非常认真谨慎，就不会出现太多的错误。拼音输入法造成的错误是字形输入法的9倍的说法令人费解。

另外，关于“提笔忘字”，这是使用电脑输入汉字以后容易出现的普遍现

象，并非单单由拼音输入法造成。即使使用五笔字形，由于五笔字形省略了汉字的很多部分，录入人也不会记住整个汉字的写法，也会造成忘记汉字写法的现象。因此，这种说法也不能成立。

从汉字的结构上看，汉字的一些部件当然反映了思维的载体，但是在历史的长河中，汉字经过多次变化，很多已经失去了这种作用，只是作为一个语言的符号被使用。况且字形输入法并不一定是为汉字的文化含义而设计的，而是为了容易使用或学习而创立的。在输入汉字时，只是机械地把字分解后输入，并没有考虑汉字的写法和组成元素的意义，使用字形输入法就可以不忘记汉字的想法也是不能成立的。

对于“汉字必将因此而形神俱灭，汉字本身所固有的文化遗传基因，也将因此种输入方式大行其道而丧失殆尽。”这种说法，北京大学教授、中国语文现代化学会会长苏培成也予以反驳说：

“有着几千年悠久历史、负载着丰富汉文化的汉字，自有其生存和发展的规律。由于计算机的使用，汉字扩大了使用的范围，谁曾见到了汉字形神俱灭的景象？说穿了不是汉字受到什么威胁，而是形码感到了威胁。这其实大可不必。形码和音码是不可缺少的两种输入方式，两者相辅相成，不过哪一种都不能统一天下。就字的输入说，知道读音的字可以用音码，不知道读音的字就只能用形码。”¹⁷

笔者以为，形码和音码是互补的。虽然拼音输入有了很大的进步，变换汉字的准确率非常高，但是如果输入古代汉语，或者只输入单个的汉字，效率会大大降低。在这种情况下就要使用其他方法。特别是专业录入人员可以使用一般人难以学会但是效率高的字形输入法。因为对他们来说这是一种职业。

汉语电脑输入法的开发研究已经有30年的历史。其基本理念就是以使用26个英文字母的键盘为前提。这里考虑到了经济原因，一般用户不可能为中文输入另外购置键盘。但是就因为这个硬件方面的限制，给研究输入法的人带来很大的困难。但是，中国人的智慧终于使汉语的电脑输入这个难题得到解决。

3.0 结语

本文从汉字检索的四角号码考察开始，探讨了电脑输入的仓颉输入法、五笔字型输入法和拼音输入法的现状。可以看出，王云五创出的四角号码查字法是为提高在字典上查寻汉字而创制的，然而他对后来的电脑输入法的创立起到了非常重要的启发作用。四角号码被冷遇虽然源于许多原因，其中重要的原因笔者认为是不愿意在学习一种汉字的查字法上花太多的精力。科学技术的发展使人们大多希望得到速成的东西。这是整个社会的一种倾向，也许是最主要的原因。

让电脑来处理上万的汉字一直是很多中国人的“科学之梦”。无论是字型输入还是拼音输入，汉字进入电脑的问题已经基本上得到了解决。电脑开始处理汉字以来，输入法经过了几次变化，由原来的五笔字型席卷天下到现在的拼音大流行。这种变化不一定起因于输入法本身的优劣，而是与社会的其他因素有着密切的关系。中国改革开放后经济的高速发展，政治体制的改革，电视广播媒体空前发达，中国的年轻一代几乎都可以说相当标准的普通话，这些都给拼音输入法奠定了坚实的基础。80后几乎都使用拼音输入法就是一个证据。

随着科技的进步和社会的变化，IT（信息技术）在不断革新，以前由电脑处理的信息，现在已经完全可以由一台小小的手机来处理。在手机上如何输入汉字也是一个新的课题。除了使用拼音输入法以外，手写输入与语音输入也得到了广泛利用。

笔者借助在中国长期考察的机会对汉字的电脑输入法作了一些研究，探讨很不深刻，只是提供一个反映现状的报告。今后随着信息技术的进步与变化，如何处理汉字问题仍然值得我们一直关注。

注

1 商务印书馆辞书研究中心修订《四角号码新词典》

商务印书馆出版 2008年1月第10版（凡例所收）

2 涂建国「号码检字法探究」《辞书研究》

- 1992年02期 p. 40 中国辞书研究杂志编辑部
3 徐祖友「王云五与四角号码检字法」《辞书研究》
1990年06期 p. 128 中国辞书研究杂志编辑部
4 韩黎新「四角号码检字法亟待重视推广」《现代语文（语言研究版）》
2007年第6期 p. 126 中国语文现代化学会会刊
5 中国大辞典编纂处编《国语辞典》（商务印书馆 1957年版影印本）
2011年7月 商务印书馆国际有限公司出版发行
6 涂建国「号码检字法探究」《辞书研究》
1992年02期 p. 44 中国辞书研究杂志编辑部
7 杨继本「汉字结数码检索法探究——兼评四角号码检索法」《湖南师范学院社会科学学报》
1978年 Z1期 p. 89-90 湖南大学出版社发行
8 王国印、刘培勇《常用的汉字输入方法》（前言）清华大学出版社
1995年1月
9 注8所引用的《常用的汉字输入方法》目录中所列举的输入法
10 此处的说明取自菅野苓著《中国語入力方法の話》p. 83-90
日本朝日出版社 1991年4月
11 资料采自前言文化编著《中老年人学拼音、五笔打字》
科学出版社 2012年5月
12 笔画字形无法印刷，采用某字的某一部分来表示。
13 采自《广辞苑第五版》岩波書店 CASIO 電子辞典 EW-V3500L 所录
14 盛玉麒编著《语言文字信息处理》p. 75-76 山东大学出版社 2006年5月
15 张普「步入信息社会的汉语和汉字（三）」《语文建设 (Language Planning)》
1993年05期 p. 40 语文出版社
16 王永民、杨桃源「警觉拼音输入法对运用汉字能力的销蚀」《光明日报》
2005年10月23日（采自《光明网》网站 <http://www.gmw.cn/>）
17 苏培成「也谈汉字输入中的形码和音码」《光明日报》2005年11月8日
（采自网站 <http://www.gmw.cn/>）