

## Annotation Remarks

All the editions in the KuiSCIMA database are results of woodblock printing, a process involving manually carving out the shapes of each individual character. Similar to modern prints, errors when copying the source material may occur, and errors already included in the source material itself may be copied. In addition, when printing, some parts of the woodblock may already be worn or the ink is not distributed evenly, causing artifacts in the final print. Since *Baishidaoren Gequ* had been forgotten for over 300 years, before a manuscript appeared in the beginning of the 18th century, all the editions available today been copied from this manuscript in Qing dynasty. At this time, the *suzipu* notation was already uncommon and *ci* 词 poetry had long turned from a musical practice into a mostly literary genre.

These factors may result in anomalous shapes or misprints of the text or the *suzipu* notation characters, such that the meaning of certain musical symbols is unclear and needs further interpretation. Some of these instances are common to all or multiple editions in this dataset, while others are restricted to individual editions.

Therefore, in this file the case-by-case decisions when making the *symbolic dataset* and the *optical symbolic dataset* are listed.

### 0.1 Symbolic Dataset

The symbolic dataset consists of the *Wu edition* preserving certain features of the original editions as close as possible, and the *normalized edition* where pitches outside the indicated mode are corrected. If not stated otherwise, the following changes only refer to the *normalized edition*.

#### *Xinghuatianying* 杏花天影

Since many modern renditions of the piece, e.g., ([1], p.61), start with the pitch of 1<sup>st</sup> character “綠” is *He* 𪛗 instead of *Che* 𪛖. Therefore, the this change is made in both the Wu and the normalized editions.

#### *Zuiyinshangxiaopin* 醉吟商小品

Here, the pitch belonging to the 1<sup>st</sup> character “又” is changed from *Gou* 𪛗 to *Shang* 𪛚, and the pitch of the 22<sup>nd</sup> character “點” from *Gou* 𪛗 to *Che* 𪛖.

### *Nishangzhongxu Diyi* 霓裳中序第一

The pitch of the 28<sup>th</sup> character “索” is changed from *Gou* ㄌ to *Shang* ㄣ.

### *Danhuangliu* 淡黃柳

All of the *Gaowu* pitches (12<sup>th</sup> character “單”, 14<sup>th</sup> character “寒”, 39<sup>th</sup> character “攜”, 31<sup>st</sup> character “小”) are changed to *Wu* ㄣ.

### *Shihuxian* 石湖仙

All of the *Gaowu* pitches (44<sup>th</sup> character “今”, 88<sup>th</sup> character “槐”) are changed to *Wu* ㄣ. In addition, the pitch of the 11<sup>th</sup> character “衍” is changed from *Gou* ㄌ to *He* ㄣ.

### *Xihongyi* 惜紅衣

All of the *Gaowu* pitches (8<sup>th</sup> character “日”, 29<sup>th</sup> character “城”, 39<sup>th</sup> character “說”, 47<sup>th</sup> character “陌”, 70<sup>th</sup> character “沙”) are changed to *Wu* ㄣ.

### *Zhishao* 徵招

All of the *Gaowu* pitches (16<sup>th</sup> character “幾”, 64<sup>th</sup> character “不”) are changed to *Wu* ㄣ.

## 0.2 Optical Symbolic Dataset

As opposed to the symbolic dataset from above, the optical symbolic dataset contains symbolic notational and textual information together with the bounding boxes as an annotation to photographic reproductions of the historical editions. This dataset is the basis for the development and evaluation of OMR methods acting on *Suzipu* notation.

The individual editions are in the folders `optical_symbolic_dataset`, and they are named after the historical editions, i.e., `lu_edition`, `zhang_edition`, `siku_quanshu_edition`, `zhu_edition`. Inside each of these folders are two subfolders, one for the images called `images`, and one for the JSON files called `json`.

### *Geximeiling* 鬲溪梅令

*Siku Quanshu* edition:

- The 25<sup>th</sup> suzipu notation *Shang Xiaozhu* 么リ has no corresponding lyrics character. In the other editions, in this place there is a blank separating the two stanzas. This is probably an error, nevertheless this is included as such in the corpus to gain an additional instance.

