**高等教育自学考试《汉字学概论》强化复习资料**

　　(一)汉字的性质

　　说到汉字的性质，我们不能不说一说文字的性质问题。文字是记录语言的书写符号系统，书面上代表语言的符号，世界上有多种多样的文字，可是从本质上看，从功能上看，其实都是一样的，即都是辅助语言进行交际的工具。

　　文字是在有声语言的基础上产生并发展起来的，因此，语言是第一性的，文字是第二性的，先有语言，后有文字，没有语言，就没有文字。语言是文字符号所代表的具体内容，如果没有语言，文字也就不可能存在。世界上任何一种符号，都是由一定的形式和一定的意义构成的统一体，文字作为记录语言的符号，也不例外，也是由形式和内容两个部分组成的。世界上有许多民族，至今只有语言，没有文字，但没有任何一个民族是只有文字没有语言的，语言并不依赖于文字而存在。可见，语言的确是第一性的，文字是第二性的，记录语言，离不开语言。世界上有各种各样的文字体系，虽然记录语言的方式不同，结构上千差万别，但依附于一定的语言而存在的道理是一样的。

　　因此，从本质上看，文字实际上就是一种符号系统，只不过这种符号是代表语言的，而且必须通过语言才能发挥作用。语言是一种符号，由形式和意义构成;文字也是一种符号，其构成也包括形式和意义两个方面。虽然都是形式和意义构成，但它们的内涵是不一样的，语言的形式是语音，文字的形式是书写单位(笔画或字母)构成的;语言的意义是一定的语音形式表示的内容，如普通话音节“shu”表示的意义是“成本的著作”，文字的意义是所记录的语言，具体地说是语言中的词或语素，包括语音和意义两个方面，例如汉字“书”这个书写符号所代表的，不单是语音“shu”，也不单是“成本的著作”这个意义，而是语音和意义两个方面。语言的语音形式和意义内容，在任何体系的文字中都是统一于一体的。世界上所有的文字，拼音文字也好，非拼音文字也好，记录语言都是既要记录语音，也要记录意义，不存在只记录语音或只记录意义的文字体系。语言是语音形式和意义内容的统一体，文字既然是记录语言的符号，当然要两个方面都记录下来，不可能剥离二者，那样文字是不可能完成记录语言的任务的，也不可能成为书面上代表语言的符号了。

　　世界上各种各样的文字体系，按其记录语言的方式的不同，一般可分为拼音文字和非拼音文字两大类。拼音文字用一定的构字符号表示语言中的各个语音单位，符号的排列组合顺序同所记录的语言单位的实际读音一致，所以能比较容易拼读出来。非拼音文字的构字符号同语音联系不紧密，从字形结构上看不出同所记录的词语的语音的联系，只能一个一个死记，不能拼读。

　　我们教材是按文字单位与语言单位的联系情况来划分类别的。文字基本单位的划分，不同的文字体系是不一样的，例如英文、法文、俄文等是字母，日文是假名，汉字是单字。语言的单位可分为语音层和符号层，在语音层可分音位(音素)和音节，在符号层可分语素、词和句子三级。把文字的基本单位和语言的基本单位相互对应起来看，根据所记录的语言单位的不同，一般把文字分为音素文字、音节文字和语素文字三大类。

　　音素文字是用字母记录音素，即在记录语言的时候，要把音节分析出辅音和元音音素，分别用相应的字母表示，阅读时再把元音和辅音拼合为音节，是比较完备理想的拼音制度。音素文字的系统比较简单，因为一般一种语言的音素，分析成元音和辅音只有几十个，所以用几十个字母就可以把语言记录下来，音素文字适用于记录音节结构类型多的语言，英语的音节结构就比较复杂，类型很多，据资料载，日本有关方面统计英语的音节不会少于8万个，所以只能用音素文字记录。现今世界上大部分文字都属于音素文字。

