

# 如何界定汉字的理论单位

——试论汉字的性质及同一性问题

朱 一 星

〈要約〉

筆者は2011年の年初に「国際漢字単位」の概念を初めて打ち出した。その後、さらに「漢字単位説」の明確化を目指すべく、「デフォルト読み」という、漢字に絶対音価のない事実と、漢字の「標意文字」性質などを手掛かりに、日本語における中国人漢字名読みの問題について議論じた。筆者の考えでは、現代漢字の理論はつねに現実の文字現象で検証可能であり、難問の解決に役立つ方法論を提供してくれるものでなければならない。本稿では、議論の更なる深化を目指し、現代漢字理論が立脚する漢字の文字論的性質とその基本単位の定義についての検証を試みる。

〈キーワード〉

标意文字 文字功能取向 默认读音 同一性 国际汉字单位

本文借鉴近年脑神经科学，认知神经科学的研究成果，尝试以结构主义语言学，功能语言观的立场来解释汉字——这一语言视觉符号系统，讨论其中若干本质性的问题，通过这些讨论希望打破长期困扰汉字研究的僵局，试图为解决一系列悬而未决的难题打开道路。

## 一、汉字对应语言的语意切分

不少语言学家经常重复这样一个观点：即文字不等于语言本身，文字只能是语音的书面表达形式，即符号的符号。这个观点阻断了语言学界许多汉字研究者的思路，直到今天，仍然认为汉字也和字母文字一样，必定是通过（语音的）语言才能实现其符号功能的“不可动摇”的观点在语言学界十分有市场。

这一观点的持有者经常引用索绪尔（F. de Saussure）《普通语言学教程》中的如下一段话：“后者（指文字符号——引用者）唯一的存在理由是在于表现前者（指语言符号——引用者）。语言学的对象不是书写的词和口说的词的结合，而是由后者单独构成的”<sup>1)</sup>。

笔者其实也无意否定索绪尔的观点，然而一百多年前，普通语言学诞生时主要观察研究的对象，是以字母文字为其书写形式的欧洲语言。站在这一立场上来描述字母文字体系和语言的关系是可以接受的，但是今天当我们思考这一观点对于汉字体系是否得当时，就很值得商榷了。

汉字和字母文字有着本质上的区别，进一步说：汉字对应语言“双层切分”的第一层，而字母文字则对应语言“双层切分”的第二层。大家知道，在马丁内（André Martinet）“双层切分”思想中，语言切分第一层的最小单位是一系列有着“两张脸”（语意和语音）的“符号素(monème)”。

第二层则失去语意，只存在“一张脸”，即只对应音节或音素的分节。第一层的语意分节对区别性要求高，各异的分节数量十分庞大；第二层分节则不超过人类发音器官的生理极限，一般对区别性

要求不高，各异的分节数量也不会很多。马尔丁内认为，双层切分机制符合语言的经济原则：用（绝对数量少的）语音符号间接地代表（绝对数量多的）语意符号。显然，第二层的语音符号就不可能越位对应第一层的语意分节，这就是（字母）文字不可能是语言符号的实质性含义<sup>2)</sup>。

然而，汉字则未必然。

同样套用马尔丁内的语言“双层切分”思想，就应该说：汉字（个体）恰恰对应的是语言的第一层（语意带语音的）切分，这才顺理成章。打比方说，单个汉字能够对应的是语言的“分子颗粒”，保留着最小的语言功能。而字母文字对应的就不是语言的“分子颗粒”了。犹如物质世界构成分子的原子不再具备分子的物理性质一般，字母文字对应的基本是不具备语言功能的语音成分。

汉字作为语言的书写符号体系，它的单位个体（符号能指）不仅能够理论上对应语意（符号所指），而且事实上的确不必由语音中介，也能对应语意分节<sup>3)</sup>。

这一认识得到当今认知神经科学研究成果的支持。读者可以参阅彭聃龄，陈宝国等在2001年以后多次的研究报告<sup>4)</sup>。这些报告中的实验结果说明，大脑在汉字识读过程中，字义和字音呈现同时激活，或者字义的激活发生在字音激活之前（因汉字使用频率的高低不同）。这一现象在该领域同行中也得到了确认。

汉语常用字里有为数不少的“多音多义”现象，比如像“都”[dū / dōu]、“还”[hái / huán]、“行”[xíng / háng]之类的字。经验告诉我们：当你念书时，必须在确认了字义之后才可能读准字音，比如“长头发”这个短语如果没有语境，可能让人无从开口。对于这些多音汉字，是很难想象必须先念出它的声音后才能懂得字义那样的理解机制的。所以，上述实验结果符合我们日常生活中“见字知义”的实际体会，甚至并不让人感到意外。值得注意的是，一系列实验并没有报告测试对象的汉字因其属于所谓象形、指事或是形声而有差异，说明汉字的造字理据并不构成对汉字功能的影响。

