Chapter 3

Understanding *Xiaoshuo* From Warring States to Late Imperial China with the Aid of Digital Tools

Introduction

In their study on mind-body relationship in ancient China through textual big data, Edward Slingerland et al. point out that scholarly understanding of the relationship is misshapen by over-emphasis on the *Mengzi*'s 孟子 (*Master Meng*) explicit definition of it rather than how it is implicitly described in a large number of early texts. They argue that to draw conclusions from a few explicit definitions in famous texts is to "mistake an explicit claim for a background assumption." In the case of the *Mengzi*, the text is "making an argument, which he [the author] no doubt expects to be surprising or counterintuitive," which should not be conflated with an indication of how this issue is understood by the general public (990). To demonstrate how the *Mengzi*'s explicit definition differs from general understanding, Slingerland et el. use a combination of word collocation, hierarchical clustering, and topic modeling to analyze the relationship between *xin* 心 ("mind") and *ti* 體 ("body") in 96 texts totalling 5.7 million characters from Donald Sturgeon's Chinese Text Project (988), which reveals that these two terms are usually used in conceptual opposition, contradicting the *Mengzi*'s claim that they are a unity.

As discussed in Chapter 1, existing scholarship generally focuses on explicit definitions in famous early texts (pre-Qin and Qin-Han) to understand the meanings of

xiaoshuo.[11] This is problematic because *xiaoshuo* acquired different meanings throughout history; relying on philosopher or historian-bibliographers' words on the meaning of *xiaoshuo* is also concerning because their primary consideration is *xiaoshuo* as a rhetorical or bibliographical concept, which differs from its usage in other contexts. Moreover, as in the case of the *Mengzi* mentioned above, those who offer explicit definitions of *xiaoshuo* are likely making an argument about its nature and status, which may not reflect general understanding.

When we seek to understand how a common term is understood and used throughout history, which is an essential aspect of literary and historical scholarship, the sheer quantity of its usage poses an immense challenge for human reading, as it is difficult to virtually impossible to locate all cases of usage without digital help. Due to the availability of a large quantity of digitized text, it is now possible to detect the usage of *xiaoshuo* throughout ages from a wide variety of literature, which can remedy our bias built upon a small number of explicit definitions.

In *Macroanalysis: Digital Methods and Literary History*, Matthew Jockers points out that the abundance of digitized texts in the humanities is comparable to the big data revolution in the sciences and social sciences, as "massive data sets are allowing for investigations at a scale that reaches or approaches a point of being comprehensive," which alleviates problems associated with observations based on small sample sizes (7). In this chapter, I study the meanings of *xiaoshuo* throughout premodern Chinese history and its relationship with fiction with the aid of digital tools, which alleviates the potential

[11] See Chapter 1's discussion of Ming Dong Gu, Sheldon Lu, and Luo Yuming.

biases derived from focusing on a handful of famous texts by allowing the machine to find usages of *xiaoshuo* without making assumptions about what texts are more important.[12] I found that whereas *xiaoshuo* could denote fiction since as early as the Tang Dynasty (618-907), fiction will not become a dominant aspect of its significations until the late imperial era, especially the Qing Dynasty. Moreover, the meanings of *xiaoshuo* had accumulated since the Warring States period; as it developed new meanings throughout the ages, its older meanings did not fall into obscurity or disuse, which renders its nature and definitions more and more complex as time progresses.

Chronological Database of Chinese Literature (CDCL)

In this chapter, my research on primary sources relies on performing keyword searches in Donald Sturgeon's Chinese Text Project and the Chronological Database of Chinese Literature (CDCL); moreover, I performed collocation analysis on the Song, Yuan, Ming, and Qing corpuses of the CDCL. I created the CDCL, which is a fully machine-accessible digital database that consists of almost 2,000 titles from the Three Kingdoms period (220-280) to the Republican era (1912-1949). The premodern part of the CDCL, which excludes texts from the Republican period, contains 1,865 titles (totalling 154,711,312 characters) sourced from my scraping of Wikisource and Paul

[12] It is important to note that although the machine does not make assumptions about what texts are more important in a corpus, the constitution of the corpus can reflect human biases regarding what is important and what to include in the corpus. In my case, what I included in the corpus is largely based on what is available in a digitized format, which reflects the human biases of those who decide to prioritize the digitization of certain texts.

Vierthaler's digital *Siku quanshu* collection. All texts in the CDCL are plain text files that are not formatted for human reading but are suitable for computational processing.

The CDCL collection is built based on the catalogue of the *Guoxue baodian wangluo ban* 國學寶典網絡版 ("Treasured Index of Chinese Studies, Internet Edition," hereafter the Treasured Index), the largest online digital library of premodern Chinese texts that encompasses the Zhou period through the Qing Dynasty. [13] In addition to texts that fall into the *jing* 經 ("classics"), *shi* 史 ("history"), *zi* 子 ("philosophy"), and *ji* 集 ("collection") categories, the Treasured Index includes content that cannot be characterized by these four traditional divisions, such as novels and dramas. In total, the Treasured Index contains 6,003 titles, out of which more than 5,800 are from the Three Kingdoms through the Qing period.[14] The CDCL represents about 32 per cent of the Treasured Index collection from the Three Kingdoms era through the Qing period. Although the Treasured Index is one of the most comprehensive online collections of premodern Chinese texts and highly representative of the entire extant traditional literary tradition, it is not in plain text and cannot be freely used for computational analysis, which necessitates the creation of the CDCL. I did not include pre-imperial and early imperial material in the CDCL because Donald Sturgeon's Chinese Text Project covers this period to a highly comprehensive extent.

---

[13] See http://www.guoxue.com/cp/gxbd_ml01.htm for a comprehensive catalogue of the Treasured Index.

[14] The Treasured Index does not provide statistics on its holdings based on time periods. "More than 5,800" is based on my manual counting.

Whereas the Treasured Index cannot sort content by time periods, one of the CDCL's key features is the chronological arrangement of its texts. All texts in the CDCL are categorized by the dynasty of origin, which enables synchronic and diachronic analysis. Each text's dynasty of origin is determined in accordance with information provided by the Treasured Index catalogue.[15] In cases where a dynasty is divided into two segments, like the Northern (960-1127) and Southern Song (1127-1279), the sorting is based on my own research, as the Treasured Index catalogue does not always acknowledge chronological divisions within a dynasty.[16] To ensure the accuracy of analysis based on historical periodization, I have, to the best of my abilities, manually deleted parts of texts that were written in another period, such as a preface or postface composed in a later dynasty; this is to prevent texts categorized in a dynasty to include content from another period. Though it can be used for a variety of purposes, the primary intended function of the CDCL is the study of the development of literary and linguistic properties across different eras. For this dissertation, the CDCL is used to study the meanings of *xiaoshuo* throughout history. Currently, the CDCL does not have

---

[15] Although the Treasured Index does not have mechanisms to sort texts based on the dynasty of origin, it provides a catalogue that indicates each text's dynasty of origin. I used this catalogue to sort the texts in the CDCL into different dynasties. See the catalogue here: http://www.guoxue.com/cp/gxbd_ml01.htm.

[16] The vast majority of texts that require me to sort into sub-dynasties are those from the Song dynasty. Fortunately, I am very familiar with Song literature and history during the Northern-to-Southern transition, so I was able to sort them into Northern and Southern Song with relative ease. Texts that cannot be determined to be either Northern or Southern Song are kept in a separate category. I wanted to distinguish Northern and Southern Song texts because these two periods saw important changes in literary culture, especially in the realm of popular and vernacular literature; the ability to distinguish these periods can be critical in literary studies.

mechanisms to sort texts based on non-chronological criteria, such as by bibliographical category or author.

The following section provides a detailed break-down of the CDCL catalogue.

By Dynasty

Three Kingdoms: 7 titles

Western Jin: 9 titles

Eastern Jin: 13 titles

Southern and Northern Dynasties: 23 titles

Sui: 2 titles

Tang: 165 titles

Five Dynasties: 18 titles

Northern Song: 201 titles

Southern Song: 266 titles

Yuan: 251 titles

Ming: 378 titles

Qing: 481 titles

Proportion of Each Bibliographical Category[17] in Title Count[18]

Jing: 5.2% of all titles

Shi: 7.8% of all titles

Zi (Confucian and Daoist): 2.5% of all titles

Ji: 13.7% of all titles

Biji 筆記 ("brush notes"): 45.5% of all titles

Wenlun 文論 ("literary discourse"): 8.7% of all titles

Xiqu 戲曲 ("theater and tunes"): 8.9% of all titles

Xiaoshuo (vernacular and classical language): 7.7% of all titles

Proportion by Character Count[19]

Pre-Tang: 6,466,320 characters; 4.2% of all characters

Tang: 9,470,301 characters; 6.1% of all characters

Five Dynasties and Song-Yuan: 50,366,521 characters; 32.6% of all characters

Ming: 24,654,026 characters; 15.9% of all characters

Qing: 63,754,149 characters; 41.2% of all characters

---

[17] The bibliographical sorting and naming of the bibliographical categories are in accordance with the Treasured Index catalogue (http://www.guoxue.com/cp/gxbd_ml01.htm) and does not represent my personal views. The titles categorized to be *xiaoshuo* are not necessarily fictional; they are designated as *xiaoshuo* by the Treasured Index catalogue. This percentage break-down accounts for only the premodern portion the CDCL and excludes the Republican section.

