

报纸、广播电视、网络(新闻) 字母词语使用状况调查

在汉语中,字母词语的使用已有较长历史。近年来,字母词出现的数量越来越多,使用的频率也越来越高。字母词的使用给人们的交际带来了方便,同时也带来新的问题。究竟什么是字母词,哪些是字母词,哪些不是字母词,这些问题在学界还有争议;而字母词在语言运用中究竟占多大比例,都有哪些情况,至今还很少有人在大规模语料中作出量的测算和描述。为了掌握字母词语使用的实际情况,国家语言资源监测与研究中心在2006年语言生活状况调查中,对字母词语的使用状况作了调查统计。

一 调查对象的界定

字母词是由字母单独,或字母与汉字、数字、符号等一起构成的词。^①这里的字母包括汉语拼音字母。字母词来源复杂,形式多样,用途广泛,有典型字母词与非典型字母词之分。在典型字母词、非典型字母词与非字母词(主要指外语词)之间存在一定的中间地带,界限比较模糊,我们主要调查的是典型字母词,也顺带考察了非典型字母词以及带字母的分词单位的使用状况。

所谓典型字母词,指的是在两个方面都具备一定条件的字母词:

一是形式上要典型,最典型的字母词要符合四个条件:(1)与汉字组合;(2)读字母名称音;(3)形体大写;(4)缩略。如“CT机”,四个条件都具备,是形式上最典型的字母词。符合三个或两个条件的,是比较典型的字母词;只符合一个条件的,是不典型的字母词;一个条件也不符合的,不是字母词,如“copy”。

二是语言运用上要典型,即能否作为“语言词”。最典型的语言词要符合三个条件:(1)具有较强的概括性,指称的往往是一类事物,而非专指某一个事物;

^① 本文在字母词的界定和某些术语的使用上参考了邹玉华《现代汉语字母词研究》(中国传媒大学博士学位论文,2006年7月)和郭熙等《字母词使用规范方案研制报告(2006修订稿)》。

(2)具有一定的公众性,是使用该语言的大多数成员所掌握词语的一部分;(3)在公众传播中具有较高的使用频度。如字母词中“T恤衫、卡拉OK、IT”等,尽管在形式上可能不是最典型的,但符合这三个条件,是比较典型的语言词,也是典型的字母词;而像“粤K45468”这样的车牌号,尽管在形式上看符合典型字母词的条件,但不符合语言词的条件,因此不是典型的字母词。

另外,需要说明的是,典型字母词与非典型字母词并不是简单的二元对立,而是一个连续统。典型字母词不一定具备上面所有条件,但具备条件越多越典型。如“IC卡、VIP贵宾”,符合所有的条件,是最典型的字母词;而“NBA、WTO”,虽不符合“非专指某一事物”和“与汉字组合”这两个条件,但符合其他条件,因此也划归在典型字母词之内。

二 调查所使用的语料

调查语料源自国家语言资源监测语料库(包括平面媒体、有声媒体、网络媒体),共计1 311 749个文本文件,1 170 367 879字符次,978 994 406字次,596 080 949词次。其中报纸语料399 488 842字次,239 316 328词次;广播电视语料53 029 167字次,33 374 440词次;网络语料526 476 397字次,323 390 181词次。

三 调查方法

1. 首先使用中国科学院自动化研究所研制的分词标注软件将所有语料进行自动分词标注,然后用词频统计软件将所有的分词单位回收。

2. 在回收的词表中提取出所有包括字母(含大小写)的分词单位。首先对这些分词单位进行过滤,排除英文单词、网址、邮箱地址等非字母词,以及由于分词错误造成的非词单位(如“山西VS”、“全国卡拉OK厅”等),然后对所得结果进一步排查,根据形式和语言运用方面的条件区分典型字母词和非典型字母词,以典型字母词为核心形成2006年报纸、广播电视、网络(新闻)媒体字母词表。