日语是目前世界上仅存的，同时使用汉字和字母文字（假名）的语言。长期以来，日本的脑神经科医生就十分注意对大脑损伤者失读，失写病症状进行临床观察和纪录。医生和研究人员积累了丰富的第一手资料，有力地说明了日语母语的失读或失写患者对于汉字和假名——这两种不同性质的文字呈现不同的临床症状。最早发现这一事实的，据说是一名叫三浦谨之助的医生。他在1901年第一次纪录了一名念不出假名但却能念出汉字的失语症患者的病例。1975年，脑神经科医生山鸟重（Yamadori Atsushi）首次确认了一名左角回脑损伤后患有失读失写症的患者虽读不出假名，却仍保持完好的汉字念诵能力。这一发现促使岩田诚（Iwata Makoto）等进一步做了大量临床调查，并在1984年提出“汉字·假名读写双通道”假说。经过多年的临床采样证明，这些假说如今已经成为该领域的常识性见解<sup>5)</sup>。

尤其给人留下深刻印象的，是1996年东京电机大学的小谷诚（Kotani Makoto）和NTT基础研究所的今田俊明（Imada Toshiaki）共同领导的一个实验：使用122通道全头型脑磁图扫描系统记录了日语单词识别时的脑磁图。6名男性受试者的测试结果表明，假名单词在大脑呈现活跃的语音活动，而汉字单词的语音活动相对不明显。在激活日语词义的测试中，汉字词语所费的时间比假名词语更短，只需要假名的三分之一时间。对于这一结果，小谷诚认为是因为假名词语必须分两步完成：先激活语音，然后通过语音才能通达词义，因此耗费一些时间。而汉字词语则一步完成，由视

觉信号直接通达词义，不必以语音为中介，所以速度快得多。这一实验结果在当时就受到了媒体的关注，被认为证明了汉字能直接表达语意<sup>6)</sup>。

## 二. 文字语言的时代

语言学的讨论对象只能是（语音）语言这一“正统”观念还从另一个侧面受到挑战。手语，作为一种“纯正”的视觉语言已经在国际上广泛被确认为属于“自然语言”之一。这也是经过脑神经科学以及脑功能障碍研究证明的<sup>7)</sup>：尽管听障者依靠手、指、眼神、表情等多种视觉手段进行交流，但是他或她赖以思维的大脑深层区位却是与依靠听说进行语言交流时大脑所使用的区位一致的。也就是说，听障者和非听障者虽然使用不同的知觉系统（听觉和视觉），却都是使用相同的大脑语言中枢接受、记忆、和思维，并进行“语言”交流的。

比如 Howard Poizner 的研究团队就曾经详尽地描述了几例以美国手语（ASL）为母语的听障者大脑受损伤后并发失语症的临床观察，证实美国手语母语患者同普通英语母语患者具有许多相同的失语症病理机制。

据说当乔姆斯基（A. N. Chomsky）定义语言“是语音和语意的固有对应”时，曾有人问他手语是否被排除在语言之外？对此，乔姆斯基回答说：“我所说的语音指的是‘信号’，应该说（语言）是信号和语意的对应。语音是否具有决定性意义是个悬而未决的问题。语音也许是决定性的，但支持这一论断的佐证尚不充分。”<sup>8)</sup>

如果说汉字有别于手语的话，那就只能是汉字比起手语更靠近语音语言。在这里，语言的符号能指不只限于语音。这与我们在第一节讨论过的，汉字对应语言的第一层切分的观点紧密关联。由此看来，将汉字视为语言的视觉符号，象索绪尔言及的“对汉人来说，……文字就是第二语言”<sup>9)</sup>的观点，是否也应该如同手语那样，成为语言学界的共识？

在此，我们不可能回避一个汉字争论中十分初级的，但却往往非常迷惑人的问题：汉字对于使用者来说，虽然是可以“见字知义”的。但是对于非使用者就不可能“看出”个所以然了。不用说对于非汉字圈阅读者如此，即使对于汉字使用者来说，从未学过的汉字也是“看”不出意义的（根据上下文推测不在此列）。即便是对于所谓“象形”或“指事”的汉字，也不会特殊照顾。人们通常说汉字“承载”或“表现”语意，这种说法只是在作为比喻时才说得过去。不论这种观点的持有者是否悟到这一点，汉字的偏旁部件既不是，也无法成为准确描写语意的手段。这一点，在任何时候都是不可改变的既定事实。