　　音节文字用一定的字母记录语言中的音节，符号和音节之间有固定的联系，适合记录音节类型比较少的语言，日文中的假名是音节文字的典型代表。

　　语素文字以单字记录语言中的语素，字与语素基本上相互对应，语素文字最大的特点是字符多，由于要相互区别，又使得结构非常复杂，汉字就是典型的语素文字。由于语素文字着重于字形同语素意义的对应，不要求找共同的标准音点，因此特别适合记录方言复杂，方言间语音差异大而语法规则一致、语素意义一致的语言。

　　三种文字和所记录的基本单位对应如下：

　　音素文字――音素

　　音节文字――音节

　　语素文字――语素

　　汉字的基本单位是字，而每个汉字对应的语言单位，大体上是一个语素，所以，从文字单位同语言单位的对应关系看，汉字是语素文字。语素是最小的语音语义结合体。例如汉语的“祖国”是由“祖”和“国”两个语素构成的。每个语素都有音和义：“祖”的音是zǔ，意思是先代：“国”的音是ɡuó，意思是跟“国家”相对。汉字的一个个的单字记录的是汉语里的一个个语素，因此汉字是语素文字。汉语语素有同音的，也有不同音的。不同音的语素一般用不同的汉字来表示。例如汉语里的“人”和“民”是两个不同音的语素，书写时就用“人”和“民”这两个不同的汉字。而同音语素也常常用不同的汉字来表示。例如“一、依、衣、医、伊、铱、漪”这七个语素都读yī，而书写时用七个不同的汉字，从而在形体上把不同的语素区别开来。形态各异的汉字可以有效地区别同音字。

　　每个汉字除了有形体和读音外，还有意义。这和英文字母、日文假名很不相同。语素是最小的音义结合体，是有意义内容的，因此，汉字记录语素，它的音和义来自它所记录的语素的音和义。例如我们用“书”这个字表示汉语里“书”这个语素，因为语素“书”读shū，意思是“成本的著作”，于是汉字“书”就读shū，意思就是“成本的著作”。汉字的形和它所表示的语素的音和义紧密结合在一起，形成了一个形音义的统一体。汉语语素大多数是单音的，但也有一些是多音的。多音语素有的是连绵词和叠音词，例如“蜘蛛、琵琶、秋千、茫茫、渐渐”等;有的是是译音词，例如“吉它、巧克力”。记录多音语素的每个汉字，只有形体和读音，而没有意义，必须几个字合起来才有意义，因而不是形音义的统一体。这样的汉字数量不多，不反映汉字的本质特点，不影响汉字是语素文字的论断。

　　关于汉字的性质，历来有种种不同的说法，比较典型的有表意文字、意音文字、词素-音节文字(或语词-音节文字)几种说法。

　　汉字是表意文字，这个说法的出发点是汉字不限于同一种读音挂钩，从读音看汉字缺乏体系性，而从意义角度看汉字显得非常有规律性，而且字形结构反映了字义特点，例如“桦、桉、柳、桂、椅”等字以“木”表示字义类属，体现出表意特点。不过，由于汉字经过隶变和简化后，字形结构发生了很大变化，字形结构与字义的直接联系远不如古代汉字那样明显，而且表意与否本身也是比较模糊的，所以近年来采用这种说法的已经越来越少了。如果说汉字绝大多数都有意义，因此从这样的角度说汉字是表意文字也是可以的。

　　意音文字的说法着重于汉字80%以上都是形声字这一客观事实。但如我们教材分析的那样，现代汉字中真正称得上是形声字的已经很少了，大部分形声字或者意符不表意，或者音符不表音，从现代汉字的角度看，这样的字不应该再当作是形声字，所以真正意义上的形声字不多，到底有多少，笔者尚没有看到这方面的确切资料，但比较肯定的数据是，现代形声字，音符和字音完全相同的只有全部形声字的20%，如果这20%的形声字再用意符和字义的联系情况加以考察，估计又得排除一些形声字，这样所剩就无几了。由此可见，把汉字称为意音文字并不合适。