[18] The percentage is calculated by dividing the number of titles in each category by 1865, the total number of titles in the premodern portion of the CDCL.

[19] The percentage is calculated by dividing the number of characters in each period by 154,711,312, the total number of characters in the premodern portion of the CDCL.

Warring States Through the Six Dynasties

Lü Hailong 呂海龍 points out that it was not until the Northern Song when *xiaoshuo* began to predominately denote narrative literature (94). Prior to that, *xiaoshuo* referred to narrative and non-narrative discourse. In the Chinese Text Project's pre-Qin corpus, two mentions of *xiaoshuo* are in the *Zhuangzi* and *Lüshi Chunqiu* 呂氏春秋 (*The Annals of Lü Buwei*). As mentioned in Chapter 1, *xiaoshuo* in the *Zhuangzi* is contrasted with *dada*:

> (But) if the prince had taken his rod, with a fine line, and gone to pools and ditches, and watched for minnows and gobies, it would have been difficult for him to get a large fish. Those who dress up their small tales [*xiaoshuo*] to obtain favour with the magistrates are far from being men of great understanding [*dada*]. (Legge 134)

In the *Lüshi Chunqiu*, it is used in parallel to "slight errors" and described as something that causes "monumental calamity": "An incompetent ruler brings about a catastrophe because of some slight error. Bao Si ruined the state by causing King You to indulge some trivial pleasure [*xiaoshuo*] that would lead to monumental calamity" (Knoblock and Riegel 573-574). In both cases, *xiaoshuo* seems to be a derogatory term that refers to inferior knowledge, pursuit, or discourse that is potentially harmful. This notion of *xiaoshuo* is also seen in the Eastern Han (25-220) text *Zhonglun* 中論 (*Balanced Discourse*):

> A lord's greatest weakness is none greater than being too attentive to petty things and negligent of the great Way; he [a weak lord] is aware of what is near him but

> oblivious to distant plans. Thus, since antiquity until present day, it has never been that [the lord] is like this and there is no chaos; it has never been that [the lord] is like this and [the state] does not collapse. Those who are attentive to petty things and aware of what is nearby can be said to be hearing the harmony of zithers, flutes, songs, and chants with their ears, seeing the colorful patterns of carvings with their eyes, voicing the words of debates, riddles, and poetic compositions with their mouths, comprehending the texts of short stories and minor discourse [*xiaoshuo*] with their hearts, learning the tricks of archery, horse-riding, calligraphy, and math with their hands, and pursuing the appearance of bowing and dancing with their bodies. (Xu 463)
>
> 人君之大患也，莫大於詳於小事，而略於大道；察其近物，而闇於遠圖；故自古及今，未有如此而不亂也，未有如此而不亡也。夫詳於小事，而察於近物者，謂耳聽乎絲竹歌謠之和，目視乎琱琢采色之章，口給乎辯慧切對之辭，心通乎短言小說之文，手習乎射御書數之巧，體騖乎俯仰折旋之容。

Here, *xiaoshuo* is held as an example of a *xiaoshi*, "petty things," that distracts one from *dadao*, "the great Way," which causes the destruction of the state when excessively indulged in. Other than this, two mentions of *xiaoshuo* in the Han period are from the *Hanshu*, where Ban Gu 班固 (32-92) characterizes it as talks from the streets as mentioned in Chapter 1. This particular notion of *xiaoshuo* is also seen in the Eastern Han history *Qianhan ji* 前漢紀 (*History of the Former Han*): "There is also the school of

xiaoshuo, which generally originates from the talks of the streets and discussions of the alleys" 又有小說家者流 蓋出於街談巷議 (Xun 247).

The *Hanshu*'s "*Yiwen zhi*" 藝文志 ("Record of Art and Literature") section contains the earliest extant[20] catalogue of *xiaoshuo jia* writing. This catalogue is dominated by titles that do not appear to be narratives, such as *Fengshan fang shuo* 封禪方說 (*Discourse on Worshipping Ceremonies*) and *Daizhao chen raoxin shu* 待詔臣饒心術 (*The Art of a Bountiful Heart of a Minister Awaiting an Edict*) (Ban Gu 1744-1745), which seem to be discursive and instructive in nature. Wang Qizhou 王齊洲 observes that the majority of these titles are *fangshu* 方術 ("magic") writings (22-23), which suggests that *xiaoshuo* was closely associated with non-canonical knowledge, but not necessarily related to narratives. This is in accordance with the study by Hellmut Wilhelm mentioned in Chapter 1, which argues that the Zhou texts in the *Hanshu*'s *xiaoshuo jia* catalogue are likely discursive and rhetorical in nature. Regarding Ban Gu's characterization of *xiaoshuo* as gossip on the street (cited in Chapter 1), Wilhelm notes that it does not seem to match the nature of the texts included in Ban's own *xiaoshuo jia* category:

> [. . .] it emerges that the content of the books [the titles in the *xiaoshuo jia* category] could not possibly have been popular lore; rather, they must have been expository writings with political intent. In other words, they were of exactly the

[20] The earliest, not earliest extant, catalogue of *xiaoshuo* writings is likely in Liu Xiang's 劉向 *Qilue* 七略 (*Seven Overviews*), which is lost. The "*Yiwenzhi*" section of *Hanshu* derives heavily from *Qilue*'s catalogue of titles. See Wang Qizhou.

> same type referred to in the Chuang-tzu and Hsun-tzu passages. His definition to the contrary, a substantial part of the *hsiao-shuo* for him, also, constituted expository prose. Some characteristic other than content and pragmatic use must have induced Pan Ku to include them zin the category of the *hsiao-shuo.* (252)

As far as I am aware, no other scholar has drawn attention to the contradiction between Ban Gu's explicit definition of *xiaoshuo* and the fact that texts in his *xiaoshuo jia* category seem to be discursive. Wilhelm speculates that Ban's definition could be motivated by an "attempt to reduce [. . .] a literary school to the functions of a particular office" (252). This contradiction exemplifies the problem of relying on explicit definitions noted by Slingerland et al.

In the CDCL pre-Tang corpus, which covers the Three Kingdoms period through the Sui, *xiaoshuo* only appears twice. The *Sanguozhi zhu* 三國志注 (*Commentary on the History of the Three Kingdoms*), a commentary on the *Sanguo zhi* 三國志 (*History of the Three Kingdoms*) compiled in the Liu Song Dynasty (420-479), quotes a fragment from the *Weilüe* 魏略 (*Concise Account of the Wei*), a lost text from the Three Kingdoms period: "[He] then bared his head and upper body, performed the barbarian dance 'Five Anvils,' juggled balls, swung swords, and recited *xiaoshuo* by performers until it reached some thousands of numbers" 遂科頭拍袒，胡舞五椎鍛，跳丸擊劒，誦俳優小說數千言 (Chen 449). *Xiaoshuo* here has a substantially different meaning, as it denotes oral performance. Since the Three Kingdoms period and the Eastern Han neighbor each other chronologically, it is possible that this notion of *xiaoshuo* informed Ban Gu's characterization of *xiaoshuo* being from talks on the streets, which addresses the

aforementioned concern raised by Hellmut Wilhelm that the expository titles in the *Hanshu*'s *xiaoshuo jia* category do not reflect Ban's claim. This *Weilüe* fragment indicates that *xiaoshuo*'s denotation of narratives, specifically oral narratives, significantly predates the Song period, which subverts the understanding proposed by Lü Hailong that the Song was the beginning of *xiaoshuo*'s denotation of narratives. The second appearance of *xiaoshuo* in the pre-Tang corpus is in the *Nanqi shu* 南齊書 (*The Book of Southern Qi*), compiled in the Liang Dynasty (502-557): "Those who talk must say that debased arts like brush notes cannot be used to handle matters of life and death; minor discourse [*xiaoshuo*] like *kaiquan* cannot be depended on to make judgements" 議者必雲筆記賤伎，非殺活所待；開勸小說，非否判所寄 (Xiao 894). Here, *xiaoshuo* seems to be a debased, unworthy pursuit that cannot be depended upon for critical judgement; its usage is similar to that in Warring States literature as discussed above.

Tang

The next extant major catalogue of *xiaoshuo* writings is found in the *Suishu*'s 隨書 (*The Book of Sui*) *Jingji zhi* 經籍志 ("Record of Classics and Books") section, compiled in the Tang Dynasty, where the *xiaoshuo jia* category is filled with writings related to speeches, like *Bianlin* 辯林 (*Forest of Debates*), *Za duiyu* 雜對語 (*Miscellaneous Conversations*), and *Suoyu* 瑣語 (*Trifling Talks*) (1011-1012). While this reflects a change in *xiaoshuo*'s denotation as compared to that in the *Hanshu*, the notion of narrative or fiction is not apparent. In the CDCL's Tang corpus, *xiaoshuo* appears

fourteen times, among which five signify book titles: *Xiaoshuo* and *Liangwu xiaoshuo* 梁武小說.[21] These two titles—both no longer extant—can be found in the *xiaoshuo* bibliographical section in the *Suishu*, though there is no description concerning the exact nature of them (1011-1012). In the *Beihu lu* 北戶錄 (*Records of the Northern Gate*), composed by Lu Xisheng 陸希聲 (d. 895), *xiaoshuo* is explicitly described to be a literary genre predominately characterized by fabrication:

> There have been many *xiaoshuo*-writers in recent years. They generally write about absurd and baseless things like ghosts, gods, transformations, and fantasy. Otherwise, they use comedy and humor to inspire laughter and joy. Aside from these two, they also force stories out of words, which all defame wise men of previous ages and are taken as facts by the unsuspecting. (6)
>
> 近日著小說者多矣，大率皆鬼神變怪荒唐誕妄之事。不然，則滑稽詼諧以為笑樂之資。離此二者，或強言故事，則皆詆訾前賢，使悠悠者以為口實。

As far as I am aware, this is the earliest instance of *xiaoshuo* denoting written narratives predominately characterized by intentional and/or apparent fabrication (i.e. fiction).