当伍铁平再版他主编的《普通语言学概要》（第二版）时，终于撤销了把汉字描写成“表意文字”的习惯说法。他在“第二版后记”中对此专门做了论述：

本书不采用“表意文字”这一术语……

当代中外许多语言学家反对用“表意文字”给“汉字”命名，还因为“表意文字”易被人利用来宣传一种错误观点：汉字可以见形知义，无须通过语言。

虽然“第二版后记”中论述的许多理由，直至某些基本观点与笔者并不一致（本文正在讨论之中），然而和“表意”挥手告别这一点，笔者认为合理的，科学的，面向未来的。

然而，恐怕连伍铁平也未必认识到的是，“表意”这个说法之误，误在一个“表”字，而不是“意”（或“义”）字。因为如果我们承认汉字为符号系统（这一点估计不会有人愿意反对），那么按照索绪尔的符号理论，我们就必须承认符号的第一原则是它的任意性。汉字对应语言第一层切分，从符号理论的观点来说，是不具有任何必然性的。汉字对应语意的具体方法，可能会有许多种。但是任何一种都不会是无从选择的，绝对的，或永恒的。刚才我们讨论过，汉字对于从未学过的人来说，字义是怎么也“看”不出来的，就是这个道理。我们能够认读某个汉字，除了通过课堂传授或字典说明等手段以外别无途径。学习过程，就是将汉字同我们的某个“语意心像”（大脑意识到的知觉活动）建立对应关系<sup>10)</sup>，当你下一次看到同样的汉字时，就能够凭借字形激活（想起）该汉字所对应的“语意心像”。如果同一个汉字对应几个截然不同的“语意心像”的话，学习的过程也毫无疑问地必须分别进行。

在世界上有不少学者认为阿拉伯数字(0~9)才是名符其实的“表意”文字，理由是阿拉伯数字似乎不依靠语言，能够超越语言“以形示义”。这里姑且不论阿拉伯数字能不能属于严格意义上的“文字”，更重要的是这些数字本身亦无“表”可言。它们必定是经过学习才得以理解的，只不过有的人忘记了这一幼年时代的事实而已。

如此看来，符号的任意性原则决定汉字对应语意的方法不是“表示”（=自身具有决定权），而是“标示”（=自身不具有决定权，符号能指和符号所指的对应关系来自于约定俗成）。字形和字义的关系恰如语音和语意的关系一样，不是天然的，而是被赋予的；不是积极的，而是消极的。

说明这一点的最有力的旁证，就是现代中国大陆和日本实行的汉字简化。两个政府对汉字分别实施的大胆改革，历来被谨慎的文字学家们视为大大削弱了汉字的“理据”，有悖于汉字的表意性。然而，多年的汉字运用实践表明，字形上的改变，虽然抹去了一些字形中隐含着的久远的记忆（喜好问题权且不论），并没有在实质性意义上削弱汉字的语言符号功能。

笔者赞成把汉字称做“标意文字”，因为这既体现了汉字对应语意的语言符号性质，又能避免把人们引入认为汉字能够通过不通过学习就能“见形知义”的误区。

细心的学者也许能够感觉到，在近几十年的汉字讨论中已经有不少学者不同程度地意识到了这一问题“症结”所在。要不然为什么那么多论文作者有意无意地避开“表示”而使用“标示”的说法呢。可是，大胆地直称汉字就是“标意文字”的，据调查还不多，陆锡兴和张树铮或许是论及这一核心问题的仅有的两位<sup>11)</sup>。张树铮对此明确地指出：

如果往深一步说的话，那“表意文字”也是约定的一种符号，……换句话说，表意文字和表音文字一样，字形都是“标”而不是“表”语言中的义或音。因此，准确的称名应该是“标音文字”和“标意文字”。<sup>12)</sup>

此话言之有理，但同时又不得不指出的是：上述两位探索者都未能在文字和语意对应方式的原理层面上追根刨底，因此，也就未能将“标意”之说的理论意义最大化<sup>13)</sup>。

应当明确地断言：汉字，在成为纪录语言符号的那一刻起，作为符号的“本质功能”就是“标意”的。这样，有关汉字的定义，就可以描述为：

汉字是对应语言第一层（语意）切分的语言符号。汉字单独地，或组合起来通过大脑激活与之相应的语意心像，随之激活相应的语音心像，属于“标意文字”。

### 三、“标意”是一种文字的功能取向

需要进一步说明的是，如同上文表述的那样，标意的汉字绝不是和语音毫不相关的文字，汉字在对应语意的同时，正如心理活动的“语意心像”必定对应“语音心像”一样，汉字也同时对应一定的语言音。认知神经科学的实验也说明，不论高频汉字还是低频汉字，都是能够“自动”激活语音的<sup>14)</sup>。

