Keeping It Classy: Sinitic Classifiers and Their History in Literature

Will Evans

This thesis is presented in partial fulfillment of the requirements for
the degree of Bachelor of Arts in Linguistics

Haverford College
Spring 2022

Abstract

The purpose of this study is to chart the history of nominal classifiers in Sinitic (Chinese) languages. The particular focus is the forms in which classifiers appear throughout the written record, and to aid this analysis data was gathered from a corpus of literary works spanning from the very earliest complete works of literature written in the 5th century BC to full-length vernacular novels written in the 18th century AD. The study finds that classifier phrases gradually began to overtake other methods of counting beginning around the 5th century AD, but oddly count phrases that do not utilize classifiers persisted in the literature at least as far as 1740 AD, which should not have been possible at least in the spoken language. Two solutions are presented to account for this co-occurrence of what should be complementarily distributed structures. The first being a prosodic solution, as detailed by Feng (2012), and the second being one that focuses instead on extra-linguistic aesthetic concerns that may have artificially preserved syntactic structures that were seen as more "literary" even though they were no longer found in the spoken language. Ultimately, the study is inconclusive as to which, if either, account is better suited to explain the discrepancies observed in the data, but the importance of considering extra-linguistic factors in particular is emphasized.

Acknowledgments

I would like to acknowledge and thank Professor Shizhe Huang for acting as my advisor for this project, as well as an advisor to me in general all throughout my four years at Haverford. Her dedication and invaluable insights have been a boon to my learning on a level I struggle to find the words to describe. Without her I would hardly be half the scholar I am today. 老师，大恩不言谢. I would also like to thank Professor Brook Lillehaugen for acting as my co-advisor and helping me keep on track and manage the hectic schedule required to complete this paper, as well as my classmate and student reader Liam Santry for his suggestions.

Table of Contents

1. Introduction

1.1 *Purpose of the present study and overview*

With respect to any historical study of language change and variation, the Sinitic language family, most often for convenience's sake referred to together as 'Chinese,' occupies a particularly privileged position on account of the sheer magnitude of written records that exist from the modern day to the earliest centuries of Chinese civilization. Indeed, as one of the few societies to independently invent writing (Senner, 1991), China and the Chinese languages share this position with only a select few other languages, most of whose writing systems have all but died out. This written record allows for historical investigations of considerable detail that can provide great insights into past forms of Chinese that might otherwise be more difficult to reconstruct. Relying on written data, however, presents its own set of challenges. To say nothing of the difficulties of developing any understanding of early Chinese phonetics and phonology (Jacques, 2017), the discrepancies between the syntax recorded in written works and the syntax of the language actually spoken by those who composed them produce a murky view of how early Chinese proto-languages, such as Old or 'Archaic' Chinese, were actually structured syntactically. Though one can question whether the extant written records are reliable references for the spoken syntax of earlier Chinese languages, and indeed this study will, they are still by far the most efficient means by which to reconstruct the linguistic history of China. Such are the goals and methods of the present study, to analyze syntactic structures found in pre-modern Chinese literature and draw conclusions about historical spoken Chinese syntax and syntactic change.

The present study will maintain a primary focus on the origin and development of classifiers as evidenced from writing.The specific questions under consideration were: how do earlier usages of classifiers differ from modern Chinese classifiers from a syntactic-semantic standpoint, and what, if anything, can the textual evidence tell us about the change and variation of classifiers in the spoken language. A study of such scope as to incorporate a large scale quantitative study of literature for the specific comparison classifier in vernacular and non-vernacular texts as well as a philological discussion on the veracity of said sources has not, to the present author's knowledge, previously been undertaken. Similar works such as Peyraube (1991) cover the most basic aspects of the history such as when certain syntactic structures first appear and some examples of them, but provide little further analysis. Similarly, Wong & Lee (2019), though it includes a large scale quantitative study, is limited in scope to the Buddhist canon and does not provide not a very extensive analysis of the data. It is possible that non-English language, particularly Chinese, studies have pursued similar questions with a similar methodology, but such studies have yet to be brought to light.

The study begins by first giving an overview of the syntax and semantics of modern Chinese classifiers followed by an analysis of of data from two other studies on similar subject matter though with a smaller scope, Peyraube (1991) and Wong & Lee (2019), which is the only other study found that incorporates a large scale quantitative analysis though it does not focus solely on classifiers. The data gathered from various vernacular and non-vernacular texts is then considered in the works surveyed with respect to their classifier distribution, the number of unique classifiers that are found in a work, and classifier density, the proportion of classifiers to overall characters in each work. This section also attempts to apply the prosodic-syntactic concepts found in Feng (2012) to explain the data, which is in turn followed by an alternative explanation that takes extra-linguistic factors into account in order to explain discrepancies in the

data, namely the overlap of classifiers with other numerical count phrases, and introduces the idea of an aesthetic-syntactic interface.

The conclusions made by the study unfortunately include little that had not already been posited by other works, but its value lies in its scale and methodology. This study confirms other ideas, such as the general timeline of the emergence of classifiers as found in Peyraube (1991) and Feng (2012), that being that classifiers initially developed in the post-nominal position and the earliest occurrences were strictly measure words, classifiers that denote units of measurement, containers, etc. Non-measure word or 'count' classifiers (the distinction will be clarified later) then developed also in the post-nominal position after which both measure word and count classifiers begin to occur more frequently in the post-nominal position. Alternative count phrases, those numerical phrases that don't include a classifier, occur with regular frequency throughout the entire written record.

1.2 *Matters of terminology*

An unfortunate reality when approaching historical studies of Chinese languages is the impreciseness of terminology with respect to the languages in question. Put simply, when one refers to any historical variety of 'Chinese,' a term that is already less than ideal for describing any modern language of the Sinitic family, it is often difficult to further specify the exact language being referred to. As such, before continuing the present study, a discussion of terminology is in order. The first and foremost question in this respect is what is 'Chinese' in a historical sense. 'Chinese' is of course a word used to refer to any number of modern languages within the Sinitic family, and in that sense it is not particularly useful for the present study as it is neither specific to any one language nor can it be meaningfully applied to historical Sinitic languages. As such, the present study will utilize a model much more well suited to a textual

discussion of historical Sinitic languages, the paradigm of Literary Sinitic and Vernacular Sinitic, as utilized by Victor Mair in his *Columbia History of Chinese Literature*. Literary Sinitic refers to any written variety of a Sinitic language composed in a prestige form, what is typically referred to as 'Classical Chinese,' most often found in the context of official histories, documents, and essays. Vernacular Sinitic, in contrast, is essentially any form of writing not adhering to the conventions of Literary Sinitic and is found in a number of texts, most notably the vernacular stories of the Ming dynasty and the vernacular novels of the Ming (1368-1644 AD) and Qing (1636-1912 AD) dynasties . Vernacular Sinitic will be the main focus of this study, as the syntax of classifiers in Literary Sinitic experiences little variation over time, but Literary Sinitic data, particularly with respect to very early documents, will be used for reference. Usages of other terms such as 'Old/Archaic Chinese' and 'Middle Chinese' will be used specifically in reference to reconstructions of Sinitic proto-languages done by the likes of Baxter & Sagart and Pulleyblank.

2. Structure of (modern) classifiers

2.1 *Syntax of classifiers*

The primary focus of this analysis is to chart the history of classifiers in the Sinitic family as evidenced by written records. As such, an overview of Sinitic classifiers themselves is in order. As is the case in other classifier languages, classifiers in Sinitic languages play a mediating role in counting and referring to an NP. In modern Sinitic languages, classifiers must come between a number and the noun it is counting or between a noun and a demonstrative, as in the classic example of (note that in this and all subsequent sections, any translation that does not include a citation was done by the present author):

(1a) 一本书
yi ben shu
one-CL book
'One book'

(1b) 这/那本书
zhei/nei ben shu
this/that-CL book
'this/that book'

In this sense, the role of a classifier, or CL, is to make an instance or number of instances of the entity denoted by the noun countable and able to be referred to. This combination of a CL and an NP, with an optional node for a number before the CL, form the phrasal category of 'classifier phrase.' The syntax of classifier phrases is a widely discussed topic, but the phrase-level syntax is not the focus of this study. Instead this study will focus on classifiers themselves as lexical items, though the broader phrasal structure of classifier phrases will be discussed in reference to the properties and function of CLs (see sections **4.3** and **4.5**).

A distinction is often made between 'count classifiers' (Cheng & Sybesma 1998), also known as 'sortal classifiers' (Sybesma, 2017), and 'measure words.' Whereas count-classifiers, aside from the general classifier 个 *ge* which is used for nouns without a more specific classifier, categorize the NP by some quality it possesses (e.g. 张 *zhang* being a classifier for flat things that is used to count pieces of paper, desks, etc.), measure words denote that the NP is defined by predetermined units of measurement such as 斤 *jin* or 'half kilogram' or by spatial terms that are not unique to the denotation of the noun, such as 块 *kuai* 'piece; chunk'. A more thorough discussion of this distinction can be found in Sybesma (2017), but since the two behave the same syntactically, for simplicity's sake 'classifier' will herein be used to refer to both count-classifiers and measure words, unless disambiguating the two is relevant. It is worth noting that the distribution and interpretations of phrases with classifiers is not uniform among all modern Sinitic languages, but further discussion is unnecessary at this time (for further explanation see Cheng & Sybesma, 2005). While the question has been raised as to whether Sinitic languages, specifically Mandarin, observe a count-mass distinction or instead implicitly treat all nouns as

mass nouns (Chierchia, 1998), the literature yet lacks a definitive conclusion in favor of an all mass model. As such, Sinitic nouns will be treated as observing a count-mass distinction. Indeed, there is one aspect of Sinitic classifiers that is evidence towards a clear count-mass distinction, measure words are the only acceptable form of CL for counting or referring to nouns that are, at least more overtly so, mass nouns, while count-classifiers, can only be used for count nouns. This aspect of their distribution suggests that there is some property of N that establishes a distribution requirement for CL, allowing some CLs while disallowing others. For example, in the following, (2a) is acceptable while (2b) is not:

(2a) 一杯水	(2b) *一个水
yi bei shui	*yi ge shui*
one-cup water	one-CL water
'A cup of water'	'One water'

Of course, the translation of (2b) is also unacceptable in English, but the example points to a clear divide in the distribution of classifiers and nouns, that is some classifiers cannot be used with some nouns. Interestingly, this distinction is somewhat one-sided. While mass nouns can never be used with count-classifiers, count nouns are allowed to occur with measure-words, provided the measure word is pragmatically acceptable, as in:

(3a) 一本书	(3b) 一箱书	(3c) *一杯书
yi ben shu	*yi xiang shu*	*yi bei shu*
one-CL-book	one-box book	one-cup book
'One book'	'One box of books'	'One cup of books'

As per the above, the count noun *shu* 'book' can be classified by both count-classifiers and measure words but only those measure words that make sense. That is, books do not normally come in cups, so (3c) is not felicitous. This section has largely focused on the internal mechanics of classifiers, much of which might not seem immediately relevant, but what is most important with respect to the structure and function of classifiers is that there is no other way in

Sinitic languages to directly instantiate and count or refer to of N other than the structure of (Num/Dem+)CL+NP.

The notion of classifiers being required for countability or referentiality however, will be challenged by textual evidence, as direct combination of numerals and nouns is found throughout the entire textual tradition and is largely the only acceptable structures for counting in Literary Sinitic. The acceptability of Num+NP phrases, or NP+Num phrases in some cases, suggests that either the nature of Sinitic nouns ought to be reanalyzed or that there was a fundamental shift in the structures and mechanics behind countability and referentiality between pre-modern and modern Sinitic languages. This question will be explored generally throughout the following sections and more specifically in section **4.4**.

2.2 *Semantics of classifiers*

Much has already been alluded to about the semantics of classifiers and is worth further discussion. Firstly, the relationship between classifiers and individuation. As previously mentioned, the primary semantic role of classifiers is to instantiate N, that is to manifest a specific atomic unit N from the amorphous semantic entity denoted by N. This function is the so-called 'individualizing function' (Croft 1994), and is greatly impactful for the semantics of counting in Sinitic languages. For example, as Cheng & Sybesma (2005, p.13) argues:

> Closely related to the countability function is the 'individualizing function' (Croft 1994, 162) of classifiers….We may rephrase this and state that the classifier singles out one entity from the plurality of entities provided by the semantic representation of the noun in the lexicon; it picks out one instance of what is denoted by N. This is also represented in the Chinese way of saying ‘two books’: it uses the individuating classifier as is shown in (32), not the plural one, as the ungrammaticality of the phrases in (33) asserts.

(32) san ben shu — Mandarin
saN bun zhu — Southern-Min
saam bun syu — Cantonese
three CL^{volume} book

(33) *san xie shu — Mandarin
*saN se zhu — Southern-Min
*saam di syu — Cantonese
three CL^{L} book

This description provides a solid definition of the syntactic-semantic functions of CLs, but CLs possess some aspects worth investigating that are purely semantic. It would be a mistake to view CLs as semantically bleached, when in fact they are in some cases quite the opposite. For example, in 1a. above the semantic content of 本 *ben* in 一本书 *yi ben shu* 'one book' can reasonably be said to be redundant, as *ben* is the classifier for books and *shu* means book. To take the meaning of these two lexical items at their face value would render something along the lines of a 'booky book,' that is 书 *shu* 'book' is classified under the category of, unsurprisingly, books. It is often the case however, that CLs do not have such a clear link between CL and N where the set of nouns classified by CL is not as small as in *ben,* which, as previously stated, classifies *shu* 'book' as well as other bound volumes such as magazines or encyclopedias. Indeed N often has seemingly no connection at all with CL. Take for instance the following (adapted from Zhang, 2007):

(4a) 一颗豆 — (4b) 一颗原子弹
yi ke dou — *yi ke yuanzidan*
one-CL bean — one-CL atom bomb
'One bean' — One atom bomb'

(4a) and (4b) both take the same classifier but the meanings of N are radically different. How can such be the case? These two cases are classic examples of how a CL's set of nouns expands through analogy. The CL *ke* is generally used for small, round objects, as in (4a) with *dou* 'bean.' At some point, *ke* became associated with bullets, denoted by 弹 *dan* 'bullet'. *Dan* itself became morpheme used to denote weapons that are launched or dropped, and so it was affixed to *yuanzi* 'atom' to form *yuanzidan* 'atom bomb.' Throughout this word formation process, *dan* retained its classification under *ke*, and so any word that includes *dan* is also classified by *ke*.

This example and many more are covered in greater detail by Zhang (2007) but is included here to exemplify how the semantic connections between CL and N are not always apparent but do in fact exist. This discussion is all to say that the semantics of CLs, specifically count-classifiers, should not be overlooked, and their semantic history is quite important to understanding the development of CLs, as unlike measure words, count-classifiers themselves were not arbitrarily created as units of measurement or exclusively for the role of classifiers but rather were likely nouns in their own right that were reanalyzed as classifiers as CLs became increasingly required by the syntax for any instance of counting. It is possible that this pattern of development for count-classifiers, their origins as nouns unrelated to units of measurement, could help explain the noticeable differences in the rate of adoption of count-classifiers as opposed to measure words, a topic that will be discussed later in section **4.4**.