　　也有人将构成汉字的符号分为意符、音符和记号三类，而根据符号所能表示的语言结构层次来看，汉字有些字符只跟语素这个层次联系，有些字符则起音节符号的作用，因此把汉字看作是语素-音节文字，而且认为这个名称适用于古代汉字和现代汉字。不过，音节文字现在一般专指拼音文字的一种，用来指称汉字容易引起混淆。尽管汉字是代表音节的，但是其结构中并没有严格意义上的专职的表音成分，因此用音节文字这种名称是否合适值得斟酌。

　　以上种种说法，大家可以了解，也可以深入研究。但从我们课程的教学实际出发，弄清楚汉字是语素文字的立脚点，能作必要的解释说明就可以了。

　　(二)汉字的特点

　　汉字有六个方面的特点，这些特点，有的是优点，有的是缺点。

　　一、汉字和汉语基本适应。

　　汉字和汉语相适应，主要是汉字的单音节形式和汉语的语素的单音节形式比较一致，用汉字记录汉语语素往往有一种对应的关系。汉字和汉语基本上相适应，具体可从以下几个方面来认识领会：

　　第一，一个汉字代表音节，一个语素基本上也是一个音节，因此汉字和语素是相对应的，用汉字记录汉语语素非常合适。不过，汉字同语素对应，并不是说汉字与语素相等，仔细分析一下，其实汉字和语素之间的关系也是比较复杂的，毕竟汉字是文字单位，而语素是语言单位。汉字和语素间的复杂关系表现在，一个汉字可能表示几个不同的语素，如“米”表示大米的米和表示长度单位的米，“站”表示“站立”或“车站”的意义;也可能是几个汉字表示一个语素，像外来词“沙发、吉它、马来西亚”等是受外来词读音的影响用汉字转写，其中的汉字没有独立的意义，这样的词语可以不考虑，汉语传统中还有一些联绵字，都是几个汉字记录一个语素的，如“窈窕、秋千、琵琶、蜘蛛、犹豫、踌躇”等。当然，大部分汉字是一个汉字表示一个语素的，如“山、河、灯、人、看”等。

　　第二，汉语是没有形态变化的语言，语法意义、语法关系主要通过词序和虚词表示，而不像印欧语那样要用音素表示形态变化，所以汉语的特点决定了它用汉字记录是合适的。比如英语的动词do(打)有doing、did、done等形态变化，一个词(尤其是动词)的词形在不同的句子中是不一样的，其音素要发生变化，这样的语言用音素文字记录比较合适，而动词类似的含义，汉语是用一个个独立的语素(词)“正在、曾经、已经”等表示的，这种变化用写法固定的汉字记录就比较合适。又如英语的复数要在名词后加词尾s表示，这个语素也是在词内部增加音素(即内部屈折)显示出来的，而汉语，复数是增加一个语素“们”表示，非表人名词主要在前面加数量词体现，不是在词内部发生变化，这种情况用一个独立的汉字表示就非常方便，而把汉字用来表示类似英语那样的形态变化，就会困难重重。

　　汉字与汉语相适用，我们可以通过汉字借用来记录朝鲜语、日语的情况反证。日本、朝鲜曾经先后借用汉字记录自己的语言，但是由于这两种语言与汉语在语法规则上有很大的差别，所以汉字只能记录它们中的固定的部分，而对于有形态变化的部分，写法固定，不能分解的汉字简直一筹莫展，所以日语、朝鲜语用汉字记录不完全合适，需要用字母表示形态变化的部分，正因为如此，朝鲜在汉字的基础上创造了谚文，日本在汉字的基础上创造了假名。而且历史上，汉字曾经被周边的少数民族借用，但是并没有被直接借用来记录本民族语言，而是在其基础上加以改造，使之适合自己民族语言的需要，例如契丹文、西夏文、女真文等。