As mentioned previously, none of the four traditional bibliographical categories can accommodate literary genres predominately characterized by apparent and/or intentional fabrication. Lu Xisheng's characterization of *xiaoshuo* is the only instance in the Tang corpus that leaves *xiaoshuo* outside the four traditional bibliographical categories. Aside from the *Beihu lu*, other mentions of *xiaoshuo* seem to contain it in the

[21] In *Beihu lu juan* 1 and 3 and *Suishu juan* 34.

zi or *shi* bibliographical category. In Liu Zhiji's *Shitong* 史通 (*Historical Perspectives*), *xiaoshuo* is discussed in the context of *shi* writings:

> As for the likes of the *xiaoshuo* of various masters, chronicles, and miscellaneous records, such as Wei Zhao's *Records of the Cave* and Tao Hongjing's *Yearly Calendars of the Imperial Era*, they were made according to memorials and created as texts. They are not in the same tier as stately histories. Thus, although they are extant, I will not discuss them. (77)
>
> 若諸子小說，編年雜記，如韋昭《洞紀》、陶弘景《帝代年曆》，皆因表而作，用成其書。既非國史之流，故存而不述。

In this passage, *xiaoshuo* seems to denote an inferior type of *shi* writings that is unworthy of being taken seriously. In *Sui-Tang jiahua* 隋唐嘉話 (*Fine Tales of the Sui and Tang*), Liu Su 劉餗 (fl. 742-756) reinforces this interpretation: "Ever since I was a little child, I have often heard stories of the past. They are not worthy to be grand records, so I have appended them at the end of the minor discourses [*xiaoshuo*]" (Manling Luo 1443).

There is also evidence that *xiaoshuo* was used by historians in the process of compiling history:

> Jun's talent and knowledge are sufficiently broad and grand. However, they cannot fully explore the profundity of Biao and Qiao nor comprehensively capture that of Ban and Ma. Hence, he favors the minor talks [*xiaoshuo*] of the alleyways and devotes his mind to vulgar short stories. This can be described as hard work without merit and unfitting exhaustion of the mind. (Liu Zhiji 222)

以峻之才識，足堪遠大，而不能探賾彪、嶠，網羅班、馬，方復留情於委巷小說，銳思於流俗短書。可謂勞而無功，費而無當者矣。

Here, a historian is criticized for being frivolous due to his adoption of *xiaoshuo* in his craft. The association of *xiaoshuo* with talks of the alleyways is reminiscent of Ban Gu's statement in the *Hanshu*. This notion is also repeated in the *Suishu*'s definition of *xiaoshuo* in its bibliography of *xiaoshuo* texts: "*Xiaoshuo* is the talks of streets and alleyways" 小說者，街說巷語之說也 (1011). Another example of *xiaoshuo* being consulted by historians is found in Li Yanshou's 李延壽 (fl.679-680) *Shang Nan-Bei shi biao* 上南北史表 ("Memorial of Offering the *History of the Southern and Northern Dynasties*"): "I examined the titles of chapters; there are many fascicles of histories. They all express hearings and witness accounts with an excessive amount of similarities and differences. *Xiaoshuo* and short stories are prone to being lost; they miss parts or have withered and have no means of being verified" 考之篇目，史牒不少，互陳聞見，同異甚多。而小說短書，易為湮落，脫或殘滅，求勘無所 (946). Li Yanshou seems to bemoan that the pitiful state of *xiaoshuo*'s preservation is what causes it to be not dependable for historiographical purposes. This is a markedly different characterization from that of Lu Xisheng where *xiaoshuo* is described to be predominately fabrication, which cannot be used for history-writing regardless of its condition of preservation.

The *Shitong* contains a passage that justifies *xiaoshuo*'s status as a proper historical genre:

> The texts of ancient emperors and kings and the more recent records of various lords span across generations and are used as models of virtue. The remaining unofficial biographies include [ones that record] Shen Nong tasting medicine; hence there is the *Compendium of Materia Medica*. Yu of the Xia tended to the land; in fact, he authored the *Classic of Mountains and Seas*. The *Book of Generations* differentiates family names; it was written by the Zhou royal house. *Sayings of the Kong Family* documents speeches; it was passed down from various descendants of the Kong family. Thus, we know that side records and *xiaoshuo* form their own school. It can be mixed with official histories because of its ancient origin. (Liu Zhiji 454-455)
>
> 上代帝王之書，中古諸侯之記，行諸歷代，以為格言。其餘外傳，則神農嘗藥，厥有《本草》；夏禹敷土，實著《山經》；《世本》辨姓，著自周室；《家語》載言，傳諸孔氏。是知偏記、小說，自成一家。而能與正史參行，其所由來尚矣。

Although this passage asserts that *xiaoshuo* is comparable to official history because it has ancient origins, some of the texts it mentions are traditionally understood to be *zi* writings, like the *Bencao gangmu* 本草綱目 (*Compendium of Materia Medica*) and *Kongshi jiayu* 孔氏家語 (*Sayings of the Kong Family*). Moreover, this seems to contradict the dismissive attitude toward *xiaoshuo* seen in the *Shitong* passage cited above. The *Shitong* contains another instance of attributing value to *xiaoshuo* and associating it with *zi* writings:

> The talks of the streets and discussions of the alleyways are sometimes worth looking into. *Xiaoshuo* and debased sayings can still be beneficial when they are in the past. Thus, the gentlemen who are interested do not abandon them all. Works like Liu Qingyi's *A New Account of the Tales of the World*, Pei Rongqi's *Forest of Sayings*, Kong Sishang's *Record of Sayings*, and Yang Jiesong's *Marsh of Discussions* are called trifling talks. (Liu Zhiji 459)
>
> 街談巷議，時有可觀，小說厄言，猶賢於已。故好事君子，無所棄諸。若劉義慶《世說》、裴榮期《語林》、孔思尚《語錄》、陽玠松《談藪》。此之謂瑣言者也。

As discussed previously, such records of speech are classified in the *zi* category in the *Suishu*. Although the *Shitong* is inconsistent in its evaluation of *xiaoshuo*'s value, all of its statements regarding *xiaoshuo* attempt to contain it within the four formal categories of bibliography, namely the *zi* and *shi* categories, to give it a place in documented literature.

In summary, there are five ways to understand *xiaoshuo* before the Song period (960-1276): 1) an unworthy pursuit or inferior discourse, 2) a type of inferior or unofficial *shi* writing, 3) a type of *zi* writing, 4) a type of oral performance, and 5) written fiction intended for entertainment. Among these, 1) to 3) significantly outnumber the rest in quantity, as 4) and 5) each only occur once. All five of these meanings established by this period will remain important aspects of how *xiaoshuo* was used throughout premodern history. Although the following sections on the Song-Yuan and Ming-Qing (1368-1911) periods will focus on *xiaoshuo*'s new identities in the realm of popular

performance and vernacular literature, it is important to keep in mind that its meanings as an unworthy pursuit or a type of *shi* or *zi* writings will not be eroded until the modern era.

Song-Yuan

In the Northern Song, Ouyang Xiu's 歐陽修 (1007-1072) *Xin Tang shu* 新唐書 (*A New History of the Tang*) took a decisively different approach to categorizing *xiaoshuo* by filling the *xiaoshuo jia* category with titles that appear to be narratives, such as *Guishen liezhuan* 鬼神列傳 (*Various Biographies of Ghosts and Deities*), *Yuanhun zhi* 冤魂志 (*Record of Wronged Ghosts*), and *Da-Tang qishi ji* 大唐奇事記 (*Extraordinary Affairs of the Great Tang*) (16601). Notably, many of these narratives are catalogued in *shi* 史 categories in the *Suishu*; Ouyang made the effort to reassign them into the *xiaoshuo* category. To scholars like Lü Hailong and Sheldon Lu, this redefinition signifies a turning point for *xiaoshuo*. Sheldon Lu states, "Beginning in the Sung, attitudes toward the nature of fictional biography and *hsiao-shuo* in general began to change, partly because of the vast output of fictional writings produced in the T'ang dynasty"; to prove this claim, he points to "the nature of the titles listed in the *hsiao-shuo* section of the *Hsin Tang-shu*, which come close to the modern conception of fiction" (132). Although some of the texts in the *Xin Tang shu*'s *xiaoshuo jia* category can be read as fictional, we must not assume that *xiaoshuo* in the Song-Yuan period was equivalent to fiction.

