关于小鹤音形输入法在专业环境中是否需要增加大词库的简单分析

1. 有人提出问题，说小鹤音形不适合打专有名词。先来看一下以下小鹤双拼-小鹤音形群的部分内容（中括号内文字为群昵称）

|  |
| --- |
|  |
| 【妖言】Rea(122789929) 11:22:37  偶尔打一些专用名词的时候 还是用搜狗挂接飞扬打的快 比如  C:\Users\德智\Documents\Tencent Files\1045865146\Image\Group\Image1\4`V0SW5YK7VWYAZ$4TI5ZY6.png |
| 【妖言】Rea  飞扬只能单蹦了（“飞扬”指小鹤音形） |
|  |
| 【活跃】默默练习水文  经常用的专业名词，加词库好一些 |
| 【小鹤作者】散步的鹤1  千万别说偶尔，因为输入法的关键是频率 |
| 你偶尔的东西可能一辈子打一次，慢点又何妨 |
| 【evi】酥梨  如果写论文占平常打字很大一部分的话 |
| 【小鹤作者】散步的鹤  你列举的论文的例来看看 |
| **可以统计里面的用字范围是否超越规范** |
| 这样更有说服力 |

以上对话中的一个明显的问题是：

1. **（因为这就是小鹤音形的收字范围，另小鹤音形也收录了一些该表以外的增补字，所以如果小鹤音形收字不出错的话，按照《通用规范汉字表》的标准的论文中的字应该都能用小鹤音形打出来。**

另两个不明显的问题是：

1. **小鹤音形应该安逸于官方词库而打单字还是加上专业词库？**

那么我就来找几篇专业论文吧。

希望下列图文能给读者们带来形象直观的展示。

噔地一下从知网上下载了四种类型的论文：



**四篇医学论文**



链接：[①](http://210.45.242.91/kcms/detail/detail.aspx?recid=&FileName=ZDYS201732001&DbName=CJFDLAST2018&DbCode=CJFQ&uid=M3c3UzlxTUV1cE9mOVJpR20xYXV1ZVFHeGhnT3JwK2hhSVFOTnllV1dkY1RDMXZh)、[②](http://210.45.242.91/kcms/detail/detail.aspx?recid=&FileName=ZNZK201704008&DbName=CJFDLAST2018&DbCode=CJFQ&uid=M3c3UzlxTUV1cE9mOVJpR20xYXV1ZVFHeGhnT3JwK2hhSVFOTnllV1dkY1RDMXZh)、[③](http://210.45.242.91/kcms/detail/detail.aspx?recid=&FileName=YXZS201724014&DbName=CJFDLAST2018&DbCode=CJFQ&uid=M3c3UzlxTUV1cE9mOVJpR20xYXV1ZVFHeGhnT3JwK2hhSVFOTnllV1dkY1RDMXZh)、[④](http://210.45.242.91/kcms/detail/detail.aspx?recid=&FileName=ZXJH201727024&DbName=CJFDLAST2018&DbCode=CJFQ&uid=M3c3UzlxTUV1cE9mOVJpR20xYXV1ZVFHeGhnT3JwK2hhSVFOTnllV1dkY1RDMXZh)

**三篇物理论文：**



链接：[①](http://210.45.242.91/kcms/detail/detail.aspx?recid=&FileName=DXWL601.002&DbName=CJFD9498&DbCode=CJFQ&uid=M3c3UzlxTUV1cE9mOVJpR20xYXV1ZVFHeGhnT3JwK2hhSVFOTnllV1dkY1RDMXZh)、[②](http://210.45.242.91/kcms/detail/detail.aspx?recid=&FileName=ZKJD200809007&DbName=CJFD0608&DbCode=CJFQ&uid=M3c3UzlxTUV1cE9mOVJpR20xYXV1ZVFHeGhnT3JwK2hhSVFOTnllV1dkY1RDMXZh)、[③](http://210.45.242.91/kcms/detail/detail.aspx?recid=&FileName=JGXK201104023&DbName=CJFD1112&DbCode=CJFQ&uid=M3c3UzlxTUV1cE9mOVJpR20xYXV1ZVFHeGhnT3JwK2hhSVFOTnllV1dkY1RDMXZh)