3. 调查没有区别全角和半角的字母,在统计之前,统一处理作半角,然后进行统计。

4. 调查没有区分同形词与多义词,一个形式不论有几个意义,都作为一个

单位处理。如“HD”,既是高清晰度(即高保真)的意思,源自英文 high definition 的缩写,又是英文 hard disk 的缩写,表示“(计算机)硬盘”,在网络用语中还是“厚道”一词的汉语拼音缩写。尽管这三个意义之间毫无关系,我们在形式上也不加区分。

四 调查结果

在上述语料中,包含字母的分词单位有 238 019 条,占全部语料分词单位种数的 9%;这些分词单位共出现 4 889 904 次,占有所有语料总频次的 0.8%。

排除其中英文单词 77 945 条,396 454 次;网址 7337 条,53 168 次;邮箱地址 2 449 条,7 023 次;非词单位 10 467 条,42 705 次,得到的结果是:

字母词种数:139 821 条,占带字母分词单位总数的 58.74%,占全部语料分词单位种数的 5.3%。

字母词次数:4 390 554 次,占带字母分词单位出现总频次的 89.78%,占全部语料分词单位总次数的 0.7%。

其中包括人名、地名、机构名、火车车次、航班号、车牌号、车型号、手机型号、电视型号、各种仪器和武器型号、品牌名称等非典型字母词 138 202 条,占字母词种数的 98.84%。

典型字母词:1 619 条,占字母词词种数的 1.16%。典型字母词的具体数据见附表 6。

各种类型字母词的具体情况见表 2-1。

表 2-1 字母词类型表

类 型	词种数	所占比例(%)	词次数	所占比例(%)
人 名 ^①	31	0.02	91	0.002
地 名	8 469	6.05	31 940	0.73
机 构 名	27 535	19.69	28 064	0.64
列车车次、航班号、车牌号 以及各种代号、型号等	102 167	73.07	2 112 598	48.11
典型字母词	1 619	1.16	2 217 861	50.51
总 数	139 821	100.00	4 390 554	100.00

① 这里的人名仅指“阿 Q”、“阿瑟·M.塞克勒”这类的,纯英文人名不在其中。

就统计结果进行初步分析,可以看出:

1. 字母词已经成为汉语词汇的一部分

从上面统计数字可以看出,随着我国对外交往日益频繁,人们文化素质、外语水平的不断提高,也由于现代科技的发展,字母在电子媒介中使用的方便快捷,使用字母形式已不是我们语言生活中的偶然现象,字母构成的词语已成为汉语的一部分。带字母的分词单位主要可以分为两种情况:一种是英文词语,这属于语言交际中为了达到某种交际目的而使用的语码转换。这些英文词语并没有进入汉语词汇系统,在一定时期内也不可能大批进入汉语系统。汉语中高频使用的英文单词不多,本次调查中频次在 500 以上的有 187 条,具体数据见附表 7。还有一种是字母词,字母词语中,表示机构名、品牌名和各种代号、型号的言语词语占了大部分。但也应看到,一些典型的字母词已经进入汉语词汇系统,成为人们日常表达不可缺少的一部分,如“VS、NBA、GDP、AC、IT、MP3、QQ、AMD、DVD、CEO”等已进入前 5 000 高频词语行列。从总量上来看,字母构成的分词单位虽然种数较多,占全部语料分词单位种数的 9%,但使用频次很低,仅占全部语料总频次的 0.8%;典型字母词数量不多,仅占全部字母词的 1.16%,但使用频度比较高,占字母词使用总频次的 50.51%。

