TEXTCOURT: Developing a Digital Approach to Chinese Court Drama[*]

Tian Yuan Tan,[1] Ewan Macdonald,[2] Ming Tak Ted Hui[3]

Abstract

Court theater was a core part of Chinese court and performance culture for centuries, flourishing in China's last two imperial dynasties, Ming (1368–1644) and Qing (1644–1911). Dramatic performances took place on multiple occasions such as imperial birthdays, weddings, state rituals, and seasonal festivals. Despite its significance for understanding Chinese court and performance culture, Chinese court drama texts have never received adequate attention. There are several reasons for this: the vast quantity of texts and the difficulty of accessing them; the low status of these texts in Chinese literary history as "authorless" performer's texts and "meaningless" court panegyrics; and the lack of any existing analytic framework to study anonymous Chinese performance texts at such a large scale.

Although some of the Qing court dramatic texts are reprinted in recent large-scale publication projects, many of these texts are still kept in separate archives and depositories around the world. To bring together such a vast amount of materials and develop an approach that will allow us to probe into this valuable resource, the TEXTCOURT project is creating the first digital archive dedicated to Chinese court drama. This paper will provide an overview of the project focusing on the creation of the database, namely the transcription of the dramatic scripts and the production of structurally marked-up texts following the guidelines of the TEI (Text Encoding Initiative). It will also provide suggestions on how the database could be used to pursue new types of research questions and present some of the preliminary findings based on the application of computational methods.

Keywords: Chinese court drama, digital humanities, Text Encoding Initiative (TEI)

---

[*] This article is part of the "Linking the Textual Worlds of Chinese Court Theater, ca. 1600–1800" (TEXTCOURT) project that has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement 819953).

[1] Shaw Professor of Chinese, University of Oxford.

[2] Junior Postdoctoral Researcher, University of Oxford.

[3] Postdoctoral Research Officer – TEXTCOURT, University of Oxford.

1. Introduction

Court theater was a core part of Chinese court and performance culture for centuries. The Office of Music Instruction in charge of songs and drama for court performances was set up as early as the Tang dynasty (618–906). Court theater continued to flourish in China's last two imperial dynasties, Ming (1368–1644) and Qing (1644–1911). Dramatic performances took place on multiple occasions such as imperial birthdays, weddings, state rituals, and seasonal festivals. At its height, during the reign of Emperor Qianlong (1736–1795), the imperial court was well known for its extravagant theatrical productions, vast number of court musicians and actors, and grand theater stages. Recent studies have provided us with rich accounts of these theatrical practices and institutional organizations in the late imperial Chinese court.[4] Yet the voluminous body of drama texts produced on these occasions have never received adequate attention despite its significance for understanding Chinese court, performance, and literary cultures (Tan 2022: 10-13). There are several reasons for this: the vast quantity of texts and the difficulty of accessing these court drama scripts; the low status of these texts in Chinese literary history as "authorless" performer's texts and "meaningless" court panegyrics; and the lack of any existing analytic framework to study anonymous Chinese performance texts at such a large scale.

Although some of the Qing court dramatic texts are reprinted in recent large-scale publication projects, many of these texts are still kept in separate archives and depositories around the world. Identified records and archival catalogs suggest that there is altogether a minimum of over 20,000 texts related to Chinese court drama. To bring together such a vast amount of materials and develop an approach that will allow us to probe into this valuable resource, the TEXTCOURT project is creating the first digital archive dedicated to Chinese court drama. This paper will provide an overview of the project focusing on the creation of the database, namely the transcription of the dramatic scripts and the production of structurally marked-up texts following the guidelines of the Text Encoding Initiative (TEI).[5] It will also present some of the preliminary findings based on applying computational methods to the literary data collected in the archive.

---

[4] In addition to the fast-growing body of Chinese and Japanese scholarship, there are two English-language monographs that examine Qing court theater specifically. See Ye 2012 and Chen 2021.

[5] https://tei-c.org/guidelines/

2. Creation of the First Digital Archive for Chinese Court Drama

Early studies of court drama are often disjointed in that they are usually conducted by librarians and researchers affiliated with various repositories and hence mostly dealt only with local texts, collections, and sources, making little or no references to, and more critically, having no knowledge of similar texts (sharing titles or content) in other archives. The disjointed nature of previous scholarship can be exemplified in Wu Xiaoling's discovery in 1982 of "a rare court drama of six scenes" in Princeton University's Gest Collection.[6] Wu is one of the most senior bibliophiles in the field of Chinese drama, but given the lack of a single master catalog listing all court drama collections (even now, WorldCat only shows one copy of the text in question), he could only speculate that the text may be the only extant copy, based on his personal knowledge (Wu 1986: 1). Researchers must comb the catalogs of individual libraries and archives. However, consulting catalogs alone still does not solve a more fundamental issue: Chinese court drama texts are titled very casually and therefore one cannot rely only on catalogs. In fact, in our research we have stumbled across two more copies of the same text kept in other archives in China under a different title. One copy, titled collectively as *Qianlong Andian ben* 乾隆安殿本, will be made available on the TEXTCOURT website from June 2022. Such problems demonstrate the extreme lack of tools for researchers facing this vast corpus. There is a critical need to link this vast amount of research material across archives and regions in order to arrive at a comprehensive knowledge and understanding of the material.

Not only are some court drama scripts titled differently in various collections, they also survive in varying forms, and have diverse characteristics: the majority of these scripts come in manuscript, with a small number in print; some in fragments and others in their entirety; short plays of only a single scene and bumper court plays of 240 scenes; some were copied for reading purposes and others, in the nature of prompt books, mainly used for staging and performance. For instance, in one of the many dramatic scripts gathered by the renowned scholar Wu Shuyin 吳書蔭, one can find a play entitled *Zaochun chaohe* 早春朝賀 (Congratulations to the Court in Early Spring). The script is in manuscript form with traces of amendments and changes included in the play (see Figure 1). Sometimes, scholars may produce collated modern typeset versions of court drama manuscripts to facilitate better comprehension of these texts. Yet, unlike a printed version of court drama, a manuscript filled with revision

[6] These "scenes" are in fact independent plays, organized into a collection and perhaps to be performed on a specific occasion.

details requires a different treatment since scholars interested in the editing process of a play may hope to study how texts are emended. A typeset version of the text may mask over these editorial details.

To bring together such a diverse group of materials from multiple sources and to connect the disjointed textual worlds of court drama, the TEXTCOURT project is building the first digital archive of court drama scripts. By digitizing manuscripts and printed editions, the project is producing an electronic textual corpus that will be accessible on the Internet for all researchers and the general public.