但是，尽管汉字具有“两张脸”，即代表意和音两个方面，笔者仍然认为把汉字称为“标意”文字更妥，而不是“标意-标音”之类。其理由，首先当然是因为汉字作为符号具有标音文字不具有的标意性。同时，还因为下述理由。

第一，汉字的使用历史，是一部多地区，多语言使用的历史。汉字同时对应不同方言区语言、甚至不同语种（比如韩语日语等）的语音。可以说汉字随时随地共时地对应着“一群”语言的不同语言音。过去许多学者把这个现象错误地解释成“汉字可以不依靠语言（的声音），主张汉字可以“脱离”语音而存在，这显然是一种认识或说明上的失误。真实情况是，各地区的使用者是以“自己的”语言音来读汉字的。笔者把这种“我行我素”式的汉字读音现象，称为“默认读音”。从历史上看，汉字的默认读音就无所不在。在日语母语的失读症患者中，就有把汉字的读音读错但仍然理解其意的临床病例。因为日语汉字的读音一般会有好几种，根据组词或上下文决定读音，所以念成其他发音却正说明患者理解汉字标示的语意不依靠特定字音<sup>15)</sup>。所以可以说，汉字是一种在理论上不存在绝对音值的文字。理解这一点，对于我们理解汉字的标意性也是很有必要的。

第二，尽管汉字在使用时，的确有假借（即借音=仅利用汉字的标音功能）的用法。但这在汉字功能中是辅助性的。而且往往是创制新文字的一个常见的过程。由于这个过程很长，从历史上的某一个“点”来看，就不免让人觉得是纯粹“标音”了。然而，动态地考察汉字的发展，许多假借的现象又常常分化独立成字，所以不能断定假借只是单纯为了标示语音<sup>16)</sup>。

当然，现代汉语词语的一些后缀，比如“-子”和“儿”化音等，的确不是特定的语意心像所对应的。另外外来词的音译也都可以说只是语言音的一部分。可是当我们以文字的功能取向来看的话，就能比较清楚地看到：汉字系统的文字功能取向始终是“标意”，而不是“标音”。

文字“功能取向”的观点，也可以用于解释标音文字不发挥标音功能的现象。近年来在原本标音的字母文字语言中，出现许多譬如 FTA、GTP；U-turn, O-ring；L10N、I18N<sup>17)</sup> 等缩写词或造词。这些词用字母标示了语音语言了吗？非也。如果不存在文字，这些并不为描写语音而造的词是没有人能想得出来或说得出的。这种场合，就可以说字母对应“语意”甚至图形，是一种类似“标意”的用法。但是我们仍然称字母文字是“标音”的文字，理由无非是因为字母文字系统的“功能取向”仍然是“标音”，并没有转化成“标意”。

### 四、汉字单字的“标意性”和汉字组合的“表意性”（绝对任意性和相对任意性）

我们在确认汉字文字功能的标意性的同时，也不能忘记，一旦掌握一定数量的汉字，就能通过

已知汉字的组合表达和理解一些未曾学过的词汇。据说字母文字的“优越性”和“合理性”就是当你学会了几十个字母，懂得了基本的拼写原则，就能读懂所有书籍，其实这只能说对了一半。现代社会的许多知识，难度和数量上都已经无法全凭语音语言传达，所以即使字母文字的使用者，也可能无法“念懂”自己从没有“听”到过的词。然而对于一个汉字使用者，则很有可能“看懂”一些自己从没有“听”到过的词语。

岩田诚把他开创的学说用英语命名为“neurogrammatology”。这个用17个字母组成的造词，连大型英语词典里都不一定找得到，会英语的看到这个词，若不是专业领域的，即使能尝试着念，也可能无法把握词义。这说明即使你知道26个英文字母，也不等于在日常生活中会读懂那些有可能对你至关重要的词语。一方面，岩田诚同时又用汉字命名了他的学说叫“神经文字学”，尽管原文是日语，可是对于懂汉字的人来说，这个造词能让人立刻领会，似乎毫无深奥难解之处。关于汉字化解让人望而生畏的英语学术名词的作用，铃木孝三（Suzuki Takao）曾有更详细的描述<sup>18)</sup>。

这说明以汉字组合的词语具有较强的“表意性”。这一文字系统的特性对于汉字使用者无疑是最大的利益所在。虽然可以有许多例子说明汉字的词组并不一定是单个汉字字义的简单“叠加”，但是更多的例子仍然能够显示出组字词语中的汉字具有丰富多彩的“表”词义倾向<sup>19)</sup>。

索绪尔除了反复对语言的任意性原则加以描述以外，在他的讲课记录里，他的学生还记下了一段和任意性原则不同的，有关“相对任意性”的说明：

符号任意性的基本原则并不妨碍我们在每种语言中把根本任意的，即不能论证的，同相对任意的区别开来。只有一部分符号是绝对任意的；别的符号中却有一种现象可以使我们看到任意性虽不能取消，却有程度的差别：符号可能是相对地可以论证的。<sup>20)</sup>