2. 字母在汉语中最经常、最主要的作用是指代、排序

在我们的语言生活中,经常需要给同类事物内部作一些临时性的区别,如一片楼房中需要给每一栋楼一个区别性的代号,企业需要给不同型号、不同性能的产品一个区别性的名称,飞机的航班、火车的车次、汽车的牌号,都需要一个区别性的代号。这种代号性的东西,没有很强的内涵意义,在具有区别性的前提下只求方便、简洁。而字母的无确定含义性,书写的简洁性,记忆的方便性和天然具有的位序性,使得它非常适合承担这个任务。“A 区、B 楼、C 座、记者 D”显然比“第一区、乙楼、第三座、第四个记者”更简便、更便于交际。因此,在需要表示区别意义的指代、排序等场合,字母和阿拉伯数字共同联手,在汉语的交际系统中起到了非常重要的、不可替代的作用。这种起指代、排序作用的字母词,在全部字母词种数中占 73.07%,在使用频次上占字母词总频次的 48.11%。

3. 典型字母词的造词方式主要是简缩

简缩是典型字母词的主要造词方式。字母词语表中简缩造成的字母词占 74%。简缩的方式主要有:

(1) 由一个外语词内部字母缩写而成(如“TV、kg”);

- (2)由一组外语词语的首字母缩写而成(如“CT、NBA”);
- (3)由原外语词语中前词的首字母和后词组合而成(如“E-mail”);
- (4)由汉语词的各个音节的首字母简缩而成(如“RMB、FB”);
- (5)由汉语原词语中各个词的首字母简缩而成(如“HSK、GB”);
- (6)由汉语词首字母和英文简缩组合而成(如“BTV、SDTV”。

除此之外,字母词的造成方式还有:字母表示顺序位次(如“A 楼、B 座”);字母表示类型或标志(如“B 股、A 型”);借字母的“形”表义(如“T 型台、O 形腿”)等。还有的字母词是综合多种手段而成(如“卡拉 OK”)。字母缩略形式加“汉字义标”(如“GPS 全球定位系统”)也是字母词构造的一种形式。^①

4. 异形字母词量大,使用频率不同

所谓异形,指的是意义完全相同而词形不同。字母词中存在着大量的异形现象,它们产生的主要原因是:

(1)字母大小写不确定,其中有些可能是由录入错误造成的,这里未加甄别。如“MP3”和“Mp3、mP3、mp3”,“SOHO”和“Soho、SoHo、soho”等。

(2)是否搭配汉字不确定,搭配什么汉字不确定。如“DVD”和“DVD 盘、DVD 光盘、DVD 碟、DVD 光碟、DVD 影碟、DVD 激光视盘”;“GPS”和“GPS 定位仪、GPS 定位系统、GPS 导航仪、GPS 导航系统、GPS 卫星定位系统、GPS 卫星导航系统、GPS 系统、GPS 全球定位系统”,等等,这样分别形成了以“DVD”、“GPS”为核心的一个词族。

(3)是否用连接符等不确定。如“HIP-HOP”和“HIPHOP”,“E-mail”和“Email”,“US”和“U. S.”等。

(4)是否省略不确定。如“Windows98”和“Win98”,“WindowsXP”和“XP”等。

这些异形词的使用频率并不相同,一般来说,由简缩而来的纯字母词往往是大写字母使用频率高,所以大写字母是这些字母词的基本形式。“MP3、SOHO”在使用中的具体数据见表 2-2。产品名称一般是首字母大写,其余字母小写,具体样例见表 2-3。某些计量单位、域名等是小写字母形式使用频率高,具体样例见表 2-4。

^① 参见沈孟璁《浅议字母词的入典问题》(《辞书研究》2001 年第 1 期),刘涌泉《关于汉语字母词的问题》(《语言文字应用》2002 年第 1 期),王吉辉《非汉字词语研究》(《南京师大学报》1996 年第 2 期)。

表 2-2 异形字母词使用频次样例表(1)

词 形	MP3	mp3	Mp3	mP3	总频次
频 次	20 827	578	130	2	21 537
比例(%)	96.7	2.7	0.6	0.01	100.00

词 形	SOHO	soho	Soho	soHo	总频次
频 次	1 384	42	19	1	1 446
比例(%)	95.7	2.9	1.3	0.07	100.00