3. Investigating textual data

3.1 *CL Structures present in literature*

Unlike modern Sinitic languages, Literary and Vernacular Sinitic do not have very restrictive requirements for the syntax of counting and the structure of classifier phrases, when they are used. Earlier writing in particular utilizes a diverse set of structures that perform essentially the same function as modern Sinitic classifiers. Given that they do not all contain a classifier or similar lexical item, however, such phrases will instead be referred to as count phrases given their unified purpose in enumerating N. Peyraube (1991) provides a comprehensive overview of the types of structures found from the earliest records, the oracle bone inscriptions of the Shang dynasty (1600 - 1046 BC) to the Late Medieval period (7th - 13th century AD). Given the wide scope of his paper, Peyraube provides more summation and overview than analysis and is more concerned with semantics than syntax, but the data provided

and the thoroughness of its presentation form a solid foundation for any discussion of pre-modern Sinitic count phrases. The list of structures as identified by Peyraube with some examples from his data are as follows (note that Peyraube does draw a distinction between classifiers, CL, and measure words, MW):

(5) Num+Noun
一牛
yi niu
one ox
'One ox' (14th - 11th c. BC)

(6) Noun+Num
狐一百六十四
hu yi-bai-liu-shi-si
fox one-hundred-six-ten-four
'One hundred sixty four foxes' (14th - 11th c. BC)

(7) Noun 1+Num+Noun 2
羌十人
qiang shi ren
Qiang (people) ten people
'Ten Qiang' (14th - 11th c. BC)

(8) Noun+Num+MW
贝廿朋
bei nian peng
shell twenty coupled string
'Twenty coupled strings of shells' (14th - 11th c. BC)

(9) Num+MW+Noun
一簞食一瓢饮
yi dan shi yi piao yin
one basket food one bowl drink
'One basket of food, one bowl of drink' (5th - 3rd c. BC)

(10) Noun+Num+CL
骑马三匹
qi ma san pi
ride horse three-CL
'Ride three horses' (2nd c. BC - 3rd c. AD)

(11) Num+CL+Noun
二百五十条牛
er-bai-wu-shi tiao niu
two-hundred-five-ten CL ox
'Two hundred fifty ox' (3rd - 6th c, AD)

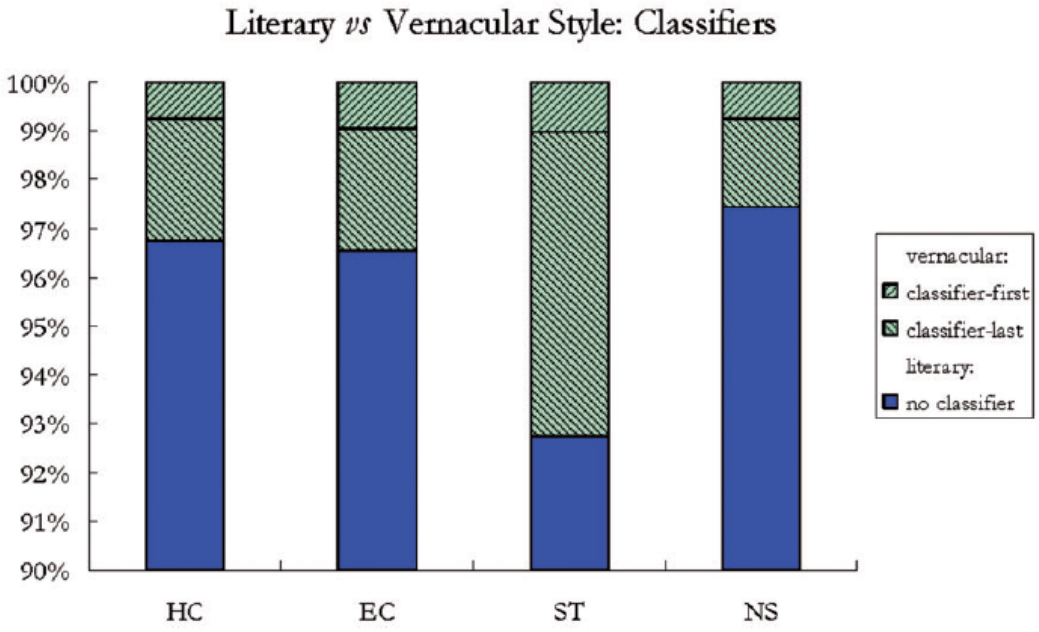
As per this data, it's clear that count phrases exist in considerable diversity throughout the written record and furthermore coexist with one another with seemingly no issue. After the first appearances of measure words, measure word phrases and then later count-classifier phrases exist alongside phrases without a classifier, with non-CL phrases making up at least a plurality of instances for the majority of history (until around 7th - 13th c. AD when CLs become more widespread), which would suggest that the language allowed for them to coexist, in contrast to modern Sinitic languages where only (Num+)CL+NP is acceptable. This trend in the distribution of pre-modern count phrases begs the question of to what extent they are syntactically and semantically similar and what forces led to the loss of alternative structures.

3.2 *Wong & Lee (2019)'s quantitative analysis*

Before discussing the potential forces and trends at work in the origins and developments of count phrases in Sinitic languages, a larger pool of data is needed. For this purpose, the study will turn to Wong & Lee (2019) as well as data procured specifically for this study. The purpose of Wong & Lee's study is closely related to the present study, as it performs a quantitative analysis of what they consider to be elements indicative of Vernacular Sinitic in literature, in this case the Buddhist Canon, which include the presence of Num+CL+N and N+Num+CL classifier phrases. The former is referred to in the graph as 'classifier first,' as, ignoring the number, the classifier comes before the noun, and the latter is referred to as 'classifier last,' but the remainder of this study will refer to them as pre-nominal and post-nominal classifiers respectively. The

graph of their data on CLs is included below, which incorporates both classifier phrases (dashed lines, right-slanting for pre-nominal CLs, left-slanting for post-nominal CLs) and non-classifier count phrases (blue), and is presented in chronological order starting from the Han-Jin (3rd c. BC - 4th c. AD) period, rendered below as 'HC' on account of the study using Wade-Giles romanization which renders 'Jin' as 'Chin,' followed by the Eastern Jin or 'EC' (4th - 5th c. AD), the Sui-Tang, 'ST' (6th - 10th c. AD), and finally the Northern Song, 'NS' (10th - 12th c. AD).

(3.2a)



The data depicts only a slight progression in the inclusion of elements of Vernacular Sinitic from the Han-Jin period to the Sui-Tang period. In the case of both pre-nominal or 'classifier-first' phrases and post-nominal or 'classifier-last' phrases the transition from the Han-Jin period to the Eastern Jin period sees an increase of only a small fraction of a percent. Following that, however, the Sui-Tang period saw an enormous increase in the use of specifically

N+Num+CL, that is post-nominal, CLs followed by an abrupt decrease in usages of all CLs in the Northern Song period, with the instances of usages falling back to even lower than the Han-Jin period. The sudden decrease is explained by Wong & Lee as likely a result of literary 'Stylists' in the Song courts who were in charge of polishing translations of Buddhist texts, often removing vernacular elements in the process, essentially forcing it into the framework of Literary Sinitic. What then explains the massive jump in Vernacular Sinitic CLs from the Eastern Jin to the Sui-Tang period? While largely speculation, it may be the case that the Sui-Tang increase in Vernacular Sinitic CLs is at least in part a product of the birth and spread of Chan Buddhism around the same time, which put particular emphasis on directly communicating the words of prominent monks in text without altering their words to fit the standards of Literary Sinitic (Mair 2001, p. 168). As a result of the proliferation of Chan texts, text within the Buddhist Canon rapidly transformed into something that greatly privileged vernacular writing over the more opaque styles of Literary Sinitic, that is until the usage vernacular styles fell out of favor with official translators in the Song courts. Wong & Lee's quantitative work is quite useful for any study of the history of count phrases in Sinitic languages, but it is not without its faults. Namely, the temporal scope is quite lacking and the selection of texts, while vast in itself, is limited with respect to style and authorship. By focusing entirely on Buddhist texts, the data says little about the language of texts in the secular literary world, which at the same time produced some at least semi-vernacular texts such as *Shishuo Xinyu* (5th century AD). CLs, of course, did not disappear from the language on account of the decisions of court Stylists during the Song. Rather, their harsh decline in usage within the data is a product of the data being limited to Buddhist texts that have always been at the mercy of translators and editors. To get a broader view of the historical usage of CLs, one must venture beyond Buddhist texts and into works that are indiginously Chinese and, in the case of those works of Chinese authors written in Vernacular Sinitic, more

representative of a spoken language than works that were either translated or edited to fit the standards of Buddhist practices and preferences. Still, the influence of the so-called Stylists points towards a very interesting point to be considered when looking at written data, the influence of extra-linguistic factors, namely stylistic concerns pushed by a centralized literati such as the Stylists, on the inclusion or exclusion of certain linguistic elements in texts. This idea is developed further in section **4.5**. Filling in the gaps left by Wong & Lee's data is where the present study's own data comes into play, as it includes its own quantitative analysis of the appearances of CLs and other count phrases. Instead of being limited to the Buddhist canon, however, its sources are works purposefully written in Vernacular Sinitic. These sources include the great vernacular novels of the Ming and Qing dynasties, as well as other vernacular stories, with earlier, non-vernacular texts, included for comparison.

4. New quantitative analysis

4.1 *Methodology*

As already noted, though Wong & Lee's study is a great starting point for a quantitative study of CLs in literature - even more so if one is specifically interested in Buddhist literature and its connection to Vernacular Sinitic writing - the study's presentation lacks depth and further discussion of how CLs appear in context, which is not helped by the lack of explanation of what criteria was used to designate a data point as a CL. The present study aims to provide more transparency with respect to the methodology utilized to gather and collate the data as accurately and efficiently as possible, though a number of difficulties arose in the process. This section will explain how the data was collected.

The primary question posed before developing the methodology for data collection was where do CLs appear in the written record and in what diversity and density do they appear (i.e.

how many distinct CLs appear in a work and how many instances are there of a given CL), and to help elucidate these two questions a simple function was programmed Python that utilizes a Regular Expression to search for any instance of Num+CL and recorded how many occurrences there were of each CL from a list of all known CLs (this approach was not without its flaws, as will be discussed candidly in the next section). The answer to the former question is that they appear sparsely at first but become increasingly commonplace. Such a conclusion is unsurprising given that precursors to CLs developed very early in Old Chinese before becoming fully fleshed out and extensively grammaticalized by the time of Middle Chinese (Peyraube, 1991). The very earliest records suggest that other count phrases, such as Num+N, N+Num, and N1+Num+N2, largely predate any instances of count-classifier usage, though not measure words, as a small number of measure words are found even in the oracle bone texts in the N+Num+MW structure, By the time of the the writing of the classics during the Spring and Autumn and Warring States periods (8th-3rd centuries BC) CLs, or at least what appear to be CLs, had developed, though in very limited numbers. By then, however, they had not replaced other count phrases in any noticeable capacity. That is to say the most widely used count phrase remained Num+N or N+Num, a fact echoed by Wong & Lee (2019). As for the latter question of diversity and density, early texts, though they might make use of a relatively diverse list of CLs, tend to only include a very small number of instances of any given CL, often only one.

4.2 *Critique of methodology*

While the approach taken to this study managed to gather a considerable amount of data, its preciseness leaves quite a bit to be desired. Firstly, perhaps the greatest issue is in how the data itself is divided. In programming the Python function that would skim each work using a Regular Expression in order to find combinations of numerals and classifiers, the Regular

Expression constructed for the task was less than ideal because it did not differentiate between post-nominal and pre-nominal CLs as, for example, Wong & Lee's study did. Furthermore, the function did nothing to detect N+Num and Num+N phrases, and so one of the most common count phrases throughout all of the works surveyed was left without a proper quantitative analysis. There are no doubt ways to solve both of these problems, but given that the present author's grasp of programming is rudimentary at best, no such solution could be devised.

Another issue lies in instances in which irrelevant data points were detected and recorded when they should have been ignored. The most notable instances of this issue relate to the problem of the number 一 *yi* 'one'. It is often the case that 一 *yi* with a classifier is not actually denoting one instance of a noun but is rather a common component in a number of set phrases. Take for example the common adjective 一点 *yidian* 'a little.' As exemplified by the following minimal pair, when 点 *dian* has the sense of a small amount of something, 一 *yi* is the only acceptable number it can occur with:

(12) 我下课以后要休息一点
wo xiake yihou yao xiuxi yidian
我下课 以后 要 休息 一点
I end class after will rest a little
'After class I will rest a little

(13) *我下课以后要休息三点
wo xiake yihou yao xiuxi sandian
我下课 以后 要 休息 三点
I end class after will rest three little
*'After class I will rest three little'

From (12) and (13) the way in which 点 *dian* is distinct from a classifier in certain contexts is clear, and yet 点 *dian* still must be included in the data, because in some cases it is a classifier. For example, 点 *dian* is also a classifier for matters of discussion as in 三点建议 *sandian jianyi* (three-CL proposal) 'three proposals'. Another similar example that might be even more confounding given that the usage of the classifier in question is much more general is

the case of 眼, which can either mean ‘eye’ and by extension ‘glance,’ or can be a classifier for large hollow objects like pots or wells. Its usage as a word for ‘glance’ however sometimes has it combined with a number in the context of certain verbal expressions such as 瞄一眼 *miao yiyan* ‘shoot a glance at’ or 看一眼 *kan yiyan* ‘have a look.’ Similar to 一点 *yidian*, such phrases are confounding to the data as they suggest instances of the classifier for large hollow objects when such is not the case.

How exactly one would get around this issue is unclear and would no doubt require stronger programming abilities than was available for this study. Any further attempts to perform a study of this manner should certainly take instances of ‘fake classifiers’ such as the above into account when developing a methodology to detect relevant data within a body of texts, but doing so will require a more nuanced approach to the search methods utilized than those of the present study. Thankfully however, even accounting for the possibility of such confounding data points, the data collected presents a clear trend with respect to the proliferation of classifiers, and many specific instances have been pulled from the texts to illustrate this trend.

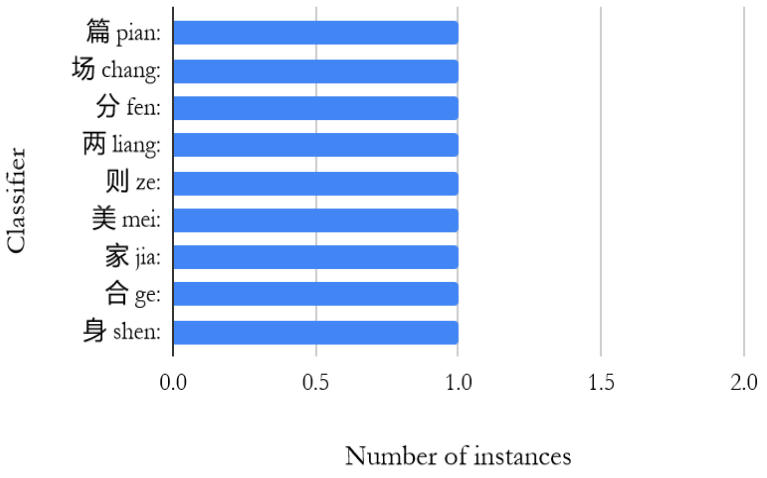
4.3 *Data from Literary Sinitic Sources*

For the Literary Sinitic data, the following sources were used: the *Analects* (5th-3rd century BC), the *Art of War* (5th century BC), the *Dao De Jing* (4th century BC), the *Mengzi* (late 4th century BC), the *Han Feizi* (mid-3rd century BC), the *Zhuangzi* (3rd century BC), *Records of the Grand Historian* (c. 94 BC), *Shishuo Xinyu* (mid-5th century AD), and *Liaozhai Zhiyi* (1740 AD). As evident from the dates given, the purpose of choosing these works in particular was to cover various points in the history of Chinese literature to ensure that a wider variety of Literary Sinitic styles were included in the data. There is of course a considerable focus on works from the Spring and Autumn and Warring States periods, that is the pre-Qin

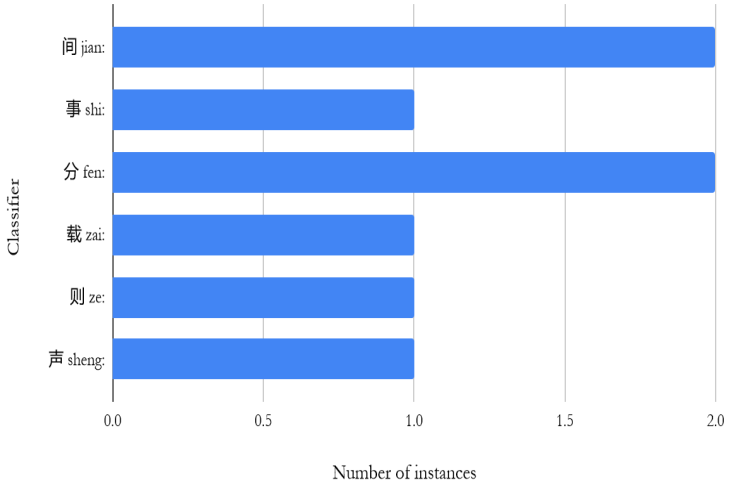
dynasty (221-206 BC) period, as these works, being the earliest works of literature, have been the subject of study for millenia and their foundational influence on Chinese literature can hardly be understated (Mair 2001, Introduction, ch. 3, 4). This style is defined by its terseness and reliance on allusion rather than directly stating one's meaning, a practice that may well have both material as well as aesthetic or philosophical origins. To quote Victor Mair, 'From the very earliest stage of writing in China, there was such an emphasis on concision that many parts of spoken language were omitted. This may initially have been due to the intractable nature of the media (bones and shells that had to be scratched with a sharp instrument) and the highly particularized function of the texts...Anaphora and elision were invoked to such an extreme degree that it became a sort of game between author and reader; the challenge was to see how much the former could leave out without losing the latter' (Mair 2001, p. 27) This style would then be perpetuated throughout history in the form of the *Five Classics*, which became the basis of formal education (Stordalen, 2007), and therefore literacy and literary style, during the Han dynasty. In this sense, the classical literature of the Spring and Autumn and Warring States periods and their styles have had an immense influence on all of Chinese literature. Still, three works are included beyond the pre-Qin period, the monolithic *Records of the Grand Historian*, which is considered one of the greatest works to come out of the Han dynasty, the *Shishuo Xinyu*, which was written during the tumultuous and transitional period of the Northern and Southern dynasties, and *Liaozhai Zhiyi*, which is a very late work in the timeline of Literary Sinitic literature, which would all but die out less than two centuries after the publication of *Liaozhai Zhiyi*. These three later works should also provide insight into how CL usage evolved beyond the classics, if it did indeed at all. The data collected, though cumbersome in their presentation, are worth looking at directly, as a number of interesting and unexpected trends can be seen within them.

