



基于六书理论的 计算机语言中文输入法研究与改进^{*}

庞景晔

【提要】计算机中文输入法领域中，五笔输入法应用汉字形体结构以输入便捷高效等优势在 20 世纪计算机初步发展时掀起群体性热潮。随着时代进步与技术发展，五笔输入法随意拆解汉字，有碍教学与文化传播等弊端逐渐显现。且伴随拼音输入法的快捷传播，五笔输入法日益边缘化。本文针对现行五笔输入法在拆解与编码汉字方面的不足进行讨论研究，同时借用六书理论对汉字拆解做出科学的解释，以使汉字编码更具条理与逻辑性。

【关键词】《说文解字》 六书理论 计算机语言 中文输入法

一、理论与应用研究概况

（一）计算机中文输入法应用

现行的计算机中文输入法主要有拼音输入法、五笔输入法、手写输入与语音输入等类别。其中，拼音输入法作为大众生活中使用率最高的输入法，以声码为汉字编码基础，通过拼音系统编译、输出汉字。而作为 20 世纪最受欢迎且普及程度最高的五笔输入法因多重因素日益边缘化，但其以形码为汉字编码基础，应用汉字形体结构的编译思路对汉字文化的传播与教学具有可借鉴的意义。

五笔输入法为王永民在 1983 年根据汉字的笔画和字形特征所创造的一种中文输入法，在汉字输入领域具有里程碑意义。该输入法以汉字自身的字形作为基本字根，通过字根的搭配组合汉字。它在提升输入效率、降低学习成本等方面具有明显优势。但伴随时代进步与技术发展，五笔输入法的部分设计涉及随意拆解汉字部件且尚未得到合理优化而使其缺陷显露。

从教育教学以及汉字构型学的角度来看，五笔输入法已不符合现代汉字教育需求。五笔输入法的字根拆分并不完全符合科学的汉字构形依据，对字根的划分与许慎在《说文解字》中对汉字的构字方法研究及偏旁部首的确定存在差异，部分字根只遵

循了汉字输入时的便捷原则，而忽视了汉字的成字规则和字形所具有的意义。这使得五笔输入法虽具有强大的经济效益，但并不利于汉字的研究和文字学的发展。因此，针对五笔输入法编码逻辑的改进与研究有待进行。

（二）《说文解字》六书理论释源

针对现行五笔输入法中随意拆解汉字的缺陷，东汉许慎（1963）《说文解字》中的六书理论可以作为改进输入法实际应用的科学理论。《说文解字》是由东汉许慎编著的语文工具书著作，也是中国最早的系统分析汉字字形和考究字源的语文辞书。许慎是首个具体阐述六书理论，并将六书理论具体应用到对文字的说解中的学者。他阐释的“六书”为传统语言文字学奠定了理论基础。

许慎对六书理论的具体阐释如下。“一曰指事，指事者视而可识，察而见意，上下是也。二曰象形，象形者画成其物，随体诘屈，日月是也。三曰形声，形声者以事为名，取譬相成，江河是也。四曰会意，会意者比类合谊，以见指撝，武信是也。五曰转注，转注者建类一首，同意相受，考老是也。六曰假借，假借者本无其字，依声托事，令长是也。”

及至清代，语言学家戴震将其六书理论总结为“四体二用”，即象形、指事、会意、形声四种造字法和转注、假借两种用字法。六书理论归纳了汉字的构造方法和形式，是最早的关于汉字构造的系统理论。

二、研究意义与价值

（一）实际价值：科学完善中文输入法

以六书为理论依据改进后的五笔输入法，能够补充完善现行以形码为主体的五笔输入法，同时与以声码为主体的拼音输入法构成更加多元且科学的中文现代输入法系统。

1. 完善以形码构成的现行五笔输入法

使用六书理论对汉字进行进一步的分类研究，

^{*} 本文系“玉馥筠声——基于六书理论的计算机中文输入法研究与改进”（编号：202411117231Y）的研究成果。

增强汉字拆分与归纳的规律性和系统性，从而深化汉字构成部件本身所具有的意义和汉字整体字形所代表的意义。进一步简化部分复杂字根，尽量使用最少的字根通过六书理论满足汉字输入的需求。