Xiaoshuo's denotation of narrative literature assumed different forms in the realms of literati writing and popular entertainment. Despite Lu Xisheng's usage of

xiaoshuo to denote written fiction in as early as the Tang period, in the realm of literati writing, there was a general expectation that *xiaoshuo* should be factual and reflect real historical experiences. Ouyang Xiu explains the role of *xiaoshuo* in history-writing:

> *History of the Ten Kingdoms*, compiled in the previous era, was a copy to be presented to the Emperor and needed to have many fascicles. Now, if it were to become standard history, it is appropriate to edit and cut out [some of the writings] and preserve what is central and important. As for trivial and petty matters, though they can be recorded, they are not relevant to the core content. They can be preserved in *xiaoshuo*; they are not worthy enough to fill up standard history. (*Ouyang Xiu Ji*, 537)
>
> 前歲所作《十國志》，蓋是進本，務要卷多。今若便為正史，盡宜刪削，存其大要，至如細小之事，雖有可紀，非幹大體，自可存之小說，不足以累正史。

According to Ouyang, *xiaoshuo* is made from historical records of a trivial nature that are not important enough for more serious histories. Sima Guang 司馬光 (1019-1086) also describes the role of *xiaoshuo* in compiling the history *Zizhi tongjian* 資治通鑑 (*Comprehensive Mirror in Aid of Governance*), where he "comprehensively read previous histories and collected *xiaoshuo* on the side" 徧閱舊史, 旁采小說 (9). This seems similar to the process of *xiaoshuo* being consulted for historiographical purposes in the Tang Dynasty as discussed above.

The Song-Yuan period saw the rise of the *biji* genre among the literati. In her study of *biji* writings, Cong Ellen Zhang characterizes the content of such literature as "the

authors' real-life experience as measured by the information they gathered from direct observation and hands-on investigation" concerning "court and capital life, famed political and literary figures, and strange occurrences [. . .] regional conditions, everyday material culture, local practices and customs, and interesting personalities [. . .]" (44-45). Despite the late imperial habit of cataloguing fictional tales in the category of *biji xiaoshuo*, which might lead some to perceive *biji* as a fictional or semi-fictional genre, in the Song-Yuan period, *biji* writings emphasized factual knowledge acquired via real-life experience. The late imperial conflation of the terms *biji* and *xiaoshuo* is not without precedent in this period as contemporary evidence suggests that *xiaoshuo* can be used to denote *biji* writings. See the following passage by the late Northern Song writer Li Xianmin 李獻民 (dates unknown):

> The spread of *xiaoshuo* in the world is extensive. In our dynasty, Yang Yi[22] is widely read through *Garden of Talk*, Ouyang Xiu[23] is widely read through *Record of Returning to the Field*. After them are Shen Kuo's[24] *Brush Talk of Mengxi*[25] and Shi Dan's *Miscellaneous Notes*. They all collect the affairs of their time [. . .] (1)
>
> 夫小說之行世也多矣。國朝楊文公以《談苑》行，歐陽文忠公亦以《歸田錄》行，其次則存中之《筆談》，師耽之《雜記》，類皆摭一時之事[. . .]

[22] Yang Yi 楊億.
[23] Ouyang Xiu 歐陽修.
[24] Shen Kuo 沈括. Cunzhong 存中 is his courtesy name.
[25] *Mengxi bitan* 夢溪筆談.

Here, *xiaoshuo* is used to refer to famous *biji* writings like Ouyang Xiu's *Guitian lu* 歸田錄 (*Record of Returning to the Field*) and Shen Kuo's 沈括 (1031-1095) *Mengxi bitan* 夢溪筆談 (*Brush Talk of Mengxi*). In the *Xin Tang shu*'s "*Yiwen zhi*" section, a number of Tang dynasty writings that are similar to Song-Yuan *biji* in style and content are included in the *xiaoshuo jia* category, such as Fan Shu's 範攄 (dates unknown) *Yunxi Youyi* 雲溪友議 (*Discussions Among Friends at Yunxi*), Wei Xuan's 韋絢 (b. 796) *Liu Gong Jiahua lu* 劉公嘉話錄 (*Record of Liu Gongjia's Talks*), and Zhang Gu's 張固 (dates unknown) *Youxian guchui* 幽閒鼓吹 *(Flute-Blowing and Drumming in Seclusion and Leisure*) (Ouyang, *Xin Tang shu Renshou ben ershiliu shi* 16601). In the realm of Song-Yuan literati writing, *xiaoshuo* referred to narratives about real-life experiences and real-world events and was not primarily a space of fictional creation, though one cannot rule out the possibilities of authors experimenting with fictionality in *biji* writings, especially those related to supernatural occurrences.

Xiaoshuo in the realm of popular entertainment had considerably different denotations. The Song, especially the Southern Song, saw the escalation of urbanization; combined with the commercialization and popularization of the printing press, the Song-Yuan period left us the first substantial collections of records of urban popular entertainment. The following account in the early Yuan (1279-1368) text *Mengliang lu* 夢粱錄 (*Dreaming Over a Bowl of Millet*) describes oral storytelling in the Southern Song capital Lin'an 臨安, in which *xiaoshuo* is a performative category:

Oral storytelling was called “tongue-discourse.” It had four subsects; each had its own following. “Minor tale” (*xiaoshuo*) is called “silver character.” [Its story content is] like romance, supernatural, and tales of the extraordinary [. . .] History-tellers told [stories from] *Zizhi tongjian* and writings, history, literature, and biographies of the Han and the Tang and each dynasty; they told the affairs of prosperity, ruin, conflicts, and wars. There were Scholar Dai, Jinshi-Degree-Holder Zhou, Miss Zhang, Miss Song, Qiu Jishan, and Preacher Xu. There was also Sir Wang Liu, who was originally a story-teller in the imperial court and paid like an imperial guard. He thoroughly knew all of history. In the *Xianchun* reign period, he performed *Chapter of Restoring the Central Plain* and biographies of famous generals of the Restoration period.[26] Listeners were plentiful. Indeed, his speech was truly elegant; his memory and knowledge and their sources were very vast. However, he feared those who told minor tales the most. The minor tale tellers could tell the past affairs of a dynasty or period and instantly make up [stories]. It was like making up stories during drinking games.[27] They occupied their own places [in the realm of entertainment]. (Wu Zimu 170-171)

說話者謂之「舌辯」，雖有四家數，各有門庭。且小說名「銀字兒」，如煙粉、靈怪、傳奇 [. . .] 講史書者，謂講說《通鑒》、漢、唐歷代書史文傳，興廢爭戰之事，有戴書生、周進士、張小娘子、宋小娘子、邱機山、徐宣教，又有王六大夫，元系禦前供話，為幕士請給講，諸史俱通，於鹹淳年

[26] *Zhongxing* 中興 refers to the period of restoring the Song court in the South.
[27] *Qiling suiling* 起令隨令 is a drinking game where players made up poems and stories.

> 間，敷演《復華篇》及中興名將傳，聽者紛紛，蓋講得字真不俗，記問淵源甚廣耳。但最畏小說人，蓋小說者，能講一朝一代故事，頃刻間捏合，與起令隨令相似，各占一事也。

Here, *xiaoshuo* refers to a subset of oral storytelling about romance, the supernatural, and the extraordinary. At the end of this passage, *xiaoshuo*-telling is held in stark contrast with history-telling as *xiaoshuo*-tellers could instantly make up stories. It is apparent that *xiaoshuo*-telling was an art of oral fictional narratives. This is further supported by another contemporary account of urban performance in Lin'an found in the Southern Song text *Ducheng jisheng* 都城紀勝 (*Records of Sights from the Capital City*):

> Dangling wire puppet show started as "Chen Ping Breaking Sieges with Six Wonders." Cane-head puppetry, water puppetry, and flesh puppetry used children and young people. When puppets performed romance and supernatural stories or the kinds of warfare and disputes, their scripts were like mixed theatre or lyrical story-telling: in general, they were mostly fictional and rarely true, like the kinds of the Juling deity or the great immortal Zhuji. (9)
>
> 弄懸絲傀儡起於陳平六奇解圍。杖頭傀儡、水傀儡、肉傀儡以小兒後生輩為之。凡傀儡敷演煙粉靈怪故事、鐵騎公案之類，其話本或如雜劇，或如崖詞，大抵多虛少實，如巨靈神朱姬大仙之類是也。

In this account, *yanfen* 煙粉 ("romance") and *lingguai* 靈怪 ("supernatural") stories are described as "mostly fictional and rarely true." In the previous account, *yanfen* and *lingguai* are said to be major genres of *xiaoshuo*-telling, which again points to the

fictional nature of *xiaoshuo*-telling. For another contemporary indication of *xiaoshuo*-telling's fictional nature, see Hong Mai's 洪邁 (1123-1202) following comment: "Though *xiaoshuo*-telling and theatre plays were fictional with ghosts and the like, they had twists, turns, thoughts, and nuances" 雖小說戲劇，鬼物假托，莫不宛轉有思致 (34). Such *xiaoshuo* are certainly very different from the kind consulted by Sima Guang to compile the *Zizhi tongjian*.

