

有关中文大数词问题的缘由和建议

刘 群

(中国科学院计算技术研究所,北京 100080)

中图分类号: N04 01 文献标识码: A 文章编号: 1673-8578(2013)01-0032-06

一 背 景

2010年7月16日,计算机学会理事长李国杰院士给本人发来电子邮件,提议对有关大数词的问题进行讨论。邮件中提出:

计算机的速度已经超出了中文常用的计数范围,目前在关于超级计算机的新闻报道中常出现一些不规范的计数,如百亿亿次,百万亿次等,如果计算机速度达到 10^{24} 次方,真不知中文该如何表达。造成此困境的原因是万以后中国采用万进制,万,亿(万万),兆(万亿),京(万兆),下面是垓、秭、穰等,也有采用兆兆为京的“上法”。但在信息领域人们已习惯将百万称为兆,如兆赫兹、兆字节等,如果将百万亿次计算机称为百兆次计算机,很容易当成100M主频一样看待,即认为是亿次机。“万亿”变成没有单位名称,10只好读成“百万万亿”之类的奇怪数字。国际上都采用千进制,从K开始,M、G、T、P、E、Z、Y(10^{24} 次方),每个单位都有标准的译名,如兆、吉、太、拍、艾、泽、尧等。按国际标准,所谓“百万亿亿次计算机”应称为“艾次计算机”(10^{18} 次方)。但由于数字单位的标准读法没有统一,文献中很少人采用“艾”作数字单位。统一规范大数字的读法将越来越成为大众的

需求。中国是有几千年文明的古国,连十几位的大数都没有一个标准统一的读法,实在说不过去。

李院士也在多个不同场合提到汉语的大数词问题。

针对这一问题,2011年12月10日,全国科技名词委组织召开了一次“数词定名研讨会”,来自各个领域的专家就汉语数词定名问题展开了热烈的讨论,最后大致有四个方面基本达成一致:

(1) “兆”的用法不用改,数词用“百万”,词头用“兆”;

(2) 整数部分的数词不采用“三位进制”,遵从中文数词习惯;

(3) “ 10^{12} ”的数词究竟用什么,建议就用“万亿”,避免与“兆”冲突;

(4) 对于大于 10^8 的进制问题,采用八位进制,“亿亿”定为“京”,“亿京”定为“秭”。

此次会议所达成的共识是非常有积极意义的,对于目前大数词使用上的一些问题给出了基本合理的解决方案,既不会对现有的数词使用情况造成太大的变动,也对今后一段时间内在数词使用中的一些做法提出了合理的规范。不过,从长远看,这仍是一个过渡性的方案,一些根本性的矛盾仍然没有解决。本文希望对这一问题进行更深入的探讨,

收稿日期: 2012-12-26

作者简介: 刘群(1966—),男,中国科学院计算技术研究所研究员,计算机科学技术名词审定委员会副主任,研究方向为自然语言处理、机器翻译、信息提取。通信方式: liuqun@ict.ac.cn。

并试图给出一个合理的解决办法。

二 中文大数词问题的历史渊源与现状

大数词应用在计算机速度方面的困扰主要涉及以下几个问题:

(1) 计算机速度的表示,是应该采用中国传统的万进制表示(万、亿、兆、京、垓、秭、穰[读作rǎng]等),还是应该采用国际单位制表示(兆、吉、太、拍、艾、泽、尧等)?

(2) 如果采用中国传统的万进制表示,“亿”以上的单位是否应该采用“兆、京、垓、秭、穰”等数词表示方法,而不是现在常用的“百亿”“万亿”“百万万亿”之类的表示方法?

(3) 如果采用中国传统的万进制表示,其中“兆”的含义为“万亿”,与目前采用国际单位制表示的“兆”(含义为“百万”)不同,这个矛盾如何解决?