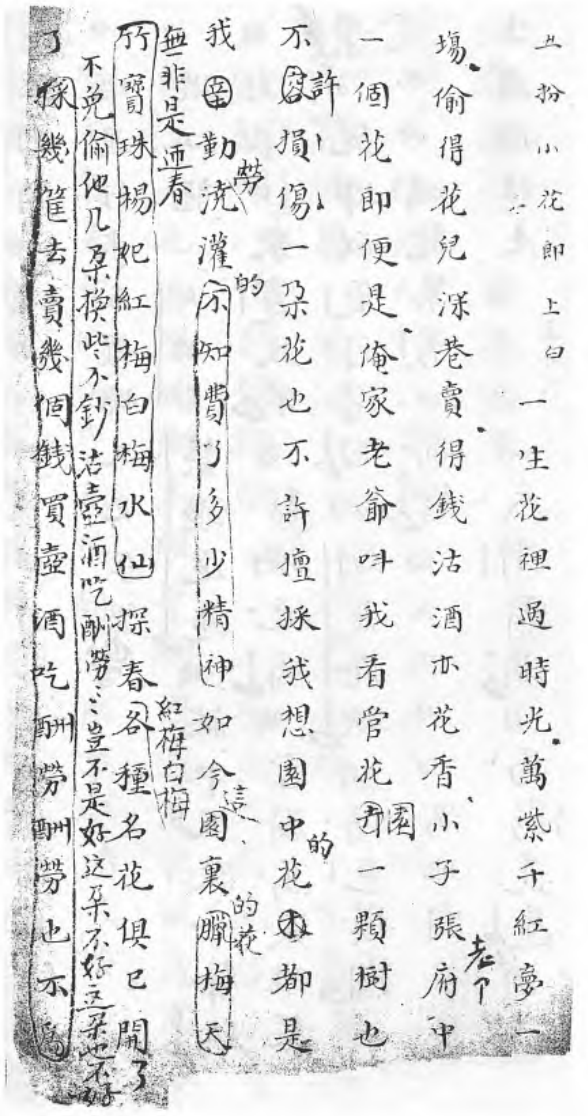

Figure 1. A Photocopy of the Manuscript of the Play *Zaochun chaohe*

Source: Wu Shuyin ed. *Suizhong Wushi cang chaoben gaoben xiqu congkan*. Beijing: Xueyuan chubanshe, 2004.

2.1 Collection and Preparation of Drama Scripts

The project began with locating the sources of the Chinese court dramas printed in several collectanea. For Ming dynasty court plays, we consulted the extant Ming inner palace editions (*neifuben* 內府本) of drama manuscripts in the Maiwangguan collection that can be dated to the 1610s.[7] There are, according to the study of Sun Kaidi, altogether 95 plays that are linked to the performances in the imperial court (Sun 1953: 77–85). Apart from these plays, there are also other plays composed by other literati to celebrate the Ming dynasty rulership.

As for the plays produced in the Qing dynasty up till the 19th century, there are a series of reprints which serve as a base for the archive:

A. *Gugong zhenben congkan* 故宮珍本叢刊 (Collectanea of Precious Books from the Forbidden Palace) published in 2001
B. *Suizhong Wushi cang chaoben gaoben xiqu congkan* 綏中吳氏藏抄本稿本戲曲叢刊 (Collectanea of Drama Manuscripts and Drafts in the Holdings of Scholar Wu of Suizhong) published in 2004
C. *Higashi Ajia zenpon sōkan* 東アジア善本叢刊 (Collectanea of Rare Books of East Asia) published in 2009
D. *Fu Xihua cang gudian xiqu zhenben congkan* 傅惜華藏古典戲曲珍本叢刊 (Collectanea of Precious Books of Ancient Dramatic Texts in Fu Xihua's Holdings) published in 2010
E. *Zhongguo guojia tushuguan cang Qinggong Shengpingshu dang'an jicheng* 中國國家圖書館藏清宮昇平署檔案集成 (Documentaries of the Qing Court Institution Shengpingshu in the Holdings of the National Library of China) published in 2011
F. *Gugong bowuyuan cang Qinggong Nanfu Shengpingshu xiben* 故宮博物院藏清宮南府昇平署戲本 (Drama Texts from the Qing Court Theatrical Institutions Nanfu and Shengpingshu in Palace Museum's Holdings) published in 2015-17

[7] This is the *Maiwangguan chao jiao ben gujin zaju* 脈望館鈔校本古今雜劇 (*Zaju*, Old and New, Copied and Collated, from the Studio of the Transformed Bookworm). For an introduction of these plays from the Ming inner palace, see Tan 2016: 96–107.

These materials cover a range of drama scripts produced in the Qing court for the imperial family. Most of these texts are anonymous and some are likely to have been adaptions from earlier court plays. In addition, there are a series of local plays written by the Qing elites for the imperial entourage when the emperors embarked on inspection tours. For instance, the project includes scripts composed by the renowned literati Cao Yin 曹寅 (1685–1712), Li E 厲鶚 (1692–1752) and Jiang Shiquan 蔣士銓 (1725–1785), who were all commissioned to write for the emperor. Some of these scripts were never reprinted in any of the aforementioned large collectanea, and can only be located via library catalogs. With the help of our partners in Beijing Foreign Studies University (BFSU) and the assistance of the library at Oxford, the project has managed to locate some extra items that have never been digitized before. After locating all these materials from various sources, the project proceeded to record the metadata of all these scripts in a catalog. To date, thanks to the help of our partners, the project has managed to locate 2,000 drama scripts all around the world.

In the process of organizing these scripts into a digital catalog, the team has assigned each play a unique code to aid in disambiguation. Among the plays the team collected, a number of these plays share similar titles: while some of them are different editions of the same play, some have completely different content. On the other hand, there are also cases when multiple versions of a same play are given a different title. The unique code thus helps the team, and other researchers, in navigating the complex relationships between different plays.

As mentioned above, the project has obtained the metadata and images of 2,000 scripts. In order to retain the bibliographical information of each text and allow the team members to preserve all the additions and deletions shown in the manuscript, we have selected a sample of 200 plays (as our Core Database) to convert into XML format by encoding them according to the Text Encoding Initiative (TEI) guidelines.

2.2 Preparation of XML Files

The foundation of the project is the first-ever digital archive of court drama scripts. While many court drama texts are available as scanned images, these do not contain searchable text, and so the first stage in the digitization process is to work with our partners at institutions including SOAS University of London to scan texts without existing available digital images. The project team then transforms the images of the script into a digital text format. Although over the years, more advanced systems have been developed to recognize Chinese printed texts (see, for

instance, Sturgeon 2018), the drama scripts gathered by the project team have been so diverse that they demand different treatment. In general, the team uses an Optical Character Recognition (OCR) program to obtain a rough digitization when the script gathered is a reasonably clear printed text. In general, these rough copies still require extensive manual post-editing to ensure that the digitized text is representative of the original source. As for the handwritten manuscripts in our core database, OCR is not suitable and so the entire text has to be transcribed manually. A "diplomatic transcription" method has been adopted to preserve the original presentation of the text – this means we seek to copy the sources as seen with minimal editorial intervention. In the context of this project, the main implication of this approach is that all character variants are reproduced as written in the source materials, although very rare handwritten variants not included in Unicode are silently amended to their standard equivalents.

