百种语文小丛书 曹先擢 主编

汉字与计算机

HANZI YU JISUANJI

顾小凤 著

语文出版社
http // www. ywcbs. com

百种语文小丛书 曹先擢 主编

汉字与计算机

HANZI YU JISUANJI

顾小凤 著

语文出版社
http // www. ywcbs. com

图书在版编目(CIP)数据

汉字与计算机/顾小凤著 . - 北京:语文出版社,2000.7
(百种语文小丛书)
ISBN 7 - 80126 - 607 - 2 / H·168

Ⅰ.汉… Ⅱ.顾… Ⅲ.汉字信息处理-基本知识 Ⅳ.TP391

中国版本图书馆 CIP 数据核字(1999)第 50354 号

百种语文小丛书
HANZI YU JISUANJI
汉字与计算机
*
语文出版社出版
100010 北京朝阳门南小街 51 号
Email:ywp@public2.east.net.cn
新华书店经销 北京印刷二厂印刷
*
850 毫米×1168 毫米 1/48 1.5 印张 38 千字
2000 年 7 月第 1 版 2002 年 6 月第 2 次印刷
印数:2,001 - 5,000 定价:3.00 元

《百种语文小丛书》编委会

说　明

在信息时代、知识经济时代，知识的普及工作比以往任何时候都更为重要和迫切。语言文字是信息的载体，是人类文明发展的不可须臾缺少的基本手段，在信息时代其发挥的作用更为巨大。语言文字知识既有传统的，也有新的。语言文字知识的普及工作有待加强。我们编辑这套《百种语文小丛书》，正是想在这方面尽绵薄之力。小丛书的篇幅较小，只有几万字。我们力求少而精，以便于读者购买和阅读。

语言文字知识的内容十分广泛，你如果看一下《中国大百科全书·语言文字卷》的分类目录，便会有一个大概的了解。我们这套小丛书，不是对语言文字知识作全面系统的介绍，而是偏重于人们在学习、生活中会遇到的各种语言文字知识。

书的质量很大程度上取决于作者水平，取决于作者对相关知识了解和掌握的程度。我们希望这套丛书是高质量的。因此，在约稿的时候采取的办法是，谁对某方面问题有研究，就请谁撰写某方面的书

稿。作者素有研究，自然能够驾轻就熟，保证所介绍的知识的科学性，而有的还具有知识的前沿性。把最新的研究成果用普及的形式介绍给读者，也是本丛书编纂目的之一。这只是个总的思路，做起来并不那么简单。作者在写书时，要考虑整个丛书的编写原则，写什么，怎么写，从谋篇到动笔，是一个新的创作过程。本丛书是为普及语言文字知识服务的，然而书中有些内容事属专门；有的见解为作者一家之言，与社会通常的见解有的不完全一致。我们认为这些内容，对读者广见闻是有裨益的。小丛书的作者都是学有专长的语文工作者，其中有的是享有盛誉的著名学者。正是由于得到了这些作者的大力支持，小丛书才能够顺利编辑出版，奉献给社会公众。

本丛书的出版，采取成熟一批出一批的方式，一时难以完全显示学科的系统性。我们打算在出版到一定数量后，再来分类配套，如语音的、语法的、词汇的、文字的，等等。敬希读者谅察。

《百种语文小丛书》编辑委员会

目　录

第一章　引言

作为本书的开头，让我们首先了解一下汉字的特点、计算机的特点以及这两者的关系。

1. 汉字的特点

我们先从语言文字说起。

什么叫语言？一般人都容易理解。语言可以用嘴说出来（这时也叫口头语言或口语，英语叫spoken language），也可以用手写出来，譬如我们说“这本书的语言写得生动而流畅”（这时也叫书面语言，英语叫 written language）。总而言之，语言既可说也可写。先有说的语言，后有写的语言。英语如此，汉语也如此。

那么什么叫文字呢？英语里一般不用“文字”这种概念，也没有“文字学”这种学科，只有汉语这类语言才有。有人把文字解释为书面语言，更通常的解释则把一个个汉字叫作文字。本书所讲的“汉字”就是这种意义下的文字。

那么汉字有哪些主要的特点呢？

(1)汉字是形、音、义三者的结合体,有很强的表义性

汉字由象形文字演变而来。象形文字就是用一个形象的图形符号来描绘某一事物的形态、性质、语音等特征,使人们一看到这个图形符号就会直接联想到它要表达的意义。它不同于拼音文字。拼音文字从字母的形状并不能立即想到它们要表达的意义,而必须通过拼写得到读音,再间接地通过有声语言联想到这种读音所要表达的意义。象形文字演变到今天,字形已经发生了很大的变化,但仍然可以看出象形文字的一些痕迹。就是说,一个汉字通过一定的形状或一定的表音符号(用来表达语音的符号)来表示一定的意义;或者说,一定的意义写出来是一个字,念出来是一个音节,说与写达到高度的统一。所以汉字是形、音、义三者的结合体。