**三篇汉语言论文：**



链接：[①](http://210.45.242.91/kcms/detail/detail.aspx?recid=&FileName=ZXWS200405020&DbName=CJFD0305&DbCode=CJFQ&uid=M3c3UzlxTUV1cE9mOVJpR20xYXV1ZVFHeGhnT3JwK2hhSVFOTnllV1dkY1RDMXZh)、[②](http://210.45.242.91/kcms/detail/detail.aspx?recid=&FileName=BDZK200301016&DbName=CJFD0305&DbCode=CJFQ&uid=M3c3UzlxTUV1cE9mOVJpR20xYXV1ZVFHeGhnT3JwK2hhSVFOTnllV1dkY1RDMXZh)、[③](http://210.45.242.91/kcms/detail/detail.aspx?recid=&FileName=CSZY201507065&DbName=CJFDLAST2015&DbCode=CJFQ&uid=M3c3UzlxTUV1cE9mOVJpR20xYXV1ZVFHeGhnT3JwK2hhSVFOTnllV1dkY1RDMXZh)

**两篇化学论文：**



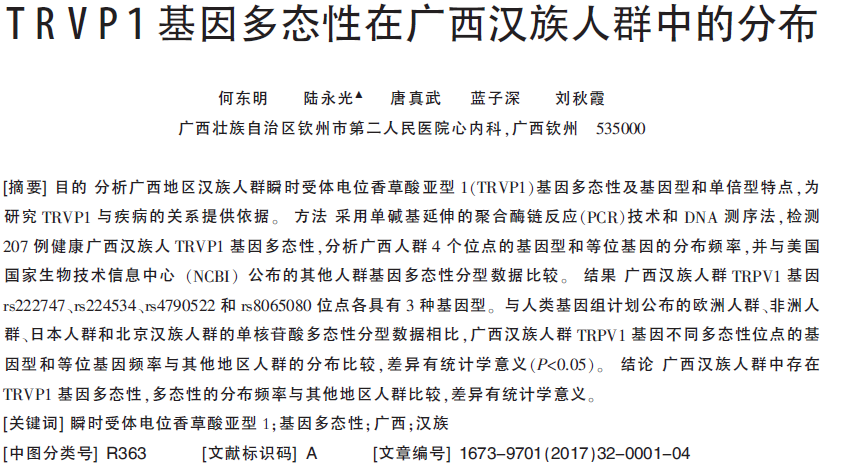
链接：[①](http://210.45.242.91/kcms/detail/detail.aspx?recid=&FileName=HXJZ601.001&DbName=CJFD9498&DbCode=CJFQ&uid=M3c3UzlxTUV1cE9mOVJpR20xYXV1ZVFHeGhnT3JwK2hhSVFOTnllV1dkY1RDMXZh)、[②](http://210.45.242.91/kcms/detail/detail.aspx?recid=&FileName=XDYQ200901040&DbName=CJFD0910&DbCode=CJFQ&uid=M3c3UzlxTUV1cE9mOVJpR20xYXV1ZVFHeGhnT3JwK2hhSVFOTnllV1dkY1RDMXZh)