　　第三，汉字记录汉语可以非常有效地区别同音词。汉语音节结构简单，声韵母配合只有400来个音节，声韵调配合只有1200多个音节，加之词形简短，以双音节词为主，单音节词依然十分活跃，可以说是汉语语汇的基础，这样同音语素和同音词就非常多，很容易混同，而形态各异的汉字就起到了分化同音词的作用，所以汉语同音词虽然很多，但用汉字记录，在书面上就把同音词巧妙地区别开了，这是拼音文字所做不到的，例如“公式-公事-攻势-宫室-工事”、“形式-形势-行事”、 “国是-国事-国势”、“期中-期终”等同音词，口头上说出来，语音完全一样，很容易混同，可是书面上用汉字记录下来，形体各异的汉字很自然就把它们区别开来，根本就没有混同的可能了。下面我们看几组读音完全相同而字形各异的汉字，对汉字的这种区分同音词的作用就一目了然了：

　　ju(阴平)：居、鞠、拘、狙、疽、驹、苴、掬、琚、趄

　　shi(去声)：是、市、式、事、示、视、世、势、适、试、室、蚀、誓、逝、柿

　　li(去声)：力、立、沥、呖、例、利、里、粒、励、笠、历、莉、荔、栗、丽、隶

　　yi(去声)：意、义、艺、易、翼、异、奕、翳、亿、抑、屹、役、逸、肄、臆、谊

　　上面4组汉字，每组声韵调完全相同，可是汉字字形差别非常大，这样就非常方便地把不同的语素区别开来了。我国著名语言学家赵元任先生曾经利用汉字的这个区别功能，用声韵完全相同的汉字编写了一个《施氏食狮史》的故事，抄录在这里供大家欣赏：

　　石室诗士施氏，嗜狮，誓食十狮。氏时时适市视狮。十时，适十狮适市。是时，适施氏适市。氏视是十狮，恃矢势，使是十狮逝世。氏拾是十狮尸，适石室。石室湿，氏使侍拭石室。石室拭，氏始试食是十狮尸。食时，始识是十狮尸，实十石狮尸。试释是事。

　　假如用拼音文字记录这些语素，按拼音文字的规则，读音相同的语素，其字形也应该相同，那这些语素就根本无法从字形上区别开来。汉字从古代的表意字发展到形声字，为什么没有再进一步跨进拼音文字的行列，这同汉字的这个有效区别同音词的功能不无关系，而且这也是不少人反对汉字改革，反对汉字拼音化的一个重要的理由。的确，拼音文字从目前看是无法解决汉语同音词问题的。

　　二、汉字是形音义的统一体。

　　因为汉字所记录的是语素，而语素是音义的统一体，每个语素既有声音又有意义，加上汉字的形体，所以汉字是形音义的统一体。一个汉字，它可能代表好几个语素，可能表示好几个音节，但是独立地看，每个汉字必定是表示一个音节的，必定是表示一个语素的，而且必定是有形的，所以汉字是形音义三位一体的，是形音义的统一体，三者之间密不可分。这一点与音素文字和音节文字是很不相同的，例如英文的字母a、b、c、d、e等，日文的あ、い、う、え、お等，它们有形体，有自己的读音，但是没有意义，因此不是形音义的统一体，因为它们记录的是语言中的音素、音节，没有独立地和语素对应，因此单独看都是没有意义的。

　　正因为汉字的这个特点，所以我们初学汉字有一定困难，但是如果学习了一定数量的汉字以后，进而学习词就不困难了，例如我们学习了“电”字，可以和别的有意义的汉字组成“电话、电灯、电线、电视、电影、电流、电机、电表、电动”，而对这些个新词，我们基本上不用重新学习记忆，很容易理解掌握，因为我们学会了一个一个的汉字，实际上就等于学会了一个一个的构成词的语素，掌握了这些构词成分的形、音、义，再来学习理解这些个词语，自然就非常容易了。反观音素文字，人们在学习字的时候非常容易，比如像英文，只有26个字母，几节课就可以学会，日文有50音图，初学也很容易，但到了学习词的阶段，就比较困难了，因为每个词都有不同的拼写形体，记忆起来也并非易事。