While processing the OCR texts or transcribing the scripts, the team members follow a set of guidelines to ensure consistency and facilitate subsequent processing of the files. For example, the tune titles mentioned in each script are always marked with block brackets, and the stage instructions are always marked with round brackets. All deletions and additions are also marked in the digital transcriptions to ensure that all similar features in the manuscripts are captured in a standardized way. In the last step of the conversion process, the transcriptions of these scripts are then converted into a digital edition made available as TEI XML for download from our repository in Figshare (a platform where users can publish and share their research data), or access on the internet through our project website.

2.3 Tagging: A Balance between Diplomatic Transcription and the Ability to Search

The vast amount of material and the nature of this research project make it a highly suitable subject for the development of a digital archive. Automated information extraction, full text search, textual comparison, parallel passage identification, and the automatic highlighting of silent citations facilitate close readings of the texts on a scale far larger than would be possible through relying on individual reading and analysis alone; these technologies also help reveal characteristics of the texts that would otherwise remain hidden. As a complement to such digital-assisted close reading techniques, the practice of "distant reading," developed by Franco Moretti and involving the aggregation and analysis of large amounts of data, is a powerful additional route to approach the vast corpus of court drama scripts. The database created by the project will underpin the project's combination of digitally-assisted close reading and distant reading approaches.

Yet, a prerequisite for applying these computational methods to the scripts is to make sure the computers understand the text and identify the patterns within the data. This could be rather difficult given that all the scripts in the TEXTCOURT archive are transcribed in a manner where all the variant characters and scribal errors are kept. One good example here would be the name of the immortal Lü Dongbin 呂洞賓. His name appears multiple times in various plays yet the characters used often vary: very often the character *lü* 呂 is written without a stroke in the middle (*lü* 吕); similarly, the character *bin* 賓 is frequently written as *bin* 賨. In this case, given that our archive reflects the manuscript as accurately as possible, we will have at least four possible combinations of Chinese characters all referring to one single person in our archive. Without further intervention this would constitute a huge problem for the users of our digital archive and prevent us from applying any digital methods onto the drama scripts. In order to solve this problem, we assigned a unique code to represent Lü Dongbin (and every other entity) in the XML files, to indicate to the computer that the many different ways of writing "Lü Dongbin" (呂洞賓, 吕洞賓, 呂洞賨 or 吕洞賨) all refer to a single person. Here we will turn to the development of the TEI coding in our project.

3. Developing a Coding System for Chinese Drama Scripts

TEI is a consortium aiming to maintain an international standard for the representation of texts in digital form. Although the TEI guidelines have dedicated a specific chapter to the encoding of performative texts,[8] the project seeks to also use the tags provided by the TEI to relink the fragmented worlds of Chinese court theater and enable the users to navigate the plays more efficiently. Hence, while the project team follows the tag sets and rules specified by the TEI guidelines, we are also using the tags to facilitate research on four areas, namely people, places, occasions, and objects. We believe that while having only a small selection of enriched texts does not necessarily reveal anything that would not be apparent with traditional close reading, when we scale the corpus up and group entities thematically, we will be able to draw links between court theater texts that would remain invisible without digital methodologies. In addition to facilitating the known research interests of project members, we hope that the encoded database will stimulate new and as-yet unknown research questions from other researchers in the future.

[8] https://tei-c.org/release/doc/tei-p5-doc/en/html/DR.html

In this section, we will outline how we design our markup schema and tag different entities in our archive. The markup process is divided into two sections: (1) structural markup, the process with which the main structural features of a play are tagged; and (2) content markup which focuses on tagging the four types of entity mentioned above.

3.1 Structural Markup

Many existing applications of TEI focus on encoding lexical and grammatical divisions to facilitate automated linguistic analysis: in our project we are most interested in content markup, and so we keep structural markup to a bare minimum. In our database, the main structural features tagged are act divisions, stage directions, line breaks, and *qupai* (tune titles), alongside other less common features such as marginalia, additions, deletions, and illustrations. As the OCR transcription process followed a consistent format throughout, we can apply these structural tags to the XML files automatically, using regular expression search and replace functions. In TEXTCOURT, the aim of the structural markup is to create a valid TEI file and support the research enabled by the content markup.

The chief challenges in the structural tagging process are related to the appropriate encoding of structural features of Chinese drama in TEI. The scripts we are dealing with in our project span a period of two centuries and are drawn from a diverse range of sources – most are manuscripts, and many have complex amendments, marginalia, and other textual features added over the years. As the scribes wrote down these scripts, they enjoyed almost complete flexibility in how the texts were laid out and written down. As we digitize their work today, however, the hierarchical nature of TEI XML is rather inflexible in contrast.

Compounding this difficulty is that although the TEI guidelines contain a module for drama, this has been designed with Western drama in mind and does not always make for a natural fit with Chinese drama. Despite this, it is still the best choice for our project as it constitutes a universal standard with a degree of compatibility with other projects using encoded dramatic texts. Moreover, features such as acts/scenes,[9] stage directions,[10] and dialogue are universal, and as such are simple to tag using existing TEI elements from the drama module. However, there are certain unique features in Chinese drama which are much more complicated to encode.

[9] https://tei-c.org/release/doc/tei-p5-doc/en/html/DR.html#DRDIV
[10] https://tei-c.org/release/doc/tei-p5-doc/en/html/DR.html#DRSTA

For example, arias are one of the key features of Chinese drama, but these can be awkward to encode in TEI. Arias can be conceptualized as a "text within a text" – a semi-independent unit within the drama. Therefore, we would ideally need to tag the aria lines to differentiate them from the surrounding dialogue and stage directions and tag the tune title so that this is recognizable as a heading. Let us take a look at the aria below:

【得勝令】	(To the tune *Desheng ling*):
呀	Alas!
白蘇堤畔起風清	A clear wind arises near the Su and the Bai Causeways.
恰遇着佳日正晴明	By chance, this is met with the brilliant sun shining bright.
娟娟的花如笑圍幽徑	The graceful flowers encircle the secluded paths with a smile,
嚦嚦的鶯似歌隔翠屏	While the chirping orioles sing behind the azure screen.
輕盈	Nimble and light –
好時光人同興	This fine time is celebrated by everyone.
清寧	Clear and serene –
好春光天作成	This exquisite springtime is created by Heaven.[11]

The "line group" <lg> element in TEI is designed for roughly this purpose. In this example, we can encode each line with the <l> element and the tune title would be tagged as the heading. To distinguish these tune titles from other headings, we uses the attribute "type" as an indicator. The aria above would thus be encoded as follows:

```
<lg type="aria"> <head type="qupai">【得勝令】</head>
    <l>呀。</l>
    <l>白蘇堤畔起風清。</l>
    <l>恰遇着佳日正晴明。</l>
    <l>娟娟的花如笑圍幽徑。</l>
    <l>嚦嚦的鶯似歌隔翠屏。</l>
    <l>輕盈。</l>
    <l>好時光人同興。</l>
    <l>清寧。</l>
    <l>好春光天作成。</l></lg>
```

[11] This aria is taken from the play *Wanshou tu* 萬壽圖 (A Picture of Eternal Life). The play is now made available in the TEXTCOURT website: https://textcourt.orinst.ox.ac.uk/database/scripts/Q00193_01_A/?tab=info. All English translations are mine unless otherwise specified.