索绪尔随后列举了若干个法语词的例子：说法语的“二十”是无法论证的，然而法语的“十九”“二十九”“十八”“七十”等“因为它会使人想起它赖以构成的要素和其他跟它有联系的要素”（“十”和“九”；“二十”和“九”；“十”和“八”；“七”和“十”），所以这些词不是完全“不可论证”的。索绪尔其后又举了许多“相对地可以论证的”例子。同时告诫说：

即使在最有利的情况下，论证性也永远不是绝对的。这不仅因为可以论证的符号的各个要素本身是任意的……，而且因为整个要素的价值永远不等于各部分的价值的总和。<sup>21)</sup>

语言中并不是所有要素都是“绝对任意的”，有些要素会似曾相识，符号中含有这些要素就“可能是相对地可以论证的”了。这就是索绪尔“绝对任意性和相对任意性”的观点。

可以看到，索绪尔对他表述的“各个要素”，还没有琢磨透是个什么概念。但是我们知道，这属于语言的第一层切分单位，也就是相当于马丁内的“符号素”。我们前面已经讨论过，这个保留着语意的语言第一层切分，也正相当于汉字所对应的语言单位。那么就可以说，具有“绝对任意性”的“要素”单位在汉语文字语言中就相当于“汉字”单位。若如此，汉字的组词，就属于“可能是相对地可以论证的”，具有“相对任意性”的要素组合了。

其实，不限于语言的第一层切分，字母文字在语言的第二层切分中，也同样体现出单位本身（代表音素的字母）的“任意性”，和单位组合（由字母组合拼读音音）的“相对任意性”的关

系。当然，这里的任意性和相对任意性都是体现在语音模写层面上，而不是语意层面上的。但是因为“相对任意性”不是人为赋予的，而是依照独立于字母的拼写原则，所以字母文字组合不一定是必须“学过”才能拼读的。毫无疑问，这也正是字母文字的优越之处。

无论第一层切分也好，第二层切分也好，文字的组合才是我们通常所说的“表示”。表示语意，或表示语音。这样，有关文字单体和文字组合的功能差异也许可以表述如下：

对应语言双层切分的第一层或第二层的文字，都是由一群“文字单位”组成的。

当文字单位对应最小的语言切分单位时，文字功能显示“任意性”原则，亦可称“标示性”；

而最小单位的组合则表现为“相对任意性”特征，使组合文字具有“表示性”。

在这里，刚才被搁置一边的“表示”，可以有一个新的安身之处了：我们不赞成单个汉字在理论意义上“表示”语意，但是无法否定汉字的组合有时恰恰是能够“表示”词义的。这时候的字组往往就具有“表意性”了<sup>22)</sup>。

## 五、差别性和同一性

语言中只有差别。

这句名言，也是索绪尔语言观的核心思想之一。他认为语言就是围绕着符号的差别性和同一性的问题。当语言学讨论的对象还是语音的时候，我们讨论语言的同一性，主要是就语音来说的。但是，当我们把汉字看作语言的视觉符号时，我们应该如何描述汉字符号的同一性，又该如何描述单位之间的差别性呢？

我们面对着一个汉字符号理论的核心问题：什么是汉字的“单位”？怎样才算“同一个”单位（字）？这里将要讨论的单位（字），显然和传统汉字学（小学）关注的汉字具有完全不同的性质。因为注重字源的传统文字学常常要面对汉字的演变、分化、统合，这是历时研究的领域。可是汉字符号更注重共时层面的系统性，关注共时层面的“同一性”和“差别性”。

索绪尔说：“语言机构整个是在同一性和差别性上面打转的，后者只是前者的相对面。因此，同一性的问题到处碰到，但是另一方面，它跟实体和单位的问题部分一致，只不过是后一个问题的富有成效的复杂化”。索绪尔列举了前后两天的两列“晚上八时四十五分日内瓦—巴黎”快车，以及“被拆毁后重新建筑起来”的街道作为例子，力图说明同一性首先不是“物质上的”，“跟它的偶然的材料毫不相干”。而是“以某些条件为基础”的<sup>23)</sup>。他说：

共时语言学研究同一个集体意识感觉到的各项同时存在并构成系统的要素间的逻辑关系和心理关系。

历时语言学，相反地，研究各项不是同一个集体意识所感觉到的相连续要素间的关系，这些要素一个代替一个，彼此间不构成系统。<sup>24)</sup>

我们把汉字符号作为一个系统来讨论，就必须强调“同一个集体意识感觉到的各项”要素，意识到“构成系统的”各个单位之间的“差别”和“同一性”。当我们说“汉字系统只有差别”时，就同时要决定什么是表现差别的汉字符号的“单位”。

