京昆字频统计及入声，尖团字统计

京剧和昆曲讲究一个字正腔圆，尖字团字，平上去入。学唱最开始就是要把字念对这些对于那些分尖团，入声的地方的人还算容易。但是对北方人来说就总是两眼一抹黑，尖团不分，就常常闹出笑话，像什么一箭射死盗令箭的公主（我不知道说的是谁，真不知道）。

我做这个统计的目的很单纯，因为我自己入声字还懂一些，尖团字实是分不清，于是想统计一下最常见的些尖团字，用心记一下。也是因为最近实在是闲的没事干，顺带着学一学Python的爬虫。

先从昆曲说起，昆曲分南北曲，南曲有入声，北曲入声则派入三声，像现在北方话里常见的‘学（xiáo）’，‘得（děi）’，‘落（lào）’等就是入派三声的例子。但是因为我~~又菜又懒~~能力不够，北曲里原是入声字的并没有按照入派三声统计，只一并算在入声字里面。一些多音字比如‘言谕人使从’的‘说（shuì/ʂui）’和‘训也，述也’的说（shuō/ʃyuɛʔ(t)）也没有进行区分1。同样，另外，曲谱中有一些繁简并用和异体字，除了一些很常见的合并字之外（如：發、髮，只、隻，雲、云，幹、乾、干，等），我把它们都转换为简体字后再进行统计。有些无法转换的罕见异体字因为本身很少见，对统计结果并不会造成显著影响。

昆曲里面最常见的20个入声字是：不，一，吓，的，得，说，日，只，出，月，著，玉，格，没，什，若，哈，勿，十，着。 这里面‘吓’和‘哈’在韵书上本有入声，但是在唱中时用作语气词，‘格’，‘勿’时用于苏白，‘说’和‘著’是多音字，并不能一概而论，需要视上下文来确定是否为入声。

昆曲里最常见的20个尖字是：小，相，请，心，将，姐，就，情，前，想，酒，亲，且，千，仙，清，先，西，新，进。

最常见的20个团字则是：吓，去，下，家，今，见，香，行，起，金，教，几，君，叫，向，休，间，军，气，喜。 同样，‘吓’多用于语气词，但是也有表‘惊吓’的团字，须视上下文确定。

对京剧来说，没有了入声的困扰，但同样的，入声字的读音也与普通话不尽相同，需要注意。 比如常见的‘说’，‘日’，‘白’，‘却’，‘百’，‘贼’等等。因为这些上口字并没有一个标准的字表，因此就不列出来了。

京剧里面最常见的20个尖字是：将，请，前，就，小，心，相，想，进，先，亲，姐，千，且，情，酒，谢，信，西，全。

20个团字则是:去，下，家，见，今，军，叫，兄，起，行，吓，金，君，贤，计，讲，江，气，驾，岂。

在做统计之后，又发现了一些有趣的事情，可以从字频的角度来对京剧和昆曲做一些剧情上的比较。

除了昆曲中多用于语气词的‘吓’，两个剧种字频前十的字是相同的。但是值得注意的是京剧中‘我’字排名第一，出现的频率是1.79%，而在昆曲中是1.23%，排名第三。 这个差别一部分来自于‘俺’字，昆曲中频率比京剧高出了0.13%，一倍的差距。另外也部分来自于昆曲中存在更多自称如‘小生’，‘老身’。

我们常说京剧是‘金戈铁马’、‘君臣父子’，而昆曲多是‘才子佳人’、‘卿卿我我’，这点在字频上也有所反映。 京剧中‘马’的字频为0.33%，而昆曲出行则多靠两条腿，‘马’出现的频率只有0.12%。同样，京剧中‘将’，‘王’，‘兵’，‘命’，‘国’，‘帅’的频率也要高出昆曲中许多。从字频差距上面也可以看出，京剧多是‘老爷们’的故事，‘兄’，‘弟’，‘爷’出现的频率比昆曲高，而昆曲中则是‘娘’，‘姐’出现较多。

再来看看京剧中的三国戏，出现最多的是‘张’，‘曹’，‘关’，这几个意义很多的字，而‘汉’，‘吴’，‘魏’，‘蜀’则分别出现了3855，2119，1640，和384次。这统计虽很粗略，也可以看出京剧中的确是演蜀汉更多。再来看看村夫专属的‘诸葛’二字。‘诸’出现了1853次，当然也包括‘诸位’，‘诸君’的用法。除去为数不多的‘葛登云’和‘瓜葛’二词，‘葛’则更能反应诸葛村夫的出场次数，京剧剧本中‘葛’字一共出现了850次，而孔明的‘孔’则出现了959次。 更有趣的是三顾茅庐的‘顾’字出现了877次，果然诸葛亮每次出场都要把这件事念叨一遍。

再来看看昆曲中比京剧中多了哪些字，首先差别最大的是‘花’，‘娘’，‘香’，‘春’和‘风’，频率差都超过了0.15%， 同时‘小姐’和‘小生’的频率也远远高出京剧。而拜没事就做梦的汤显祖所赐，‘梦’字在昆曲中比京剧也高出了0.14%。如我们之前所说，苏白里的些常用字‘勿’，‘格’，‘哉’，‘介’在昆曲中出现的频率也很高。

字频方面也可以反映出多成稿于明清两代的昆曲和清末的京剧在用字方面的差别，昆曲中‘吓’，‘阿’，‘吖’，‘咳’，‘呵’，‘﻿嗄’，‘咦’常用，而京剧中‘啊’，‘哎’，‘吧’，‘哈’，‘哦’，‘哇’比较常用。另外，昆曲中不常区分‘那’和‘哪’，而京剧中加以区分，这也可以从字频的统计中看出来。

另外，从姓氏上面也能看出一些京剧特有的剧目，比如之前所说的三国中‘曹’，‘刘’，‘孙’，‘张’，‘关’，‘诸葛’，还有像隋唐中的‘秦’，‘李’，杨家将的‘杨’，三侠五义中‘包’在京剧里面出现的频率较昆曲中要大。而昆曲则因剧目所限，并没有特别突出的姓氏，况且很多姓氏本身也有其他含义。硬是要说的话可能也只是柳梦梅的‘柳’和杜丽娘的‘杜’吧。

同时也要注意到一些技术上的问题，因技术所限~~就是菜~~，京剧剧本是选取了全部，其中自会有许多不同版本的同一剧目。而昆曲剧本则是挑选了唯一版本，这样本的差别自然也会导致对字频统计的误差。

注：

1. 韵典网，洪武正韵，说，https://ytenx.org/tcyts/sieux/1421/，https://ytenx.org/tcyts/sieux/2055/

京剧剧本取自中国京剧戏考，<https://scripts.xikao.com/>

昆曲剧本取自昆曲工尺谱数字化，<https://gongchepu.net/>

尖团字表取自乱语，尖团查询，<https://lsngb.com/ch/jiantuan/query-jian/>

以上已经取得了网站作者授权。

尚未取得授权的有：

入声字表:程序员小辉，入声字自动标注工具,<https://www.xiaohui.com/rsz/>

jqx声母汉字:在线汉语字典, <http://xh.5156edu.com/>

感谢以上所有的资料来源，同时也对录入剧本的众多志愿者们表示感谢。

源代码及数据可以前往Github获取，<https://github.com/WesleySiqi-96/jingkun-zipin-stats>。

所有数据均可以用于非商业用途，商业及学术用途请联系本公众号或直接发送邮件至 [siqitang96@gmail.com](mailto:siqitang96@gmail.com)

作者：**汤思奇**，博士在读，胸无大志，腹无经纶，除了正经事什么都懂一点。