Although some of the words may lie outside the prescribed metrical structure of the tune pattern (i.e. padding words 襯字), in our design of the schema, they are still treated as part of the aria. While this encoding principle may be applied when the aria is delivered as a discrete unit, this encoding does not work where the aria is interspersed with spoken dialogue and/or stage directions. In our archive, we can often find instances like this:

【勝哉薑】 (To the tune *Shengcai lu*):

扮着箇治國齊家專務本 I pretend to be one who regulates my family,
governs the state,
and directs my aims to what is fundamental

（殿頭官云）老太監若是到於西洋海夷向化，小官奏知聖人，自有重賞加官也。

(*The court official says:*) If you, the senior eunuch, went to the Western Seas to cultivate the barbarians across the ocean, let me, this petty official, report back to the Sagely Emperor. Certainly, you will be highly rewarded and bestowed a high office.

（唱） (*He sings:*)

又不圖重職管軍民 Yet I am not craving for a high office
to supervise the armies and civilians,

則要俺君王聖意欣 I just hope that my Sagely Ruler will be delighted.

也不怕江翻海沸 Neither am I scared of the startling rivers and turbulent seas,

波濤千尺 Nor the swelling waves which surge for a thousand feet.

憑着我忠義應天真 This is because my loyalty corresponds with the truth of Nature.[12]

Here, the hierarchical structure of XML prevents an easy solution. In a comparable project, the Chinese Classic Drama Database (CCDDB) from Keio University, Japan, the encoders have chosen to treat interspersed text as part of the aria hierarchy.[13] As can be seen in the example here, in the HTML encoding used by CCDDB, each aria is encoded with a <p class="曲"> tag; stage directions and interspersed dialogue are both encoded with tags and distinguished by differing “class” attributes. This solution has the advantage that it is still possible to retrieve only the lines of verse in any particular aria by filtering out the unwanted tags.

[12] This excerpt is taken from the Ming play *Feng tianming Sanbao xia xiyang* 奉天命三保下西洋 (By Imperial Command, Zheng Sanbao Sets Sail for the Western Oceans). Punctuations are added for clarity. For the details of the play, see https://textcourt.orinst.ox.ac.uk/database/scripts/M0004_01_A/?tab=info.

[13] See here for an example: http://ccddb.econ.hc.keio.ac.jp/wiki/index.php?title=%E5%85%83%E6%9B%B2%E9%81%B8/%E9%B4%9B%E9%B4%A6%E8%A2%AB/02%E7%AC%AC%E4%B8%80%E6%8A%98/08%E6%9B%B2&action=edit

However, CCDDB is based on HTML and TEXTCOURT uses TEI; TEI has particular rules on what tags may be placed within <lg> elements. Therefore, while we can nest stage direction tags within the <lg> element, there are no elements appropriate for dialogue that can be inserted. Nor can we employ a similar solution to CCDDB and add a “type” or other attribute to specify that the <l> tags actually enclose dialogue, as TEI does not permit this. The solution provided for in the guidelines for cases such as this is slightly more complex, involving the creation of two separate <lg> elements. These are then given a unique identifier by means of the @xml:id attribute and highlighted as partial arias by means of the @part attribute. The two parts of the aria are then linked by a subsequent <join> element:

```
<lg xml:id="AR1" type="aria" part="Y"><head type="qupai">【勝哉蘆】</head>
<l>扮着箇治國齊家專務本</l></lg>
<stage>殿頭官云</stage>
老太監若是到於西洋海夷向化小官奏知聖人自有重賞加官也
<stage>唱</stage>
<lg xml:id="AR2" type="aria" part="Y">
<l>又不圖重職管軍民</l>
<l>則要俺君玉聖意欣</l>
<l>也不怕江翻海沸</l>
<l>波濤千尺</l>
<l>憑着我忠義應天真</l></lg>
<join type="lg" target="#AR1 #AR2"/>
```

In view of the workload involved in creating unique identifiers for each partial aria, however, and due to the specific requirements of the TEXTCOURT project for tagging at scale, the guiding principle of diplomatic transcription, and the project’s focus on content markup, we have taken the decision to initially tag only the tune title (*qupai*) with empty line group elements, like this:

```
<lg type="aria"> <head type="qupai">【勝哉蘆】</head><l></l></lg>
```

This approach creates a framework within which the individual lines of the arias can be tagged in future as an enrichment to the database. The reasoning behind this decision primarily stems from the principle of diplomatic transcription dictating that we do not add punctuation marks where these are not present in the source materials. While some plays in our corpus include these, the majority do not, and even where they are present there is no distinction between

punctuation marks indicating the end of a verse line and those indicating the end of a sentence of dialogue. This means that there is no consistent marker in the edited files indicating where one line ends and the next begins that can be leveraged as a basis for automated tagging. The scale of the project requires the use of automated structural tagging as far as possible; manual tagging of all aria lines in the corpus is currently impractical due both to time constraints and the fact that the time so invested would not contribute directly to the core research goals of the project. It is possible to envisage a system that could tag some lines automatically when provided with the tune title and the metrical pattern and rhyme scheme associated with it. However, given all the complexities of interspersed text, filler words, and natural inconsistencies/variations in the metrical patterns, the development of such a system with an acceptable degree of accuracy would be a significant project in of itself.

We hope that our approach to the structural TEI tagging of Chinese court drama texts may be applied and refined by future projects related to the digitization of Chinese and other drama texts, especially those from non-Western traditions which may have structural features that are challenging to encode in TEI.

3.2 Content Markup

The requirement for scale and the implications of diplomatic transcription also guide our approach to tagging the entities in the scripts. This tagging underpins the functionality of the core database and is aimed at allowing us to reveal links between the people, places, objects, and occasions mentioned in court drama scripts; the four types of entity in the project mirror these four links. To provide more flexibility in how users access the information in the database, these four types are each further divided into categories: for example, the Places type contains the categories China, Asia, Europe, and Mythical. Each entity is assigned a unique internal code to serve as an authority control for cases where one entity has multiple alternative names. The first step in content tagging, however, was gathering a list of entities to mark up. The project consists of a large collection of specialized and diverse fictional texts; this means that they do not make for a natural fit with existing external authorities and tools. For example, the online tool MARKUS enables the automated markup of texts based on a selectable range of external databases, and is highly effective way of performing such markup on non-fictional texts. However, when in trials we used MARKUS to mark up entities of the People type based on the China Biographical Database (CBDB), we found that as most of our entities are fictional characters we ended up with a large number of false positive matches. Fictional characters were

tagged as historical figures that they shared a name with. Often there were multiple possible CBDB entries found for one name, requiring a large amount of manual post-editing to work out which if any was correct for the entity.

More importantly, relying on authorities such as CBDB, which are based on historical materials, would result in anonymous characters being missed out entirely despite these holding significant research interest. For example, the figure of the humble pedlar (*huolang dan* 貨郎擔) would not be included in any external biographical database, but they and other humble townsfolk form a significant potential avenue of research into court drama.