The accounts of Song-Yuan *xiaoshuo*-telling cited above illustrate the awareness that *xiaoshuo* is self-consciously fictional. Sheldon Lu argues that by the Tang, literati's interpretative strategies toward writings that are possibly fictional can be divided into two types: historical or allegorical. In other words, such writings were either read as (defective) history or as allegories. By "allegories," he means *yuyan* 寓言 stories like the ones found in the *Zhuangzi* (93-128). To him, it is not until the Ming-Qing period when people increasingly started to recognize that fictional writings were "self-consciously non-historical and ostensibly creative and [. . .] ought not to be judged and read as defective history and quasi-history but to be understood on its own terms" (134). While I agree that there is a drastic difference between Tang and late imperial interpretative strategies toward fiction and the Ming-Qing period saw the full maturity of the awareness of the nature of fiction, Lu's brief treatment of the Song-Yuan period overlooks the important developments in the realm of popular entertainment that are fundamental to the Chinese fictional consciousness.

Hong Mai, the author-compiler of the *Yijian zhi* 夷堅志 (*Record of the Listener*), is a Southern Song scholar who is vocal on the rather silent issue of fictionality. The

Yijian zhi is a collection of strange stories, allegedly based on what people have seen and heard. It is written in the *biji* style: short anecdotes written in literary Chinese placed next to one another around a loose theme, often preceded by the author's preface.[28] Whereas Song *biji* writings tend to stress their factuality, Hong's preface toys with the ideas of fictionality and factuality:

> When the first installment of [*Yijian zhi*] was complete, it was circulated among gentlemen and scholar officials. Today it has been published in Fujian, Sichuan, Wuzhou as well as Lin'an. Every household has a copy.
>
> Due to my interest in the extraordinary and veneration of the strange, people from far and wide send me details whenever they hear of such a story. Therefore, the amount of material I have received these last five years is comparable to what I had previously collected. And so I compiled it all under the name of Yi zhi. In total, both the two books and Jia and Yi comprise of six hundred stories and all manner of strange and uncanny stories found throughout the world have been included therein.
>
> As for the anomalies of Qi Xie and the reciprocity of Zhuangzi, they are but illusive and insubstantial and cannot be questioned. Moreover, Gan Bao's *Record of the Search for Spirits*, Niu Sengru's *Anomalies of the Recondite*, Gu Shenzi's *Broad Expanse of the Extraordinary*, the *East of the River*, the *Record of the Dark Chamber*, the *Examining Spirits*—these works cannot all be without some allegorical content. My book, however, having come about within a cycle of no

[28] For extensive analysis of *Yijian zhi*, see Alister D. Inglis.

> more than sixty years, has utilized both my eyes and ears—and the stories within are all based on factual sources. If one does not believe me, they may go to Mr. Nobody and ask him. (Inglis 24)

This brief preface, when read in the light of a rising Song consciousness of fictionality, contains significant commentaries on the issues of fictionality and factuality that have been overlooked. The statement that stories in the *Yijian zhi* are "all based on factual sources" and deriving from "eyes and ears" (as opposed to imagination) is a nod to the Song custom of stressing factuality and real-world evidence in *biji* records of anecdote as demonstrated in this preface to a Northern Song *biji* collection of supernatural tales:

> Things that are not extraordinary are not worth being passed down; affairs that are not strange are not worth being recorded. Because of my leisurely days, if my eyes have seen something, my heart does not forget it; if my ears have heard something, I certainly chant it in my mouth. I observe spirits and go after [beings of] the other realm; I search for deities and collect the strange. When I run into something, I erect my brush and record it right after. I name my writings *The Secret Record of Searching for the Supernatural*. I open discussions and debates and extensively collect the ill and auspicious omens. I do not differentiate or grade them. My writings are not literary or decorated. I do not lie or avoid [certain subject matters]. My nature is carefree and unrestrained; I cannot make sure that nothing is left out. (Zhang Bingwen 593-594)
>
> 雖然物之不奇，不足以為傳也，事之不異，不足以為記也。予因暇日，苟目有所見，不忘於心，耳有所聞，必誦於口。稽靈即冥，搜神纂異，遇事直

筆，隨而記之，號曰《搜神秘覽》。每開談較議，博採妖祥，不類不次，不文不飾，無誕無避。性多踈曠，不能無遺。

Moreover, Hong's assertion of factuality is also an acknowledgement of similar practices found in earlier collections of supernatural tales referred to in the preface, notably the *Soushen ji* (*In Search of the Supernatural*) 搜神記 from the Jin 晉 period (226-420). In the author's preface, it is stated that the purpose of this collection of supernatural tales is to "make clear of that the Way of the gods is not a fabrication" 發明神道之不誣也 (Gan 5). Such emphasis on real-world evidence falls into the long literary tradition where recorders of anecdotes and extraordinary tales, as observed by Victor Mair, painstakingly "tell us exactly from whom, when, where, and in what circumstances they heard their stories" (22), which calls for a historical and factual, rather than fictional, reading of the stories, however strange they may seem.

Now, is Hong asking for a historical and factual reading of the *Yijian zhi*? Alister Inglis suggests that the preface attests to Hong's "obsession with achieving a historically accurate record" (26). If so, how would one explain Hong's ending statement "if one does not believe me, they may go to Mr. Nobody and ask him"? As explicated by Inglis, *Wuyou xiansheng* 烏有先生 ("Mr. Nobody") in this sentence is found in Sima Xiangru's 司馬相如 (179-117 BC) "Zixu fu" 子虛賦 ("Rhapsody of Master Illusion") and literally means "this person does not exist," which is a play on words (141). In "Zixu fu," both *Zixu* ("Master Illusion") and *Wuyou xiansheng* are made-up entities, as their names suggest. After painstakingly assuring his readers that his stories are factual, Hong tells his

readers to ask a Mr. Nobody for the validity of the stories. Does this not render his previous assertion of factuality ironic and sarcastic? If we recall the above-cited comment made by Hong elsewhere that "*xiaoshuo*-telling and theatre plays were fictional with ghosts and the like," it is apparent that Hong is explicitly aware of the fictionality in stories of "ghosts and the like"; read in this light, this preface seems to implicitly give readers permission to read the stories as fiction whose validity can only be verified by a fantastic Mr. Nobody. In my reading, Hong pretends to be serious about asserting the factuality of his stories until the end of the preface, where the readers are suddenly thrown off by Mr. Nobody.

There is one Southern Song source that describes *xiaoshuo* in a unique way titled *Zuiweng tanlu* 醉翁談錄 (*Record of Talks from a Drunken Old Man*). It is a collection of short stories prefaced by the following statement on *xiaoshuo* by Luo Ye 羅燁 (dates unknown):

> The tradition of *xiaoshuo* originated from the office of secretive remonstration. Then, it was assigned to administer the records of the hundred offices. Thereafter, there are persuaders travelling throughout the four seas and galloped across the hundred schools. They use the elusive and profound writings of the high antiquity to differentiate and illuminate today's debates and discourses. It [*xiaoshuo*] is also called "history-telling," "joined-origin," "tongue-sowing," or "pick-and-dodge." They all have evidence and do not dare to tell lies. They speak of the wise men of previous generations as teachers; they single out the fools of recent ages to caution. Their speeches are not without basis and are beneficial to hear. (2)

> 小說者流，出於機戒之官，遂分百官記錄之司。由是有說者縱橫四海，馳騁百家。以上古隱奧之文章，爲今日分明之議論。或名講史，或謂合生，或稱舌耕，或作挑閃，皆有所據，不敢謬言。言其上世之賢者爲師，排其近世之愚者可謂戒。言非無根，聽之有益。

There are several claims about *xiaoshuo* in this passage that contradict other more well-known accounts. First, the claim that the *xiaoshuo* tradition originated from a secretive remonstration office is nowhere else to be found. It significantly deviates from the aforementioned *Hanshu*'s statement about the origin of *xiaoshuo* as well as that of Ouyang Xiu: "As for biographies, *xiaoshuo*, as well as regional speech, geography, records of offices, and family trees, they all originated from the tradition of official historians" 而傳記、小說，外暨方言、地理、職官、氏族，皆出於史官之流也 (Ouyang, *Xin Tang shu* Vol.5 1421). Secondly, the *Zuiweng tanlu* claims that "history-telling" (*jiangshi* 講史) is an alternative name for *xiaoshuo*. In all other extant contemporary accounts of urban performances, including the *Mengliang lu*, *Ducheng shengji*, *Wulin jiushi* 武林舊事 (*Old Affairs of the Martial Forest*), and *Dongjing menghua lu* 東京夢華錄 (*Record of Dreams and Extravagance of the Eastern Capital*), history-telling is described as a different type of oral performance than *xiaoshuo*; none of these sources state or imply in any way that *xiaoshuo* is highly factual or equivalent to history-telling. Without a manifesto of authorial intention, we can only speculate that perhaps the author is trying to elevate *xiaoshuo*. As mentioned above, this passage is a preface to a collection of short stories. Although it is apparent that *xiaoshuo* as used in

this passage refers to the oral performative kind ("Their speeches are not without basis and are beneficial to hear"), it is not implausible to speculate that the preface implies that the written short stories that follow are also called *xiaoshuo*. However, as the preface stresses the factuality of *xiaoshuo* stories, we cannot say that the author uses *xiaoshuo* to denote written fiction, though the stories can certainly be read as fiction. Although the Song-Yuan period shows considerable fictional consciousness and the rising usage of *xiaoshuo* to denote oral fiction, we will not see a clear relationship between *xiaoshuo* and written fiction until late imperial times.