为了更好地了解这些问题,有必要追溯一下现代汉语中所使用的大数词的历史渊源,理清目前这些复杂问题表象后面的来龙去脉。

1. 《国语》《数术记遗》和《五经算术》所记载的中国古代数词体系(下数、中数、上数)

《国语》是记载中国西周时期王室和一些诸侯国历史的史书。《国语·郑语》第十六记载“计亿事,材兆物,收经入,行垓极”,这里顺序用到了“亿、兆、经(通‘京’)、垓”这几个数词,但应该都是概数,表示很多的意思,并不是具体的数字。不管如何,这至少说明中国在西周时期就有“亿、兆、经(后写作‘京’)、垓”这样的计数单位了。三国时期吴国学者韦昭在注解《国语·郑语》时,对这一句加注称“计算也;材,裁也。贾唐说皆以万万为亿,郑后司农云:十万曰亿,十亿曰兆,从古数也。”可见这时对“亿”已经有“万万”和“十万”两种解释了^[1]。

《数术记遗》^[1-3]和《五经算术》^[1-3]对大数记法的表述是一致的:

按黄帝为法,数有十等。及其用也,乃有三焉。十等者,谓“亿、兆、京、垓、秭、壤、沟、涧、正、载”也。三等者,谓“上、中、下”也。其下数者,十十变之。若言十万曰亿,十亿曰兆,十兆曰京也。中数者,万万变之。若言万万曰亿,万万亿曰兆,万万兆曰京也。上数者,数穷则变。若言万万曰亿,亿亿曰兆,兆兆曰京也。

具体解释就是:万以上的大数,自古有十个等级,分别是“亿、兆、京、垓、秭、壤、沟、涧、正、载”。而这十个等级的具体含义,有三种进位体系,分别叫作下数、中数、上数:

(1) 下数体系:采用一位进制(十进制),即每一个等级相差十倍,即十万为亿(10^5)、十亿为兆(10^6)、十兆为京(10^7)、以此类推;

(2) 中数体系:采用八位进制(万万进制),即每一个等级相差一万万倍,即万万为亿(10^8)、万万亿为兆(10^{16})、万万兆为京(10^{24})、以此类推;

(3) 上数体系:采用平方进制,每一个等级是上一个等级的平方,即万万为亿(10^8)、亿亿为兆(10^{16})、兆兆为京(10^{32})、以此类推。

现代汉语中的大数词基本都来自这十个等级中的数词,但这三种计数体系目前都已经不再被采用。

2. 华严经的数词体系

佛教重要经典《华严经》^[2]中,给出了一个严格定义的大数体系,其中每一个数词对应于序列 $10^{7 \times 2^n}$ 中的一个值, n 的取值从0一直到123,各个等级之间采用平方进制。每个数都有明确的名称,最小的是“俱胝”(10^7),最大的数是 $10^{7 \times 2^{123}}$,这个数被认为是佛所能知道的数的极限,其数值之大,即使是现代科学中似乎也还没有哪个实际的物理量达到了这个数量级别。当年佛教的思想家们对纯数字的理解和想象力确实令人惊叹。

佛教对中国文化产生了重大的影响,包括对计数的认识。前面提到《数术记遗》中“上数体系”的平方进制就被认为是借鉴了《华严经》的数词体系。《数术记遗》中的很多术语也来自于佛教经

书。《数理精蕴》的数词体系也有很多数词来源于佛教。

3. 清代《数理精蕴》的数词体系

《数理精蕴》^[1-2] 是中国清朝康熙雍正年间由清政府组织编纂的《律历渊源》的第二部分。《数理精蕴》中有一套专用数字名称: 个、十、百、千、万、亿、兆、京、垓、秭、穰、沟、涧、正、载、极、恒沙河、阿僧祇、那由他、不可思议、无量数, 其中个至万为十进制, 从万以上采用万进制。比个位小的数, 按十分之一递减, 依次为分、厘、毫、丝、忽、微、纤、渺、漠、模糊、逡巡、须臾、瞬息、弹指、刹那、六德、虚空、清静。