***Xinghuatianshuying*** 杏花天影

Lu and *Siku Quanshu* editions:

- The notation of 21<sup>st</sup> character “風” indicates something resembling マ久, it is therefore annotated as マ.
- For the 35<sup>th</sup> character “燕”, the notation suggests 久久, which is annotated as 久フ.

Zhang edition:

- The notation of 4<sup>th</sup> character “拂” resembles フㇿフ and is annotated as *Gaowu Dingzhu* ㇿフ.
- Regarding the notation of 21<sup>st</sup> character “風”, see the remark for Lu and *Siku Quanshu* editions.
- Regarding the 37<sup>th</sup> character “舞”, the notation suggests 久久, which is annotated as *Liu Dingzhu* 久フ.
- For the 42<sup>nd</sup> character “人”, the notation is unidentifiable and annotated as *Yi Zhe* 一ㇿ.

Zhu edition:

- The notation of 21<sup>st</sup> character “風”, see the remark for Lu and *Siku Quanshu* editions.

Shanghai MS:

- The notation of 21<sup>st</sup> character “風”, see the remark for the Zhang edition.
- Regarding the 37<sup>th</sup> character “舞”, see the remark for Lu and *Siku Quanshu* editions.

***Zuiyinshangxiaopin*** 醉吟商小品

Zhang edition:

- The 3<sup>rd</sup> character “是” is corrupted, it is annotated as *Shang* 么.
- The notation of 8<sup>th</sup> character “暗” has an artifact, it is annotated as *Gong* フ.

***Yumeiling*** 玉梅令

Zhu edition:

- The notation of the 16<sup>th</sup> lyric “館” is unclear, it is annotated as *Gong Xiaozhu* フリ .
- The notation of the 33<sup>rd</sup> lyric “巳” is unclear, and annotated as *Si Zhe* マㄣ .

***Nishangzhongxu Diyi*** 霓裳中序第一

Lu edition:

- The suzipu notation of the 64<sup>th</sup> character “沈” is distorted and annotated as *Gong* フ .

Zhang edition:

- The notation of the 53<sup>rd</sup> character “亂” is distorted and annotated as *Gong Dingzhu* フフ .

Zhu edition:

- The notation of the 67<sup>th</sup> lyric “少” is noisy, it is annotated as *Si* マ .

***Yangzhouman*** 揚州慢

Lu and Zhu editions:

- The notation of the *suzipu* pair-character notation of the 73<sup>rd</sup> character “深” is unclear and annotated as *Gong Ye* フㄣ .

***Danhuangliu*** 淡黃柳

Lu and *Siku Quanshu* editions:

- The left side of the pair character notation of the 19<sup>th</sup> character “鷺” is unclear, and the annotation is chosen to be *Gong Ye* フㄣ .

Zhu edition:

- The notation of the 18<sup>th</sup> lyric “盡” and 20<sup>th</sup> lyric “黃” are unclear, they are both annotated as *Si* マ .

***Shihuxian*** 石湖仙

Zhang edition:

- The notation of 35<sup>th</sup> lyric “紅” is distorted, it is annotated as *Gong* フ .

### ***Shuying*** 疏影

Lu edition:

- The notation of 52<sup>nd</sup> character “花” is unclear, see the remark below for the *Siku Quanshu* edition.

*Siku Quanshu* edition:

- The 35<sup>th</sup> character “但” has an unclear notation, it is annotated as *Shang* 么 .
- The notation of 52<sup>nd</sup> character “花” is corrupted and annotated as *Gong Ye* フ .

Zhu edition:

- The notation of 97<sup>th</sup> character “曲” is unclear, it is annotated as *Shang Dadun* 么ㄅ .

### ***Xihongyi*** 惜紅衣

Zhang edition:

- The notation of 26<sup>th</sup> character “誰” is unclear, it is annotated as *Fan* リ .