Ming-Qing

According to Sheldon Lu, the late imperial period saw a rise in general understanding of the nature of fiction:

> In the Ming and Ch'ing periods, people increasingly realized that much of fictional narrative is self-consciously non-historical and ostensibly creative and that fiction ought not to be judged and read as defective history and quasi-history but to be understood on its own terms. Many commentators no longer regarded *hsiao-shuo* as something that needs to be faithful to history. (134)

Lu argues that Ming-Qing discourse explicitly and implicitly suggests an understanding of the fabricated and fictional nature of certain narratives. He points to terms commonly used to analyze novels in this period, such as *bizhen* 逼真 ("realistic"), *moxie* 摹寫 ("imitative writing"), *mohua* 摹畫 ("imitative drawing"), and *ruhua* 如畫 ("drawing-like"), as indications of an acknowledgement of "the fictionality of literary texts [that]

point to their artistry and artifice" (134-135). Moreover, he quotes a passage in Xie Zhaozhe's 謝肇淛 (1567-1624) *Wu zazu* 五雜俎 (*Five Miscellanies*) that explicitly admits to the fabricated nature of *xiaoshuo*: "In regard to fiction [*xiaoshuo*] and dramatic compositions, there should be a mixture of the fictive and the real. Then they become writings that capture the essence of literary games" (Sheldon Lu 135). Wang Wei 王煒 points out that in the Ming-Qing period, it was common to refer to vernacular novels like the *Shuihu zhuan* 水滸傳 (*Water Margin*) as *xiaoshuo* (85).

While it is true that *xiaoshuo* denoted written fiction by the Ming-Qing period, its older meanings, as noted by Li Zhongming 李忠明, did not disappear and were still in usage; Li points to the *Siku quanshu*'s 四庫全書 (*Complete Library in Four Sections*) usage of *xiaoshuo* as a bibliographical category that includes non-fictional works as an example (11-12). In fact, in *Wu zazu*, Xie also uses *xiaoshuo* in a way that refers to its earlier meanings as a catch-all category of a wide range of writings and a bibliographical division:

> Aside from the schools of *ru*, *dao*, *yinyang*, *ming*, *mo*, *zongheng*, *xiaoshuo* and *nong*, there is the school of *za*. It is said that its writings generally originated from the office of remonstration. It combines [the teachings] of *yinyang* and *mo* and merges [the teachings of] *ming* and *fa* [. . .] The school of *xiaoshuo* came from the petty officials of the court. They are creations by those who engaged themselves in idle talks in the streets and alleys and by those who heard gossip and rumors on the way. The two schools are different on these grounds. Ban [Gu] says there are nine schools that are worth looking into; he means to debase [the school of] *xiaoshuo*.

> In later ages, *xiaoshuo* has been extremely popular. It encompasses everything; in this way, *xiaoshuo* is similar to [the school of] *za.* (1)
>
> 儒、道、陰陽、法、名、墨、縱橫、小說、農之外有雜家。云其書蓋出於議官，兼陰陽墨合名法 [. . .] 小說家出於稗官，街談巷語，道聽途說者之所造。兩家不同如此，班言可觀者九家。意在黜小說。後代小說極盛，其中無所不有，則小說與雜相似。

In this passage, Xie draws a link between the school of *xiaoshuo* as documented in the *Hanshu* and *xiaoshuo* of "later ages," which likely includes fictional writings that contain a "mixture of the fictive and the real." This linkage, which disregards the difference between early discursive *xiaoshuo* and later narrative and fictional *xiaoshuo*, is also seen in Lu Xun and Ming Dong Gu's evolutionary discourse. Though the notion of evolution is a modern product, the perceived connection between these two distinct categories of *xiaoshuo* has a premodern origin.

In the CDCL Ming-Qing corpus, *xiaoshuo* is often used to denote vernacular episodic novels, which were widely understood to be fictional. In *Tongsu bian* 通俗編 (*Collection of Common Customs*), the Qing bibliophile Zhai Hao 翟灝 (d. 1788) refers to vernacular novels like the *Shuihu zhuan* and *Sansui Pingyao zhuan* 三遂平妖傳 (*The Three Sui Quash the Demons' Revolt*) as *xiaoshuo* (655-656). Due to the sheer magnitude of *xiaoshuo*'s prominence in late imperial discourse and the size of available literature, the relationship between *xiaoshuo* and fiction can be better illustrated through machine reading and computational analysis.

Collocation Analysis: Methods

As the CDCL is compiled in plain text, the data can be easily used for machine reading and analysis. I used the relevant modules in the Natural Language Toolkit (NLTK), a Python platform for natural language processing, to perform collocation analysis. The NLTK's collocation function is based on Christopher Manning and Hinrich Schütze's *Foundations of Statistical Natural Language Processing* (hereafter *Foundations*).[29] In *Foundations*, collocation is defined as "an expression consisting of two or more words that correspond to some conventional way of saying things" (151). According to this definition, phrases like "nuclear weapon" and "United Arab Emirates" are collocations. Collocation analysis is often used to detect fixed expressions that contain words that frequently appear next to each other, which enhances machine understanding of natural human language. However, in digital humanities, the usage of collocation analysis can significantly differ from that in computer science. For example, in "The Eurocentric Fallacy. A Digital Approach to the Rise of Modernity, Civilization and Europe," Joris Van Eijnatten and Ruben Ros search a corpus of nineteenth- and twentieth-century Dutch newspapers for the co-appearances of words like "Europe," "modern," and "civilize" to study the association between the conception of Europe and the ideas of modernity and civilization (714-716). In this case, the digital humanists are not necessarily looking for set, conventional expressions; rather, they are interested in the intensity of the association among these concepts to understand the cultural perceptions

[29] For the NLTK collocation module's theoretical basis in Foundations of Statistical Natural Language Processing, see the documentation: https://www.nltk.org/api/nltk.html?highlight=collocation#module-nltk.collocations.

regarding their inter-relationship. To understand the development of *xiaoshuo*, I attempt to delineate the textual contexts of its usage throughout history by searching for characters with which it most frequently co-appears in literature of different eras. In such cases, it is better to understand collocation as **a co-appearance of words or characters that is meaningful and significant for a particular investigative purpose.**

According to the NLTK documentation, collocation-finding entails "calculating the frequencies of words and their appearance in the context of other words"; to filter out meaningless co-appearances that occur by mere coincidence, they are "scored according to some association measure [. . .] to determine the relative likelihood of each [co-appearance] being a collocation" ("NLTK 3.4.5 documentation"). Depending on the association measure, which estimates the validity of attributing significance to a co-appearance by calculating its likelihood of being purely coincidental (and thus having no meaning), the results can be significantly different. There are four association measures offered by the NLTK that are explained in *Foundations*: the t test, chi-square test, likelihood ratio, and mutual information (163-183). As mutual information is specifically designed to more accurately identify fixed, conventional phrases and expressions (Manning and Schütze 178-183), it is not appropriate for this study and will be excluded from consideration. The t test, chi-square test, and likelihood ratio all measure, with different statistical reasoning processes, the likelihood of words in a co-appearance having a dependent relationship; if a co-appearance scores low, the words or characters likely reside near each other due to pure chance. Out of these three, as pointed out by Manning and Schütze, the likelihood ratio is the most suitable for sparse data and works well with words that rarely appear in a corpus (172-175). The dataset used in this study is

considered sparse because the targeted keyword, *xiaoshuo*, occurs in a very small portion of the whole corpus. Moreover, according to Manning and Schütze, the likelihood ratio is more interpretable and easier to understand compared to the t test and chi-square test, as it is "simply a number that tells us how much more likely one hypothesis is than the other" (172). The hypotheses here refer to two possibilities: 1) the co-appearance has significance and its words or characters have a dependent relationship and 2) the co-appearance has no significance and its words or characters have no dependent relationship because they co-appear due to pure chance.

The tokenization is performed using *jieba*, a Python text segmentation tool for modern Chinese. Compared to the conventional way of tokenizing Classical Chinese, which assumes the basic unit of meaning to be mono-syllabic and segments texts into single characters, *jieba* has the advantage of being able to recognize and work with bi-syllabic words like *xiaoshuo*.[30] As this study heavily focuses on the Song-Yuan and Ming-Qing periods and involves a large quantity of vernacular literature, the ability to detect bi-syllabic words is essential. Nonetheless, it is important to keep in mind that

[30] Under the guidance of Professor Paul Vierthaler, I have experimented with tokenizing the corpus via an algorithm that segments the corpus into non-overlapping bigrams. This algorithm produces many more false positives than *jieba*, especially with the Qing Dynasty corpus, in the sense that a much larger portion of the bigrams it identifies are not real words. In other words, *jieba* is much better at find real bi-syllabic words in the CDCL, probably because bi-syllabic words largely overlap in modern and premodern Chinese. Although *jieba* proves to be the most suitable method of tokenization for this particular investigative purpose, it is important to keep in mind that tokenizing and segmenting Classical Chinese is an ongoing difficulty in digital research; there is no one-size-fit-all solution as of now. For a summary of currently available ways to segment Classical Chinese, see Shilei Huang and Jiangqin Wu, "A Pragmatic Approach for Classical Chinese Word Segmentation."

jieba's imperfect segmentation of Classical Chinese can affect the calculation of likelihood ratios in unpredictable ways.