《数理精蕴》的数词体系保持了古代数词体系中的十个大数词“亿、兆、京、垓、秭、壤、沟、涧、正、载”并且保持了原来的等级顺序。但该体系没有采用古代的下数、中数和上数体系, 而是采用了新的万进制。这个数词体系还增加了“极、恒沙河、阿僧祇、那由他、不可思议、无量数”这些个来自于佛教的大数词, 另外还增加了比个位数小的数词体系, 按十分之一递减。一些极小数词的名称同样来源于佛教。

《数理精蕴》的数词体系与现代汉语的数词体系是基本一致的, 其中“个、十、百、千、万、亿”和“分、厘、毫、微”等数词现在仍然在使用, 而且与现在使用的含义完全一致。“兆”这个数词的使用后来出现了分歧。可以说, 《数理精蕴》为现代汉语数词体系的建立奠定了基础。

4. 英文的数词体系

英文的数词体系与中文不同, 通常采用三位进制或六位进制, 如表 1 所示。

简单说来, 就是英文中采用的数词等级依次是: million、billion、trillion 延伸下去, 一直到 centillion。英文数词分为英美两种体系, 这两种体系中, 所采用的数词等级顺序相同, 但进制不同。两种体系中 million 的含义都是一百万, 但其他数词的含义不同。美国体系采用的是三位进制(千进制), 也就是相邻等级的两个数词之间是相差一千倍, 而

英国体系采用六位进制(百万进制), 相邻等级的两个数词之间相差一百万倍。

表 1 英文数词英美不同含义

英文数词	美国含义	英国含义
million	10^6	10^6
billion	10^9	10^{12}
trillion	10^{12}	10^{18}
quadrillion	10^{15}	10^{24}
quintillion	10^{18}	10^{30}
...
centillion	10^{303}	10^{600}

在现在的科学技术论文中, 数词的使用通常都采用美国的计数体系。

5. 民国时期中央研究院的第一次院士会议确定的数词体系

1948 年 9 月在南京举行的中央研究院第一次院士会议提出了度量衡中的大数计量问题的提案^[4]。

现在看来, 当时的院士们在大数词定义的问题上是比较激进的, 试图改变长期以来汉语中的万进制体系, 而采用西方的千进制体系。其基本想法是:

- (1) 弃用“亿”;
- (2) 定“兆”为百万;
- (3) 从“兆”开始, “京”“垓”采用千进制体系(三位进制体系)。

总体上, 弃用“亿”, 保持“兆”“京”“垓”的顺序, 与中国古代的数词体系的顺序基本上还是相容的。如果这套方案能够在全国强制推行下去也不失为一个可行的做法, 未必不会形成新的中文数词标准。不过, 这套方案仅仅作为院士大会的一套提案提出, 并没有以政府法律或者规范的形式加以发布并实施。因此这套方案总体上对中国的数词使用并没有造成很大影响。

6. 国际单位制(SI)规定的英文词头

在英文中, 除了有一套大数词体系以外, 还有一套在科学技术中经常使用的数词词头体系^[1-2],

这两套体系的用词是不同的。比如说,表示一千的数词是 thousand,而词头是 kilo-,表示一百万的数词是 million,而词头是 mega-,表示十分之一的数词是 tenth,而词头是 deci-。

所谓词头,就是词语的前缀。前缀是不能单独作为词语出现的,数词词头要附加在表示科学技术单位的名词前面,合起来构成一个新的单位名词。数词词头大部分来自希腊语或者拉丁语。

在国际单位制中,对表示科学技术单位的前缀(词头)进行了明确的定义。整数单位词头依次是 deca-(十)、hecto-(百)、kilo-(千)、myria-(万)、mega-(百万)、giga-(十亿)、tera-peta-exa-zetta-和 yotta-,从 mega-开始,均采用三位进制(千进制),下一个词头是上一个的一千倍。分数词头分别是: deci-(十分之一)、centi-(百分之一)、milli-(千分之一)、myrio-(万分之一)、micro-(百万分之一)、nano-pico-femto-atto-zepto-yocto-,从 micro-开始,逆向采用三位进制(千进制),下一个词头是上一个词头的千分之一。