表 2-3 异形字母词使用频次样例表(2)

词 形	Vista	VISTA	vista	Vlsta	ViSta	总频次
频 次	7 329	144	48	4	1	7 526
比例(%)	97.4	1.91	0.6	0.05	0.01	100.00

词 形	Windows	windows	WINDOWS	Wlndows	总频次
频 次	13 058	361	296	12	13 727
比例(%)	95.1	2.63	2.15	0.09	100.00

表 2-4 异形字母词使用频次样例表(3)

词 形	cm	CM	Cm	总频次
频 次	9 834	627	3	10 464
比例(%)	93.97	5.99	0.02	100.00

词 形	gov	GOV	Gov	总频次
频 次	292	33	2	327
比例(%)	89.29	10.09	0.61	100.00

某些字母词在使用中有同义叠加现象。本来字母已表达了一个相对完整的意义,但人们为了更明确地表义,还要在它后面加上汉语的释义,即“汉字义标”,构成一个词。如“VIP”,本来就是英文 very important person 的缩写,但在使用时,除字母形式外,又加上表示相同意义的汉字义标,说成“VIP 贵宾”。

纯字母形式与字母加汉字形式相比,前者形式简单,所以在使用中占了主导地位。比如在语料中,“DVD 视盘”和“DVD 激光视盘”都只出现 3 次,“DVD 压缩碟”7 次,“DVD 盘”32 次,“DVD 光碟”47 次,“DVD 碟”和“DVD 影碟”各 125

次,“DVD 光盘”339 次,而纯字母词“DVD”则出现了 15 357 次,占总数的 96%。当然,这也与该字母词的知晓度有关,如果是人们不太熟悉的,加汉字义标的比例就会高一些。如以“GPS”为核心的这一词族中,“GPS 卫星导航器”出现 6 次,“GPS 全球定位导航系统”14 次,“GPS 卫星导航系统”15 次,“GPS 定位仪”30 次,“GPS 导航仪”39 次,“GPS 全球卫星定位系统”51 次,“GPS 全球定位系统”77 次,“GPS 导航系统”81 次,“GPS 卫星定位系统”138 次,“GPS 定位系统”146 次,“GPS 系统”240 次,而“GPS”出现次数最多,5 177 次,占总数的 86%。

是否加连接符等符号,从统计数据来看,一般以加的居多。在总语料中,带连字符的各种形式的“HIP-HOP”共出现 1 105 次,而不带连字符的各种形式的“HIPHOP”只有 162 次,前者几乎是后者的 7 倍。“U. S.”出现 458 次,而“US”只有 292 次,前者是后者的 1.5 倍。在广播电视语料中,“E-mail”出现 59 次,而“Email、email”两种形式共出现 14 次,前者是后者的 4 倍多。但也有不同情况,“XO”出现 226 次,而“X. O”只有 17 次。

在 1 619 个字母词中,异形的有 461 个,占总数 28.5%。这类字母词多数是大写字母形式使用频次较高,如“NBA、IBM、CEO、MP3”等;也有少数是小写字母形式使用频率较高,如一些计量单位“cm、ml、kg”及表示域名的“mobi、com、org、gov、ca”等;不同形体使用频度相当的只是个别的,如“KHz”(50 次)和“kHz”(53 次)。

针对上述异形字母词的情况,宜根据使用频度及其他原则,确定一个基本形式,有意识地引导人们使用。

5. 同形字母词量大,有时会影响交际

同形字母词是指书写形式相同而意义不相干的字母词。造成“同形”的主要原因是缩略。字母词大多是英文缩写,本来是不同的英文单词,缩写后变成了相同的形式。如一个“PM”,至少有下列不同的来源和含义:

- ①下午。(英文 post meridiem 的缩写)
- ②产品市场。(英文 product marketing 的缩写)
- ③首相,总理。(英文 prime minister 的缩写)
- ④项目管理。(英文 project management 的缩写)
- ⑤项目经理。(英文 project manager 的缩写)
- ⑥排版软件。(英文 page maker 的缩写)
- ⑦拍马(网络用语)。(汉语拼音“pāi mǎ”的缩写)