　　三、汉字有较强超时空性。

　　汉字的超时空性，可以从两个方面来认识：一是古今一致性比较强，二是具有沟通方言的特殊作用。汉字的超时代性，就是古今可通，上千年前的文献中的汉字，今天我们照样认识，其中看不懂的，不是汉字本身，而是古今语素意义的变化。虽然汉字已经有数千年的历史，但是由于古今汉字结构方式、结构体系在本质上是一致的，同语言中的基本单位的对应关系相同，即都是记录语素(在古代同时又是词)，没有发生本质的变化，所以就具有了通贯古今的特殊作用了。比如我国历史上宋、元、明、清时代的小说，我们今天阅读起来基本上没有太大的困难，而同时代的用拼音文字记录的文献，今天一般人就很难看懂了，只有专家才能看懂，因为拼音文字要随词语读音的变化调整自己的拼写(当然不是绝对一致，比如英语单词know中的k今天就不发音了，但拼写没有调整)。

　　汉字的超空间的特性，就是具有特殊的沟通方言的作用。汉语方言复杂，有七大方言，如果细分，方言更多，而且各方言间的差别特别大。口头上，不同方言区的人根本不能进行交谈，如果谈话，就好像是在同外国人说话，互相不知所云，除非用双方共通的中介语(如普通话)，否则就不能进行沟通。可是书面上用汉字笔谈，就出现了非常奇妙的情况：无论你是上海人、广东人，还是湖南人、福建人，互相之间毫无障碍。假如没有汉字，不同方言间的交往怎样进行真是令人难以想象!当然，汉字的这种沟通方言的作用，根本之处在于汉字记录的是汉语，虽然汉语方言众多，但各个方言的语法结构特点(词的特点、句子的特点)是一致的，语素意义差别小，不然，同样也是看不懂的。用汉字记录的日语，虽然每个字都认识，但是却不能看懂其内容，有人夸张地说学习日语的人往往是“笑着进去，哭着出来”，就是这个道理，而学习汉语方言，是不会出现这种情况的。

　　从上面的分析可以看到，学习汉字，需要我们认识汉字的超时空作用，也要注意到，汉字这个特点的根源还是汉语的语法、词汇系统古今、方圆方面的共通性，汉字的这个沟通古今和沟通方言的作用，其实仍然是汉字基本上适应汉语这个特点的反映。

　　四、汉字字数繁多，结构复杂，缺少完备的表音系统。

　　汉字系统的字数，数量之多是非常惊人的。从古到今，汉字的总字数已达到8万多个，据资料统计，现代汉语用字在1万左右，国家公布的《现代汉语通用字表》收录汉字7千个，《现代汉语常用字表》收录汉字3千5百个。

　　汉字字数众多的根源在于汉语的特点。由于汉语的音节结构比较简单，音节数量比较少，和语言中的语素的数量相比较，简直不成比例。比如现代汉语声母和韵母拼合组成的基本音节只有4百来个，再配上四个声调也只有1千2百多个，而现代汉语语汇系统中的语素，多达数千个，在这种情况下，读音相同的语素就非常多，加之含此词形简短，以双音节此为主，但因借此也非常多，而且往往是常用词，假如书面上不用形体各异的汉字把这些同音的语素和词区别开来，就会造成混淆，造成使用上的混乱，影响交际使用，影响人际沟通，严重的甚至有可能早引起误会，造成损失。汉字字形分化字义的作用，我们可以从一些汉字的分化过程看出来。比如“云”，本意就是云彩的云，后来增加了“说话”的意义，为了区别，便另外造了一个“雲”來区別：“其”本来就是簸箕的意思，后来产生了表示第三人称的意义，于是人们便在院子基础上增加一个竹字头，构成“箕”：“莫”，从构形上看是太阳淹没在草丛中，就是傍晚黄昏的意思，后来引申出否定的意义，表示没有人，为了区别，于是在原字形的基础上另加一个“日”来区别，这就是“暮”字。类似的例子很多，例如“昏-婚”、“然-燃”、“止-趾”、“采-採”、“益-溢”等等。可见，汉字字数繁杂，有其不便于学习掌握的一面，但也有其便于区别意义的一面。