这些词条进入国际单位制的时间是不同的,大部分词头在 1960 年已经被确定进入国际单位制,1964 年确定 femto-和 atto-两个词头进入国际单位制,另外 peta-和 eta-两个词头是 1975 年正式确定进入国际单位制的,还有 zetta-yotta-zepto-和 yocto-四个词头是 1991 年才被正式确定进入国际单位制的。

在英文中,数词体系和数词词头体系的使用并行不悖,互不干扰。这是因为不仅它们的用词不同,在构词上所起的作用也完全不同。二者又都采用三位进制,存在单纯的一一对应关系,互相转换非常方便,没有任何问题。

但在汉语中,由于没有“词缀”的概念,数词和数词词头并没有明确的界限,这就给汉语的数词体系带来了新的困扰。

7. 《中华人民共和国法定计量单位》规定用于构成十进倍数和分数单位的词头

1984 年公布的《中华人民共和国法定计量单位》确定了以国际单位制单位为基础的我国法定计量单位,规定了“用于构成十进倍数和分数单位

的词头”。其中整数单位依次是:十(deca-)、百(hecto-)、千(kilo-)、兆(mega-)、吉[咖](giga-)、太[拉](tera-)、拍[它](peta-)和艾[可萨](exa-);分数词头依次是:分(deci-)、厘(centi-)、毫(milli-)、微(micro-)、纳[诺](nano-)、皮[可](pico-)、飞[母托](femto-)和阿[托](atto-)。还明确规定:“ 10^4 称为万, 10^8 称为亿, 10^{12} 称为万亿,这类数词的使用不受词头名称的影响,但不应与词头混淆。”1986 年中华人民共和国国家标准 GB 3100—1986《国际单位制及其应用》^[5] 正式发布,其中包括了有关上述词头的规定,明确了上述词头的法律地位。1993 年,GB 3100—1993《国际单位制及其应用》^[6] 增加了泽[它](zetta-)、尧[它](yotta-)、仄[普托](zepto-)、幺[科托](yocto-)四个新词头。

“十、百、千、分、厘、毫、微”这几个词头与《数理精蕴》中定义的数词含义是完全相同的,而除了这些词头和“兆”以外的其他所有词头,都是从英文国际单位制(SI)词头翻译过来,以前在汉语中也并没有作为数词使用,因此也不会给我们带来太大的困扰。唯一的麻烦就是“兆”的定义。把“兆”定义为“百万”,不仅仅与已经被广泛接受的《数理精蕴》数词体系不符的问题。最不合理的地方在于,当时我国的主流媒体与日常生活中已经广泛使用的数词“亿”来表示“万万”的含义。在这种情况下,同时又定义“兆”为“百万”,严重违反我国自古以来非常稳定的大数词等级顺序,也与目前海外华人地区或者使用汉字的地区(台湾、新加坡、日本和韩国)不符。这是我国目前数词体系使用上的最大困扰^①。

随着科学技术的迅速发展,《中华人民共和国法定计量单位》中定义的很多词头已经被广泛使用。比如作为“纳米”科技的“纳”也已经被百姓所熟悉。同样,“兆”作为词头,在物理领域有“兆赫兹”表示频率、“兆瓦”表示功率,计算机领域有“兆字节”表示存储容量,等等。由于各种电子设备和计算机设备的广泛使用,作为词头的“兆”的使用范围已经不仅仅限于计算机科学,而是进入了人们的日常生活。由此可知,虽然“兆”定义为表示“百万”的词头是非常不合理的,但如果要对“兆”的含