Collocation Analysis on the CDCL Corpus

Other than words and characters, the segmentation of time can also affect the quality of this study. Though considerable in size, the CDCL does not have the same amount of texts in each historical period and has drastically fewer texts for earlier periods. As a small amount of text likely will not provide enough occurrences of the word *xiaoshuo* for meaningful computational analysis, it is necessary to combine some periods into one corpus. The entire pre-Tang segment of the CDCL only contains two appearances of *xiaoshuo*; thus, the pre-Tang dynasties cannot form individual units of analysis and must be combined with other periods. Even though the NLTK likelihood ratio analysis is supposed to be able to handle sparse data, it failed to produce any result based on two data points. We can experiment with using Three Kingdoms through the Tang as one historical unit of analysis and using texts from this period to form a pre-Song corpus. The Tang corpus contains fourteen appearances of *xiaoshuo*; in total, the pre-Song corpus contains sixteen mentions of *xiaoshuo*, which is enough to generate some results. However, sixteen is still a very small number of data points and the analysis did not generate any co-appearance that contains information valuable to this study. Since the pre-Song corpus does not contain enough mentions of *xiaoshuo* for the collocation algorithm, we can only use collocation analysis on the post-Tang corpuses.

In the Song-Yuan corpus, *xiaoshuo* occurs 214 times, among which 196 are in the Song corpus. The Ming corpus contains 200 mentions of *xiaoshuo*. The Qing corpus

contains 2,350. The vastly higher number of *xiaoshuo*'s appearances in the Qing corpus is not purely a function of the Qing corpus being larger. To illustrate, the Qing corpus is about 2.5 times larger than the Ming in character count, but 2,350 is significantly beyond twice or thrice of 200. The following are the top collocations, ranked by likelihood ratio, that contain *xiaoshuo* in each corpus (the numerical value is the likelihood ratio as measured by the NLTK):

Northern Song

雜家 小說 43.63

小說 漢武帝 31.86

小說 元后 25.14

小說 晉習 25.14

小說 晉陸士 25.14

Southern Song

稗官 小說 60.92

小說 尋之經史 25.86

小說 私記則 25.86

小說 載毅 25.86

小說 類定 25.86

Yuan

稗官小說 97.50

妙靜 小說 28.28

小說 演史 24.47

小說 私史 22.54

小說 寓目 22.01

Ming

稗官 小說 118.31

唐人 小說 106.94

古今 小說 97.73

小說 補附 49.01

野史 小說 31.28

Qing

稗官 小說 520.64

英文 小說 475.94

唐人 小說 289.95

明人 小說 94.99

小說 家言 85.23

戲曲 小說 72.71

章回 小說 61.18

The output data is modified by a stop-word list that filters out common characters that have no value for our purpose and highly associated with many words due to their quotidian nature, like *yi* 矣, *yan* 焉, *yue* 曰, *zai* 載. For example, *yue* ("to say") is highly associated with *xiaoshuo* to form the compound *xiaoshuo yue* ("*xiaoshuo* says that. . ."); since this compound provides no insight for our purpose, it is disregarded as noise in the data and not included in the lists above. However, even with a stop-word list, the output data still contains significant non-sensical data and noise, like *si ji ze* 私記則, due to the imperfect word segmentation performed by *jieba*, which is inevitable since there is currently no word segmentation tool that can accurately decipher what constitutes a word in Classical Chinese. Nonetheless, the amount of Classical Chinese words recognized by *jieba* as shown in the lists is still impressive considering that *jieba* is designed for modern Chinese.

In the Qing corpus, *xiaoshuo* is highly associated with *xiqu* 戲曲 ("theater and tunes") and *zhanghui* 章回 ("episode"). As discussed above, since the Song-Yuan period, theater was perceived to be largely fictional; this association between *xiaoshuo* and theater reflects a prominent relationship between *xiaoshuo* and fiction. "Episode" here is indicative of the rise of the episodic novel, a literary genre widely understood to be fictional in the late imperial era; this collocation also ties *xiaoshuo* to fiction. Whereas there are instances where *xiaoshuo* explicitly indicates fiction since the Tang Dynasty, it is not until the Qing period that the connection between *xiaoshuo* and fiction is prominent

enough for the relevant collocations to rise to the top of the rankings. Based on my manual reading of *xiaoshuo*'s usage throughout history, I believe the machine's analysis on this issue accurately reflects that *xiaoshuo* did not overwhelmingly denote fiction until the Qing Dynasty.

The collocation *yingwen xiaoshuo* 英文小說 ("English novel") curiously occupies the second place in the list of the Qing corpus. In this particular case, the high likelihood ratio is probably due to the fact that *yingwen* is a rare compound in premodern literature and most of its occurrences are next to *xiaoshuo*, which tells the machine that it has a highly dependent relationship with *xiaoshuo*; it does not necessarily indicate that people in the Qing era regularly discussed English novels. It is important to keep in mind that the likelihood ratio calculates how likely that two words form a meaningful expression and the intensity of their association; the frequency of their appearances is only one factor in this calculation. If a word rarely appears but when it does, it is almost always next to *xiaoshuo*, the machine reckons that they must have a special relationship beyond merely neighboring each other due to pure chance. Among thirty-one appearances of *yingwen xiaoshuo* in the Qing corpus, most are from one text: *Chushi Ying-Fa-E guo riji* 出使英法俄國日記 (*Diary of My Diplomatic Missions to England, France, and Russia*) by Zeng Jize 曾紀澤 (1839-1890). In the diary, Zeng repeatedly documents himself reading English novels (99-101). Although this collocation does not shed light on widespread phenomena, it points to a finding that I would not have otherwise noticed: *xiaoshuo*'s capacity to denote foreign novels in as early as the nineteenth century. The significance of this to the emergence of a modern East Asian discourse on fiction will be discussed in

the next chapter. Another nineteenth century instance of *xiaoshuo* denoting foreign novels is in a text by Wei Yuan 魏源 (1794-1857) in 1842, *Haiguo tuzhi* 海國圖志 (*Maps and Records of Oceans and States*), in which he discusses the culture of Aden (in modern day Yemen) and refers to the *Arabian Nights* (*Yiqian ling yi ye* 一千零一夜, or "one thousand and one nights," in Chinese) as *xiaoshuo* (773).

The collocations *Tangren xiaoshuo* 唐人小說 ("*xiaoshuo* of Tang people") and *Mingren xiaoshuo* 明人小說 ("*xiaoshuo* of Ming people") in the Ming and Qing rankings also deserve attention. Since the Song period, *xiaoshuo* is regularly discussed in the context of *xiaoshuo* from a previous dynasty; this phenomenon started with Song literati's fondness of invoking Tang *xiaoshuo*. This is important because it reflects the fact that the meanings of *xiaoshuo* are accumulative throughout the ages. As mentioned before, although *xiaoshuo* acquired new significations throughout history, its previously established meanings persisted and would not be erased until the modern era; by late imperial times, not even one of *xiaoshuo*'s earlier significations discussed above fell into obscurity and disuse. In the Tang period, literature concerning "absurd and baseless things like ghosts, gods, transformations, and fantasy" was denoted as *xiaoshuo*; this meaning of *xiaoshuo* carried over to later dynasties and can be invoked as *Tangren xiaoshuo*. In the Song period, *biji* writings and tales of the supernatural were called *xiaoshuo*; these meanings were passed on to the later ages and became *Songren xiaoshuo* 宋人小說 ("*xiaoshuo* of Song people"). The immensely persistent nature of *xiaoshuo*'s layers of meanings built on top of each other over the course of Chinese history is what

renders it one of the most puzzling and complex concepts to decipher. In the next chapter, we will examine yet another layer of meaning added to it in the twentieth century.

The most prominent, and also the most puzzling, pattern shown in the collocation analysis is the rise of *xiaoshuo*'s close relationship with *baiguan* 稗官 ("petty official") starting from the Southern Song period. A collocation analysis performed on *baiguan* also demonstrates its strong relationship with *xiaoshuo* since the Southern Song period. Due to the scarcity of *baiguan*'s appearance before the Southern Song (seven in the Northern Song corpus and two in the entire corpus prior to the Northern Song), the entire corpus prior to the Southern Song is combined together to form a pre-Southern Song corpus:

Pre-Southern Song

庶繼代洪烈 稗官 29.46

旁收 稗官 29.46

逮傳記 稗官 29.46

稗官 所采 24.96

秘方 稗官 22.76

Southern Song

稗官 小說 60.92

稗官 虞初 49.08

稗官 小 30.57

稗官 帝之學 26.39

稗官 街談巷 26.39

Yuan

稗官 小說 97.50

之燕談 稗官 52.30

稗官 小 43.88

稗官 之紀 38.36

下極 稗官 26.08

Ming

稗官 小說 118.31

不博覽 稗官 26.31

旁史 稗官 26.31

海誌 稗官 26.31

碎之言 稗官 26.31

Qing

稗官 小說 520.64

稗官 之傳記 486.82

考諸 稗官 446.18

稗官 野乘 125.75

稗官 家 49.52

The collocation analyses on *xiaoshuo* and *baiguan* reveal that the close association between these two is a distinctly post-Northern Song phenomenon. The lists of collocations as shown above are not complete, as the complete ones are very long and include collocations with likelihood ratios as low as less than three. In both analyses, pre-Southern Song corpuses do not contain the collocation *baiguan xiaoshuo* at all, not even at the bottom of the lists with a low likelihood ratio.