Since this study gathered a considerable amount of data, presenting it all at once would be far too cumbersome, so in this section and the next graphs will be presented as required, and the full list of graphs will be included in the appendices. Furthermore, the number of characters in each graph are simply too many to fit in alongside a full gloss, so definitions will be provided when a specific character is referenced.

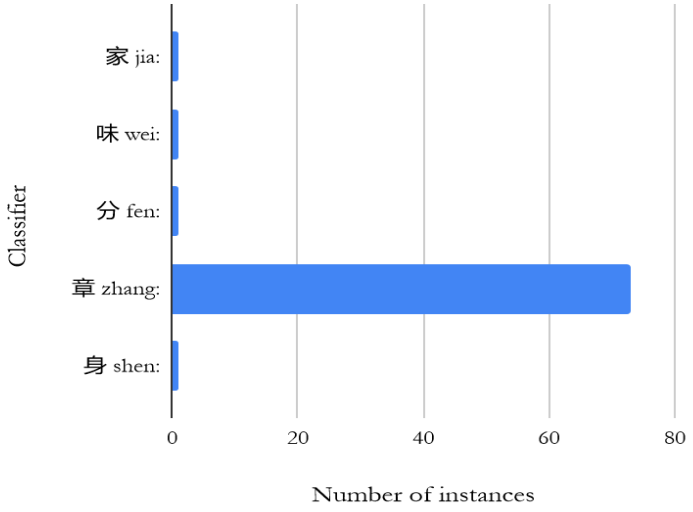
4.3a Analects CL counts



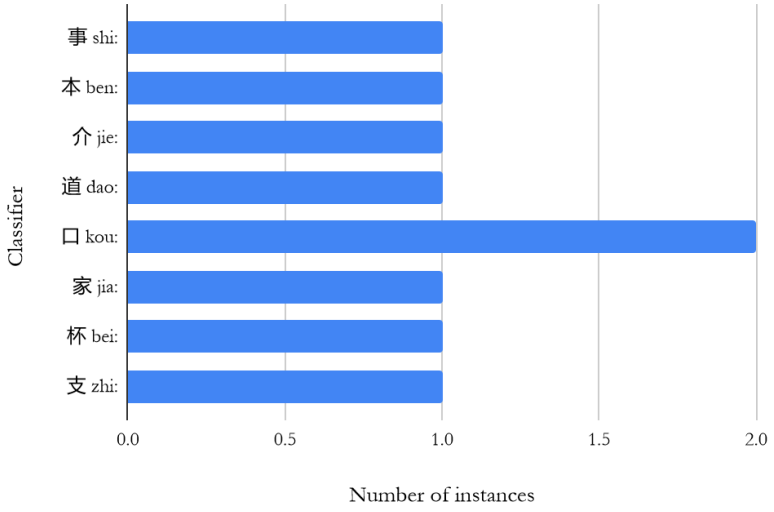
4.3b Art of War CL counts



4.3c Dao De Jing CL counts



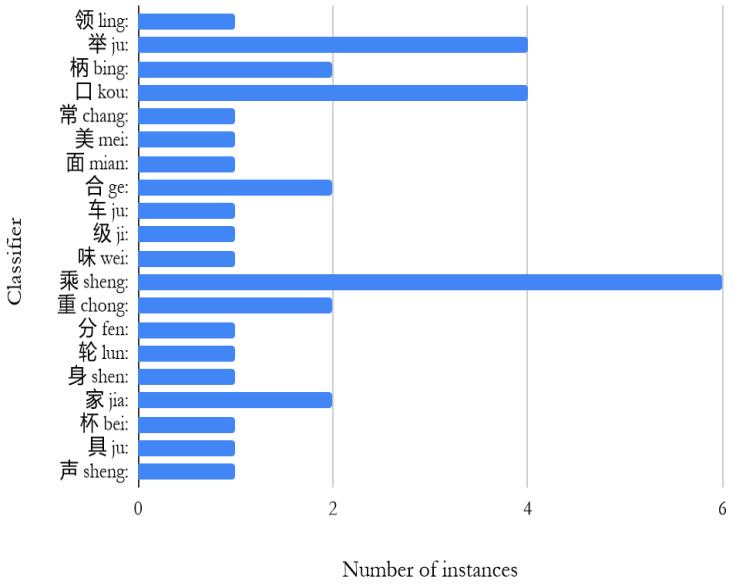
4.3d Mengzi CL counts



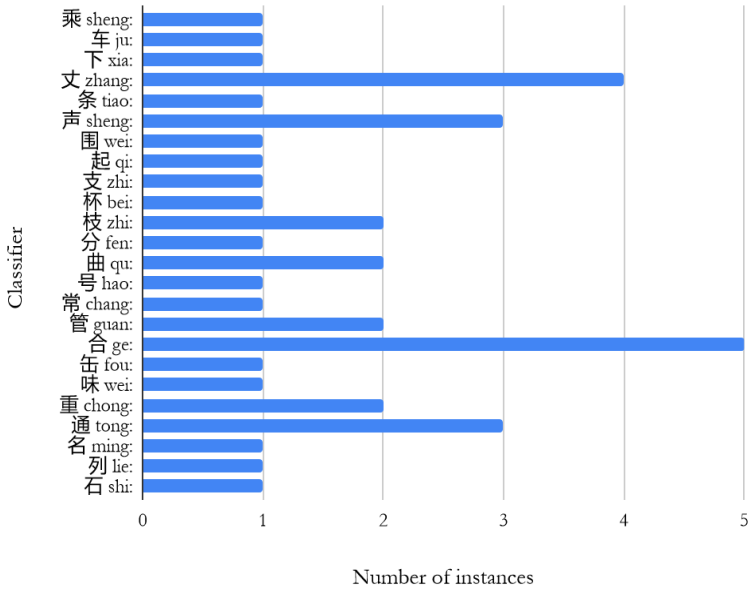
Firstly, the earliest works covered, 4.3a to 4.3d present a very noticeable lack of both diversity and density in their usage of CLs. The *Analects* have only nine instances of a CL, each

only appearing once in what is not exactly a short work, as the *Analects* is over 21,000 characters long. The pattern is repeated in 4.3c and 4.3d with the *Dao De Jing* and the *Mengzi*. It is clear that the earliest works of Chinese literature had little use for CLs, but there is of course one glaring exception, namely *Dao De Jing* in 4.3d. The *Dao De Jing* has an overabundance of the character 章 *zhang* 'chapter,' which no doubt resulted from the Python function picking up the headings for each chapter that denoted the chapter number. These usages of 章 *zhang* are not CLs but rather nouns. They were included, however, because there are some instances of *zhang* in the body of the text itself (although after looking into the text itself, it seems that the other instances are probably not CLs either). Aside from 章 *zhang* in *Dao De Jing*, the other early works seem to almost uniformly have a single instance of each CL, which suggests that CL usage at this point in literature was still very rare and what few examples are found in early works more than likely still retained their full meaning as nouns, and therefore their usage was exclusive to Num+N or N+Num count phrases (the same issue is dealt with in Peyraube, 1991, p. 110). Additionally, such examples as found in these four and other early works could also be instances of ‘fake classifiers’ as warned against in section **4.2**.

4.3e Han Feizi CL counts



4.3f Zhuangzi CL counts



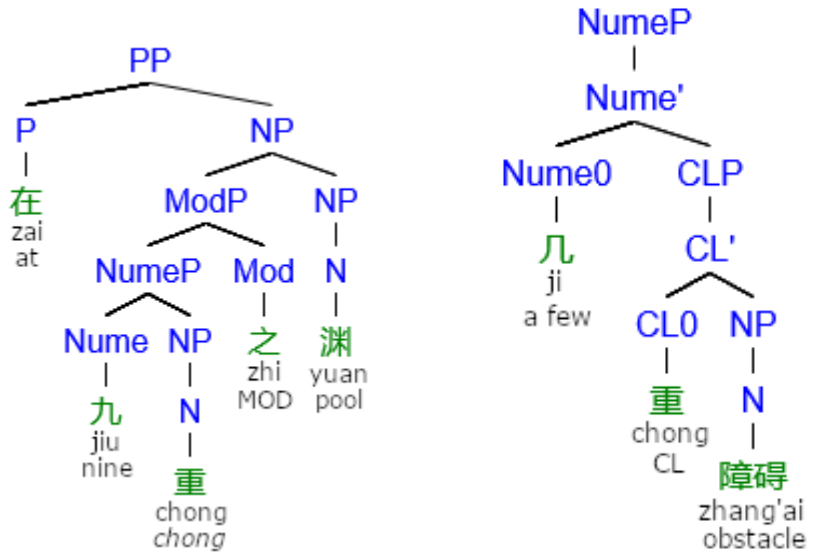
The last of the pre-Qin and Han texts, 4.3e and 4.3f more or less follow the same patterns of distribution found in earlier texts but with a noticeable increase in both diversity and density. In these works a number of measure words, or at least characters shared by measure words, seem to rise to prevalence, which could be the result of an increase in their distribution in the language or simply due to the context of the work. It is tempting to claim that these works include a number of count classifiers that would be recognizable to any modern reader of Sinitic writing, such as 口 *kou* (found four times in 4.3e *Han Feizi*), which in modern usage is used to count a number of living things such as members of a family or pigs, and 门 (found three times in 4.3g *Records of the Grand Historian*, see Appendix I for full data), which in modern usage is used to count abstract concepts such as subjects of study or languages. Instances of these CLs, however, can not be taken to be proof of count classifiers in early literature, however. One major issue with how the data were collected is that they make no distinction between instances of a CL and instances of a noun that shares the same character. Consider the difference between the following (count phrases/classifier phrases are bolded in the gloss just for clarity):

(14) 夫千金之珠, 必在九重之渊而骊龙颔下
fu qian jin zhi zhu, bi zai jiu chong zhi yuan er li long ling xia
夫 千金 之 珠, 必 在 九 重 之 渊
MDL thousand gold ATTR pearl must at nine zhong ATTR deep pool
而 骊龙 颔 下
and black dragon chin under
'So a pearl worth a thousand pieces of gold must have been within a pool nine *chong* deep and under the Black Dragon's chin'
(*Zhuangzi - Miscellaneous chapters - Lie Yu Kou,* line 982)

(15) 写毕业论文的时候我遇到了几重障碍
xie biye lunwen de shihou wo yudao le ji chong zhang'ai
写 毕业论文 的时候，我遇到了 几 重 障碍
write thesis DE time I encounter-PRF a few CL obstacles
'When writing my thesis, I encountered a few obstacles'

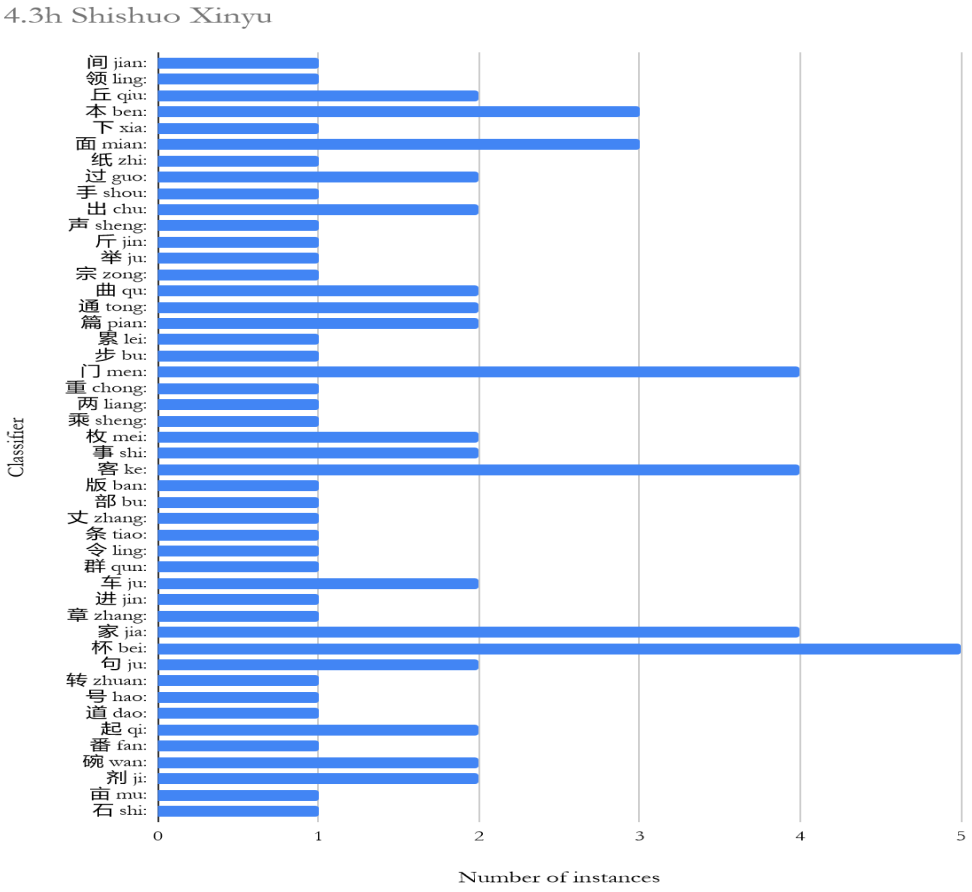
In (14), an excerpt from the *Zhuangzi* (4.3f), we see the character 重 *chong* used in a very similar context to its usage in (15), a sentence written in modern Mandarin, that is as a part of a noun phrase that is an object of a PP in the former case and the object of a VP in the latter case. The key difference between the two usages of 重 *chong*, however, is the presence of the attributive marker 之 *zhi*, whose modern equivalent, 的 *de* can be seen in (15), in (14). In Literary Sinitic, *zhi* marks the preceding phrase as being an attribute of the following phrase, or in the case of (14) it connects two noun phrases that are in some way related to one another. That is to say, 重 *chong* in (15) is a classifier, but in (14) it is a noun, evidenced by the fact that it is followed by the attributive particle *zhi*, which can only connect two nouns. Considering the two sentences structures makes the matter even clearer:

(14a) (15a) (Adapted from Cheng & Sybesma 2005)



As made clear by the syntax trees (14a) and (15a), the structure of the two sentences are, unsurprisingly, quite different. While the 重 *chong* in (15) forms the head of the CL, with the NP as its sibling, in (14) it is the head of its own NP within the numeral count phrase 九重 *jiu chong*, which is sister to the modifier particle 之 *zhi*. The treatment of CLs as full nouns by the syntax is perhaps the greatest difficulty in accurately defining where exactly CLs first arose in the written record. This same problem is also mentioned by Peyraube (1991), in support of his argument that true count-classifiers did not develop until at earliest the period between 2nd c. BC and 3rd c. AD. According to Peyraube, earlier examples of what appear to be classifiers still retain too much of their original meaning and function as nouns to be considered true classifiers (Peyraube, 1991, p. 112-113). Still, whether a supposed instance of a classifier is too noun-like to be considered a CL or not is difficult to empirically determine, particularly during a transitionary period where some nouns are actively undergoing a process of grammaticalization, so this study suffers from the same uncertainty with respect to the exact time period of the emergence of CLs as other studies. Regardless, the above does help flesh out the timeline of the development of CLs in earlier periods, and given both the syntactic evidence above and the similar line of

argumentation given by Peyraube (1991) in this case it is more likely than not that CLs found in literature up to the first century BC and the writing of *Records of the Grand Historian* are not CLs but rather Num+N or N+Num count phrases in which the N is a noun that would eventually develop into a CL.