义重新定义,不仅将会给人们的生活造成极大的混乱与不便,而且社会成本和经济成本也都是难以接受的。

8. 台湾、新加坡、日本和韩国使用的数词体系

中国台湾地区也采用国际单位制,与《中华人民共和国法定计量单位》的规定大部分相同,主要的区别在于:

(1) mega-在大陆定名为“兆”,在台湾定名为“百万”;

(2) tera-在大陆定名为“太[拉]”,在台湾定名为“兆”;

(3) 除了上述两个词头以外,台湾地区定义的其他词头基本与 GB 3100—1986 保持了一致,并且为四个新进入国际单位制的词头确定了台湾地区所使用的中文词头名:皆(zetta-)、佑(yotta-)、介(zepto-)、攸(yocto-);

(4) 在台湾地区,一些中文常用数词,如“亿”“万”“百分”“百万分”“十亿分”也被纳入了可以使用的范围。

可以看出,台湾地区定义的数词词头体系除了“兆”以外,与大陆地区基本上没有什么不同,“兆”是唯一的麻烦。

在新加坡、日本和韩国,“兆”的含义都是“万亿”,而不是“百万”。在这些国家,“兆”最常见于货币单位“兆元”,指的是“一万亿元”。国内新闻界在引用海外新闻的时候,也经常直接采用它们的数词,把“兆”理解成万亿。2011年,日本富士通与日本理化学研究所共同开发的超级电脑被命名为“京”,也是取其速度达到了每秒1亿亿次,与《数理精蕴》中对数词“京”的规定相符。另外,新加坡受英文影响较大,也经常把“万”称为“十千”,“十万”称为“百千”。

三 关于中国数词使用规范化的建议

1. 关于数词与数词词头两套体系并存的建议

严格说来,在汉语中是没有“数词词头”的说的。所谓“词头”就是词的“前缀”。在汉语中,前缀是很有限的,如“老王”“老李”的“老”,“非理性”“非现实主义”中的“非”才是词头,而汉语中并

没有使用数词作为词头的传统。

但由于在科学技术领域,国际单位制(SI)词头已经成为被广泛接受的标准,数词词头作为该标准的一部分,在中文里面如果拒绝使用,会给人们带来非常大的不便。比如“21km”要翻译成“2万1千米”,“233Kbyte”要翻译成“23万3千字节”,不仅转换麻烦,与国际交流也很不方便。

而在日常生活中的数词采用万进制的做法,目前也很难改变,主要是“亿”的用法和含义已经被广泛接受,要废弃“亿”而采用其他数词是非常困难的。

为此,笔者建议如下:

(1) 在中文中保留两套数词体系,采用国际单位制(SI)的数词词头标准并翻译成对应的中文数词词头,日常生活中的大数词(京、垓以上)最好也以国家标准的形式确定下来。

(2) 两套数词体系可以并行,前提是没有冲突,即不允许同一个数词在两套体系中表示不同的含义;但有时同一个数值可以用两套体系分别表示,两种做法应该都是可以接受的。比如计算机速度是每秒1京次,也可表示为每秒10拍次,具体使用哪个用法取决于具体应用场合和老百姓的接受程度。

(3) 两套数词体系的使用场合尽量加以区分,数词词头体系只在科学技术领域以及与国际交流的必要场合使用,其他场合都使用传统的数词体系。

(4) 数词词头严格限制只能与单位名词结合使用,不单独使用。比如可以说硬盘容量是“50太字节”,但不要说是“50太”。计算机运算速度可以说是每秒“1艾次”,或者说是“艾次级计算机”,但不要说每秒“1艾”,或者“艾级计算机”。单独说数词的时候,只能说“20亿”和“3000万亿”,不能说成是“2吉”和“3拍”。

2. 关于数词“兆”的建议

由于“兆”作为“百万”的含义目前已经被广泛接受,马上要改变这个数词的含义是不现实的,不仅会给社会带来更大的混乱和困扰,也会导致巨大的社会成本和经济成本。但长远看,“兆”解释为