语言学界比较普遍地认为,语言中具有完整意义,能够独立运用的最小单位是词。[①]例如“枇杷”是一个词,表示生长在南方的一种水果。如果把这两个字拆开,单独的“枇”和“杷”没有完整的意义,所以它们是字不是词。也有很多汉字本身符合词的定义,如“我”、“大”、“木”等,这叫做单字词。按此类推,“枇杷”是二字词,还有三字词、四字词等,统称多字词。

英文字母可以组成英文的词,方块汉字可以组

成汉语的词，那么可不可以把汉字在汉语中的地位比作英文字母在英语中的地位呢？这种类比是不可以的，原因在于我们绝不可以忽视汉字的强大的表义性（即汉字表达语义的这种特性）。

英文字母从一开始就是用来拼写英语单词的一种声音符号，有音没有义。只有两个字母可以作为独立的词使用（即“a”作不定冠词，大写的“I”表示“我”）。而我们的祖先当初造字时，每个汉字都有形、音、义，都表示一种独立的意义。也就是说，一个汉字就是我们现在所说的一个词。只是到后来社会发展了，语言丰富了，单独一个汉字有时难以表达一种比较复杂的意思，于是才出现了多字词。即使这样，在现代汉语中，单字词仍然起着十分重要的作用。据统计[②]，尽管在日常用语中单字词的数量只占全部词数的12%左右，但由于单字词大多为高频字（高频即频繁出现的意思），因而在9000个常用词中，它的出现频率约为三分之二。在古汉语中这个比例肯定还要高。即使是构成多字词的每个汉字，一般也都有某种确定的字义。像我们前面所举的“枇杷”的例子，在汉字中毕竟只占少数。

正因为如此，语言学界有人认为汉语的基本结构单位是字而不是词。[③]

(2) 汉字的字数总量大

英文靠26个字母可以拼写出各种不同的句子，而每个汉字只能表达一种或几种语义，想用少数汉字的组合来表达丰富多彩的客观世界和思想感情是不可能的。因此，汉字的字数总量很大。究竟有多少？没有确切的统计数字。《康熙字典》收字47035个（一说为47043字[④]），《汉语大字典》收字约5万6千个，都不能说已经包括了全部汉字。但是常用的汉字并没有那么多。1981年5月国家标准局公布了《信息交换用汉字编码字符集·基本集》（GB 2312—80）共收一级汉字3755个，二级汉字3008个，合计6763字。1988年国家语言文字工作委员会（以下简称国家语委）、新闻出版署联合公布的《现代汉语通用字表》共收7000字，其中包括《现代汉语常用字表》所收的3500字（2500个常用字和1000个次常用字）。以上的6763字或7000字都是不包括繁体汉字的现代规范汉字。

(3) 多数汉字是由部件组合而成的结构字

以“铜”字为例，它由左右两个偏旁构成。左边的“钅”称为形旁，它描述事物的形态和属性，在这里表示一种金属。右边的“同”称为声旁，表示这个字的读音与“同”一致。在汉字字典中常常把属于同一偏旁的字归为一类，编成索引，以便于检索。每类称为一个部，把代表该类的偏旁置于每

部的开头，叫做部首。“铜”属于“钅”部，因此“钅”又称为部首。现在人们倾向于把偏旁和部首统称为部件。

部件之间是按一定的位置关系组合成汉字的。“铜”是左右结构；“雷”是“雨”和“田”两个部件按上下结构组合成的。除了这两种结构外，还有包围结构（如“国”、“困”）、半包围结构（如“匡”、“幽”）、左中右结构（如“辩”、“做”）等等。比较常见的结构有20多种。[5]

一般我们把按一定关系经常紧密结合在一起的一组笔画叫做部件，但是究竟有哪些部件，目前尚无统一的标准。《汉语大字典》采用的201个部件被国家语委建议作为部件的规范。

已经无法再分成部件的汉字称为独体字（如“凹”、“木”、“火”等），非独体字则称为合体字。在常用字中独体字不到5%。[6]

像“铜”这样由形旁和声旁构成的合体字称为形声字。形声字的声旁可以用来表音（即表示读音），形旁可表义（即表示字义）。但大约只有四分之一的形声字能够正确表音[7]，其余的只能近似地表音，甚至完全不能表音。相对来说，形旁的表义功能要强于声旁的表音功能。但这种表义往往是比较模糊的而不是精确的。例如“口”旁的字大致与嘴有关，“木”旁的字大致与植物有关。正确地认

识形声字的这种特点，可以帮助我们学习汉字并避免谬误。

(4) 汉字的语音特点

汉语普通话有 6 个基本元音音素，21 个辅音音素，用它们可以构成 21 个声母，35 个韵母。声母与韵母结合构成音节。汉字是单音节字，即一个汉字一个音节。有的音节只有韵母没有声母，具有这种音节的汉字也叫做零声母汉字。