One work remains of the Literary Sinitic works surveyed to be discussed in this section, *Shishuo Xinyu* as seen in chart 4.3h (Given the fact that it was published long after the other Literary Sinitic works and much closer to the Vernacular Sinitic works, a discussion of the very last work among the Literary Sinitic works surveyed, *Liaozhai Zhiyi* will be saved for the following section, where it will be used for comparison against the Vernacular Sinitic works). *Shishuo Xinyu*, written during the 5th century AD, shows a considerable progression compared to earlier works with almost double the number of distinct CLs found in later early works and

well over double that of the earliest works. First and foremost, this work seems to include the earliest verifiable use of 杯 *bei* 'cup' where it clearly acts as a measure word, as in the following sentence:

(16) 长星！劝尔一杯酒，自古何时有万岁天子！(*Shishuo Xinyu*, line 818)
长星！劝 尔 一杯 酒，自 古 何时 有 万 岁 天子
comet exhort you **one-CL wine** from old when have ten thousand year-old emperor
'Comet, I toast you with a cup of wine! From all antiquity, when was there ever a Son of Heaven who lived ten thousand years?' (Mather et al., 2002, p. 206)

Which is essentially identical to modern usage:

(17) 昨晚我喝了一杯酒，马上就累了
昨晚 我喝了 一杯酒， 马上 就 累了
last night I drink-PERF one-CL wine immediately then tired-PERF
'Last night I had a glass of wine and immediately was tired.'

In this sense, pre-nominal CLs demonstrably existed by the 5th century AD, at least in the form of measure words. Post-nominal MWs are also found:

(18) 郗嘉宾钦崇释道安德问，饷米千斛 (*Shishuo Xinyu*, line 802)
chi jia bin qin chong Shi Dao'an de wen, xiang mi qian hu
郗 嘉宾 钦 崇 释道安 德 问 饷 米 千 斛
Chi esteemed guest respect revere Shih Tao-an virtue gift, give **rice thousand *hu***
'Ch'ih Ch'ao, out of respect and reverence for the virtuous reputation of the monk Shih Tao-an, made him a present of a thousand hu of rice'
(Mather et al., 2002, p. 203)

In (18), the unit of measurement 斛 *hu* (~fifty liters) is combined post-nominally with the noun 米 *mi* 'rice' to render '(a thousand) *hu* of rice.' It is quite peculiar that both pre-nominal and post-nominal MWs are present within the same text, but a clear reason as to why is not immediately apparent. Given that both the pre-nominal example in (16) and the post-nominal example in (18) are objects it seems unlikely that for MWs there existed some positional requirement that would require MWs in object position to be exclusively post-nominal. While both pre-nominal and post-nominal MWs exist in the *Shishuo Xinyu*, the same cannot be said about count-CLs.

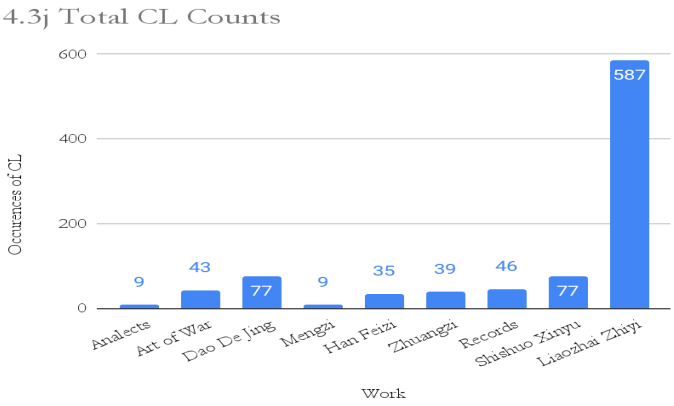
In addition to having a number of clear instances of MWs, *Shishuo Xinyu* also includes count-CLs, though to a much lesser extent, all instances of which are post-nominal. The clearest early instances of count-CLs, where the CL is neither a measure word nor a noun that shares a character with a later count-CL, can be found in *Shishuo Xinyu*'s use of the count-CL 枚 *mei*, which is a general classifier used to classify objects that lack a more specific classifier, a role that would later be taken over by the more modern 个 *ge*, which remains the general classifier in modern Sinitic languages. Though appearing only twice, both instances of 枚 are worth looking at:

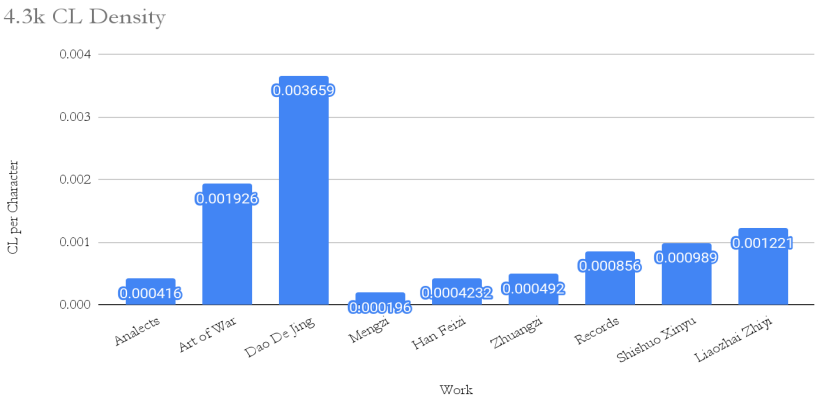
(19) 因下玉镜台一枚。(*Shishuo Xinyu*, line 2263)
因 下 玉 镜台 一枚
so deliver jade mirror stand **one-CL**
'So saying, he deposited an engagement present of a jade mirror stand'
(Mather et al., 2002, p. 481)

(20) 乃命左右悉取珊瑚树，有三尺、四尺，条干绝世，光彩溢目者六七枚。(*Shishuo Xinyu*, line 2346)
乃 命 左右 悉 取 珊瑚 树，有 三 尺、 四 尺，
Thus order attendant all fetch coral tree has three *chi* (0.33 meters) four *chi*
条 干 绝世 光 彩 溢 目 者 六七枚
branch trunk unique light color overwhelm eye NOM **six-seven-CL**
'Where-upon he ordered his attendants to bring out all his coral trees, every one of them three and four ch'ih in height., their branches and trunks surpassing anything in this world, and their luster and color overwhelming the eyes-six or seven trees in all' (Mather et al., 2002, p. 497)

As evidenced by both (19) and (20), where count-CLs appear in *Shishuo Xinyu*, they are post-nominal. These two are further evidence against the idea of a positional requirement, as the CL in (19) is an object and the one in (20), while not a subject modifies the phrase '有...者.' Therefore, the position of a CL likely has no effect on whether it is a pre-nominal or post-nominal phrase, and the lack of pre-nominal count-CLs could simply be a case case of pre-nominal count-CLs being entirely unacceptable at the time in which *Shishuo Xinyu* was written, which would result in all count-CLs being post-nominal by default rather than on account of some positional requirement. One matter that bears discussing before further data is

introduced, however, is the matter of how post-nominal CLs are actually interpreted. By all accounts, classifier phrases with post-nominal CLs seem to be predicates rather than nominal structures themselves, especially in the case of sentences like (20) where the CL follows a phrase that has been clearly nominalized by the particle 者 *zhe*. The question of whether or not post-nominal CLs should instead be treated as predicates is a difficult one to treat with, as if they were always predicates, then such would imply that post-nominal classifiers as a nominal structure never really existed. The matter is further complicated by the theory that Proto-Chinese was a head-final SOV language, under which case post-nominal CLs can be seen as a remnant of older syntax. In this sense, how post-nominal CLs ought to be read relies heavily on whether one believes Sinitic languages before Old Chinese were SOV. Evidence towards the word order shift hypothesis can be found in Feng (1996) and further discussion on how such a word order shift would have affected the reading of post-nominal CLs can be found in Feng (2012), which is discussed in detail in section **4.5**.





Finally, looking at 4.3j and 4.3k, which list the works in chronological order, we see a steady increase in the number of CLs or words sharing a character with a CL. Since instances of actual count-CLs were not found in the early data, it's likely that before *Shishuo Xinyu* what are marked as increases in CL occurrences are actually increases in the usage of nouns that share a character with or were precursors to CLs, though it is possible that some CLs might have been

missed in later early works such as *Records of the Grand Historian*. If Peyraube (1991) is any indication, however, measure words, both pre-nominal and post-nominal, developed relatively early, but were perhaps not as ubiquitous as to be found in the small selection of works covered. There is also a remarkably small increase in CL usage between *Shishuo Xinyu* and *Liaozhai Zhiyi* considering the two were written around 1300 years apart, which gives credence to the idea that Literary Sinitic writing was guided by very conservaitve ideals when it came to incorporating vernacular diction.

4.4 *Data from Vernacular Sinitic Sources*

The second half of the data gathered all come from various sources written in Vernacular Sinitic. The earliest source is the vernacular novel *Romance of the Three Kingdoms*, written some time in the 14th century by Luo Guanzhong. *Romance* is written in a distinctly vernacular style that was uncommon for the time, but is nevertheless a valuable resource for understanding the spoken language of its author and audience. *Romance* is the first part in a canon of works that would become known as the Four Great Classic Novels, which includes *Romance of the Three Kingdoms* (14th c. AD)*, Water Margin* (mid-14th c. AD)*, Journey to the West* (c. 1592 AD)*, and Dream of Red Chambers* (c. 1791 AD). Additionally are two more novels that are sometimes grouped together with the Four Great Classic Novels into another canon known as the Six Great Classic Novels, *The Plum in the Golden Vase* (c. 1610 AD) and *The Unofficial History of the Scholars* (1750 AD). In addition to the vernacular novels are three collections of vernacular short stories written and compiled by the Ming dynasty author Feng Menglong, *Stories to Instruct the World* (1620 AD)*, Stories to Caution the World* (1624 AD)*,* and *Stories to Awaken the World* (1627 AD). With all these works together, the Vernacular Sinitic data spans a period of about four hundred years, which is about the same time span as the Literary Sintic data sources,

excluding *Shishuo Xinyu* and *Liaozhai Zhiyi*. A similar timespan should allow for a better perspective on the gradual development of the language represented by the vernacular texts as is seen in the earlier Literary Sinitic texts.

The data itself (see Appendix II for a full list of graphs) points towards a clear conclusion, the prevalence of CLs in Vernacular Sinitic literature cannot be understated. It is clear that by the time of these works' writing, the presence of CLs in the language of the time, potentially Early Mandarin, as the authors, or at least alleged authors were from the east and northeast in and around the capitals of Nanjing and Beijing, had already begun to approach that of modern Sinitic languages. What is particularly worth mentioning is the overwhelming presence of the general classifier 个 *ge* in every work. As previously mentioned, *ge* is the modern general classifier that replaced the earlier general classifier 枚 *mei*. Interestingly, however, 枚 *mei* seems to still exist in limited quantities within texts where *ge* is present in great numbers. Upon further inspection, however, it seems there is no real contradiction, as 枚 *mei* has lost its usage as a general classifier and gained other specific meanings (or perhaps retained other specific meanings that had been previously overshadowed by its role as a general classifier), namely as a classifier for small round objects like dates, peaches, hair buns, etc. This denotation is especially noticeable in *Water Margin*, where instances of 枚 *mei* almost exclusively classify small fruits or objects of a small, round shape, as in the following:

(21) 头绾两枚丫髻，身穿一领青衣 (*Water Margin*, line 59)
头 绾 两枚 丫髻，身 穿 一领 青 衣
head bind **two-CL hair bun** body wear one-CL blue garment
On (the child's) head was tied two hair buns and (the child) wore a blue garment

What then of the structure of other CLs in the Vernacular Sinitic data? With other changes that bring them closer to modern CLs, such as the introduction of *ge*, have they also

moved to the modern, pre-nominal position? The short answer is yes but with some exceptions. Even at a cursory glance at each work, it's clear that a number of CLs have become Num+CL+N phrases. Take for example the following:

(22) 从褡裢中取出一面镜子来——两面皆可照人 (*Dream of Red Chambers*, line 435)
从 褡裢 中 取出 一面 镜子 来 两面 皆可 照 人
from pouch within take out **one-CL mirror** come two side all can reflect person
'(He) took out from his pouch a mirror, both sides able to show one's reflection'

(23) 到得半路，只見一簇軍馬 (*Romance of the Three Kingdoms*, line 42)
到得 半路， 只 見 一簇 軍 馬
arrive halfway just see **one-CL soldier horse**
Arriving halfway down the road, (they) saw a crowd of soldiers on horseback

Interestingly, however, some CLs are still post-nominal, as per the following also from *Dream of Red Chambers* and *Romance of the Three Kingdoms*:

(24) 春天开的白牡丹花蕊十二两 (*Dream of Red Chambers*, line 294)
春天 开 的 白 牡丹 花蕊 十二两
spring open ATTR white tree peony stamen twelve-CL.tael
'Twelve taels of pistil from the white tree peonies that bloom in the Spring'

(25) 各引哨船五只 (*Romance of the Three Kingdoms*, line 1083)
各 引 哨船 五只
each lead sentry boat five-CL
'Each leading five sentry boats'

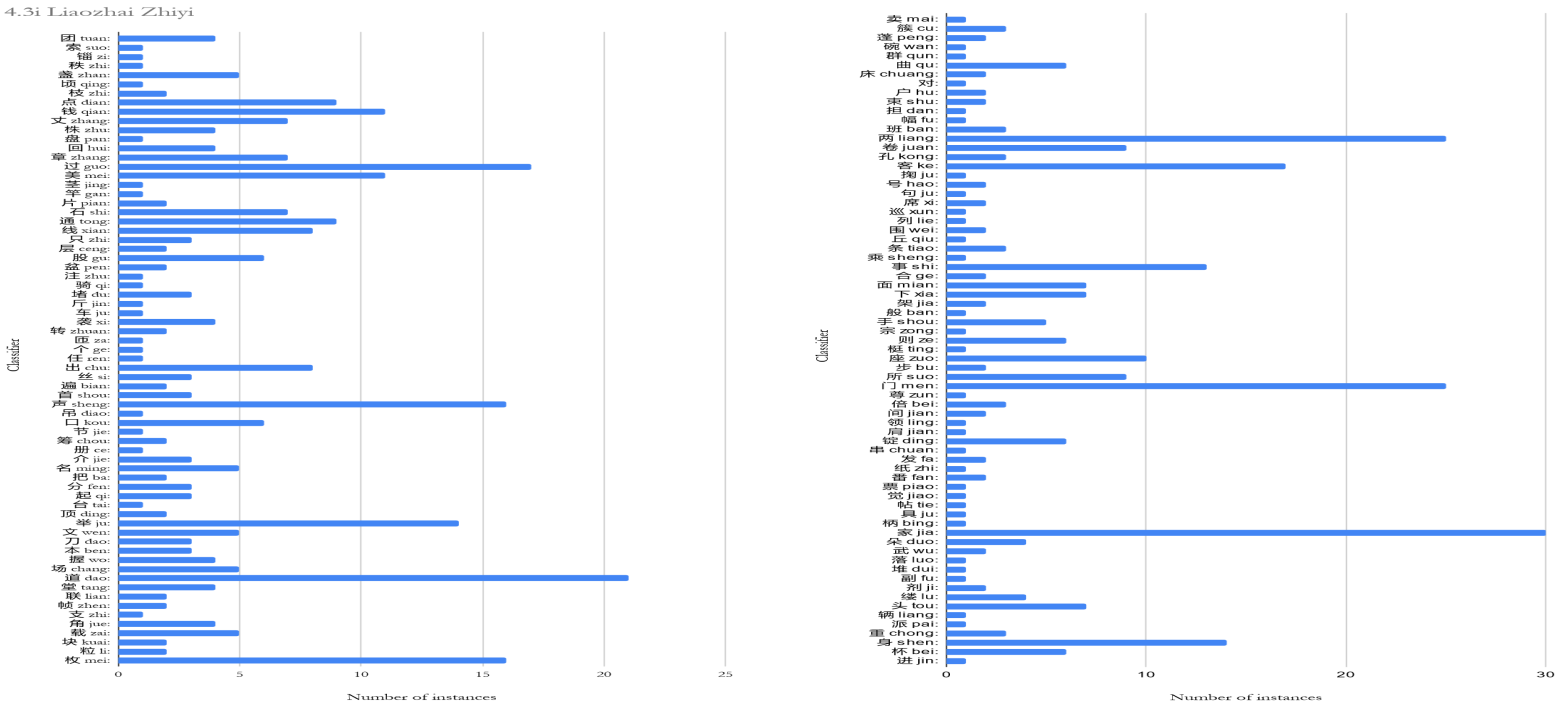
It's difficult to determine what exactly is the environment in which a CL is generated post-nominally. Indeed, there seems to be very little reason behind which CLs are post-nominal and which are pre-nominal with respect to their semantic content. Perhaps it is the case that CLs classifying military equipment in *Romance* and units of measurements in *Dream* are post-nominal by default, but a definitive conclusion would require a much more extensive textual analysis. There is another aspect of CLs found in Vernacular Sinitic works that further brings into question the degree to which there is any positional requirement for CLs at all, the classifier 枚

mei . As mentioned before, 枚 *mei* still exists in literature after its role as a general classifier had largely been replaced by *ge*, though it is often unclear whether its usage is meant to be as a classifier for small things or a general classifier. Regardless of what exactly the semantics of its classifications are, however, what is quite noticeable about 枚 *mei* is that it often occurs post-nominally even when few other CLs do. Still, when they do occur post-nominally they often are seemingly not general classifiers but rather express the same classification of small round object as in the following:

(26) 痛者内有针十个，痒者内有黑白棋子二枚 (*Romance of the Three Kingdoms*, line 1717)
痛者 内 有 针 十个, 痒者 内有 黑 白
pain-NOM inside have needle ten-CL, itch-NOM inside have black white
棋子 二枚
chess piece two-CL
'The painful ones inside have ten needles, the itchy ones inside have two black and white chess pieces' (Adapted from Brewitt-Taylor, 1925)

In this case 枚 *mei* is once again classifying a small, round object, a chess piece. Whether this usage is the same as in *Water Margin* where they appear pre-nominally is entirely unclear. If it is the case that (21) is a non-general usage of 枚 *mei* , then Vernacular Sinitic writing has no explicit requirement for the position of CL, as even the later works surveyed contain post-nominal CLs. If there is anything that can be said about the post-nominal CLs that do appear is that 枚 *mei,* in just the same structure as (23), (24), and (25), makes up a sizeable portion of them. Given that 枚 *mei* is repeatedly used pre-nominally in *Water Margin* and in a few other cases, 枚 *mei* itself has no strict positional requirement but rather simply tends to be post-nominal. 枚 *mei* itself can still be said to have a strong affinity for the post-nominal position, as outside of *Water Margin*, pre-nominal instances of 枚 *mei* are actually quite hard to find. This issue in many ways relates to the position of the classifier that replaced 枚 *mei* , the general classifier 个 *ge*. 个 *ge,* in almost every case in the data, occurs pre-nominally just as most instances of 枚 *mei* occur post-nominally. When 个 *ge* does appear post-nominally, however, it

is arguably not being used as a general classifier. Indeed, prior to its replacement of 枚 *mei* as the general classifier, *ge* was itself a word for bamboo stems and was possibly used to count arrows (Peyraube 1991, p. 113). Given its origins, it's possible that the meaning of 个 *ge* was extended through analogy, as CLs often are, to classify objects that are long, thin, and perhaps pointed, which clearly relates to its usage with needles in (26), which is one of the very few instances of post-nominal 个 *ge*. In this sense, there could actually be a positional requirement at least with respect to general classifiers, as it's entirely possible that pre-nominal instances of 枚 *mei* are specifically classifying small, round objects and post-nominal instances of 个 *ge* classify long, thin objects. That is to say, when not in their normal position (i.e. post-nominal 枚 *mei* and pre-nominal 个 *ge*), general classifiers lose their generality and are instead limited to a specific domain of classification as other non-general classifiers are. This conclusion is of course largely speculation, and without more data of pre-nominal 枚 *mei* and post-nominal 个 *ge*, it is hard to say for sure what their readings are.



Where there is at least some straightforwardness in CL position, however, is in the one Literary Sinitic work left out of the discussion in section **4.3**, *Liaozhai Zhiyi*. In *Liaozhai Zhiyi*, CLs seem to be invariably post-nominal, though in many cases, just as in the other Literary Sinitic works, what appear to be CLs are actually Num+N count phrases. Even though it was published in 1740, *Liaozhai Zhiyi* retains the Literary Sinitic tendency to avoid using CLs at all, opting instead for Num+N count phrases. Still, its usage of 枚 *mei* is undeniable, which means that though *Liaozhai Zhiyi* attempts to maintain a style reminiscent of the classics, it is much more comparable in syntax to *Shishuo Xinyu* and the Vernacular Sinitic texts. As already discussed, where Vernacular Sinitic works seem to have a strong affinity for pre-nominal CLs, though post-nominals are acceptable as well, Literary Sinitic works seem to have a requirement of post-nominal CLs. Even though early works might allow for pre-nominal MWs, as mentioned in section **4.3** with respect to *Shishuo Xinyu*, in *Liaozhi Zhiyi*, MWs seem to follow a post-nominal requirement. In this sense, it seems that post-nominal CLs were seen as being a feature of 'the Classics,' even if in reality some cases allowed for pre-nominal CLs, and as a result instances of pre-nominal CLs were largely removed from the stylistic features of Literary Sinitic. Vernacular Sinitic, on the other hand, seems to have never developed such a requirement, as even by the time of the writing of *Liaozhai Zhiyi*, vernacular works neither followed a post-nominal requirement nor had their own pre-nominal requirement. Rather, there was simply a tendency towards pre-nominal CLs, which could very well have been an effect of Vernacular Sinitic works following spoken syntax, whereas Literary Sinitic works, being the product of conservative literary strictures, follow artificial restrictions on syntax. What remains unclear, however, is just how Literary and Vernacular Sinitic works allow for seemingly free choice between using classifiers and using Num+N or N+Num count phrases.

Though it has not yet been directly discussed, in later early Literary Sinitic works and in some contexts in Vernacular Sinitic works, Num+N and N+Num count phrases, particularly those that count people seem to freely co-occur, which seems impossible as CLs supposedly replaced direct combination Num+N or N+Num phrases as the sole acceptable syntax for counting. For example, consider the following where Num directly combines with a noun to denote people. It's worth noting though that in the case of (28), Num clearly acts as a predicate and doesn't come after a noun but rather comes after a nominalized clause. (29) similarly follows a nominalized clause, but Num also precedes a noun, so it is still worth considering here. Given its lack of similarities structurally to other examples of N+Num, whether (28) is the same structure as (27) and (29) is unclear.

(27) 岂以五男易一女 (*Shishuo Xinyu*, line 158)
岂 以 五 男 易 一 女
How take five male exchange one female
Would I exchange five sons for one daughter? (Mather et al., 2002)

(28) 四方百姓, 裹黄巾从张角反者四五十万 (*Romance of the Three Kingdoms*, line 25)
四 方 百姓, 裹 黄 巾 从 张角 反者
four side common people, wrap yellow turban follow Zhang Jue oppose-NOM
四 五 十万
four five hundred thousand
'The common people from all directions, wrapped in yellow turbans following Zhang Jue in rebellion were four or five hundred thousand'

(29) 所愿去者止四五人 (*Dream of Red Chambers*, line 1894)
所愿 去者 止 四 五 人
NOM-want go-NOM only four five person
'Those who wanted to go were only four or five people'

These questions could themselves be the subjects of large-scale studies, but thankfully the work of Feng (2012) provides solid ground to develop a basic understanding of the possible reasons for what seem to be irreconcilable syntactic developments. Indeed, Feng provides a much clearer example of the free choice issue as in the following where there is seemingly no reason for one to phrase to have a classifier while the other does not:

(30) 十八铁丸 (*Fayuan Zhulin*, ch. 7, line 28)
shiba tie wan
十 八 铁 丸
ten eight iron ball
'eighteen iron balls'

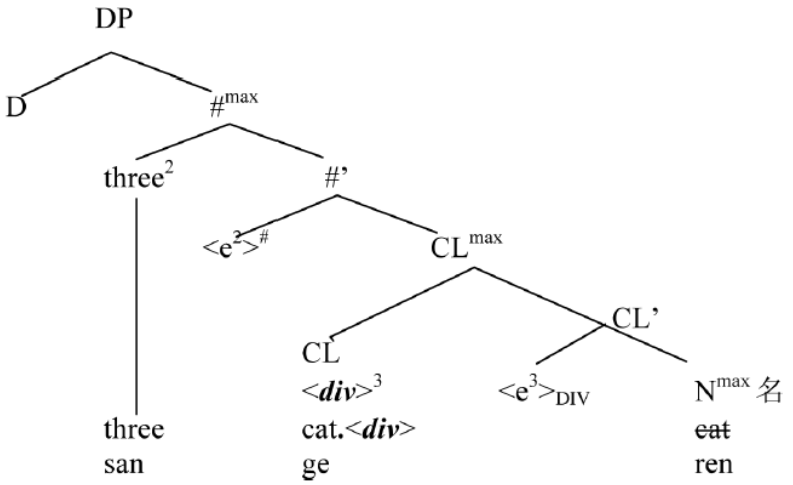
(31) 七枚热铁丸 (*Fayuan Zhulin*, ch. 57, line 22)
qi mei re tie wan
七枚 热 铁 丸
seven-CL hot iron ball

From (30) and (31) it is clear that there is some degree of free choice with respect to the inclusion or exclusion of classifiers when counting. Feng (2012) describes the free choice as being a result of prosodic factors, the argumentation behind which will be explored in the next section.

4.5 *Accounting for contradictions - Feng (2012)*

At this point, there are two major unanswered questions: Why did classifiers develop at all when N+Num and Num+N count phrases already existed to enumerate and what accounts for what appears to be free choice among N+Num and Num+N and pre-nominal and post-nominal CLs? Feng (2012) provides some answers to these questions. With respect to the first question, one aspect of Sinitic count phrases without classifiers that has been left unsaid is that they should not be able to exist syntactically. In reference to Hagit Borer's theory on the matter (Borer, 2005), Feng describes the syntactic structure of CLs as containing a feature <*div*> or 'divider.' and includes the following structure, in which the classifier phrase itself is embedded within a numeral phrase that is sister to D. Within the classifier phrase, the head of CL provides a space for <*div*>, which is instantiated by the CL itself.

(32)

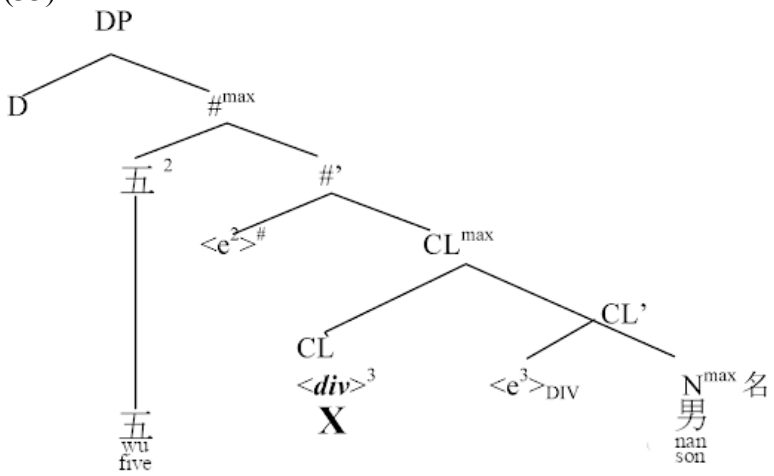


This structure is quite similar to (15a) in **4.3**, which itself was adapted from Cheng & Sybesma (2005), but is more suited for this discussion in that it contains feature nodes that assign a structural place to <*div*>. As can be seen from the structure the CL head itself carries the <*div*> function, which Feng describes as working in a similar manner to inflection in verb phrases, 'the div of noun phrases and the INFL of verb phrases may have a similar grammatical function in locating entities (for nouns) or actions (for verbs) in concrete spatiotemporal reality, respectively' (Feng 2012, p. 70). In this sense, the function of <*div*> is to divide N into separate, referable atomic units so that they can then be assigned number. The point of bringing up <*div*> is to introduce a caveat that Feng also mentions, <*div*> can be realized either through a classifier or a plural marker but never both or neither. That is, the two realizations of <*div*> are in complementary distribution and one or the other realization syntactically required, which explains why classifier languages do not have clear parallels to plural markers in non-classifier languages (consider for example the lack of an equivalent to the English plural marker -*s* in Sinitic languages). This rule, however, no doubt seems strange in the face of much of the data that appears in this very study, such as (14), (27), (28), and (29) in which neither a CL or a plural marker appear. If we were to apply the structure in (32) to any of these sentences the syntax

would lack a lexical item upon which to realize <*div*>. The second count phrase of (27), repeated below, '五男' has been applied to the structure in (32) to further illustrate this point:

(27) 豈以五男易一女 (*Shishuo Xinyu*)
豈 以 五 男 易 一 女
How take five male exchange one female
Would I exchange five sons for one daughter? (Mather et al., 2002, p. 45)

(33)



As the structure in (33) depicts, without an intervening classifier between 一 'one' and 女 'daughter,' <*div*> cannot be realized. Given that there are no other elements in the phrase that could possibly act as a CL or a plural marker, the above sentence, and indeed all similar sentences and the grammar that produced them writ large, should not exist. Obviously, such sentences do exist and in quite a great number, as seen in the Literary Sinitic data. How then did such sentences come to be?

When discussing the syntax of early Sinitic languages as evidenced by written records, one cannot take for granted that the morphology of these languages is the same as modern languages, particularly with respect to how morphology is orthographically realized. As analytic languages, modern Sinitic languages such as Mandarin lend themselves well to character based writing systems, as morphemes can easily be represented by a single character per syllable. The same cannot be said about characters in a non-analytic context, however. Were morphemes not

necessarily full syllables themselves, then a character-based system like that of Sinitic languages would struggle to express more complex synthetic systems of morphology, as the characters cannot be broken down into component parts to represent phonetic subsegments that might be added to a word through affixation. Such is all to say that were a synthetic language represented with characters as they are used today, any morphological processes that do not produce full syllables that can be assigned to a character would be essentially invisible, and without knowledge of the language's morphology, one would be entirely unaware of any subsegmental morphemes. Such is likely the case with early written Sinitic, and therein lies a potential solution to the question of how Num+N and N+Num count phrases are allowed. Indeed, at least according to the Baxter-Sagart Old Chinese reconstruction (Baxter & Sagart 2014), the earliest reconstructed Sinitic languages, Old Chinese, did in fact have a complex and synthetic system of morphology. With respect to the specific morphological process that would allow a count phrases without CLs, Feng quotes earlier work by Baxter on the Old Chinese *k- prefix:

> It is tempting to regard the functions of *k- in verbs and nouns as being fundamentally one: *k- would serve for actions and objects that are well-delimited in time and space, and hence usually concrete and countable. If so, disappearance of *k- between the Old Chinese and Middle Chinese periods deprived Chinese of a means of distinguishing between count and mass nouns. This may have been a factor in the rise of numeral classifiers in Chinese during the same period. (Sagart 1999, p. 107)

This conclusion by Sagart fits in well with Feng's analysis, particularly in his association of <*div*> with both nouns and verbs, just as Sagart suspects *k- in the context of nouns and verbs could be performing the same function. Feng and by extension Sagart's explanation is perhaps the most plausible, or at least the most obvious explanation for Num+N and N+Num count phrases in early written Sinitic sources. The affixation of *k- to N in these phrases provided a morpheme upon which <*div*> could be realized, just as -s does in the case of English plurals. With the disappearance of *-k between Old and Middle Chinese, a new system was needed to instantiate <*div*>, that of course being classifiers. Much of the development of classifiers has

already been covered, but it is worth noting some instances in which Feng's own conclusions echo those of this study. Firstly, Feng notes that general classifiers, i.e. 枚 *mei* and 个 *ge*, developed first among count-classifiers, which is in line with the earliest occurrence of a count-CL in the data surveyed. (19) and (20) in section **4.3**, the earliest count-CL phrases found in the data, are both instances of the generic count-CL 枚 *mei*. It's unclear whether 个 *ge* developed at the time and was simply used less often or if it developed later, but at the very least it can be said that the earliest count-CLs use a generic classifier, 枚 *mei* . The development of generic classifiers first according to Feng could be on account of a growing syntactic rather than semantic requirement for a classifier, which would prompt a need for general classifiers to appear first so that there would be something to fill the CL requirement. This conclusion fits in nicely with both Sagart's theory of *k- and Borer's <*div*> requirement (i.e. as *k- disappeared, something was needed to support <*div*> and so generic CLs were developed with specific ones developing later on account of semantic requirements). Feng also notes how the majority of CLs with 枚 *mei* occur post-nominally, while the majority of CLs with 个 *ge* occur pre-nominally, which is the same conclusion that was reached at the end of section **4.4**.