***（读者可通过链接自行下载完整文献）***

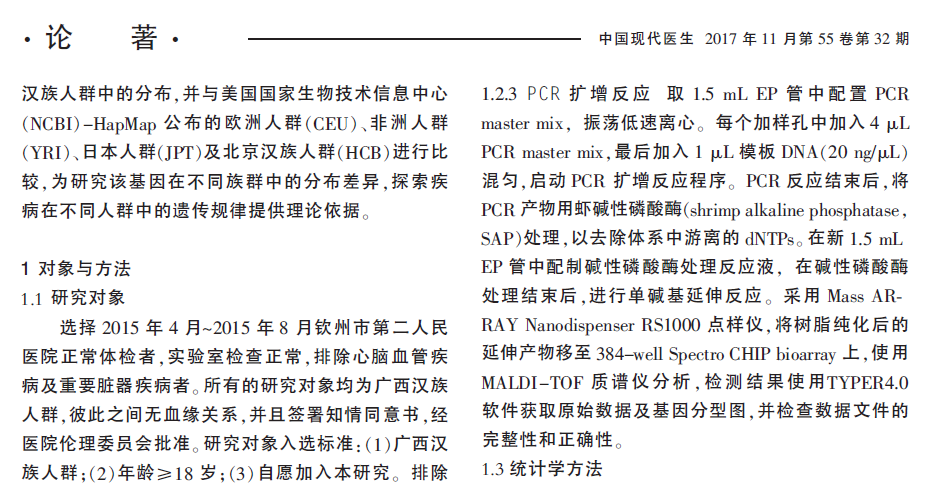
限于篇幅（实际上，我懒得展开篇幅），现取部分论文的部分内容进行展示：

读者注意里面有没有自己不认识的字即可。

我们先来看中文摘要部分：



再来看正文部分中专有名词较多的部分：



可以发现对于单个汉字而言，上面两张图的汉字我们大家都认识。

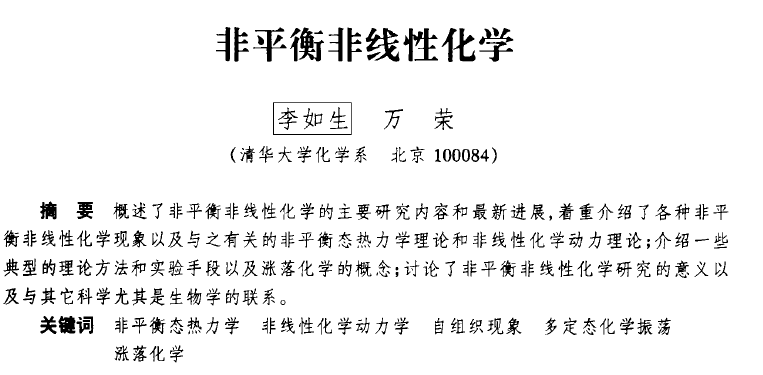
再来看一篇，我们直接看人名吧：



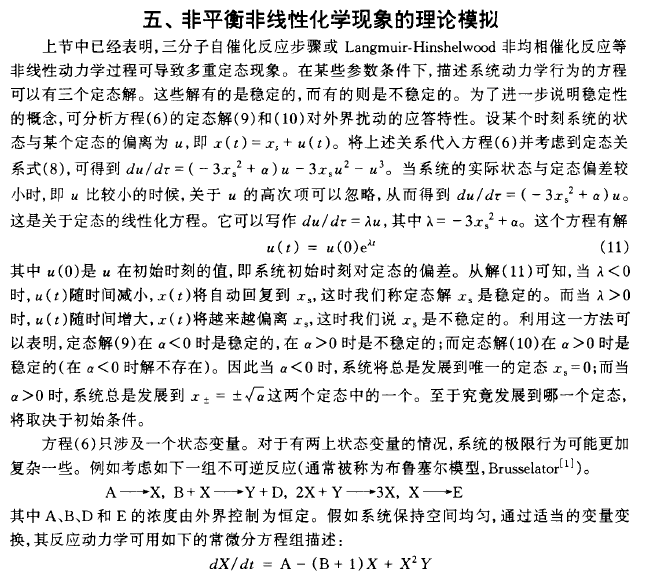
也许你会不认识这个由三个“牛”字组成的汉字，根据小鹤音形反查规则，我们反查一下：连续输入两个grave键，再连续输入两个n，再翻一页，就会看到带星号的犇字→所以这个字也是可以打出来的。而该字在论文《负性情绪对学习记忆能力的影响》中仅出现了一次。