一个“PS”,也至少有下列6个比较常用的意义:

- ①备注。(英文 postscript 的缩写)
- ②专业图像处理软件。(英文 Photoshop 的缩写)
- ③索尼公司的游戏机名称的简称。(英文 Play Station 的缩写)
- ④政治科学。(英文 political science 的缩写)

人们所熟悉的“ABC”,也至少有5个比较常用的意义:

- ①指一般常识或浅显的道理。
- ②美国广播公司。(英文 American Broadcasting Company 的缩写)
- ③澳大利亚广播公司。(英文 Australian Broadcasting Corporation 的缩写)

- ④中国农业银行。(英文 Agricultural Bank of China 的缩写)
- ⑤香蕉人,即美国出生的中国人。(英文 American-born Chinese 的缩写)

附表6字母词语表所收的1619个词中,同形字母词有288个,占17.8%。这种同形字母词区别性不强,如果语境限制不明确,往往会影响交际。

6. 字母词音、形、义之间关系复杂多样

字母词表面上根据书写形式容易判断,但其形、音、义之间关系复杂多样。主要有以下八种情况:

(1)读字母名称音,取字母所代表的英文单词的意义,这类字母词大多来源于英文缩写,是字母词的主体,如“NBA、WTO、DVD、GPS、VCD、OCR”。

(2)不读字母名称音,而是拼读,取字母所代表的英文单词的意义,这类字母词主要来源于英文缩写,如“SOHO、DINK、FAX、ROM、SARS”。

(3)一半读字母名称音,一半拼读,取字母所代表的英文单词的意义,这类字母词部分来源于英文缩写,部分是英文原词,如“E-mail、dBase”。

(4)有时读字母名称音,有时拼读,取字母所代表的英文单词的意义,这类字母词大多来源于英文缩写,如“SIM卡”。

(5)读字母名称音,取字母所代表的汉语单词的意义,这类字母词来源于汉语拼音缩写,如“HSK、WSK、PSC”。

(6)读字母名称音,取字母的形状意义、位序意义或替代意义,这类字母词大都是说汉语的人自造的,如“T型台、V字领、O形腿、B本、A股、D座”。

(7)写成字母,但读汉字音,取的是汉语单词的意义,这类字母词来源于汉语拼音缩写,大多在网络上使用,如“RMB(人民币)、LP(老婆)、FB(腐败)、MM(美

眉)、GG(哥哥)、JJ(姐姐)、JM(姐妹)、TJ(太监)、BT(变态)”。

(8)写成字母,读汉字音,但取的是字母所代表的英文单词的意义,这类字母词大多是计量单位的英文缩写和一些机构名称,如“ML、cm、HP”。

除这八种情况外,还有一些比较特殊的,如“BtoB”,“B”读字母名称音,是“Business”的缩写,也代表它的意义,但“to”是英文词,读拼读音,它是第一种和第二种的结合;而“B2B”则更复杂了,其中的“2”是“to”先谐音为“two”,再转写为阿拉伯数字“2”,因其书写方便,又有趣,所以十分流行,使用频次大大超过“BtoB”和“BTOB”。这些比较特殊的情况,在下表中归入“其他”。各类字母词的具体数据见表 2-5。

表 2-5 字母词形、音、义关系类型分布表

类型	1	2	3	4	5	6	7	8	其他
例词	NBA WTO	SARS SOHO	CDROM E-mail	SIM 卡 POS 机	HSK WJ	T 台 A 股	FB MM	HP ML	BtoB C2C
数量	1 154	137	24	13	8	163	35	38	47
比例(%)	71	8.5	1.5	0.8	0.5	10.1	2.2	2.3	2.9