On the matter of pre-nominal versus post-nominal CLs, Feng also offers a solution. In previous works (e.g. Feng 1995), Feng discusses the possibility and implications of proto-Chinese having a SOV word order rather than an SVO word order, which Old Chinese and later Sinitic languages would share. Were proto-Chinese, or 'pre-Archaic Chinese' in Feng's words, an SOV language it would be head-final, and, predictably, the CL as the head of the classifier phrase would instead occur at the end of the phrase. After the word order shift, post-nominal classifier phrases were reanalyzed as a predicate rather than a nominal structure and were allowed to remain in the grammar. This word order shift and subsequent reanalysis accounts for the co-occurence of pre-nominal and post-nominal CLs and why two separate

generic classifiers developed. That is, 枚 *mei* developed in the post-nominal structure and 个 *ge* developed in the pre-nominal structure, which might also explain the two classifiers' near-complementary distribution, though the mechanics of how 枚 *mei* maintained its affinity the for post-nominal structure are unclear, and the occurrence of pre-nominal 枚 *mei* further complicates the issue. Regardless, if one were to accept the idea of a word order shift, then it remains the most plausible explanation for the existence of both pre-nominal and post-nominal CLs. According to Peyraube (1991)'s data, the word order shift first began to occur before the earliest written records, as Num+N phrases make up the majority of count phrases from the very beginning, with N+Num becoming increasingly rare over time. When MWs are first introduced, they are post-nominal, which suggests that other nominal structures aside from Num+N were still influenced by the earlier SOV word order. The word order shift, as it relates to CLs, seems to have begun to come to fruition c. 5th-3rd century BC, the time in which Num+MW+N phrases are first seen. This timeline is roughly in line with the data surveyed, as the early Literary Sinitic sources overwhelmingly use Num+N phrases with later early sources using both pre and post-nominal MWs. Count-CLs seem to first occur at this time as well. As with this study's timeline, Peyraube, also using *Shishuo Xinyu* as a source, finds that pre-nominal MWs and pre-nominal count-CLs both first occur within roughly the same time period, although pre-nominal phrases, especially pre-nominal count-CLs, are extremely rare. With all of these developments in mind it seems to be the case that the earliest records largely lack CLs entirely, and for most of history the structure Num+N is by far the most common count phrase. It is also clear that the SOV to SVO shift began before the introduction of classifiers, as evidenced by the gradual disappearance of N+Num, but when classifiers were first introduced, they were post-nominal until they gradually underwent a shift to the pre-nominal position. While Peyraube's data presents a solid guide to when certain structures first appeared, it still does not

account for why N+Num or Num+N count phrases and CLs, two structures that accomplish seemingly the same task of counting N, are allowed to co-occur seemingly at random. For this issue, Feng proposes a prosodic solution.

The influence of prosody as argued by Feng can essentially be summarized by a matter of prosodic strength. Prosodically strong phrases are found not to require a CL whereas prosodically weak phrases require the presence of a CL. To give a clearer illustration of prosody's effect on classifiers, Feng provides examples for Yi, an SOV Lolo-Burman language that developed a system of post-nominal classifiers, similar to that of Old Chinese. Following the data, Feng notes the three possibilities with respect to prosodic structure and the inclusion or exclusion of a CL:

> (i) In the structure of [N + monosyllabic numerical form], if the NF (numerical form/number) is monosyllabic, then the NF is not acceptable;
> (ii) In the [N + monosyllabic numerical form +CL], if the NF + CL form is a disyllabic unit, then the result is acceptable.
> (iii) If the NF is disyllabic itself, the result is also acceptable (i.e., [N + disyllabic numerical form + __]) even if there is no CL' (Feng 2012, p. 89)

In this way, whether the Num in a Yi classifier phrase is monosyllabic decides if a CL is required or not. In the case of a disyllabic Num, the phrase is prosodically strong with or without a CL, so either option is acceptable. These features of Yi form the basis of Feng's prosodic description of CLs in early Sinitic texts, as the requirements are largely the same. Where the data of this survey does not agree with Feng, however, is in this treatment of monosyllabic Num+N. According to Feng, monosyllabic Num should form a prosodically deficient phrase if combined with N, and it is this deficiency that gave rise to the need for CLs, beginning first with general classifiers. And yet, (27) clearly includes not one but two monosyllabic Num+N phrases, 五男 'five sons' and 一女 'one daughter. Were these examples from an earlier text before the emergence of CLs, perhaps some other prosodic or morphological factor could explain their existence, but given that they exist within the *Shishuo Xinyu*, a work that also includes

pre-nominal MWs and post-nominal count-CLs, the association of prosodically weak phrases with the emergence of CLs and and general classifiers in particular becomes a bit questionable. That is not to say that Feng is wrong. Rather, Feng provides far more evidence towards his proposed association than there are examples such as (27). Some other factor must be at play that allows for prosodically weak phrases to exist after the introduction of general classifiers intended to fix their prosody, but what that factor is may remain a mystery. The matter of the <*div*> function still remains, however.

Given the evidence and argumentation of Feng (2012), it seems likely that <*div*> is itself prosodically triggered, as suggested at the end of Feng (2012, p.96). The wider implication of this conclusion, however, is that <*div*> is prosodically triggered differently depending on the prosodic structure of the language itself, or more specifically, <*div*> seems to hold the requirement that it is realized as a part of a prosodically strong phrase. In Old Chinese, phrases such as Num+N or N+Num, i.e. those that include the *k- affix as described by Sagart, are prosodically strong enough to stand on their own with <*div*> residing on the *k- affix. Through phonological change, however, *k- is lost and <*div*> no longer has a place to be realized. Without *k-, phrases that once included it are now prosodically weak, and so general classifiers begin to develop to bridge the prosodic gap. At this point, phrases with either disyllabic numerals or a CL are considered prosodically strong. Where exactly <*div*> is realized, however, is unclear. It is possible that in both cases with a CL and without a CL, <*div*> is realized on N but those phrases with a CL could have already shifted <*div*> to CL, as is the case in modern Sinitic languages. If one were to maintain that all instantiations of <*div*> are in complementary distribution, however, it would make more sense to assume <*div*> was still realized on N, provided it is in a prosodically strong phrase. Later, however, it seems the presence of a CL became increasingly syntactically required, regardless of prosody, and <*div*> came to be realized

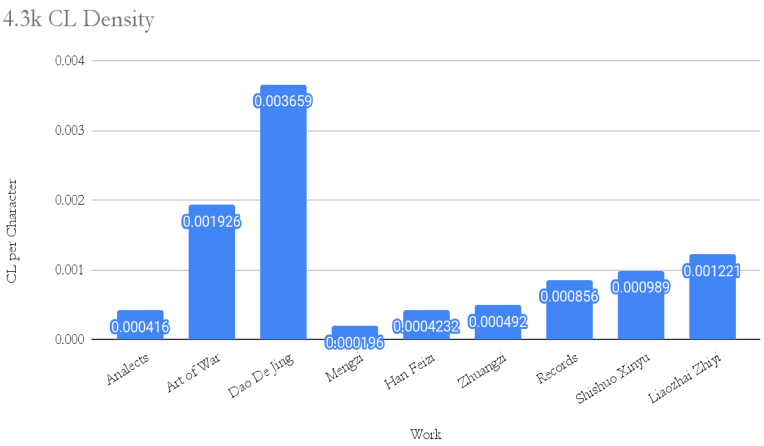
by CL. How exactly <*div*> managed to move to CL or if this timeline of development is at all accurate is unfortunately quite difficult to prove. Moreover, becoming mired in the exact developmental process from lacking CLs to having a strict CL requirement distracts from one aspect of the written record that cannot be explained by Feng (2012), that being that Num+N and N+Num count phrases continue to exist in the literature well beyond the early days of CL development. In particular, *Liaozhai Zhiyi* and other Literary Sinitic works continued the use of count phrases without classifiers and the intermingling of them with CLs long after CLs had become a staple of Sinitic syntax and count phrases were allowed only in very rare cases. In this respect, this study's attempt to fully bridge the gap between the earliest written records and modern language with respect to the development of classifiers leads to a matter more philological than linguistic. That is, what exactly is the relationship between linguistic and literary development and to what extent can written records be relied upon for linguistic analysis?

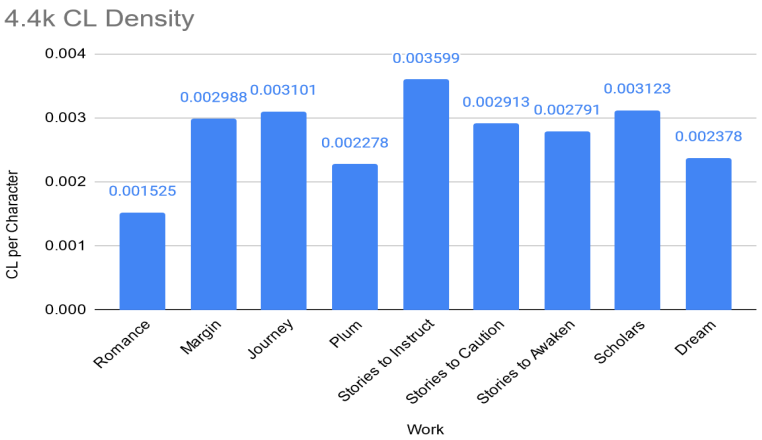
4.6 *The trustworthiness of sources and the aesthetic-syntactic interface*

Written sources, and the authorities who approve of them are often the source of prescriptivism that can interrupt or delay the natural development of language. Prescriptivism in language is often seen as a modern phenomenon (Curzan, 2014), and from a sociological perspective, it is. That is, prescriptivism as an institution and arm of authority bent on 'fixing' how language is spoken is very much a modern concept. While the effects of prescriptivism are important to study, however, so too is prescriptivism itself. What syntactic, semantic, and phonological structures are pushed by the forces of prescription and why are valuable topics for linguists to consider, but only asking these questions within the context of modern efforts to police the nature of spoken language would be a mistake. Written language has long since been

used as the source of prescriptive standards for language that were then pressed onto spoken language, but written language does not exist in a vacuum nor did it appear out of thin air. Rather, written language is a linguistic system unto itself, and one that also polices its own rules. Literature is often the product of many layers of conventions, and just as prescriptivism in spoken language attempts to halt the development of language, so too do literary conventions attempt to keep literature within certain stylistic boundaries. The issue becomes even more egregious when literature abandons, or indeed never embarks on, any effort to reflect the spoken language. Therein lies this discussion's relevance to this study.

In the context of pre-modern written Sinitic sources, one must ask themselves can the sources be trusted or is the true developmental path of a language shrouded by the need for written sources to adhere to standards. This matter could not be more relevant to the questions posed in this study. As the data has depicted, the use of CLs are very much an aspect of a vernacular style that is much more in imitation of the spoken word than Literary Sinitic styles. Consider the two of the latest works surveyed, *Liaozhai Zhiyi* (1740 AD) and *The Unofficial History of the Scholars* (1750 AD). Published just ten years apart, the two works could not be more different linguistically. As per the CL density charts in 4.3k and 4.4k, *Scholars* has almost triple the number of CLs per character than *Liaozhai Zhiyi*, which itself has only a slightly greater CL density than *Shishuo Xinyu*, a work written 1300 years before it:





Such a difference goes well beyond a matter of stylistic differences and into the realm of representing two radically different languages, only one of which represents actual speech to any meaningful degree. It is simply a fact that classifiers have become a major facet of modern Sinitic languages and N+Num and Num+N count phrases have long since become uncommon and far from their original role as the primary structure for counting, so for a work to so overwhelmingly use count phrases without classifiers it cannot be said to be written in imitation of speech but is rather in imitation of older styles that were themselves a product of an earlier language. From this example it is clear that the aesthetic concerns of a writer influence what is written well beyond diction, going as far as to shape the writing's very syntax. The influence of style on syntax is of particular concern, as it can artificially extend the lifetime of syntactic features well beyond their disappearance in speech, as is seen with the proliferation of count phrases without classifiers rather than CLs in *Liaozhai Zhiyi*. The influence of aesthetic literary concerns on the syntax of written sources is such a major feature of the linguistic history of Sinitic languages that it deserves its own name, the 'aesthetic-syntactic interface.' Such an interface exists along the borders of philology, literary history, and linguistics, and in that sense is deserving of attention, as language, both spoken and written, do not exist in a vacuum. Just as prescriptivist pressures have affected the course of the development of spoken language, so too

have aesthetic concerns influenced the fundamental structure of written language. Indeed, the aesthetic-syntactic interface, though a concept that warrants much further research, could be the answer to some of the questions not yet satisfyingly resolved by this study. For example, the co-occurence of non-CL count phrases with CLs could be an aesthetic as well as prosodic issue. That is to say, count phrases without classifiers continued to exist after the development of general classifiers not simply due to prosody but on account of aesthetic concerns. That is, for whatever reason count phrases that do not include classifiers were simply more pleasing to writers, and so they were used over CLs. Of course, prosody still has its place, as a major aesthetic concern is how a sentence sounds being read aloud. This reframing of Feng's prosodic analysis also accounts for the continued use of non-CL count phrases after CLs had fully developed. An abundance of non-CL count phrases and CLs, when they do occur, being post-nominal seem to be characteristics of Literary Sinitic. While it is possible that early works with such structures were mimicking speech, works written at the same time as Vernacular Sinitic works, which very much favor pre-nominal CLs and are themselves certainly meant to mimic a spoken language, cannot be said to be at all representative of speech. These works must be favoring aesthetic concerns over the realities of the authors' spoken language.

It is no coincidence that a great rift between Literary Sinitic, or 'Classical Chinese,' and Vernacular Sinitic has long been recognized (Evans, 2021), because for many centuries works in Literary Sinitic had refused to develop along with the language of its authors but rather remained obstinate living fossils of a language long dead. Indeed, by the late 19th century, Literary Sinitic was seen by many as a rotting corpse of a language, one completely detached not only from speech but from the reality and experiences of the average person at the time. Though attempts to reform language in the Late Qing dynasty failed, by the early 20th century, after the fall of the Qing and the establishment of the Republic of China in 1912, a new generation of literary critics

took up the same banner in support of replacing Literary Sinitic with Vernacular Sinitic as the common written form as a part of the New Culture Movement (1910s-1920s). Among them were Hu Shi, who in his essay 'Some Modest Proposals for the Reform of Literature' noted how though Vernacular Sinitic writing flourished for a time (i.e. with the vernacular novels and stories this study covers) it never achieved the stylistic dominance enjoyed by Literary Sinitic:

> 'At that time (Yuan dynasty), Chinese literature came closest to a union of spoken and written languages, and the vernacular itself had nearly become a literary language…Unexpectedly, this tendency was suddenly arrested during the Ming. The government had already been using the ‘eight-legged essay’ to select its civil servants, and scholars like Li Mengyang (1472-1529) and the followers of the ‘former seven masters’ raised ‘archaism’ as the most lofty of literary goals. So the once-in-a-millenium opportunity to effect the unity of the spoken and written languages died a premature death, midway in the process' (Denton, 1996).