　　汉字的结构复杂，可以从几个方面来看，一方面，汉字笔画很多，以7千个通用汉字为例，笔画在7画至15画的多达5千以上，最多的一个汉字笔画多达36画;另一方面，汉字的结构单位笔画和部件，在组合成子是不是按照统一的方向排列组合的，而是上下左右全方位展开，多向行进，具有平面型特点，而且结构单位的组合模式也是多种多样的，笔画组合有相离、相接、相交几种，部件组合有左右结构、上下结构、包围结构、框架结构等，而且很多子往往是几种方式的总和，结构就更复杂了。汉字不像拼音文字那样是按一个方向行进的。所以学习汉字，只知道笔画部件，不知道结构方式，依然无法把一个字写出来。

　　而且汉字结构中可以说是没有表音成分，虽然汉字系统中有数量众多的形声子，但是形声字的音符并不是专门的表音符号，很多依然是独立成字的，更主要的事，现代汉字形声字的音符绝大部分已经失去了表音作用，有的音符在一个字中可以表音，在另外一个字中不能表音，这种种情况夹杂在一起，很难判断，例如“台”在“抬胎邰”中是表音的，在“殆怠迨治冶”中是不表音的。根据现代形声字音符的表音情况，如果用音符去推断一个不认识的字的读音，大部分都有可能出错。而人们在使用汉字时，往往有这样一种错觉，认为汉字既然实行绳子，那么就可以利用其音符来推断字音，结果常常推断错误，闹出笑话来。例如下面这些汉字，之所以读错，就是人们盲目使用音符推断造成的结果：“诣、愎、菅、龋、啮、酵、涸、烨、酗”。人们常说，汉字难读，这同汉字结构上不能表音不无关系。

　　五、汉字用于机械处理和信息处理比较困难。

　　汉字用于机械化和信息化处理比较困难，我们可以从以下几个方面来认识。首先，汉字数量多，现代汉语通用字有7千个，如果涉及到专业领域的用字，数量更多，这样多的汉字机械处理当然就比较困难。其次，汉字结构复杂，大多数汉字笔画在10画左右，多的有好几十画，而且汉字笔画和部件组合的模式也非常复杂，随人我们可以给这些汉字的构造单位的组合方式归类，但一字一字观察比较，可以说是一个字有一个字的构造模式，有些字，例如“董、麝、爨、齉、冀”等，信息处理就非常困难。相比较而言，拼音的音素文字由于数量有限，机械处理就没有障碍，比如引文用拉丁字母，只有26个，加上大写字母也只有52个，加上其它字符(例如标点符号等)，总量不超过100个。所以英文的字符可以全部搬上键盘，而把7000个汉字照搬上键盘，那简直是不可想象的。所以在我国，打字成为一门技术，需要专门学习。

　　我们可以用英文和汉字做个比较，这种差异就看出来了。现在电脑处理文字信息，一般使用点阵表示，即用电子方阵来表示一个字符，每个方阵有若干个格子，其中有笔画的点亮，每有笔画的点暗，这样就实现了文字自负的显示。显示英文，最低要用7×9点表示一个字符(一般要使用9×12点)，显示汉子，最低要使用16×16点表示一个字符(一般要用24×24点)。英文的字符我们算100个，汉字的字符只算7000个，那么所占的点是：

　　英文：7×9×100=6300点

　　汉字：16×16×7000=1792000点

　　汉字字形存储量在电脑中所占的空间，是英文的284倍。差距之大是惊人的。汉字的机械化处理以前之所以落后于其他文字，因为7千个汉字字符需储存1750k位，而英文不到10k位，在计算机发展之初，根本就没有那样大内存。据资料载，美国IBM公司60年代末期生产的最先进的一种型号的电子计算机，内存为4096-131072，可以装入汉字，但是价格高得惊人，根本不可能推向市场，而其他型号的计算机，则不能装进汉字。这还是普通型汉字点阵，如果是精密型汉字点阵，比如是32×32点，或者100×100点，处理1万个字符，那所占的空间更是惊人。当然，在电子计算机储存量急速扩大的今天，这个问题已经得到顺利解决，但比较拉丁字母，汉字的局限仍然是明显的，所以有的学者认为，从长远看，汉字只有拼音化才能便于信息处理。