## 六. etic 和 emic

如果让两个人分别写同样的汉字。我们可以说这“两个字”写得如何，也可以说他们两人写的是同“一个字”。很明显，前者的“两个”是不具有理论意义的纯粹物理量，后者的“一个”才指的是文字概念。按照这样的思路，我们就有了判断问题的依据。试看下列一组字：

(1) 𠄎 (大和篆书) 𠄎 (GSNP空心体) 李 (DF娃娃体)

字组(1)形体上虽然有很大差异，但是我们仍然很容易把它们视为同一个字：“李”。因为它们是同一个字的不同风格体，它们之间不构成系统意义上的“差别”。再看下面一组字：

(2) 圖 (台湾正体字) 图 (大陆简化字) 𠄎 (日本简略字)

这一组虽然是同一种风格体，然而一眼望去就觉得笔形结构各异，成字理据不一。但它们仍然是同一个汉字概念。由于许多在大陆被称为“旧体”的汉字对于台湾来说是没有“新旧”之分的，因此我们必须承认字组(2)各个汉字之间具有“共时等价”关系。

再看下面两种不同组汉字：

(3) 不属于同一字：壬/王 (笔画方向不同) 士/土 (长短不一) 己/巳/已 (接触形态不一)

(4) 属于同一字：反/反 (笔画方向不同) 天/天 (长短不一) 花/花 (接触形态不一)

从上面两组并不稀罕的例子看，说明单从形体上的差异，是找不出区分单位的依据的。

为什么把有些只有细微的差别的字不作为“同一个”字，而把另外一些形体差异很大的字却说成是“同一个”字呢？如何辨认有些形体不同，来源各异，但具有等价关系的汉字呢？

在这里，讨论一下 etic 和 emic 的概念对我们理解问题的实质意义是不无益处的。

etic 和 emic 原本是一对源于语音学 (phonetic) 和音位学 (phonemic) 的词，由派克 (K. L. Pike) 将其术语化，用来描述如下两个概念：

**etic**：描写不具有系统意义的，任意切分的，不成单位的要素群

**emic**：描写具有系统意义的，有目的切分的，成单位的要素群

这两个概念在理论上是相对的，我们不妨拿象棋子来说明。象棋子面上的文字标记，通常有“帅/将”、“仕/士”、“相/象”、“兵/卒”之别，这些区别从文字上说不具有同一性，然而对于象棋的游戏功能来说则别无二致。在这里，文字上的区别，对于象棋游戏来说是“不具有系统意义 (etic)”的，也就是说不构成“单位”之间的差别的。同时，“帅(=将)”、“仕(=士)”、“相(=象)”、“兵(=卒)”，这4组棋子之间，以及和其他的“车”“马”“炮”之间的差别，才是构成游戏功能上“具有系统意义 (emic)”的“单位”之间的差别了。我们可以由此得到中国象棋棋子的游戏功能意义的“单位”数是“七”，而不是“十一”这一结论。我们再来判断下列汉字：

(5) 国 國 际 際 图 圖 书 書 馆 館

以不成单位(etic)的观点看，这些是相互不同的“十个”字，而以成单位(emic)的观点来看，我们就能判断出它们是两组繁简对应的“五个”汉字单位。也就是说“国”和“國”在汉字符号系统里，是具有同一性的，也就是说是一个“单位”。际(際)、图(圖)、书(書)、馆(館)则依此类推。这样，我们就得到共时层面上的，由具有同一性的汉字构成的“汉字单位”。

## 七. “国际汉字单位”

由于汉字在汉字圈是相通的，所以汉字单位也实质上是跨区域的。因此笔者提出“国际汉字单位”（Chinese Character Grapheme）<sup>25)</sup>的概念。它是**包容汉字共时层面的多种字体的抽象概念**。即包括繁体字、简化字、异体字、俗字等等。必须指出的是这是一个理想的理论模式，从汉字的现状看，目前还没有实现整个汉字系统的统一单位。理想的汉字理论单位，应具有以下几个基本特征：

**特征一：原子性。**讨论汉字的性质，描述汉字之间的理论关系，都是以方块汉字作为基本要素的。在这里，汉字以下就不具有对于整字而言的语言符号功能。无论从字义侧面来说还是从字音侧面来说，汉字的部件都无法承担汉字的功能。汉字就是“最小的”，“不可分割”的符号。

