

# 日本的中国古代文学研究与信息处理

木村 守

## 1 东京学艺大学开设的「信息处理基础演习」及「国语科与信息处理」

### (1) 「信息处理基础演习」

因为它是必修科目，一年级的学生不管理科文科都得学习。内容是电脑操作法、安装软件的方法、作成 HTML 及 Internet 利用法等等。以前我教过文科系的学生。当时，不少学生进大学後才开始用电脑，所以理科文科一起学习同一内容也没有什么问题。

但是，现在，在高中已经开始学习「信息处理」，这样，大学生的信息处理水平也更高了。因此，大学的「信息处理」课的内容应该有所改变和充实。

### (2) 「国文科与信息处理」

讲义题目是「国文科与信息处理」（其他学科有「××科与信息处理」，比如数学科有「数学科与信息处理」，音乐科有「音乐科与信息处理」等等），但是现在听课的人都是和听「信息处理基础演习」课的差不多一样。所以讲义内容也没有很大的差异。

习惯使用电脑的学生一定会用电脑做自己的研究。这是「信息处理」课本来应有的状态。使用电脑的能力是进行研究的手段，而不是目的。这样「××科与信息处理」课的作用，今后会越来越重要。

## 2 日本的中国古代文学研究

### (1) 编制自己使用的字库

近十几年来，在日本也有不少研究者开始使用电脑了。在此以前，用打字机的人也很少见。电脑或打字机里面可以用的汉字不算多。只用 JIS 汉字的话，我们研究中国古代文学的人，输入一个诗人的诗赋也会遇到汉字不够用的问题。所以，大家都自己制造规格外的字。各人都有自己编制的外字的字库。这样字库是各人随便编制，不能互相利用。各人自己编制的资料库也不能互相利用。

### (2) 利用大型汉籍文献资料库

最近，围绕电脑的环境有很大的变化，真可谓日新月异。有人说：「电脑是永远没有十全十美的」，

确实是这样。但是，电脑日日进化只会给人带来方便。现在《四库全书》电子版及《四部丛书》电子版等大型资料库已经完成了，我们利用这些资料库就能很容易地找到书籍里面所有的词语、诗句的出典等等。

日本的文学研究者（特别是年轻人）研究中国古代文学时，都使用电脑。如果不利用电脑的话，文学研究也不能进行。没有《四库全书》电子版及《四部丛书》电子版等大型资料库的时候，为了知道某个词语的意义，我们就得查词典，找到出典的书籍后，再要从卷头翻到卷末，寻找有没有那一句。找到后，才好不容易确定词语的意义。所要的时间很多很多。现在呢？要找哪个词语就可以检索哪个词语，而且转眼就找得到。不过，也有一个新的问题。检索结果不多的话比较简单，但是有的词语检索结果超过一千、一万。我们研究者都好像成了一个辞书编者，应该把检索结果一个一个斟酌后，才能确定词语的意义。

### 3 利用电脑进行研究中国古代文学上的几个问题

#### (1) 汉字数码的问题

现在 ISO10646(所谓 Unicode)已经普及，使用汉字的环境很方便了。但是，汉字就是国家的遗产。日本有日本的汉字，中国也有中国的汉字。日本的汉字就是为了写日语的。中国的汉字就是为了写中文的。而 Unicode 里既有日本的汉字，也有中国的汉字，混在一起。用字的时候应该充分注意。

还有一个异体字的问题。在 Unicode 里本来就有异体字的问题。我认为整理异体字是目前应该最优先解决的问题。比如说：「东京学艺大学」的「艺(yi)」字，用中国简体字写的话是「艺」，但用繁体字的话是「藝」，用日语常用字写的话是「芸」。这还不算麻烦，麻烦的是日语常用字的「芸」与中文的「芸(yún)」字，字形完全一样而意思完全不一样。

#### (2) 利用大型资料库的问题

上面说到的异体字的问题，在利用资料库时也会发生。现在还没有整理异体字的构造，所以在检索时即使明白各字的异体关系，也还是不够充分。为什么呢？例如，想找「藝」字，而中文简体字的「艺」或者日文常用字的「芸」，以及中文的「芸(yún)」字都会出现。就是说，连不想找的字也找出来了。因为机器程序怎么也不能区别「芸(yi)」字和「芸(yún)」字的不同。再如，「巳」「己」「巳」的异体关系，机器也不能区别。这些字在影印上原来没有什么区别。「己」字有的地方写成「巳」或「己」，「己」字有的地方写成「巳」或「己」。读者除了从上下文判断确定字的意义以外，没有别的办法。

#### (3) 为了利用大型资料库的方便

为了方便利用大型资料库，我觉得应该考虑更简便的方法。

现在利用《四库全书》《四部丛书》等的人大部分大概是研究中国古代文学的研究者或学生。但是，《四库全书》《四部丛书》等这些大型资料库并不只是为了研究中国古代文学的人。对其他学科的人来说，检索起来也应该很方便。不过实际上并不是很方便。我觉得主要有三个原因。

第一，不知道那样方便的资料库的存在。第二，虽然知道这些资料库的存在，但是不知道怎样访问那个互联网地址。第三，网址能自己找到，也会自己检查，但是看不懂检索到的原文。这种情况不限于外国人。母语是中文的人，如果不是搞中国古代文学的话，也许也会出现这类问题。

为了方便利用大型资料库，除了提高资料库的可靠性以外，现代中文、日文、英文等等的解释也应该数据化，和原文一起公开，这样做也是为了中国古代文学研究者之外的学者或学生。这样，服务就会渐渐好起来。

#### 4 「信息处理」课该教什么？

信息处理无论对哪个学科，都有较大的关系，而且担任「信息处理基础演习」的教师也是各个学科的专家。所以本来就不能讲统一内容的课。对学生或者对教师都有利的方法是，在大学不用开设「信息处理基础演习」，而应该增加「××科与信息处理」这样专门性更高的课。

这样，讲课的内容就自然明确了。对研究中国古代文学的人来说，一边解决前几节已经说到的问题，一边进行自己专门的研究。我们日本的研究中国古代文学的学者，为了发展中国古代文学研究，除了自己专业研究以外，应该进行中国古代文学文献的日译或者用日文注释。其他学科也做相应的工作，把自己研究的内容介绍给专门外的人，跨学科的研究就会越来越发展。