7. 典型字母词主要用于专业术语

字母词大多用于专业术语,主要出现在电子、医学、汽车、经济、化学等领域。具体使用情况见表 2-6。

表 2-6 字母词使用领域统计表^①

类型	电子	医学	汽车	经济	化学	体育	机构	品牌	计量	其他
词数	503	152	90	82	76	39	198	35	50	394
比例(%)	31	9.4	5.6	5.1	4.7	2.4	12.2	2.2	3.1	24.3

8. 字母词的书写错误现象比较严重

字母词不是汉语固有的,加之字母只表音不表意,数量又少,区别特征不明显,书写时容易产生错误。如“POS”指销售点的结计算机或商场的电子收款机,“POSE”指姿势,但在语料中发现很多该用“POS”的误写成“POSE”。计算机上的“USB”接口,也有很多误写成“UBS”的。非典型性肺炎“SARS”误写成“SRAS”。全球定位系统“GPS”误写成“GPX”和“GPC、GSP”的,也不是个别现象。

① 本表同形字母词的使用领域,一般依其最常用的意义统计,不同的意义没有分别统计。

9. 带字母的分词单位在不同媒体中使用频度不同

带字母的分词单位,在报纸、广播电视、网络(新闻)三类媒体中的使用频度有所不同。具体数据见表 2-7。

表 2-7 带字母的分词单位在报纸、广播电视、网络(新闻)三类媒体中的使用状况

媒 体	总词次	带字母的分 词单位词次	所占比例 (%)	词种数	带字母的分词 单位种数	所占比例 (%)
报 纸	239 316 328	1 749 607	0.7	1 609 761	100 260	6.2
广播电视	33 374 440	32 759	0.1	289 497	5 489	2.0
网络(新闻)	323 390 181	3 106 042	0.96	1 518 952	172 631	11.4
全部语料	596 080 949	4 889 904	0.8	2 640 356	238 019	9.0

数据显示,广播电视语言在字母形式的使用上频度最低,种类最少。这是因为广播电视作为政府的喉舌,语言使用须要慎重,同时,传播形式决定了有些字母词只能用于书面,如“LP(老婆)、FB(腐败)”之类。报纸对字母词的使用也比较谨慎。网络语料中字母形式的词语出现最多,所占比例最高。

五 《报纸、广播电视、网络(新闻)字母词语表》说明

《报纸、广播电视、网络(新闻)字母词语表》见 048 页附表 6,有关情况说明如下。

1. 和汉字搭配造成的异形字母词在词语表中均以词条形式列出。字母大小写不同和是否加连接符等造成的异形字母词,在词语表中只以频次最高的词形为词条,其余在备注中给出。该词条的频次为所有形式之和。这样,原来的 1 619 条缩并为 1 377 条。

2. 考虑到字母词的实际使用,一些频次在 200 以上的高频机构名、品牌名(如“NBA、IBM、TCL”)也收入词语表。

3. 为帮助理解,词语表给出了提示性释义。限于篇幅,释义尽量从简,意义不求完备。释义参考了《现代汉语词典(第 5 版)》、《字母词词典》(刘涌泉主编)、《实用字母词词典》(沈孟璎主编)以及百度网站、Google 网站、金山词霸中的相关解释。

4. 受到自动分词技术的局限,某些词条的频次会高于该词条的实际使用次数。如“TL”,是“铁路”汉语拼音的缩写,在这一词条的频次中可能包含有产品型号“TL-T29、TL-WN250”中片段“TL”的出现次数。

六 《报纸、广播电视、网络(新闻)高频英文单词表》说明

《报纸、广播电视、网络(新闻)高频英文单词表》见 128 页附表 7,有关情况说明如下。

1. 本表收录频次在 500 以上的原始英文单词,不包括缩略形式。
2. 本表没有归并由于大小写等造成的异形。
3. 像“Word、in”这种既是英文单词,在汉语中又有特定含义的,在本表和《报纸、广播电视、网络(新闻)字母词语表》分别列出,但频次没有分别计算。