**特征二：抽象性。**相当一部分汉字单位因多字形而不具有绝对形值。因为书写符号在使用时会造成变异，变异的原因本来都是随机的、偶然的。然而书写符号的**非挥发性**使其保留和固化。这些历史上的多种因素就形成大量的异体字、俗体字、通假字。这些字形群共同构成一个符号价值，共同存在于汉字使用者的头脑之中，代表汉字的理论值。汉字的理论值（单位）犹如语言符号的“语言（langue）”，而每此汉字的使用（理论值的兑现）犹如“言语（parole）”。

**特征三：共时性。**作为符号系统，汉字单位必然是个共时的概念，历史上不断变化着的汉字是不应该在同一个层面上进行描述的。共时层面的同一性，保证了地区规范字体之间的对应，保证其归属于同一个汉字单位。

**特征四：泛语言性。**汉字作为汉字圈共同的书面交流工具以及纪录手段，历史上促进了地区间的文化交流和社会的共同进步。今后这一地区将仍然在户籍人名地政方面具有互通的要求，汉字单位的“泛语言性”将进一步显现出来。人们将会理解：汉字系统的整理优化，不一定是字体上的一致，而是“汉字单位”的一致。

可以说因为汉字单位不具有绝对形值和绝对音值，不同方言区、不同语言区的人才有可能更方便地共同使用汉字系统。

## 八. 明确汉字单位概念的积极意义

尽管结构主义语言学已经问世多年，然而，运用这一强大思维工具解析汉字的工作几乎就没有起步。本文用结构主义语言观讨论了汉字符号系统的一些特点，以及汉字理论单位的界定问题。

笔者认为建立国际汉字单位的概念，就能有一个新的讨论平台，有利于弄清一些迄今为止无法弄清楚的关键问题，避免许多无谓的争议和由此造成的脑力人力资源的浪费。

历史上的，以及近代的汉字整理，可以说基本上都是凭藉学者们的文献知识和直觉。然而在电脑时代，经验主义的整理方法施展身手的余地已经不多了。王宁在言及汉字整理面临的困境时说：

（汉字的改动）“最忌在没有全面研究的基础上，灵机一动，想到什么改动什么，结果常常是改了这里，那里的矛盾又显露出来”<sup>26)</sup>。“简单地论定汉字的优劣、简繁、难易的只言片语，已经不能解决问题。……，诸如此类的问题，都要在深入、科学的汉字理论研究中来解决。一个汉字科学研究补课的时代到来了”<sup>27)</sup>。诚如所言，关注汉字未来的研究者都会感觉到，当前需要进行大胆的理论探讨何等重要。

限于篇幅，本文只能就汉字理论的一些重要问题简单地勾勒一个框架，还不能说已经完美。笔者相信关于“国际汉字单位”概念的研究决不会是终点，因为它尚还刚刚起步，今后的研究前景不

可估量。但愿看到传统文字学家，语言学家，神经认知科学家，电脑专家们运用新知识、新手法，携起手来，为客观地理解汉字、使用汉字，促进推动汉字的优化、合理化和科学化共同努力。

（本文是京都外国语大学国际语言和平研究所2012、2013年度委任研究的成果之一）

后注：

- 1) 索绪尔《普通语言学讲义》商务印书馆 47页。
- 2) 参见 André Martinet (1965)，（笔者参阅的是渡濑嘉朗翻译的日文版本）。
- 3) 人们常用“语素”定义语言的最小分节，所以赵元任称汉字为“语素文字”。然而笔者近来对此甚为犹豫，因为“语素”意味着一个“音义结合”的最小单位，一般离不开某种特定的语言音。象汉字这样毋宁说是“音义+形义结合”的文字，其符号和语言音的关系，也就是说与经典意义的“语素”是何种关系，尚有待考察。
- 4) 参见彭聃龄（2006）61-79页；150-157页。
- 5) 参见岩田诚，河村满（2007），1-4页。
- 6) 参见川胜真喜，小谷诚（1996），另见《日本经济新闻》1996年1月8日报道：“漢字認識，仮名より速く”。
- 7) 见 Howard Poizner, ... (1987)，（笔者参阅的是河内十郎等翻译的同书日文版）。
- 8) J. C. Marshall 为《What the Hands Reveal about the Brain》一书写的“序”（收于河内十郎等翻译的同书日文版）。译文责任在于笔者。
- 9) 索绪尔《普通语言学讲义》，商务印书馆（1985）51页。
- 10) 关于“语意心像”的说法，笔者参考了山鸟重（2011）。
- 11) 这一调查得到了王迈的协助。
- 12) 引自张树铮（2004）。
- 13) 陆锡兴在《从划分标准看文字类型》（《语言文字学刊》1998第一辑，收于《急就集》）能见到较多的描述，且见于论文集《急就集》（2001）序文中。按照陆的说法：“标意文字的原字是一个类义符号，是一种抽象意义的符号，字形的特点是抽象性指示。汉字的标意文字由标形文字直接发展而来”。主要是强调了汉字符号意义的抽象性，为了有别于突出汉字图示性的“标形”而使用的。所以笔者认为陆锡兴使用的“标意文字”更多的似乎起因于汉字理据的丧失，而不是符号理论层次的思考。张树铮在本文引叙述后接着又说：“当然‘标意文字’中的有些字形同时也是能够‘表’的，只不过是约定的基础上”。……这样，张树铮将某些字源意义造字理据混同于符号学意义，轻易地放弃了他已经抓在手里的真理，把“标意”看作近似于隐形“表意”的概念。另外又将“标意”同时作为“表意”和“意音”的上位概念来处理，这些矛盾的做法削弱了其主张的理论性。
- 14) 参见彭聃龄（2006）。
- 15) 参见山鸟重（1984）207-208页。另见岩田诚，河村满（2007），79页。
- 16) 比如王宁（2007）就不认为同音假借是为了标音，而是单纯的“一字多用”，而且“在古籍里，很多同音借用字又用增加偏旁的方法产生借义分化”。
- 17) L10N是Localization的缩写，因L和N之间有10个字母而得；I18N是Internationalization的缩写，因I和N之间有18个字母而得。见 Ken Lunde《CJKV日中韩越情报处理》，O'REILLY, 东京2002, 1056, 1063页。
- 18) 参见铃木孝三（1990）的第四章“汉字鲜为人知的作用”，127-164页。
- 19) 有关组字词语中单汉字丰富多彩的语意功能，可参见孙银新（2003）。