In this sense, the nature of Vernacular Sinitic works truly is that of attempts to bring about a 'union of spoken and written languages' amidst a desert of Literary Sinitic formalisms. Imagine, then, if such works did not exist. Had Vernacular Sinitic works not enjoyed moments of popularity throughout history, enough to prevent them from fading away into obscurity, the current conception of the historical development of Sinitic languages would no doubt be radically different. Such is all to say, further research into how extra-linguistic pressures have historically influenced the visibility of linguistic development is a worthwhile endeavor. At the very least, doing so would allow for more even ground that is clear of artificially preserved syntax upon which to perform historical linguistic analysis. Until more work is done on the subject, however, it is ever important that one not take the idea that writing represents speech of any kind for granted. In the context of historical work on Sinitic sources, when working with texts before Vernacular Sinitic was first conceptualized, and indeed after, one must ask what actual language, if any, does the writing represent. As Literary Sinitic can be considered a language unto itself, one that exists solely in writing, linguists must be wary when using such

writing as data lest they end up confusing actual speech with an artificial and formalistic language spoken not even by the dead.

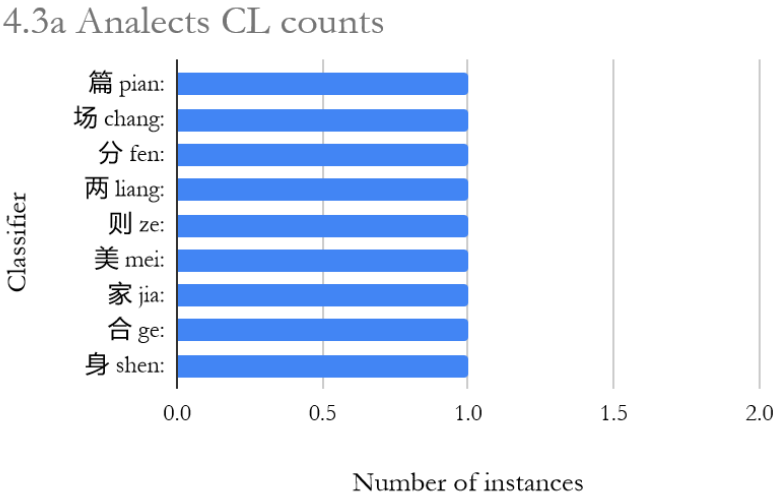
5. Conclusion

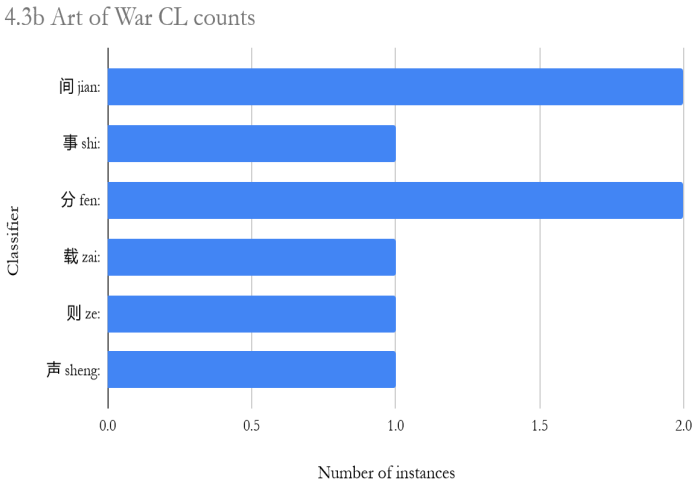
The purpose of this study was to provide a description of the emergence and development of classifiers in Sinitic languages, beginning with the earliest Old Chinese records, and to answer the question of when, why, and how they emerged. The first sections were dedicated to describing the structure and distribution of CLs in modern Sinitic languages, specifically Mandarin. A distinction was also made among the data sources surveyed between Literary and Vernacular Sinitic, as the two styles of writing are fundamentally different with respect to their syntax and distribution of certain lexical items, namely classifiers. The answer to these questions seems to be much in line with the pre-existing literature, particularly Peyraube (1991)'s account. From the earliest records, the oracle bone inscriptions, onward, Num+N and N+Num count phrases were used almost exclusively, though the latter would rapidly dwindle in its usage not long after the beginning of the written record. By the 6th to 3rd centuries BC, however, measure words developed and the first classifier phrases arguably began to appear. These instances may have still acted more closely to nouns, but by the 5th century AD both measure word and count-classifiers appear in a manner identical to their modern usage in works such as *Shishuo Xinyu*. By then, a fundamental shift also became undeniably apparent with respect to word order, as CLs of both types increasingly occurred pre-nominally until appearing almost exclusively so. Post-nominal CLs remained the most common position for CLs in Literary Sinitic works, while Vernacular Sinitc works only retained them in very rare cases, though why they occurred at all remained unclear (though in hindsight this matter too could be the result of both prosodic and aesthetic pressures). With reference to Feng (2012), the theory of prosodic motivation for the

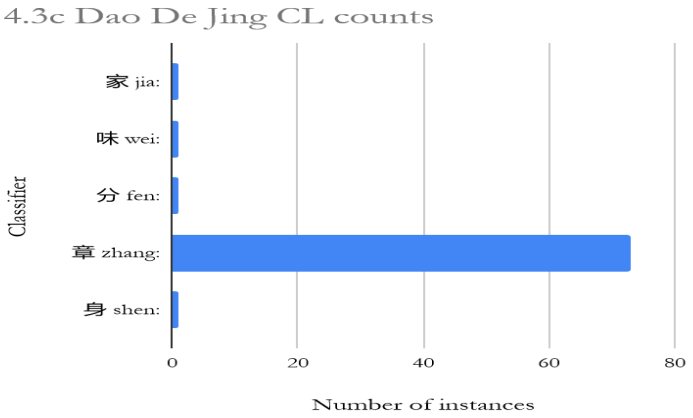
development of CLs was tested and found sufficient in a number of ways, but lacking in one particular way, that being the further development of CLs and the motivation for the disappearance of non-CL count phrases from the spoken language and their continued appearance in Literary Sinitic works going unaccounted for. Finally, a possible extra-linguistic solution was proposed, the aesthetic-syntactic interface, which amounts to the influence of literary conventions and aesthetic concerns on syntax, particularly with respect to imitating outmoded syntax that is no longer grammatical in the author's spoken language in pursuit of a certain literary aesthetic. With the possible effects of the aesthetic-syntactic interface in mind, it is entirely possible that syntactic structures such as non-CL count phrases might have died out and been replaced by CLs in speech long before such a change is observed in literature. In this sense, written sources must be scrutinized further with such extra-linguistic factors taken into account before they can be truly relied upon as evidence towards a certain developmental feature of spoken language.

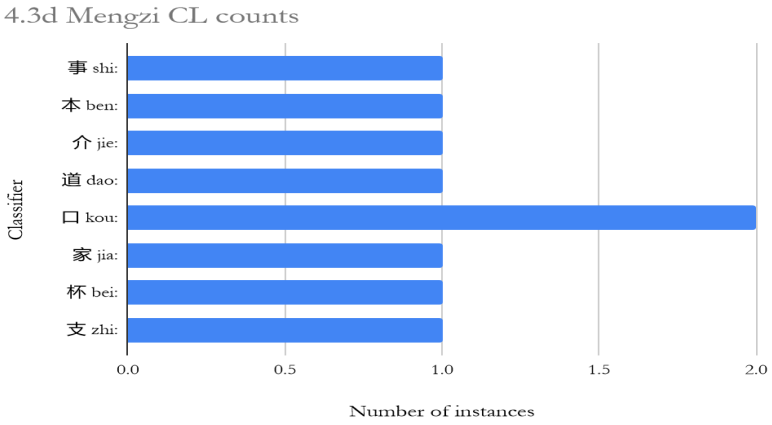
Ultimately, the results of this study were rather unsatisfying and mostly reinforced previous studies on the topic of CLs. Still, the introduction of the aesthetic-syntactic interface and the emphasis placed on considering the more philological factors that give rise to certain aspects of written language bear further investigation. Perhaps the emphasis on the uncertainty and arguable unknowability of the degree to which written language, particularly in the context of Literary Sinitic, represents spoken language is needlessly cautious, but ultimately an uncertain or inconclusive conclusion is preferable to one misguided by a failure to consider extra-linguistic factors in the development of language.