Compounding these issues is the factor of variant characters: as we have shown in the above discussion, following a diplomatic transcription practices results in having a wide range of sometimes very rare variant characters employed inconsistently throughout the corpus, to search for and tag entities we need to know not only the name of the entity, but also how it is written in each individual drama script.

To solve these issues, we decided that we had to create our own project entity lists. This step was done manually during the OCR editing process. The editor recorded the characters, places, occasions, and objects appearing in each script verbatim, to ensure that variant characters were caught. These lists were then cleaned by screening out duplicate entries and linking absolute synonyms before each entity was assigned a TEXTCOURT code. For example, entity RS00007 is Lü Dongbin. By assigning this code, we will be able to pull up all the instances when the figure appears in the digital archive no matter which variant names are used (be it written as 呂洞賓, 吕洞賓, 呂洞賨 or 吕洞賨). The code also links his personal name Lü Yan 呂嵒 as well as his religious name Dongyangzi 東陽子, so that all occurrences of this figure can be retrieved with ease. We maintain a list of all these synonyms, and once this list is complete, the markup itself can be accomplished relatively efficiently with a series of regular expression search and replace functions.

Given that the aim of the project is to reveal links between different court drama scripts, we took the further step of assigning entities to one or more "analytical groups". By doing so, we seek to provide an additional optional method for database users to sort and retrieve information, and link a greater number of scripts than is possible through individual entities. For example, we assigned Southeast Asian characters and countries to the analytical group "南洋諸國".[14] This allows us to easily retrieve all scripts in the database that contain entities

[14] See https://textcourt.orinst.ox.ac.uk/database/analyticalgroups/24/

with a connection to Southeast Asia, even if there are no specific entities in common between the scripts. We also created a broader analytical group "萬國" containing all entities linked to foreign countries, linking a greater number of scripts sharing a wider theme.[15]

While the approach we have employed meets the needs of the project, there are limitations: for example, if a variant orthography of a certain entity is not recorded, then that entity will not be found and tagged, resulting in a gap in the markup. There are also issues connected to the tagging of single character entities: without parsing the context, it is impossible to tell if any instance of a given character is one that we want to tag or not. Currently the two solutions we have available to us are to manually check each occurrence, or if this is unfeasible due to time constraints, to make an editorial choice to tag only the single character entities that are a) distinct enough that we can be reasonably sure that we can tag the right ones automatically or b) hold enough research importance that we can feasibly go through the process manually.

Fortunately, implementing entity tagging in TEI is relatively simple, using elements provided for in the standard guidelines. We use the element <persName> for people,[16] <objectName> for objects,[17] <placeName> for places,[18] and <date type="occasion"> for occasions.[19] The <event> tag, which may appear a more natural choice and cleaner choice to encode what we refer to as "occasions" in the project, cannot be used as it cannot be added into the main body of a text.

The entity codes and analytical groups introduced above are then implemented into these tags using attributes: @ref for the entity code[20] and @ana for the analytical groups.[21] Taking Lü Dongbin again as our example, his name would therefore be represented in TEI as follows:

<persName ref="psn:RS00007" ana="@八仙">呂洞賓</persName>.

Using the @ref attribute allows us to store centralized lists of people, places, objects, and occasions in files separate to the play texts- something absolutely vital in a project of our scale.

---

15 See https://textcourt.orinst.ox.ac.uk/database/analyticalgroups/23/
16 https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-persName.html
17 https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-objectName.html
18 https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-placeName.html
19 https://tei-c.org/release/doc/tei-p5-doc/en/html/ND.html#NDDATER
20 https://tei-c.org/release/doc/tei-p5-doc/en/html/SA.html#SAPU
21 https://tei-c.org/release/doc/tei-p5-doc/en/html/AI.html#AIATTS

The three characters “psn” in our example above serve as a private URI- an abbreviated link pointing to the project’s list of people. In the header, we define the path that this abbreviation points to (underlined for emphasis):

```
<prefixDef ident="psn"
  matchPattern="([a-z]+)"
  replacementPattern="../references/people.xml#$1">
</prefixDef>
```

In a similar way, the abbreviations “plc”, “obj”, and “occ” point to centralized lists of places, objects, and occasions respectively.

4. The TEXTCOURT Website

The core database will be made available through two platforms. Firstly, the TEI XML files and associated images will be made available for download from the Figshare repository. We hope that providing the original XML files will allow other interested researchers to enhance and modify our database in their own research, or allow the incorporation of our scripts into broader projects dealing with enriched drama scripts from a variety of traditions. For most users, however, TEI files do not offer a user-friendly and accessible way to use the database. Instead, the primary way of interacting with the database will be through our project website: http://textcourt.orinst.ox.ac.uk/ (see Figure 2).

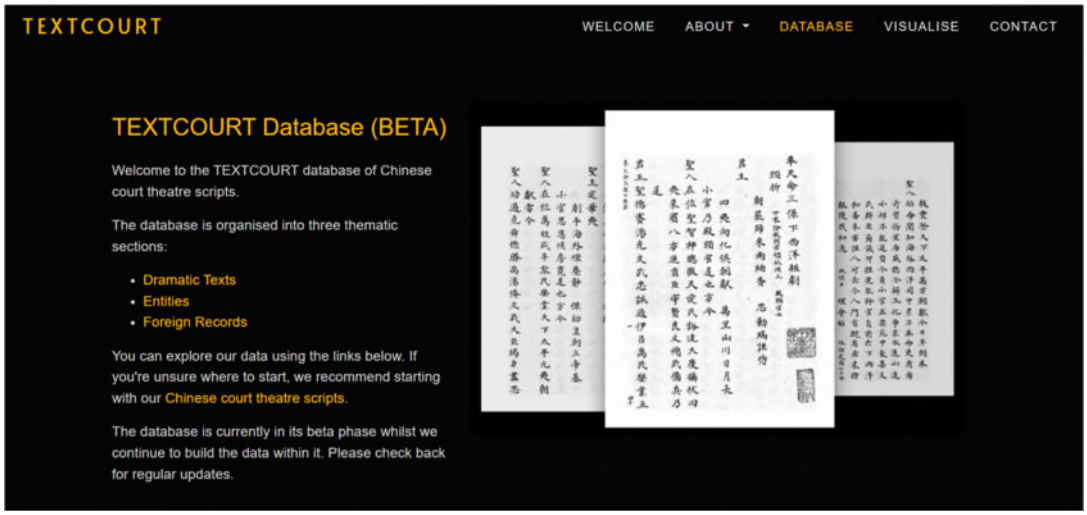

Figure 2. The User Interface developed by the TEXTCOURT Team

Source: TEXTCOURT 2022, accessed 12 April 2022, <https://textcourt.orinst.ox.ac.uk/database/>.

The database in our project website is divided into three sections: (1) Dramatic texts; (2) Entities; and (3) Foreign records. The website is currently available in beta with a limited number of scripts uploaded. The final version will be made available towards the end of the project term.