2. 建构形码、声码共同发挥作用的中文输入法系统

提升现行五笔输入法的输入效率及科学性，帮助现行五笔输入法进一步扩大普及度，让年轻大众熟悉五笔输入法。目前，以声码为主体的拼音输入法普及度明显更高，但拼音输入法在未知汉字读音的情况下无法使用。因此，提升五笔输入法的使用率便具有合理性，即使五笔输入法和拼音输入法同时发挥作用，建构形码、声码共同应用的中文输入法系统，实现两者互补。

（二）文化价值：弘扬中华优秀传统文化

1. 尊重汉字结构规律，科学拆分汉字

以《说文解字》为解构汉字形体的依据，使得输入法对汉字的拆分更加科学，也更符合汉字的构形理论。而不是仅仅将汉字看成笔画的组合物，在拆分的时候忽视对笔画组合的意义与合理性的考量。对汉字的合理拆分，能够帮助使用者在使用输入时加深对汉字组成的正确认识，帮助大众理解汉字的结构和意义，尊重汉字的存在本质。

2. 形码输入法对于发扬汉语文化的重要价值

拼音输入法以声码为核心，而汉字是表意文字，对于字形形体有一定的要求。早期的汉字因义构形，汉字依据它所记录的汉语语素的意义来构形，所以，词义和据词而造的字形是统一的，且其形包含着古代造字智慧和历史文化意蕴。用声码编译不能体现汉字的造字构字特点。相较之下，形码输入法对于发扬汉语文化的重要价值便凸显出来。

（三）路径价值：实现传统与现代的创新融合

中华文化源远流长、博大精深，要全面深入地了解中华传统文化，在此基础上对其进行创造性转化、创新性发展，这就要求将传统文化与时代相结合，与时俱进，让传统文化能够适应现代社会发展要求，在新的时代中焕发出新的生机与活力。用六书理论改进五笔输入法，正是将传统文化中的优秀理论成果应用于现代科技应用成果之中，既能够实现对传统文化的创新性发展与保护传承、弘扬和壮大，又能够在实践上优化现有成果，为当代生活服务，让大众于生活中体会中华传统文化的魅力，提升自身文化素养，坚定文化自信。

三、研究与改进路径

（一）分析现行五笔输入法的局限性

通过分析五笔输入法的编码逻辑，梳理五笔输入法的迭代历程，根据不同阶段的创新点探究五笔输入法的缺陷，从不同角度指出五笔输入法在应用和底层编码逻辑的局限性。主要包括以下几方面的内容。

第一，对汉字结构理解的局限性。传统的五笔输入法主要基于对汉字结构的直观理解，但在面对一些复杂的汉字时，这种直观理解可能无法涵盖所有的汉字构成要素。例如，一些汉字可能包含复杂的笔画结构或特殊的部首组合。这些在五笔输入法中可能没有被充分考虑。

第二，编码规则的不统一和应用的不平衡性。五笔输入法的编码规则在一定程度上是主观的，不同的五笔输入法版本或者不同的输入法软件可能会采用不同的编码规则，这增加了大众的学习成本和使用难度。

第三，对现代汉字变化的滞后性。随着汉字的发展和演变，许多汉字的形态和结构都发生了变化。然而，五笔输入法在更新和调整其编码规则方面相对滞后。这导致一些新字或变体字在五笔输入法中无法得到准确的输入。

（二）以六书理论为指导改善五笔输入法

运用六书理论对五笔输入法的不足进行更深入地分析。六书理论提供了理解和分类汉字的科学框架。以六书理论为背景改善五笔输入法，结合其特点与优势，从以下几个方面考虑。

第一，简化五笔输入法的字根。五笔输入法的应用问题之一为字根冗余，对于使用者来说难以记忆。根据六书中的象形和指事原理，可以尝试将部分字根进行合并或者简化，使其更直观，易于记忆。

第二，优化形声字根。形声字根是五笔输入法的一个特色，可以根据汉字的声旁和形旁进行拆分。以六书中的形声理论为基础，可以对形声字根进行优化，使得输入更为准确和高效。

第三，引入转注和假借类别。转注和假借是六书中的两种用字类型，虽然五笔输入法中对转注和假借已经有所体现，但仍可以进一步优化和完善。例如，对于一些常用的转注字或假借字，可以设置专门的字根或输入方式。

除此之外，六书理论由于历史局限性、汉字字形的演变等多方面影响，其中面对不适用或者无法覆盖的具体情况应灵活应变，对传统文字学理论有





所扬弃，如此实现古今交融。

(三) 现代信息化技术的启发、呈现与自我完善
结合以计算机编码和人工智能为主的现代信息化技术，针对基于六书理论改进的五笔输入法进行模拟呈现，利用现代信息化技术的优点，为改进后的五笔输入法赋能，使其具备自我革新和调整的能力，实现数字与人文的和谐发展。