一说到化学，我们非化学专业的人第一反应可能是元素周期表，里面有很多我们只会“有边读边”的字，也许化学的论文会有小鹤音形打不出来的字？那么我们来看化学的论文吧：

中文摘要部分：

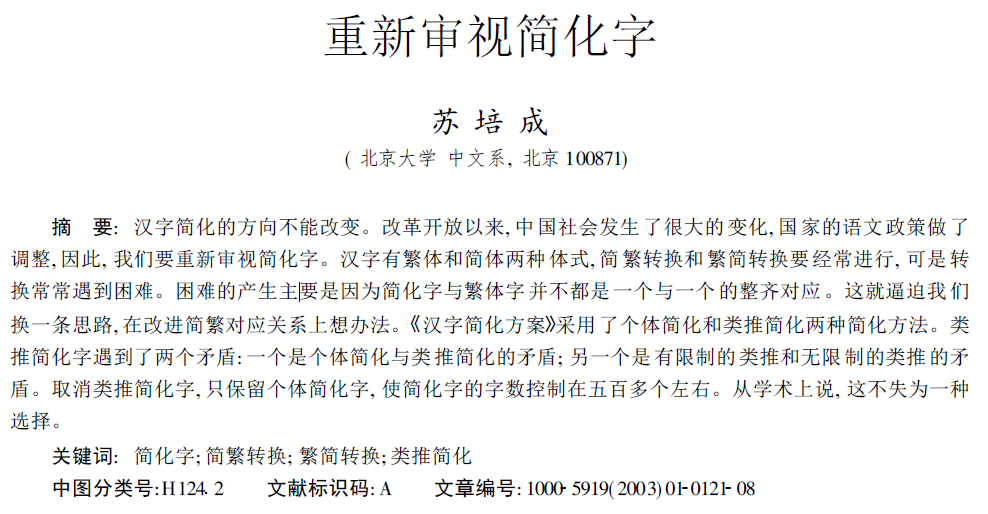


看正文中理论推导部分：

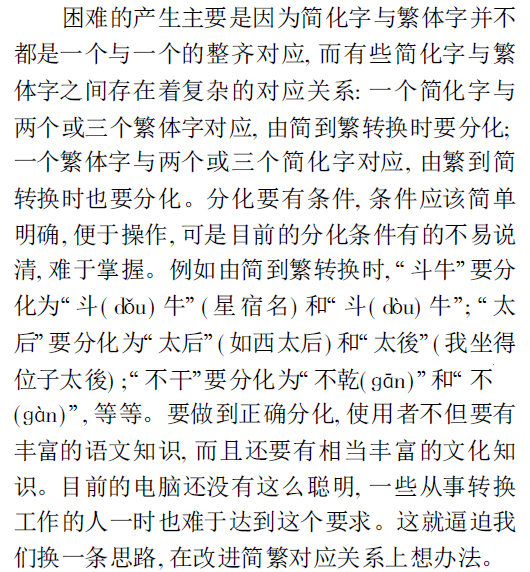


可以发现，我们还是所有的单个汉字都认识（尽管将其组合成词语则未必能看懂）。看到这里，读者们是否发现这些论文用得最多的是陈述句？

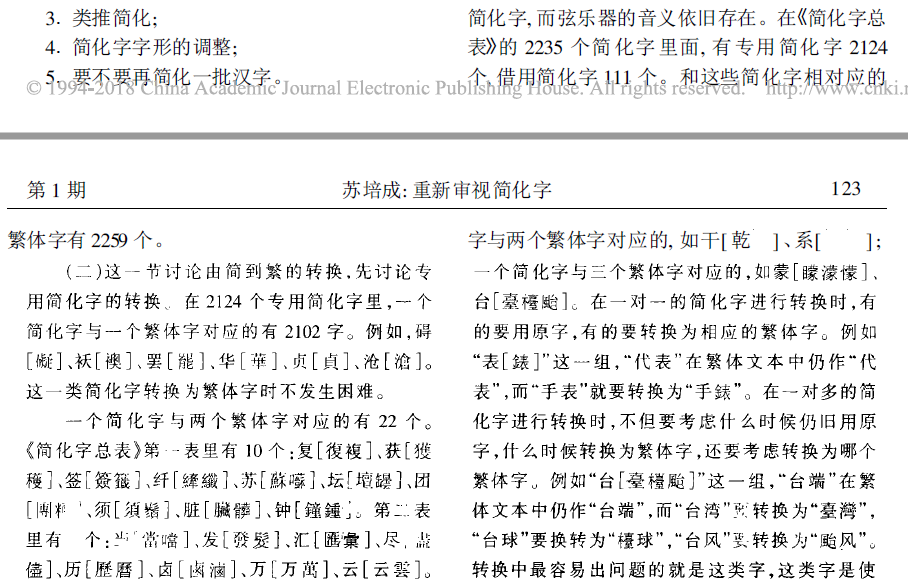
汉语言文学论文：



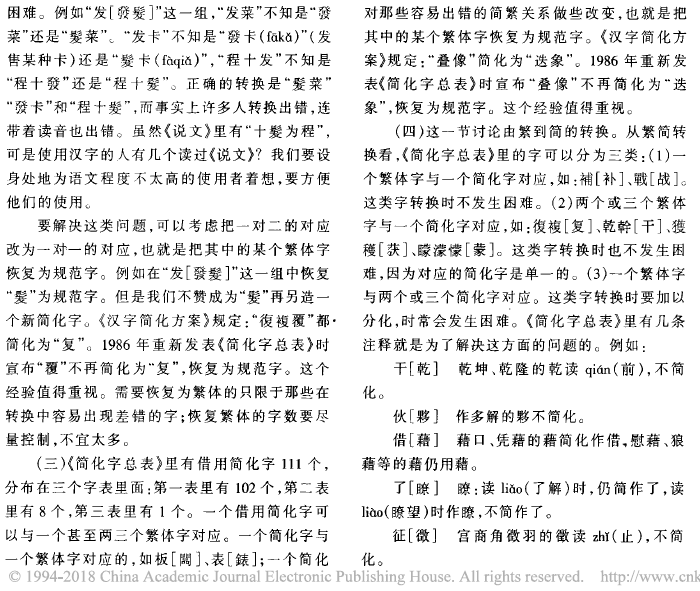
读者可能会觉得：咦，这篇文章讲简体字与繁体字，那么是否会出现很多生僻字，使得用小鹤音形打不出来呢？我们接着往下看：

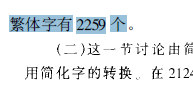


