

谈谈汉字整理工作中可以 参考的某些历史经验

裘锡圭

在汉字发展演变的过程里，有很多值得我们在今后的汉字整理工作中认真加以参考的历史经验。对这一问题作全面论述，不是这篇短文所能胜任的，下面只准备简单谈谈字数和字形变化这两方面的历史经验。

先谈字数。一般都认为汉字是不断增多的，而且增加的数量相当可观。就大型字典收字的情况来看，确实是这样。例如：公元2世纪编纂的《说文解字》收字9353个，加上重文共10516个。本世纪70年代开始编纂，去年已出版了第一卷的《汉语大字典》，全书收字56000左右，是《说文》字数的五倍多。但是大型字典里收了很多实际上已经死亡的字，如果以各个时代一般使用的汉字为考察对象，就会发现汉字的数量其实是相当稳定的。

我们所说的一般使用的汉字，就是除去很生僻的字和比较专门的专业用字以后的实际使用的汉字，跟有些人所说的“通用汉字”大致相当。在具体划定一般使用的汉字的范围的时候，人们很难取得完全一致的意见。所以即使是统计现在一般使用的汉字的数量，也很难得出一个公认的精确数字。统计历史上一般使用的汉字的数量，困难当然就更大了。不过在这方面作一些比较笼统的估计，还是有可能的。

由于掌握的资料太少，我们只能考察一下汉字发展史的上古和现代这两头的情况。先看现代汉字的情况。1965年1月由文化部和文字改革委员会联合发布的《印刷通用汉字字形表》收字6196个。1981年国家标准局发布的《信息交换用汉字编码字符集·基本集》收字6763个（其中比较常用的一级字3755个）。据郑林曦先生统计，现代书报刊物的用字数为6335个（郑林曦《精简汉字字数的理论和实践》，1981年）。据另一种统计，掌握了3800个汉字，就能阅读一般书刊内容的99.9%左右；掌握了5200个汉字，就能阅读一般书刊内容的99.99%左右（陈明远《数理统计在汉语研究中的应用》，《中国语文》1981年6期468页）。又据香港的一个统计，4份中文报在45天内所用字数为4687个（陈重瑜《试评汉字简化的一些论说》，《文字改革》1985年4期13页）。此外，从汉字构词能力的角度来看，

据中国人民大学语言文字研究所统计, 4990个汉字就“构成了《现代汉语词典》中的几乎所有词”(据上引陈明远文469页)。从上面举出的一些数据来看, 大致可以说, 现在一般使用的汉字的数量是五六千左右(一般知识分子个人掌握的字数只不过两三千左右)。

接着来看一下上古汉字的情况。在已发现的内容比较丰富的古文字资料里, 商代后期(约公元前14~前11世纪)的殷墟甲骨文字是时代最早的一批资料。出土的殷墟有字甲骨已达15万片左右(胡厚宣《八十五年来甲骨文材料之再统计》,《史学月刊》1984年5期), 总字数当以百万计, 大体上应该能够反映商代后期一般使用文字的情况。1965年出版的《甲骨文编》根据当时已出版的40种甲骨著录书的资料, 整理出单字4672个。此书《编辑序言》认为“其中有些字还可以归并; 目前甲骨刻辞中所见到的全部单字的总数, 约在四千五百字左右”(引者按:《甲骨文编》中也有应分而误合之字, 但数量不多)。于省吾先生在1973年发表的《关于古文字研究的若干问题》一文中也说:“截至现在为止, 甲骨文不重复的字约共四千五百多个。”(《文物》1973年2期33页)在此后新发表的甲骨文资料里, 出现了一些《甲骨文编》所未收的新字, 但数量不是很多。如果把商代后期一般使用的汉字数量估计为5000左右, 大概离事实不会很远。周代遗留下来的主要文献是十三经。据《十三经集字》统计, 十三经共用单字6544个(据钱玄《秦汉帛书简牍中的通假字》转引, 见《南京师院学报(社科版)》1980年3期47页)。十三经由于后世的传抄刊刻, 用字情况已不完全是原貌, 但是在单字总数方面大体上总还是能够反映周代的情况的。十三经的时代延伸得比较长, 几乎包括了整个两周时代, 而且其中还有一些秦汉时代的东西。如果有可能以周代某一段一二百年的时间范围为范围来进行统计, 当时一般使用的字数大概要比十三经使用的字数低一些。

