

突破“毕升难关”的汉语文字设想

我过去在 Adobe Systems 公司的印刷部门工作时，曾经试图推动汉语文字印刷/显示的突破，结果不了了之。这个问题还涉及汉语结构甚至汉语思维的调整，非常复杂。几年来，我虽然断断续续在思考这个问题，却没有成熟的设想，现在提出来，希望能够抛砖引玉，引起人们的关注。

Adobe Systems 以 Post Script 技术（以精确的函数计算程序而不是用一堆小麻点来显示一个符号）引发文字显示和印刷革命，奠定了网络时代的基础之一，在 Post Script 基础上演变的 pdf 格式的广泛引用就是一个例子。除了各个字符的精确显示（可以放大、缩小、旋转等而不走样），Post Script 基于西方字符的特征，还用 kerning 等方法连接字符（这个词还没有适当的中文翻译，指在如 V 和 A 之间用倾斜的小间隔分开），达到比只用传统的 tracking 方法（平均分配字块之间的矩形空间）优美的显示效果。

汉字显示的最大难关是毕升一千年前发明活字印刷术时的限制：每一个字占同样的方块空间、每一个字之间用同样的（小方块）间隔分开。这在一千年前是一场技术革命，现在却严重阻碍着中国文化的进步。先看一个简单的比较：

[1]西文句子的表示以概念（词）为单位，如“I live in California”，词之间用空格分开，易于学习，阅读速度快。

[2]日文句子由汉字、平假名和片假名组成，“私はカリフォルニアに住んでいる”，也很易于把“私”、“カリフォルニア”、“住”分出，阅读速度也快。日语的问题也在于汉字和假名字母的方块显示，特别是カリフォルニア中的オ，不管是发音还是美观，都应该用 kerning 方法与フ连接起来。

[3]中文句子：“我住在加利福尼亚”，对于初学者或电脑，要把含义丰富的五个汉字“加”、“利”、“福”、“尼”、“亚”转换成一个简单的地名概念，颇费周折。

[4]朝鲜语“캘리포니아 있는 라이브”，与中文的问题相同。

可见，汉语的学习和阅读，除了文字本身的复杂外，句子显示的这个“毕升难关”（过去甚至连标点符号也没有）也浪费了中国人世代多少时光。问题还不止于此，它还制约着我们的思维。在汉语检索的技术上，“分词”（把作为一个概念的词从一个句子的一串字符中分离出来）根本不可能完全实现。例如，我今天检索朋友“方组”，结果找不到人名，却出现“澳方组长”、“妙方组图”等一大堆无关的连接。再如，“非典”（非典型肺炎）本来是 2003 年才出现的一个概念（单位），有人却从《三国演义》中查到首出于东汉末年的曹操：“非典，吾死也！”

利用现代技术可以容易地解决这个“毕升难关”：在人名“方组”的两个字之间不插入间隔，而用 kerning 连接成为一个概念（单位），在“澳方组长”的“澳方”之间用 kerning 连接成为一个概念（单位），“组长”之间用 kerning 连接成为一个概念（单位），“澳方”和“组长”之间用传统方法隔开，并适当加大间隔；在前一个“非典”的两个字之间不插入间隔，而用 kerning 连接成为一个概念（单位），后一个“非典”之间用传统方法隔开，并适当加大间隔。这样，检索起来就不会出现上述的错误或笑话。同时，这样的解决方案，在我们的阅读过程中，极大地简化、解放了我们的思维，因为我们的思维是以概念（如“斯德哥尔摩”）而不是以没有意义或具有多种含义的字（“斯”、“德”、“哥”、“尔”、“摩”）为单位的。这种解放，对于用汉语逻辑的人，有助于把语言作为工具而往高层次思维，益处不可估量。

从美观效果上讲，汉语最悠久、自然、优美的显示，不是来自毕升的方块字，而是取自王羲之等书法家的毛笔字体，其中每一个字所占空间因笔划、字形，甚至前后连接而不同。书法还讲究“笔断意连”等。现在的技术使得我们可以继承这样丰富的特征来显示汉语文字，结束方块字“大一统”的时代。

当然，从技术上实现这样的突破，比只处理几十、上百个西语字符的 Post Script 等技术复杂的多，恐怕要全社会各相关组织的更广泛参与。除了技术开发外，还需要汉语文字工作者对汉语结构、语法等问题的探讨，还需要汉语文字使用者的国际间的合作和认同。我在这里，只能提供“异想”，需要大家来达成“天开”。

赵京，中日美比较政策研究所
2009年9月27日，美国圣拉蒙，初稿