我们再往下看：

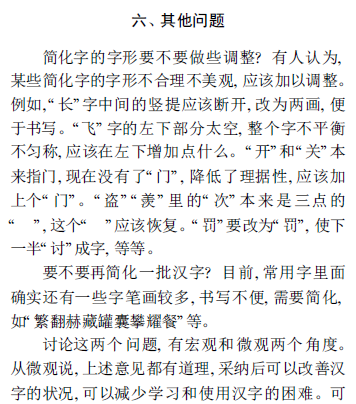


果然出现了很多繁体字，不妨再往下：



但请注意！仔细看之会发现字体突然变了，是什么原因导致的？读者下载pdf原文之后则会发现，这个突然变化的字体里的字和上面的字不同，它们无法复制，所以唯一的可能就是它们可能是直接从旧文献中截图过来的。可复制部分的字体为“繁体字有2259个”。

再往下，会发现又恢复正常字体了：

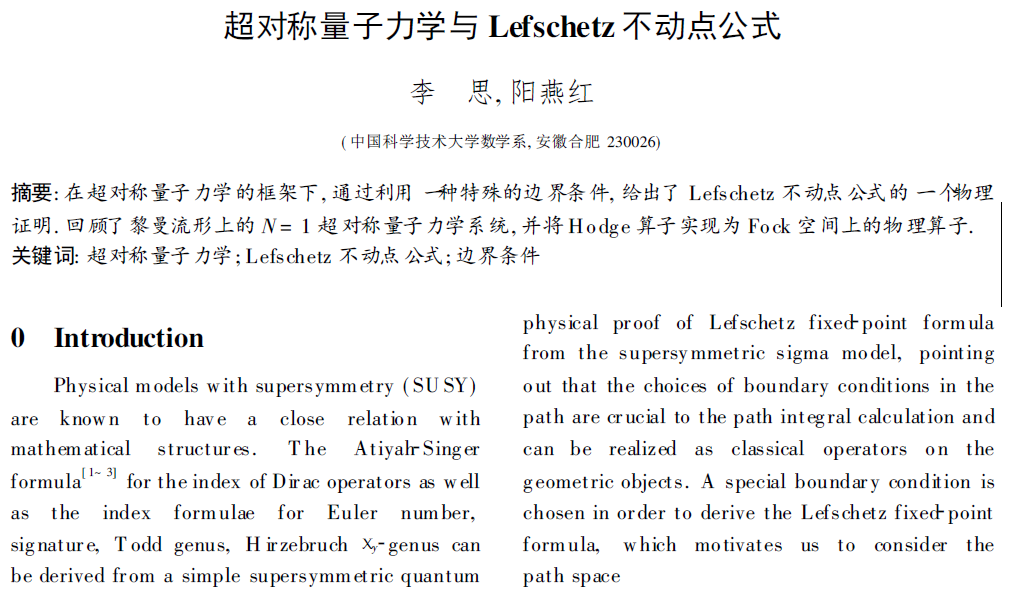


发现中间那段文字的末尾引号内的字笔画较多，将之搬运如下：

**如“ 繁翻赫藏罐囊攀耀餐”等。**

仔细一看又发现，其实这些字我们都认识。

物理论文：



注意到，上图中的中文摘要部分的名词大部分都用英文表示了，比如Lefschetz, Hodge. 这就大大避免了给名词起中文名的麻烦，除非这些定理全都是中国人发明的。