4.1 Dramatic Texts

The dramatic texts are organized in a three-level hierarchy: Scripts, Plays, and Collections. Of these, the fundamental unit is the Script. A Script is a distinct version of a Chinese court drama text relating to a single hardcopy edition: for example, the version of *Bailing xiaorui* 百靈效瑞 found in *Yingluan xinqu erjuan congshu jichengben* 迎鑾新曲二卷叢書集成本 equates to one Script in our database (Bailing xiaorui B), while the version found in *Fuzhuang jin yuefu xuan* 復莊今樂府選 equates to another Script (Bailing xiaorui C). Where a single play is available in multiple different Scripts, these are distinguished by an A/B/C suffix to the TEXTCOURT code.

A Play, in our definition, is a grouping of different Scripts: all the different versions of a dramatic text that exist in the core database. The purpose of the Plays function is to allow users to quickly view and if necessary compare all Scripts belonging to the same dramatic text in the database. The Plays on our website have no underlying XML file and function merely as a means to categorize Scripts.

Finally, we have Collections. These are defined as a group of independent court drama texts which were originally published or circulated under a common title. The utilitarian, occasion-led nature of court drama means that it is not always clear what constitutes an "independent" court drama text and what is a scene of a larger piece. Certainly, it appears that in practice this distinction was not always a rigid one, as scenes from longer plays could be performed singly depending on requirements. This perhaps reflects the prioritization of the needs of performance and ceremony over any overall creative conceptions of the piece held by the (mostly anonymous) dramatists. Nonetheless, the concept of the collection is still a useful one, especially with regards to court plays written by literati. As a rule of thumb, we regard any piece in which the constituent parts are labelled as "Scene X" 第 X 齣 as being a Play, whereas where the constituent parts are not labelled in this way we would regard the text as being a Collection. However, we are flexible to the individual circumstances of each text in the

application of this rule: therefore, there is necessarily some editorial component to what is treated as a Collection. On our website, a Collection, like a Play, is a grouping of Scripts. However, in the case of the Collection, the Scripts are drawn from different plays, as in the example of *Yingluan xinqu* 迎鑾新曲 given in this link.[22]

4.2 Entities

The site contains many powerful functions to link and organize the scripts and entities: users can readily view which scripts a given entity or analytical group appears in, and a list of all the entities in a particular play. Each entity is linked to a Wikidata search, so website users can simply and easily search all the alternative names for any given entity and retrieve ID numbers from CBDB and other databases via Wikidata. Scripts are also linked using Keywords, which act to link scripts sharing a broad theme, period in time, or textual format. All these features are designed to reveal links between the scripts in the core database that would otherwise remain obscured. The website also provides a platform for the application of more advanced digital humanities techniques which can provide further insights into the core database texts.

4.3 Foreign Records

Apart from the core database of drama scripts in the TEXTCOURT website, we have also supplemented the database with a digital archive of foreign records. Accounts of Chinese drama performances by foreign envoys who were invited to these occasions contain valuable information on the authorship and practices of Chinese court drama. For instance, a diary kept by the Korean scholar Pak Chiwŏn 朴趾源 (1737–1805), who went to China in the entourage of a Korean embassy, recorded names of 80 plays performed in 1780 to celebrate the Qianlong emperor's seventieth birthday and, most significantly, stated the authorship of one of the plays, which is not recorded in Chinese historical texts. These foreign documentaries are also rich material for researchers to explore how court theatrical culture was perceived, practiced, and circulated among the various regions in East Asia during this period. As for the European ambassadors such as those in Lord Macartney's mission to China in 1793 or the Dutch embassy in 1794–95, Chinese court drama presented a vastly different theatrical world. This project seeks to place Chinese court dramatic texts in the global context of other similar Asian and

[22] https://textcourt.orinst.ox.ac.uk/database/collections/Q90373_01_A/

European royal texts and genres such as masques, intermedia, and pageants, terms of which are often used in translation to refer to Chinese court drama.

To draw these links, we have collaborated with the Seoul National University (SNU) and the Institute of Sino-Nom Studies, Vietnam Academy of Social Sciences to gather these foreign records. To date, we have collected altogether around 200 related entries from works written in English, French, Latin and Hanja. These foreign records include texts documenting the foreign envoys' observations and early studies of Chinese drama. There are also coloured drawings of Chinese actors and comedians included in the database (see Figure 3).

Figure 3. The Image of a Chinese Comedian in William Alexander, *The Costume of China*
Source: TEXTCOURT 2022, accessed 15 March 2022,
<https://textcourt.orinst.ox.ac.uk/media/researchdata-documentaryexcerptimage/WE00034A/WE00034A%20Excerpt_1.jpg>.

5. Digital Approaches to the Dataset

The development of a digital archive enables us to adopt a computational approach to the Chinese court plays. We can make use of the entities tagged in the database to examine how individual figures are featured in various plays. The electronic texts also serve as a solid foundation for us to identify the text reuse patterns within and beyond Chinese court plays, use quantitative methods to obtain a bird's eye view into the formal textual features of certain literary phenomena, and to extract the relational data for social network analysis. In the following section, we will introduce some of the many potential uses of the TEXTCOURT digital archive.

5.1 The Use of Tags and the Study of Individual Characters

As indicated in the above section, entities in our dataset are tagged and grouped together in the TEXTCOURT archive to highlight the connections between various plays. Using the tags as a potential tool to conduct literary analysis, it is possible for us to observe how certain figures are portrayed in plays.

A good example would be the goddess Jinmu 金母 (or the Divine Mother). If we conduct a simple search in the database, all we can find is a list of plays in which the term is mentioned. Yet if we use the TEI files and refine our search by focusing only on the plays in which Jinmu is mentioned only in the stage directions, we observe a pattern across many of the plays when Jinmu is mentioned – Jinmu appears on stage only in the beginning or at the end of the play. This pattern is most obvious when we compare a simple text search of the term Jinmu with a refined text search in the Ming play *Bian Dongxuan mudao shengxian* 邊洞玄慕道昇仙 (Bian Dongxuan Embraces the Way and Ascends to Immortality):