从上述情况来看, 从商代到现代, 一般使用的汉字的数量似乎并没有显著的变化, 很可能一直在五六千左右。在这三千多年里, 新字在不断产生, 但是旧字也在不断退出历史舞台。二者相抵, 字数的变化就不大了。一般说来, 时代越晚, 新词增加得越快。但是由于复合词在汉语里越来越占优势, 音译外来词又大多数用假借字记录, 所以需要为它们造新字的新词却并不是越来越多。这是一般使用的汉字的数量有可能比较稳定的重要原因。

汉字使用意符、音符和记号三种字符(字符专指文字所使用的符号, 记号指跟文字所代表的词在语音和意义上都没有联系的字符, 参看拙文《汉字的性质》, 见《中国语文》1985年1期36页), 跟拼音文字相比, 的确有结构复杂、不易记忆的特点。这种性质的文字体系, 单字如果太多, 使用的人就会无法掌握。所以汉字里常常用假借等方法, 让一个字在不同场合代表不同的语素, 也就是用一个字代表两个以上不同的语素(对双音节或多音节语素来说, 则是语素的一个音节)。但是如果单字太少, 让一个字承担过多的任务, 或是让一个字代表两个意义容易混淆的语素, 又会影响记录语言的明确性, 使用起来也会发生困难。在汉字发展的过程里, 几乎始终同时存在文字分化和文字合并这两种相反相成的现象。前者如由“师”分化出“狮”(“狮子”在古代本作“师子”), 由“陈”分化出“阵”(军阵之“阵”古本作“陈”)等等。从汉字历史上看, 新字大部分是分化字。后者如顺治之“汧”并入迎逆之“逆”, 开耑之“耑”并入端正之“端”, 又如汉字简化中把“鬥”并入“斗”等等。进行合并是为了控制单字的数量, 进行分化是为了加强记录语言的明确性。字太多了固然不行, 字太少了也同样不行。汉字里一般使用的字数从古到今变化不大, 显然不是一个偶然的现象。

建国后，有少数语文工作者主张通过以同音借代为主的合并文字的办法，大量压缩通用汉字的数量，有的人甚至想把通用汉字砍掉一半以上。从3000多年的汉字发展史来看，这种办法显然是行不通的。如果方块汉字还要长期使用下去的话，只允许汉字合并而绝对不允许进行新的文字分化，恐怕也是做不到的。

其次谈字形变化。这里所说的字形变化是广义的，既包括某一个字的写法的变化，也包括某一个语素前后使用不同的字，即原来用甲字表示后来改用乙字表示的那种现象。斗争的“斗”这个字本来用“鬥”字，由于汉字简化改用“斗”字，就是后一种现象的一个例子。

一般都认为历史上汉字字形变化的主要趋势是简化。这应该说是正确的。但是在这个问题上也存在一些不够全面的认识。

有的人把字形简化跟减少笔画等同起来。这跟历史事实有出入。

在汉字发展史上，由古文字演变为隶书，是最重要的一次字形简化。这次简化使带有图绘性质的象形的字符变成用点、横、撇、捺等笔画组成的字符，大大方便了书写。由于这次简化的主要目的是改造那些需要描画的不便于书写的形体，所以并不斤斤计较笔画的多少。有些字经过隶书的改造，笔画甚至反而增多了。如“日”字的外框在古文字里是可以一笔画成的圈圈，在隶书里却分解成了三笔（竖、折、横）或四笔（竖、横、竖、横）。但是隶书的“日”书写起来的确比古文字的“日”方便，所以我们还得承认由古文字的“日”变为隶书的“日”是简化。

为了摆脱古文字的象形性质以便书写，隶书不得不破坏了一部分字的结构。经过隶书的改造，独体表意字（大致相当于一般所谓象形字）的字形大都丧失了表意作用，它们成了所谓“不象形的象形字”，用我们的话来说就是成了记号字。一部分图绘意味比较浓厚的合体表意字的构造也遭到了破坏。例如象二人并立之形的字，变成了字形毫无意义可言的记号字“並”（现已并入“并”字）。但是由常见的成字的偏旁所组成的合体表意字和形声字的结构，绝大多数没有遭到破坏。尤其是形旁的表意作用和声旁的表音作用都很明确的那些形声字，它们的结构几乎全都没有遭到破坏。仅仅为了减少一些笔画而破坏文字结构的情况，在古文字演变为隶书的过程里是并不多见的。所以从历史上看，不但不能把字形简化跟减少笔画等同起来，甚至能否把减少笔画看作字形简化的主要内容都还是个问题。