由于汉字是单音节字，因此，从语音的角度讲，前面所说的单字词、多字词也可称为单音节词、多音节词。

汉字有 4 个声调。如果不考虑声调，汉字的音节数只有 417 个[⑧]，加上声调也只有 1200 多个。汉字的字多而音节少，这就决定了汉字的同音字多，而且各个音节的同音字数量很不均匀。少数音节没有同音字，而同音字最多的音节，其同音字可达 100 多个[⑨]。

汉字的同音字问题对于汉字的拼音输入法以及汉语语音识别等均有很大影响。

由于汉字不是拼音文字，所以需要有一种方法给每个汉字注音。曾经使用过的注音方法有威妥玛拼法和汉语注音字母（后改名为注音符号）。1958 年 2 月我国公布了《汉语拼音方案》，采用 26 个拉丁字母作为标注读音的符号，并于 1982 年被国际

标准化组织（ISO）确定为世界文献工作中拼写有关中国的专门名词和词语的国际标准。应该指出的是，《汉语拼音方案》只是拼写和标注汉字读音的方法，并不是一种拼音文字的方案。

(5) 汉字的简化及规范化

汉字不但字数总量大，而且不少字的笔画多，最多的可达 60 多画。笔画多，特别是常用字的笔画多，给汉字的书写带来很大的不便。因此，自古以来民间甚至官方文件中就一直流传着一些简化字。可见汉字的简化并非现代才有，而是贯穿于汉字发展的历史过程之中。1964 年 5 月中国文字改革委员会（以下简称文改会，是国家语委的前身）编制出版了《简化字总表》，共收简化字 2236 字。1986 年 10 月国家语委重新发表《简化字总表》，并对表中内容作了个别调整：规定“叠”不简作“迭”，“像”不简作“象”，“囉”不简作“罗”而简作“啰”，“覆”不作简化，“瞭望”的“瞭”不作简化。全表实收简化字 2235 个。

已经被简化的繁体字在一般情况下不应该再使用，因为它们已经不属于规范汉字。那么除了繁体字以外还有哪些是不规范的字呢？

①《简化字总表》以外的自造的简化字不是规范汉字。1977 年 12 月公布的《第二次汉字简化方案（草案）》由于不够成熟，于 1986 年 10 月被正

式废止，因此该方案中的简化字也不是规范字。

②1955 年 12 月由文化部、文改会公布的《第一批异体字整理表》中的 1055 个异体字属于被淘汰的汉字。经过三次调整，恢复了 28 个字，实际淘汰 1027 个字，它们不是规范汉字。

③地名中的 15 个生僻字已改用同音常用字，这些字也属非规范字。

④根据文改会、国家标准计量局 1977 年 7 月发布的《部分计量单位名称统一用字表》，“砗”、“呎”、“浬”等 20 字已被淘汰，也属非规范字。

以上各点只讲到规范汉字的范围。汉字的规范化还涉及字形、笔顺、部件、读音等多个方面，这里不再详述。

2. 计算机的特点

早期的计算机主要用于科学和工程方面的计算。随着计算机应用领域的不断扩大，“计算”已经只占计算机全部功能中的一小部分（尽管是十分重要的部分）。计算机愈来愈多地被用于处理各种信息，这种信息包括语言文字、图像、声音等等。现在人们可以用计算机来编辑排版、检索文献资料、发电子邮件、玩电子游戏、听激光唱盘、看 VCD……，“计算机”似乎有些名不副实，因此有的专家建议改称“电脑”。但鉴于长期以来人们已

习惯了“计算机”的叫法，而且“计算”仍然是计算机的重要职能，加之与英文名称“Computer”比较一致，所以继续使用“计算机”这一名称的也不在少数。

计算机的基本职能是：把外界的各种信号转成数字化的信号，输入并贮存在计算机里，对这些数字进行加减乘除的运算和判断，再把运算结果贮存起来或转成声、光、电等各种信号予以输出。尽管计算机的功能和应用领域发生了很大的变化，但以上的基本职能并没有变。正像现代人与古代人相比，生存环境、物质条件、生活习惯、兴趣爱好等发生了很大变化，但作为人的基本特征并没有变，人还是人。

那么计算机的基本特征是什么呢？

(1) 计算机能快速运算

人脑虽然聪明，但是每秒钟最多只能作一两次加法，而计算机的运算速度现已发展到以每秒多少亿次作为衡量单位。负责整个计算机的运算和控制功能的部件叫中央处理器（英文缩写为 CPU），计算机的运算速度主要取决于 CPU 的性能。

怎样看待这种高速运算的特点呢？

①拿计算机与计算器相比，计算器的运算速度其实也不慢。当我们用按钮在计算器上按了一个算式，最后按“=”的一刹那，结果就出来了。但是