The earliest extant appearances of the word *baiguan* can be found in the excavated Qin[31] bamboo slips from Shuihudi 睡虎地 and Yunmenglong 雲夢龍. In these documents, *baiguan* refers to low officials at the level of counties and villages; it designates the administrative level, not a specific job function (Wang and Liu 68-70). In terms of received literature, a search in the Chinese Text Project's pre-Qin and Qin-Han corpus demonstrates that the only appearance of *baiguan* in this period is in the aforementioned passage by Ban Gu in the *Hanshu*, where *baiguan* is an agent who collects talks from the streets. This particular definition of *baiguan* is the one that is passed onto later literature. In the CDCL pre-Song corpus, *baiguan* only appears twice.

[31] Qin here refers to both the Qin Dynasty (221-206 BC) and the Kingdom of Qin.

The first is in the *Wenxin diaolong* 文心雕龍 (*The Literary Mind and the Carving of Dragons*), compiled by Liu Xie 劉勰 (fl. 5th century):

> Perhaps puns and parables were indispensable to literature, as 'Trivia' [*xiaoshuo*] must be accorded a place among the ten schools of philosophy; what the minor officers [*baiguan*] gathered could after all still be a useful source of information. (54)

In the Tang corpus, Liu Zongyuan 柳宗元 (773-819), in his preface to a poem, writes,

> In honesty, such unsightly words and profligate sounds are not worthy of being inscribed in metals and stones. Fortunately, due to the great contributions of each generation, *baiguan* and villagers can collect and sing them. (320)
>
> 誠醜言淫聲，不足以當金石，庶繼代洪烈，稗官里人得采而歌之。

Though not in the CDCL, Wang Qizhou 王齊洲 and Liu Fuling 劉伏玲 point out that the Three Kingdoms scholar Ru Chun's 如淳 (dates unknown) commentary on Ban Gu's passage claims that the office of *baiguan* was specifically created to collect talks on the streets:

> Discussions from the streets and talks from the alleyways are words of a trifling and piffling nature. Kings wished to know the customs and cultures of the villages and alleyways and thus created the office of *baiguan* to have [*baiguan* officials] report them. (Wang and Liu 68)
>
> 街談巷說，其細碎之言也。王者欲知閭巷風俗，故立稗官，使稱說之。

It is evident that Ban Gu's narrow definition of *baiguan* was taken de facto in later times and eroded the earlier, broader notion as found in the excavated bamboo slips.

In the CDCL's Song corpus, *baiguan* designates discourse that is unreliable, nonserious, unorthodox, or fabricated. This passage in Liu Qi's 劉跂 (fl. 1079) *Xueyi ji* 學易集 (*Collection on the Study of The Classic of Changes*) illustrates *baiguan*'s inferior status to orthodox learning:

> Being passionate about learning, he frequently and persistently [studied] the classics and histories of various schools. Reaching as low as biographical tales and the talks of *baiguan*, there was nothing he did not seek out and read. (611)
> 好學勤篤經史諸家 下逮傳記稗官之言 無不讀求。

For *baiguan*'s capacity to designate fabricated narratives, see this passage from the *Zhuzi yulei* 朱子語類 (*A Collection of Conversations of Master Zhu*) where *baiguan* denotes fabricated stories:

> The old eunuch Huang Jiefu served Emperor Hui. He said that the Daoist Lin Lingsu used magic; there is no truth in that. Talks like those of Wen Ge regarding seeing ghosts and gods are all *baiguan*. I have never seen them. (Zhu 576)
> 老內侍黃節夫事徽宗，言道人林靈素有幻術，其實也無。如溫革言見鬼神者，皆稗官，某不曾見。

In all cases of its appearances, *baiguan* seems to be highly interchangeable with *xiaoshuo*, except that it can specifically refer to the agent that collects such discourse, which *xiaoshuo* cannot. Such interchangeability partially explains these two words' link in the collocation ranking and is likely due to the fact that the very concept of *baiguan* is

defined by Ban Gu, who firmly ties it to *xiaoshuo*. After looking through each individual appearance of *baiguan* in the entire CDCL corpus, I have seen no example of this word being invoked in contexts unrelated to Ban Gu's notion of it. Like *xiaoshuo*, *baiguan* does not exclusively refer to narratives. For example, in the anonymous Southern Song text *Airi zhai congchao* 愛日齋叢抄 (*Collected Writings of the Airi Academy*), *baiguan xiaoshuo* is explicitly described to include treatises on subject matters such as magic, medicine, and ritual:

> The nine hundred forty-three pieces of *baiguan xiaoshuo* are all about magic, medicine, banquets, sacrificial prayers, and talks that are passed around in alleyways. They were collected to form this book. (691)
>
> 蓋稗官小說 凡九百四十三篇 皆巫醫厭祝及里巷之所傳言 集為是書。

Prior to the Southern Song, I have not found the exact compound *baiguan xiaoshuo*, though these two words often appear close to each other.[32] Since the compound is a distinctly post-Northern Song phenomenon, its rise is potentially connected to the Northern to Southern Song transition. As discussed above, *xiaoshuo* became significantly entangled with the performative tradition in the Southern Song. However, there is no usage of *baiguan xiaoshuo* in the Southern Song or Yuan corpus that ties it to the performative tradition. Its meanings include all of *xiaoshuo*'s meanings established by the Northern Song period and it has no specific designation that *xiaoshuo* cannot accomplish on its own, which begs the question as to why this compound even needed to exist. The

[32] The collocation analysis only account for characters and words that appear immediately next to each other.

sudden appearance of this compound in the Southern Song and its progressively increasing popularity ever since might be linked to a lost Southern Song text titled *Baiguan xiaoshuo*, which is referred to in two Southern Song texts in the CDCL. In both cases, the texts quote a specific passage from *baiguan xiaoshuo* concerning a curious worm:

> According to *Baiguan xiaoshuo*, there is a worm in the Southern Sea that has no bones and is called *ni*. When it is in water, it lives; when it leaves water, it collapses and resembles a pile of mud. (Wu Zeng 623)
>
> 按《稗官小說》南海有蟲，無骨，名曰泥。在水中則活，失水則醉，如一堆泥然。

This segment is preceded by a few sentences of similar structures in which *an* 按 ("according to") is followed by known text titles like *Hangong yi* 漢宮儀 (*Rituals of the Han Palace*) and *Beishan jing* 北山經 (*Classic of the Northern Mountain*), which renders it very likely that *Baiguan xiaoshuo* here is also the title of a text. In the other case, the context also suggests that *Baiguan xiaoshuo* is a title:

> I saw *Baiguan xiaoshuo* and learned that it says there is a worm in the Southern Sea that has no bones and is called *ni*. When it is in water, it lives; when it leaves water, it collapses and resembles a pile of mud. Then, I read *Tales of the Five Kingdoms*, which says [. . .] (Zhang Jibang 62)
>
> 予觀《稗官小說》乃得其說 云南海有蟲 無骨 名曰泥 在水則活 失水則醉 如一堆泥 然後又讀五國故事云 [. . .] 。

The matching wording of the boneless worm reference also points to the likelihood that *Baiguan xiaoshuo* here refers to a text from which exact sentences can be copied, as opposed to the general notion of minor discourse. Other than these two, I have found no other instance of this compound being the title of a text; the association between *xiaoshuo* and *baiguan* post-Northern Song is not due to this title being mentioned over and over again. Barring other potential explanations that I have yet to discover, we can only speculate that the compound *baiguan xiaoshuo* was perhaps popularized by a text bearing this title.

Conclusion

This chapter illustrates that *xiaoshuo*'s linguistic development is better understood as a process where its signifying capacity expands throughout the ages, rather than a linear trajectory where it steadily evolves toward the modern notion of fiction—a view founded by Lu Xun that will be the focus of the next chapter.[33] In my view, the study of fiction and that of *xiaoshuo* should be considered related but separate fields. Whereas the study of fiction is based on understanding premodern literature based on a modern Western concept (i.e. fiction) with no exact equivalent in premodern Chinese, the study of *xiaoshuo* concerns a traditional Chinese concept with no equivalent in English. Although the concept of fiction is modern and Western, applying it to premodern Chinese literature can avoid becoming teleological if there is no presumption that developments in premodern literature should be understood in the context of an evolution toward an end-

[33] See Chapter 1 for Ming Dong Gu's support of Lu's view.

product like the late imperial novel or the modern concept of fiction. The interchangeability of *xiaoshuo* and fiction in Lu Xun's discourse and modern scholarship is highly problematic because fiction is only one of many significations of *xiaoshuo* in some periods of history; confounding *xiaoshuo* and fiction is necessarily a teleological process because the initial equivalence drawn between them in the formative period of Chinese and Japanese modernity was motivated by a desire to help traditional East Asian fiction transition to its modern, Western-inspired reincarnation. In the next chapter, we will discuss how the traditional notions of *xiaoshuo* laid the foundation for a modern Sino-Japanese discourse on fiction founded by Lu Xun and Tsubouchi Shōyō, who transformed the meanings of *xiaoshuo*/*shōsetsu* to lead China and Japan into literary and intellectual modernity.