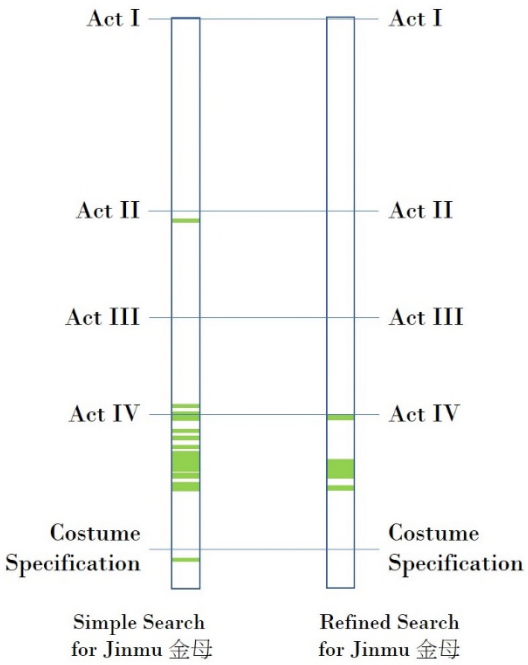

Figure 4. A Comparison between Simple and Refined Searches for Jinmu

Out of the 30 instances on which Jinmu appears in the play, only 11 occurrences come in the stage directions. And as we can see in the bar graph (figure 4), all these occurrences are clustered in the last act of the play. If we apply a similar search to other plays in the digital archive, we observe a fairly similar pattern. For instance, in the Qing court play *Siling xiaozheng* 四靈効徵 (The Four Divine Creatures Appear as Auspicious Signs), we can see the term Jinmu only appears in the stage directions of the first, penultimate and the last scenes (i.e. scenes 1, 11 and 12). This pattern allows us to conclude that the goddess Jinmu is often used as a device, but is not generally involved in the main body of the play. In fact, if we read one of the plays where Jinmu appears on stage, we would notice that Jinmu is often portrayed as a powerful goddess who will send other deities onto their missions, returning to the stage to resolve all the conflicts or to reward those who have attained spiritual transcendence. The representation of Jinmu allows us to probe into the cult of this renowned goddess in late imperial China. As a female deity who oversees the garden where the peaches of immortality grow, Jinmu is often associated with longevity and is hence recurrently used in the celebration of the birthdays of the imperial family members.[23] We can also turn our attention to the plays in which Jinmu is mentioned but never enters the stage. Conducting a search on our digital

[23] The cult of Jinmu (or, more often known as Xiwangmu 西王母, Goddess of the Western Paradise) has been extensively studied. See, for instance, Cahill (1993: 11-65).

archive, we can see that Jinmu is mentioned once in the Qing play *Bafang xianghua* 八方向化 (All Quarters of the World Turn toward Civilization). In this birthday play, one of the immortals, Han Zhongli 漢鍾離 introduces himself by stating that he has just been to pay tribute to the goddess. Although the goddess never appears on stage, this authorial religious figure is still an important signifier for the immortal realm.

Not only can computational methods enable us to study how a character is featured in various plays and visualize when a character appears on stage, but we can also use digital tools to identify the frequency of each word in the dialogue or the songs sung by a certain character. We do not have to concentrate on one single character on stage; instead, we can also apply network theory to study the relationships of each character in a play. As Franco Moretti has illustrated in his study of Shakespeare's *Hamlet*, a similar methodology can be applied to the longer plays in the archive (Moretti 2013: 211-240). It can allow us to examine how the relationships between different characters are portrayed in Chinese court plays and would allow us to compare that with other plays composed by other literati.

5.2 Towards a Quantitative Study of Chinese Court Plays

The development of the digital archive can enhance our understanding of the court drama by presenting a birds' eye view of the features of the existing corpus. By assigning keywords to every single script in the database, we can extract all the single-act birthday plays from the corpus and use that as a basis to conduct quantitative research. With the help of the archive, we can potentially calculate the word count of each of these plays and compare them with the respective performance records which may allow us to determine the time required for performing certain drama scripts. In this way, we can use this data to ascertain the time span each play may take and use that to assess how the court arranged various performances. The data would offer us a glimpse of the management of the inner court and a dataset that may suggest the variation of the length of the scripts alters with the aesthetic preference of the emperor.

Apart from comparing the word count of each single act play, the archive can also allow us to examine variations between scripts with the same title. In the process of documenting and collecting scripts for the database, the project often obtains multiple copies of the scripts sharing the same titles. For instance, we have gathered three copies of the plays with the same title *Sihai shengping* 四海昇平 (Ascendant Peace in the Four Seas): two printed in the

collection of *Guben xiqu congkan* (hereafter version A and B), and another version reproduced based on the Qianlong Andian edition. Using a computational method, we can locate parallel passages and use that as a base to locate the similarities and differences of each of these versions.

In fact, the application of digital methods in the identification of text reuse in Chinese literature is not a new concept. Donald Sturgeon has already devised a fully automated method of identifying and representing complex text reuse patterns. Using the Application Programming Interface (API) developed by the Chinese Text Project (https://ctext.org), we can find out the pattern of text reuse and explore the relationship between each version (Sturgeon 2018: 670-684). With the premise that any meaningful reuse of four or more characters is sought, we can quantify the similarity of the texts and conclude their relationships in a table as follows:

Text I	Text II	Similarity (n-gram = 4)
Version A	Version B	0.397
Version A	Qianlong Andian edition	0.00639
Version B	Qianlong Andian edition	0.0103

The calculation shows that the similarity between versions A and B is clearly higher than that of their relationship with the Qianlong Andian edition. To look closely at the texts, we can conclude that versions A and B are two renditions of the same text, while the Qianlong Andian edition of the *Sihai shengping* has a separate story. The use of calculation methods is not only applicable to the plays with similar titles, but it can also be adapted on a corpus level. If we put all the plays into the API and quantify the similarities of each of the plays in the corpus, we can visualize the text reuse pattern as a network graph and observe how certain plays in the archive may form clusters in the graph.

By applying this quantification method to early Chinese texts, Sturgeon found out that the clusters formed in the network graph correspond with the standard categorization of the texts. Applying an analogous way of analysis onto the corpus this project has collected, we may see how different texts coalesce in the graph and evaluate the method by examining how multiple factors may skew the results. In our digital archive, we have plays that share similar themes and topics (birthday plays, festival performances, plays adapted from novels and historical plays) with various lengths. The application of this quantitative method onto our core database of 200 plays will allow us to examine how certain features correspond with the

clusters formed in the graphs and shed light on a broader application of the method onto a wider pool of dramatic scripts. This provides us with the possibility to differentiate different texts within the database and identify certain clusters as points of inquiry.

5.3 Across Multiple Digital Archives: A Broader Base for Inquiry

Some of the archive's paratextual collections provide us with a glimpse of the social groups related to the composition of court plays. In 1751, Emperor Qianlong undertook a southern inspection tour. While Qianlong was received in Zhejiang 浙江, the literati Wu Cheng 吳城 (active 1735) and Li E 厲鶚 (1692-1752) both composed a play to commemorate the journey. When these plays were printed in 1895, the collection included poems written by other local writers. Using this set of data and comparing them with the local literati network in Li E's individual collection, *Fanxie shanfang ji* 樊榭山房集, we can observe that Li E rarely wrote to other local poets that were involved in receiving the emperor during the southern inspection tour. The lack of overlap between the two social groups highlight how a local literary community had been tentatively formed to receive the emperor. If we were to broaden up our study and incorporate other materials that document Qianlong's southern tours, we could gain a more comprehensive understanding of the literary field in mid-eighteenth century.