- 20) 索绪尔《普通语言学讲义》，商务印书馆（1985）181页。
- 21) 同上，183页。
- 22) 也许，从严格意义上说，两个汉字以上的组合，有时候已经越出了文字论的范畴，属于语法学家感兴趣的研究领域了，因为我们有时很难判断汉语里“词”和“短语”的界线在哪里。
- 23) 索绪尔《普通语言学讲义》，商务印书馆（1985）153页。
- 24) 同上，143页。
- 25) 朱一星（2010），2011年1月刊出。
- 26) 王宁（2007）。
- 27) 王宁（2009）。

## 参考文献

- 山鸟 重 1984 脳損傷者にとって漢字とは—漢字の神経学，  
海保博之编《漢字を科学する》有斐阁，东京，pp.192-223  
—— 2011 《言葉と脳と心 失語症とは何か》，讲谈社，东京
- 高 名凯译 1985 索绪尔《普通语言学讲义》，商务印书馆
- 龟井 孝，河野 二郎，千野 荣一 编著 1995 《言語学大辞典〈第6卷〉術語編》，三省堂，东京
- Howard Poizner, Edward S. Kima, and Ursula Bellugi 1987 *What the Hands Reveal about the Brain* ,  
The MIT Press, Cambridge, Mass., and London. (河内十郎等翻译《手は脳について何を語るか  
——手話失語からみたことばと脳》，新曜社，东京 1996 )
- 铃木孝三 1990 《日本語と外国語》，岩波书店，东京
- 川胜真喜，小谷 诚 1996 仮名单語理解作業中の脳磁図，《電気学会論文誌C》116巻6号，pp.669～675
- 陆 锡兴 2001 《急就集——陆锡兴文字论集》，中国社会科学出版社
- André Martinet 1965 *La Linguistique Synchronique* Presses Universitaires de France  
(渡瀬嘉朗翻译《共時言語学》，白水社，东京 2003 )
- 孙 银新 2003 《现代汉语词素研究》，中国文史出版社，北京
- 张 树铮 2004 汉字性质平议，《临沂师范学院学报》2004.4，pp.14-18
- 王 宁 2004 论汉字规范的社会性与科学性——新形势下对汉字规范问题的反思，  
《中国社会科学》2004.3，pp.171-179  
—— 2007 基于简繁汉字转换的平行词语库建设原则，《语言文字应用》2007.4，pp.8-12  
—— 2009 汉字研究的新时代，《语言文字应用》2009.3，pp.17-20
- 彭 聃龄主编 2006 《汉语认知研究——从认知科学到认知神经科学》，北京师范大学出版社
- 伍 铁平主编 2006 《普通语言学概要》（第二版），高等教育出版社
- 岩田 诚，河村 满 合编 2007 《神経文字学——読み書きの神経科学》，医学书院，东京
- 朱 一星 2010 漢字の国際コード規格をどう考えるべきか——漢字コード規格の根拠となる  
「国際漢字単位」の提言，京都外国語大学《研究論叢》No.76，pp.211-223  
—— 2012 続・汝の名はかくして変わる——『表意』の神話，『表音』の虚構，  
京都外国語大学《研究論叢》No.79，pp.195-207