1. 下面取几篇论文的全文或部分段落用赛码器来测理论码长

注：实际码长可能会比该码长长也可能比该码长短，具体情况取决于使用者的实力；左边是只有**六万多**条词的小鹤双拼飞扬版官方词库，右边是个人版的加了**大词库**——20万条词的小鹤双拼。









再需要注意的一点是：因为是直接从pdf中复制的字，所以文字部分应该是正常的，而符号部分则不然，符号越多，赛码器测出的码长越短，但由于公式的复杂性，不可能把公式中的每一个符号都一键打出，所以实际上符号越多的论文的码长应该是远长于理论码长的。

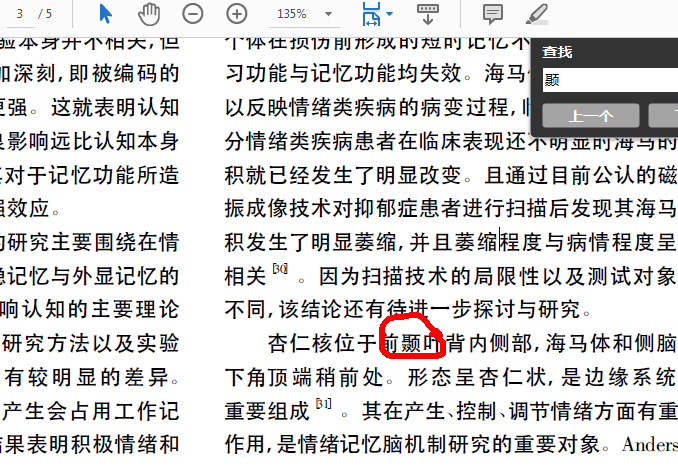
拿最后一图来说：测的是物理文献的“量子力学中的时间反演算子”中的所有文字及符号，里面的符号如左矢和**〈|** 和右矢符号 **|〉** ，需要先按shift键再按相应键。

可见如果只是单纯地加汉字类的专业词，码长收效甚微——从约6万词增加到约20万词，带来的收效仅仅是偶尔地码长比原始码长低约0.3,而这0.3的码长我用原始码表时完全可以通过提高键准[2]来弥补。有人可能会说：你所谓的大词库码表（晓岸版）加的并非全部是专业词，但请注意上面四张图也并非完全地同属一类专业论文。

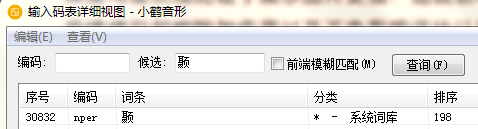
接下来我们来回答最前面的

**问题⑴：统计一下专业论文里的用字范围是否超越规范——**[**《通用规范汉字表》**](https://baike.baidu.com/item/%E9%80%9A%E7%94%A8%E8%A7%84%E8%8C%83%E6%B1%89%E5%AD%97%E8%A1%A8/9365175)**的收字范围？**

**我们从以上四张图中的选项“缺字”一栏（在右下角）可见，第二张图有缺字的情况，经查找，该字位于论文第三页，为“**颞**”**



我们用小鹤音形反查一下：



可见其实是有的，并不是缺字。那么就是赛码表的系统错误了。

那么现在可以基本肯定不存在缺字的情况，**问题⑴**解决了。

**问题⑵小鹤音形应该是安逸于官方词库而打单字还是加上专业词库而打词？**

**读者可见，专业论文一般不会是纯文字，尤其是理科类论文，而即使在加了词的情况下码长也不比官方小词库占多少优势。由于作者已经长期不用智能输入法，甚至是仅仅从感觉上也无法给出一个到底全拼输入法和纯双拼在云词库和智能算法的情况下到底会不会比小鹤音形更占优势，这点上希望能有人给出一些意见或建议。**

附录：

[1]码长：单个汉字的平均编码长度。

[2]键准：退格数按得越少，键准越高，零退格则100%键准。

附上完整的聊天记录：