还有人由于过分强调简化的重要性，而忽视了对汉字的发展起了正面作用的大量字形繁化现象的存在，至少对这种现象的重要性估计不足。

在汉字发展的过程里，一方面存在为了书写的方便而简化字形，甚至不惜因此破坏文字结构的现象，另一方面也存在为了改进文字的功能，主要是字形的表意、表音功能，不惜使字形繁化的现象。

形声化是汉字发展的一个极为重要的标志，而形声化就往往是跟字形的繁化联系在一起的。前面举过的“狮”字就是一个例子。假借字“师子”的“师”加上“犬”旁，成为形声字“狮”。对于“狮”这个词来说，它的字形就繁化了。这类例子是极其常见的（被加偏旁的字不限于假借字）。有些表意字后来被改成了字形较繁的形声字。例如象坎穴形的字，后来就改成了从“土”“欠”声的“坎”字。有时为了改善假借字或形声字声旁的功能，也不惜改用笔画较多的字。例如：表示古汉语里主要的第二人称代词的假借字，本来用“女”字，后来改用表音更确切，并且其原来字义跟假借义不会发生混淆的“汝”字。形声字“裤”本来写作“袴”，后来声旁“夸”为表音更确切的“库”所取代。“裤”字《康熙

字典》未收，大概出现得很晚，从传统文字学的观点看是一个“俗字”。“库”字未简化时比“夸”多四笔。但“夸”、“裤”不同音，“库”、“裤”同音，所以大家宁愿多写几笔，接受了这个俗字。

此外，还存在仅仅为了避免跟其他字的字形相混淆而繁化某个字的字形的现象。例如“上”、“下”二字本来写作“”、“”，为了避免相互混淆，并避免与一二之“二”相混，后来各加一竖而写作“上”、“下”。“四”字古作“”，改作“四”也是为了避免跟“一”、“二”、“三”相混。这跟数字大写的用意相同（“四”用作数字，可能是假借用法，大写数字也大都是假借字）。

上面所举的那些字形繁化的现象，大多数起了改善字形的表意、表音功能的作用。当然，改善字形的表意、表音功能，也并非一定要繁化字形。简化字形跟改善字形的表意、表音功能也有相一致的情况。例如改“孀”为“奶”，既是简化，又改善了声旁的表音功能。但是在许多情况下，简化字形对字形的表意、表音功能是起破坏作用的。汉字里的记号字大部分是字形简化的产物。这是为了书写方便不得不付出的代价。

从历史上看，对于汉字的发展，字形的简化是不可缺少的，改善文字功能的字形繁化也同样是不可缺少的。只重视前者的作用而不重视后者的作用，是不妥当的。如果古人把字形的表意、表音功能放在无足轻重的地位，绝对排斥字形繁化，一味追求字形简化，汉字里的记号字就会大大增多。记号字的字形跟字音、字义都没有联系，掌握起来要比掌握合乎规律的形声字和表意字困难得多。在我们现在使用的汉字里，记号字已经占有不算小的比例了。如果这种字大大增多的话，汉字势必成为非常难掌握，甚至根本无法掌握的一种文字。

最后还应该指出，简化字形的重要性在古代和现代是有所不同的。古人为了把象形的古文字简化成隶书而破坏一部分字的结构，是不得已的，也是值得的。今天，在比隶书更便于书写的楷书早已形成的情况下，仅仅为了减少一些笔画而去破坏某个字的结构，或打乱某一组字的系统性（如让有些从“卢”声的字从“户”），这样做究竟是不是必要，是不是值得，就大可怀疑了。有的人援引隶书改造古文字的例子，作为这类简化的历史依据。这是不合适的。在不太远的将来，随着电脑和复印机等设备的普及，书写文字的负担肯定会大大减轻。今后的汉字整理工作，究竟应该把重点放在简化上，还是放在文字结构的合理化上，这恐怕是一个需要认真考虑的问题。

自下而上的倒叠字序

· 华 明 ·

直行排列的文字，有的字序不是自上而下，而是自下而上，叫做“倒叠序列”。主要的例子是苏门答拉过去的巴塔文。巴塔文从最左一行的最低地位写起，把字母一个个向上堆叠，达到顶点以后，再写第二行，字序自下而上，行序自左而右。

这种书写序列是怎样产生的呢？原来早

期用狭长的竹片横放着书写，字序自左而右，行序自上而下。写好以后，用绳子把竹片连结起来，成为一册竹书。阅读时，两手向左右把竹书翻开，所有的竹片都直立了起来，字母一个个侧身横卧，成为自下而上的字序、自左而右的行序了。久而久之他们就习惯于“倒叠序列”了。