“百万”是不合理的,必须改变。为此,笔者提出以下建议:

(1) 目前情况下,尽量限制“兆”的使用,尤其不要扩大“兆”的使用范围。目前,“兆”的使用虽然很多,但实际上仅限于与一些科技单位名词联合使用,比如“兆字节”“兆赫兹”“兆瓦”等。在这些场合还可以使用“兆”,但尽量不要扩大使用范围,尤其是在日常生活中,不再把“兆”与其他单位名词联合使用,更不要把“兆”作为日常生活中的数词使用,比如表示房价的时候绝对不要用“二兆元人民币”表示“二百万元人民币”;

(2) 在时机成熟的时候,修订《中华人民共和国法定计量单位》取消“兆”作为表示“百万”词头的含义,但并不要给“兆”赋予任何新的含义。这时表示“百万”有两种做法:一种是像台湾目前的做法那样,不另定义词头,就用“百万”表示,第二种做法是另外定义一个表示“百万”含义的词头,建议可以跟其他词头一样,采用音译法,比如说,把“mega-”翻译成“迈”。笔者建议采用第二种做法,因为这样可以加快“兆(百万)”作为词头的退出速度,否则很多人为简短起见会一直使用使用“兆(百万)”作为词头,使得该词头很难真正退出日常使用。

(3) 在全国范围内大力推广修订后的《中华人民共和国法定计量单位》,使得经过一段较长时间的过渡后,“兆(百万)”作为词头的用法最终退出人们的日常生活;同时也可以避免因为给“兆”赋予新的含义(万亿)而造成混乱。

(4) 等到“兆(百万)”作为词头的用法完全退出人们的日常生活以后,最终是否要通过法律或者规范形式重新确定“兆”作为“万亿”的含义,笔者并不坚持,可以留待后人去讨论。

要改变一个已经被广为接受的词的含义是非常困难的,如果没有特殊理由,也不应该去尝试做这种努力。但目前数词“兆”的用法确实非常不合理,并且给人们带来了太多的困扰,笔者认为是需要

去改变的。但这种改变只能是渐进的。通过有意识的、循序渐进的微调,经过十余年甚至数十年的过渡,我认为是有可能把这个问题解决的。

3. 关于采用四位进制还是八位进制的建议

对于大于 10^8 的进制问题,这次“数词定名研讨会”上建议采用八位进制,“亿亿”定为“京”,“亿京”定为“秭”。

这个建议基本上还是与《数理精蕴》定义的数词体系相容的,只是把原来的四位进制(万进制)改成了八位进制,取消了原来“兆”和“垓”两个数词。

采用四位进制还是八位进制理论上都是可以接受的,使用上也不至于造成歧义,二者转换也很方便。但从遵从历史惯例的角度说,我觉得还是采用《数理精蕴》的四位进制更好一些,尽量减少不必要的变动。

注释

- ① 中华人民共和国法定计量单位,载于《中华人民共和国国务院公报》1984年4月1日第四号(总号:425)第135至138页,注9明确说“ 10^4 称为万, 10^8 称为亿, 10^{12} 称为万亿,这类数词的使用不受词头名称的影响,但不应与词头混淆。”可见当时亿和万亿的用法是得到承认的。

参考文献

- [1] 百度百科. [EB/OL]. [2012-12-25]. <http://baike.baidu.com/>.
- [2] 维基百科. [EB/OL]. [2012-12-25]. <http://www.wikipedia.org/>.
- [3] 互动百科. [EB/OL]. [2012-12-25]. <http://www.baik.com/>.
- [4] 郭金海. 中央研究院的第一次院士会议[J]. 中国科技史杂志, 2007, 28(1): 1-19.
- [5] 全国量和单位标准化技术委员会. GB3100—86 国际单位制及其应用[S].
- [6] 全国量和单位标准化技术委员会. GB3100—93 国际单位制及其应用[S].