These paratextual materials provide data which could contribute to a broader exploration of literary history in late imperial China. Obviously, it would be necessary to integrate data obtained from other databases. In the above example, we would have to broaden our scope of investigation by delineating the social relations embedded in Li E's collection, which is not in our digital archive. The development of a comparative corpus and the collaboration between multiple databases to cater the need of our inquiry is particularly important. While some of the data needed for a broader inquiry may be available in other digital archives, there are still limitations as to how one can obtain the relevant data from other research sites. Over the past few years, a number of digital content platform has been developed to cater to the specific needs of each scholar. DocuSky, for instance, provides the essential tools for a scholar to download materials from various websites and databases. The RISE and SHINE project developed by the Max Planck Institute also provides a set of application programs which will facilitate the exchange of textual resources from multiple databases.[24] In the future, we hope to

[24] For a detailed description of the RISE and SHINE, see https://www.mpiwg-berlin.mpg.de/research/projects/rise-and-shine-research-infrastructure-study-eurasia

further collaborate with these platforms to conduct research across various digital archives. In the meantime, the datasets, images and texts gathered in our project will all be made available in open access and will be readily available on Figshare, an online depository where researchers can download the datasets and incorporate them with data gathered from other databases.

6. Conclusion

Court plays are best known for their grand stage designs and visual spectacles, produced to suit the propaganda and entertainment purposes of the emperor. As a result, they are often perceived to be of little literary value and thus ignored by scholars of Chinese literature. Chinese literary histories are fragmented, in that they often focus largely on "elite drama" at the expense of other theatrical forms such as court drama and local or folk drama.[25] While court drama (despite its exalted audience) is assumed to be the product of an anonymous theatrical world of lower-class professional writers and actors, elite drama is associated with famed authors and canonical works. Considering that many histories of Chinese drama are largely histories of Chinese dramatic literature that are structured around renowned playwrights and their masterpieces, it is no surprise to find elite theater at the center of such histories and court theater very much marginalized. For instance, one of the early and influential histories of Chinese drama, *Zhongguo xiqu tongshi* 中國戲曲通史 (A general history of Chinese drama), contains a total of over 1000 pages in three volumes, but has only one short section of about forty pages on Qing court theater in the third volume. Not only is the discussion on court theater very limited, but more tellingly, the section title "Stagecraft of Qing Court Theater" suggests that it is the stages, props, and costumes of court theater that are deemed to be the most, and perhaps the only, desirable characteristics worthy of mention (Zhang and Guo eds. 1980-81: 278-319). The marginal position of court theater in histories of Chinese drama does not represent a complete lack of scholarly interest or attention on the subject. It does, however, suggest that for a more detailed coverage of court theater, one needs to look elsewhere and turn to studies that focus less on dramatic literature and more on the social aspects and practices of Chinese theater.[26] Studies on elite theater and court theater appear to have generated two largely

[25] For discussions of these other forms of Chinese theater, one would need to consult 'theater histories' instead, though the emphasis there is also more on their theatrical forms and performance aspects rather than on the texts of these plays. For example, in Chapter 9 of Liao and Liu's *Zhongguo xiqu fazhan jianzhi* (A short history of the development of Chinese theater), they trace the developments of Qing drama both in terms of local plays and court plays. See Liao and Liu (2006: 241-289).

[26] For example, see Idema and West, *Chinese Theater, 1100-1450: A Source Book*, and Colin Mackerras, *The Rise of the Peking Opera, 1770-1870: Social Aspects of the Theater in Manchu China*.

independent sub-fields in the study of Chinese drama. Scholarship on Qing dynasty court theater is almost entirely published in specialized monographs on stage history and performance and still not integrated with the mainstream literary or textual studies. Previous studies on Chinese court drama are therefore isolated in that no serious attempts have been made to look into its textual relations with other forms of dramatic literature and other literary genres.

Through the application of the abovementioned computational methods, this project seeks to remedy the lack of textual studies on Chinese court drama, contributing towards our understanding of Chinese theater history and world court cultures. We believe that the creation of a digital archive can facilitate a better understanding of Chinese court drama, and the methodology the project develops to study a large corpus of authorless texts could be applied to other Chinese genres such as regional theater and popular songs.

Bibliography

Cahill, Suzanne (1993). *Transcendence and Divine Passion: the Queen Mother of the West in Medieval China*. Stanford, CA: Stanford University Press.

Chen, Liana (2021). *Staging for the Emperors: A History of Qing Court Theatre, 1683–1923*. Amherst: Cambria Press.

Fu Xihua cang guben xiqu zhenben congkan (2010). Edited by Wang Wenzhang et al. Beijing: Xueyuan chubanshe.

Gugong bowuyuan cang Qinggong Nanfu Shengpingshu (2015-17). 450 vols. Beijing: Gugong chubanshe.

Gugong zhenben congkan (2001). Hainan: Hainan chubanshe.

Higashi Ajia zenpon sōkan (2009). Sendai: Kagaku kenkyūhi hojokin tokubetsu suishin kenkyū shinchō kyūtei engeki bunka no kenkyū kenkyū seikashū led by Isobe Akira.

Idema, Wilt, and Stephen West (1982). *Chinese Theater, 1100-1450: A Source Book*. Wiesbaden: Steiner.

Liao, Ben and Liu Yanjun (2006). *Zhongguo xiqu fazhan jianzhi*. Taiyuan: Shanxi jiaoyu chubanshe.

Mackerras, Colin (1972). *The Rise of the Peking Opera, 1770-1870: Social Aspects of the Theater in Manchu China*. Oxford: Clarendon Press.

Moretti, Franco (2013). *Distant Reading*. London: Verso.

Sturgeon, Donald (2018). "Unsupervised Identification of Text Reuse in Early Chinese Literature," *Digital Scholarship in the Humanities* 33.3: 670-684.

Tan, Tian Yuan (Chen Tianyuan) (2022). "Ming Qing gongting yanju de wenben shijie" (Textual Worlds of Court Theater in Late Imperial China). *Bulletin of the Department of Chinese Literature, National Chengchi University*, No. 37: 5-52.

Tan, Tian Yuan (2016). "Sixty Plays from the Ming Palace, 1615-18," in Tian Yuan Tan, Paul Edmondson, and Shih-pe Wang, eds., *1616: Shakespeare and Tang Xianzu's China*. London: Bloomsbury Arden Shakespeare. pp. 96-107.

TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange* 4.4.0 [19th April 2022]. TEI Consortium. http://www.tei-c.org/Guidelines/P5/ (Accessed: 19th June 2022).

Wu, Shuyin ed. (2004). *Suizhong Wushi cang chaoben gaoben xiqu congkan*. Beijing: Xueyuan chubanshe.

Wu, Xiaoling (1986). “Glowing Clouds in an Azure Sky: A Newly Discovered Royal Pageant.” Translated by Lindy Li Mark and Samuel H. H. Cheung. *CHINOPERL PAPERS* 14: 1-14.

Ye, Xiaoqing (2012). *Ascendant Peace in the Four Seas: Drama and the Qing Imperial Court*. Hong Kong: The Chinese University of Hong Kong.Zhang, Geng, and Guo Hancheng (eds.) (1980-81). *Zhongguo xiqu tongshi*. Beijing: Zhongguo xiju chubanshe.

Zhao, Qimei (1958). *Maiwangguan chaojiaoben gujin zaju*. Reprinted in *Guben xiqu congkan siji*. Shanghai: Shangwu yinshu guan.

Zhongguo guojia tushuguan cang Qinggong Shengpingshu dang’an jicheng (2011). Beijing: Zhonghua shuju.