1. 用户自定义

允许用户根据自己的习惯和需求进行自定义设置，包括字根、快捷键等，以提高输入的灵活性和便利性。

2. 智能化纠错

基于六书理论中的“通假”现象，可以开发智能纠错功能，对于常见的输入错误进行自动纠正，提高输入的准确性。

3. 字根顺序优化

根据六书的分类原则与五笔输入法的字根表，可以借助大数据分析对字根的排列顺序进行优化，使得用户更容易找到需要的字根，减少查找时间。

(四) 探讨例证路径

结合理论基础与实际应用，试以汉字“衣”字为例探寻五笔输入法改进路径。

1. “衣”字五笔输出

“衣”字是二级简码，五笔简码是 YE，五笔全码是 YEU。其笔顺为“丶、一、ノ、フ、ノ、”，编号为“413534”。对“衣”字的五笔字根进行拆分，第一码字根是“亠”，对应按键“Y”；第二码字根是“衣字底”，对应按键“E”；第三码是识别码，由于“衣”字是杂合型，末尾一笔为“捺”，则识别码为“三捺”。

2. 《说文解字》理论基础

许慎(1963)《说文解字》共收字 9353 个，其中“衣”部共收字 116 个。“衣”在《说文解字》衣部篇中解释为：“依也。上曰衣，下曰裳。象覆二人之形。”《段注》：“叠韵为训，依者，倚也。衣者，人所倚以蔽体者也。”其上两位文字学家皆从衣服的用途上对“衣”字进行了解释，即人们依靠衣服来蔽体。其实，“衣”字的甲骨文像衣有两袖，两片衣襟相掩之形。

3. 区块改进设计

五笔输入法对于“衣”字拆解与六书理论对于“衣”字的分析存在偏差。五笔输入法中对于“衣”字的字根拆分如下：第一码字根是：“亠”对应按键“Y”。第二码字根是：“衣字底”对应按键“E”。

第三码是识别码：“衣”字是杂合型，最后一笔为“捺”，识别码是三捺。五笔输入法依据笔画拆解字根，对一个完整的汉字、部首的拆分是无逻辑的。

拼音输入法是针对“衣”字进行声码编译，字拼音为“yī”。拼音输入法以声码为核心，而汉字是表意文字，对于字形形体有一定要求。早期汉字因义构形，词义和据词而造的字形是统一的，且其形包含着古代造字智慧和历史文化意蕴。若用声码编译则不能体现汉字的造字构字特点。

以“衣”字为例，将“衣”作为一个完整汉字形码编入新版输入法，设置对应区块键盘。“衣”也是一个部首，以“衣”为部首，连接相关汉字进行区块字设计和数据库建设。以上述过程为一周期，再针对五笔输入法字根中其他存在缺陷的部分重新设计打磨，最终形成一套完整科学的新型输入法。

四、结语

总之，现行五笔输入法对于汉字的拆解仅遵循笔画结构，忽视汉字构字法中的内在逻辑与成字的合理性，虽符合计算机中文输入法的经济效益，但对于汉字文化的考据和研究存在障碍。五笔输入法将汉字拆解为若干字根，通过不同字根的组合输入汉字，而字根的划定与许慎(1963)《说文解字》的六种构字法及偏旁部首存在差异。从现行五笔输入法对于汉字拆解的不足入手，并树立鲜明的问题意识。运用《说文解字》中的科学理论对于汉字的构字做出更清晰合理的阐释说明，从而对现行五笔输入法存在的问题加以改进，以求增进其学术理论上的科学性。

参考文献

- 段玉裁 1981 《说文解字注》，上海古籍出版社。
方东杰 2001 《“六书”初义新探》，辽宁师范大学硕士学位论文。
黄天树 2014 《〈说文解字〉通论》，北京大学出版社。
苏新春 1993 《“五笔字型”对汉字规范化的作用与影响》，《广州大学学报》(综合版)第2期。
王宁 2015 《汉字构形学导论》，商务印书馆。
许慎 1963 《说文解字》，中华书局。
袁晓光 2019 《四十年来〈说文解字〉研究综述》，天津师范大学硕士学位论文。

(通信地址：225009 扬州大学文学院)