Zhu edition:

- The notation of 21<sup>th</sup> character “碧” is distorted, it is annotated as *Si Dadun* マㄅ .

Shanghai MS:

- At the end of the piece, there is an additional notation symbol without a lyrics character. This is annotated as an empty lyrics box.

### ***Jueshao*** 角招

Lu edition:

- The first musical column contains 18 suzipu symbols, not matching the 19 Chinese lyric characters. The 12<sup>th</sup> lyric “垂” is annotated with an empty musical notation box in this corpus to establish a one-to-one correspondence between lyrics and musical notation.

Zhang edition:

- The 13<sup>th</sup> lyric “柳” has no musical notation next to it.

- The notation of 23<sup>rd</sup> character “湖” is very faint, it is annotated as *Gong* フ .
- The notation of 28<sup>th</sup> character “歸” is unclear, it is annotated as *Gou Ye* L J .

Zhu edition:

- The 8<sup>th</sup> lyric “西” has no musical notation next to it.

Shanghai MS:

- Regarding the 12<sup>th</sup> lyric “垂”, refer to the remark of the Lu edition above.

### ***Zhishao*** 徵招

Lu edition:

- The right side of the suzipu notation associated with the 17<sup>th</sup> character “何” is unclear due to noise, and it is annotated as *Liu Ye* 久 J .

Zhu edition:

- The notation of 54<sup>th</sup> character “相” has a dot-shaped artifact to the right of it. It is annotated as *Si* マ .
- The notation of 82<sup>nd</sup> character “高” is unclear, it is annotated as *Gou* L .

### ***Qiuxiaoyin*** 秋宵吟

Zhu edition:

- The notation of 51<sup>st</sup> character “銷” is ambiguous, it is annotated as *Gao Wu* ㄅ .

### ***Qiliangfan*** 淒涼犯

Lu edition:

- The 6<sup>th</sup> character “風” reads as J , and is annotated as *Gong* フ .
- The 32<sup>nd</sup> character “草” has no notation next to it.
- The 33<sup>rd</sup> character “寒” has unclear notation, it is annotated as *Shang* 么 .

*Siku Quanshu* edition:

- The 6<sup>th</sup> character “風” is unclear, refer to the remark of the Lu edition.

Zhang edition:

- The notation of 6<sup>th</sup> character “風” is unclear, refer to the remark of the Lu edition.
- The 10<sup>th</sup> character “一” is unclear, it is annotated as *Gong Dingzhu* フフ .
- The 53<sup>rd</sup> character is not recognizable, and the notation is unclear. The lyric is annotated as “小”, and the notation is annotated as *Si* マ .
- The notation of 72<sup>nd</sup> character “落” is not recognizable, it is annotated as *Shang Dingzhu* 么フ .

Zhu edition:

- The notation of 6<sup>th</sup> character “風” is unclear, refer to the remark of the Lu edition.
- The notation of 13<sup>th</sup> character “索” is unclear, it is annotated as *Shang Dazhu* 么カ .

Shanghai MS:

- Regarding the notation of 6<sup>th</sup> character “風”, refer to the remark of the Lu edition above.
- Regarding the notation of 10<sup>th</sup> character “一” refer to the remark of the Zhang edition above.

*Cuilouyin* 翠樓吟

Zhang edition:

- The notation of 65<sup>th</sup> character “戲” is too close to the remarks written next to it and therefore noisy. It is annotated as *Shang Dadun* 么ウ .
- The notation of 86<sup>th</sup> character “愁” is unusual, since it contains an extra stroke. It is annotated as *Shang Zhe* 么ㄣ .

Zhu edition:

- Regarding the notation of 86<sup>th</sup> character “愁”, refer to the comment on the Zhang edition.

## References

- [1] Xuehui, Zhongguominzu Guanxianye. *Xiao Kaoji Jiaocheng Di 1-7 Ji* 箫考级教程第 1-7 级. Beijing: Xiandai Chubanshe, 2021.