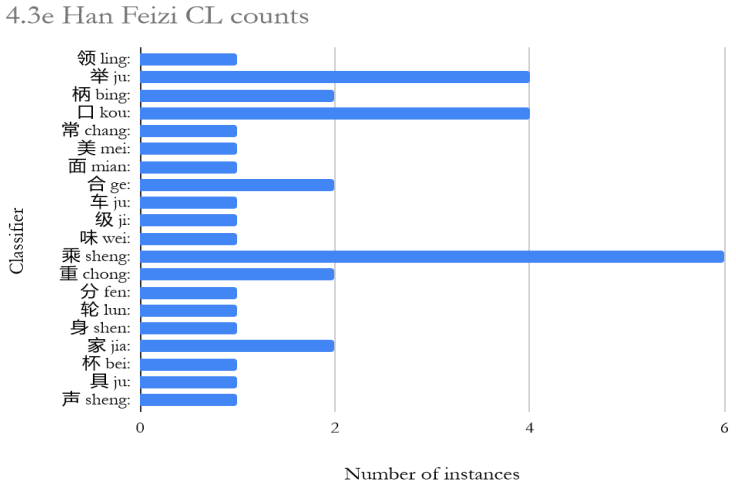
Appendix I: Literary Sinitic Data

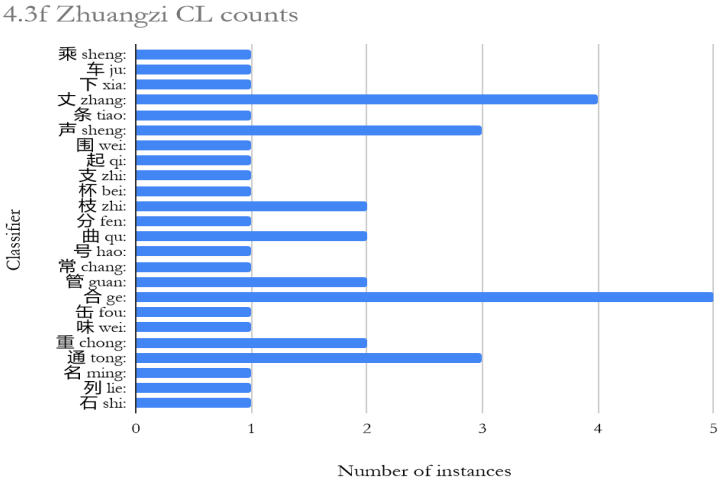




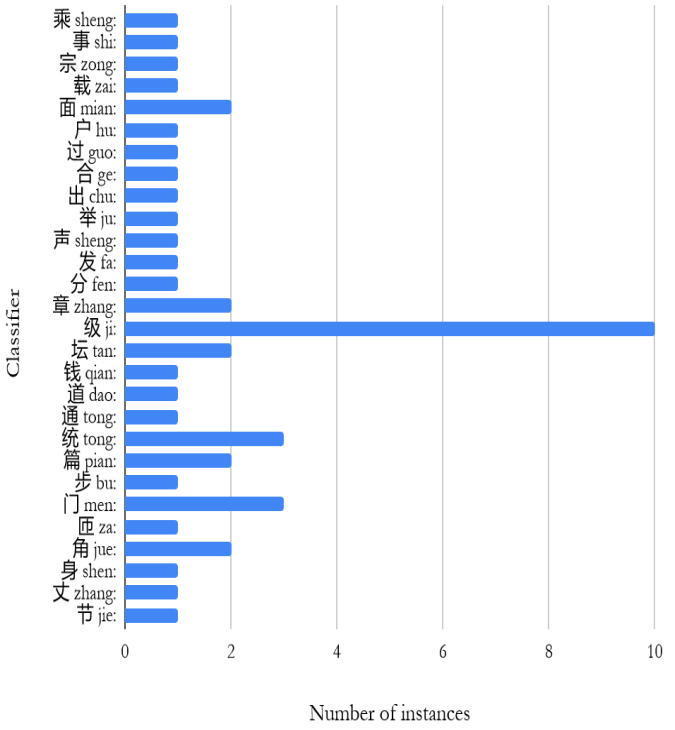




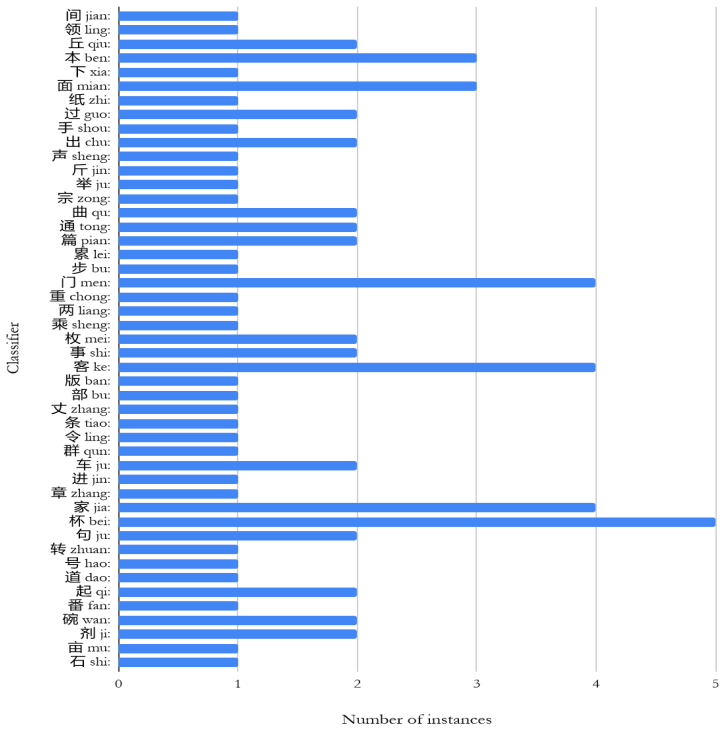




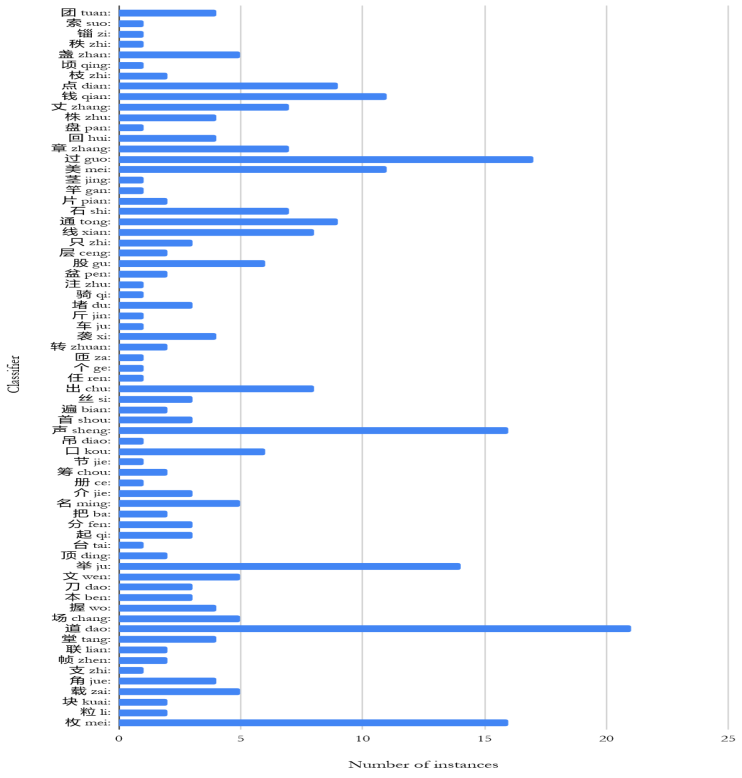
4.3g Records of the Grand Historian

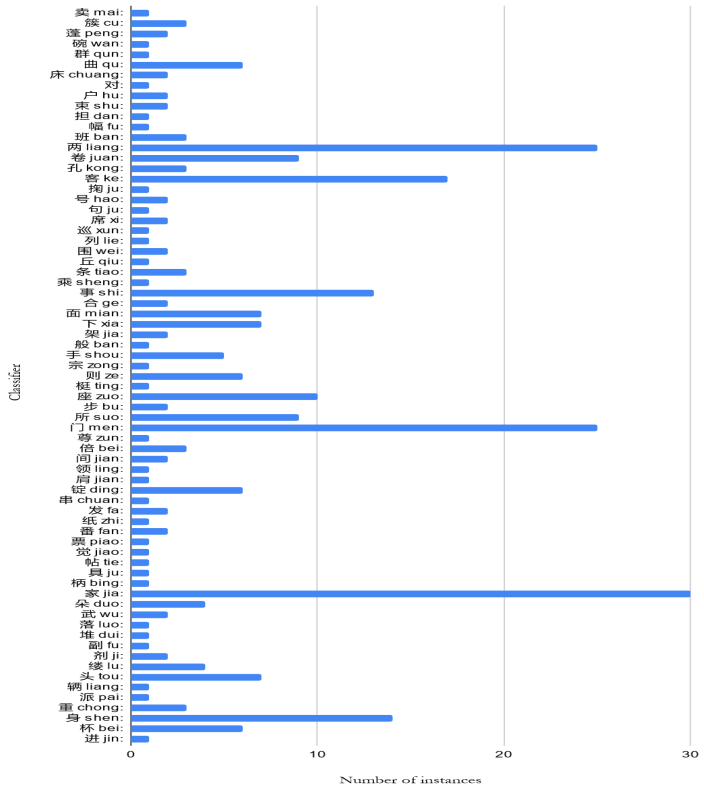


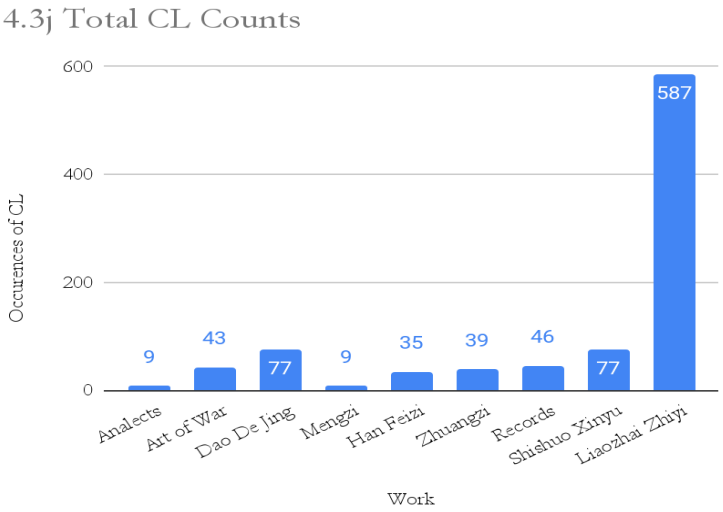
4.3h Shishuo Xinyu

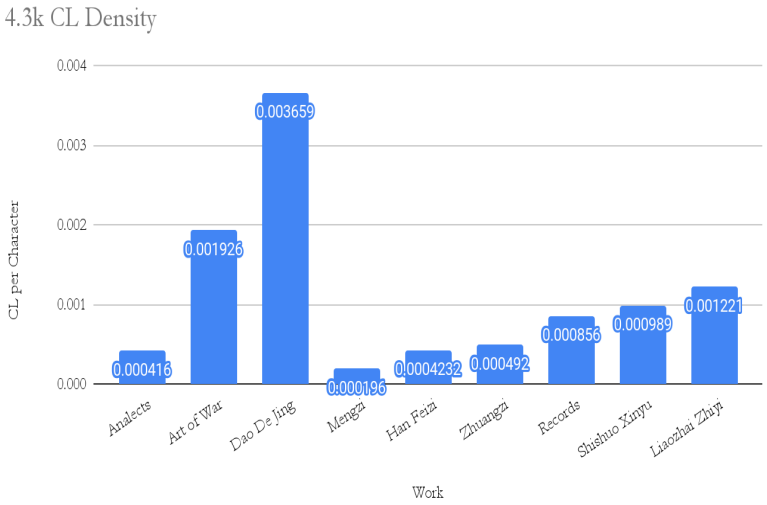


4.3i Liaozhai Zhiyi

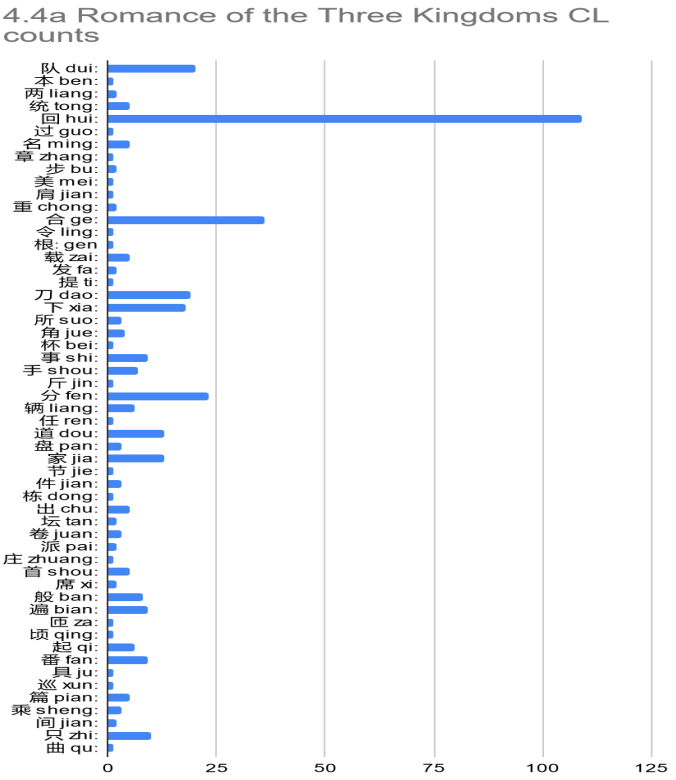


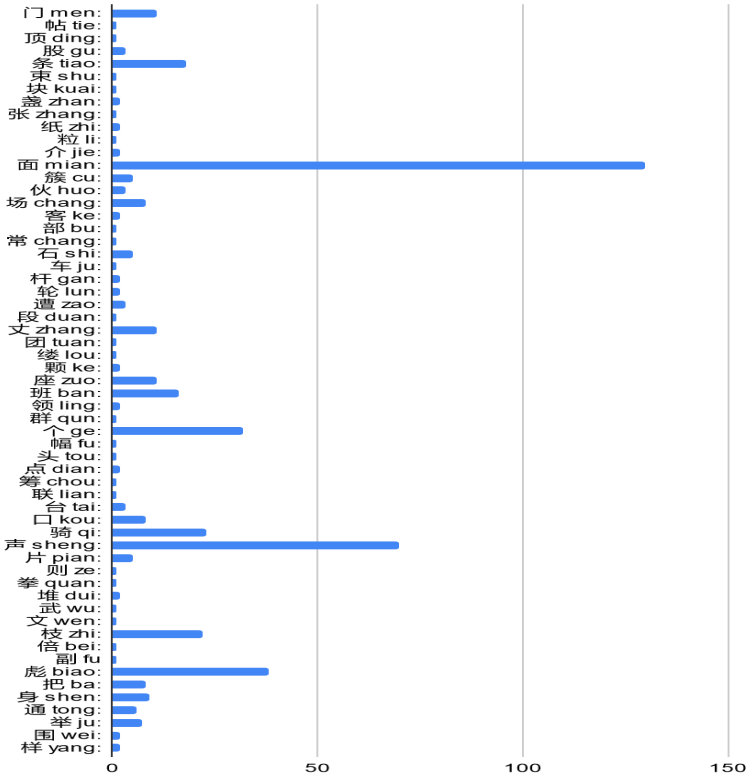




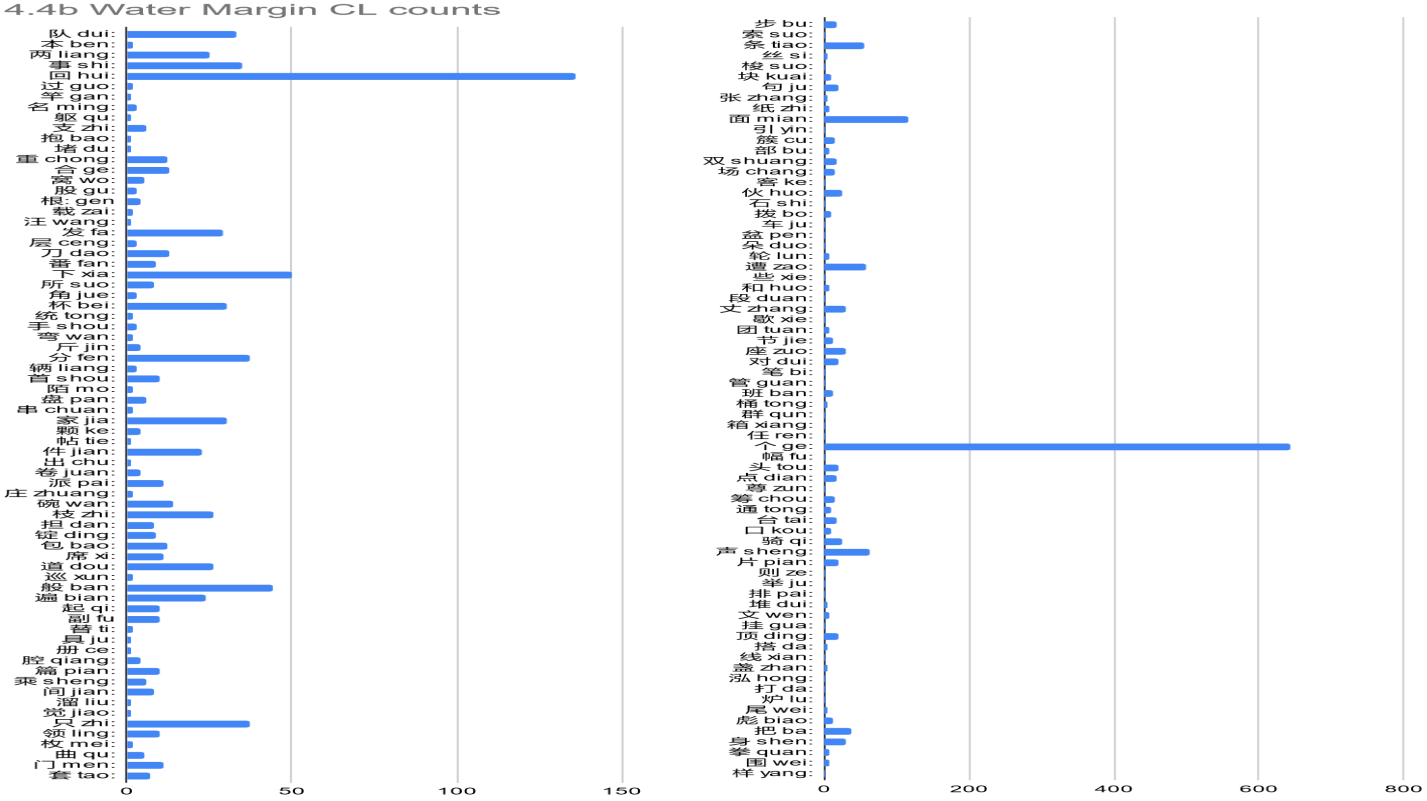


Appendix II: Vernacular Sinitic Data

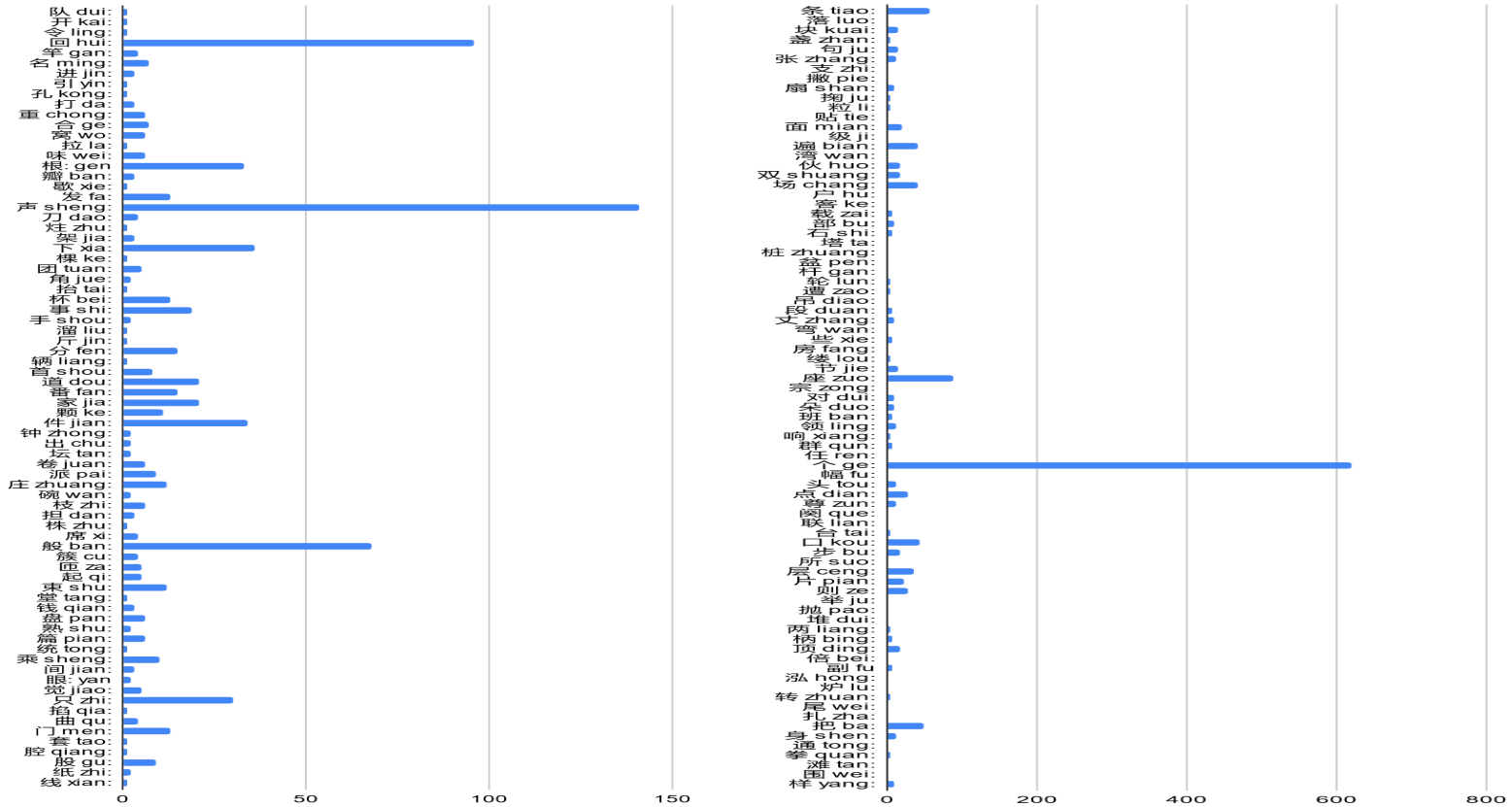




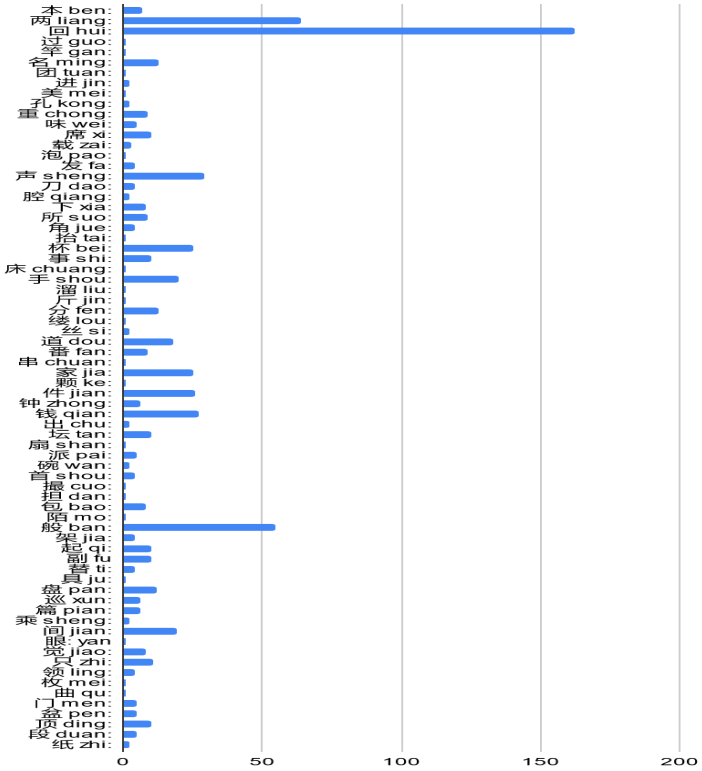
4.4b Water Margin CL counts

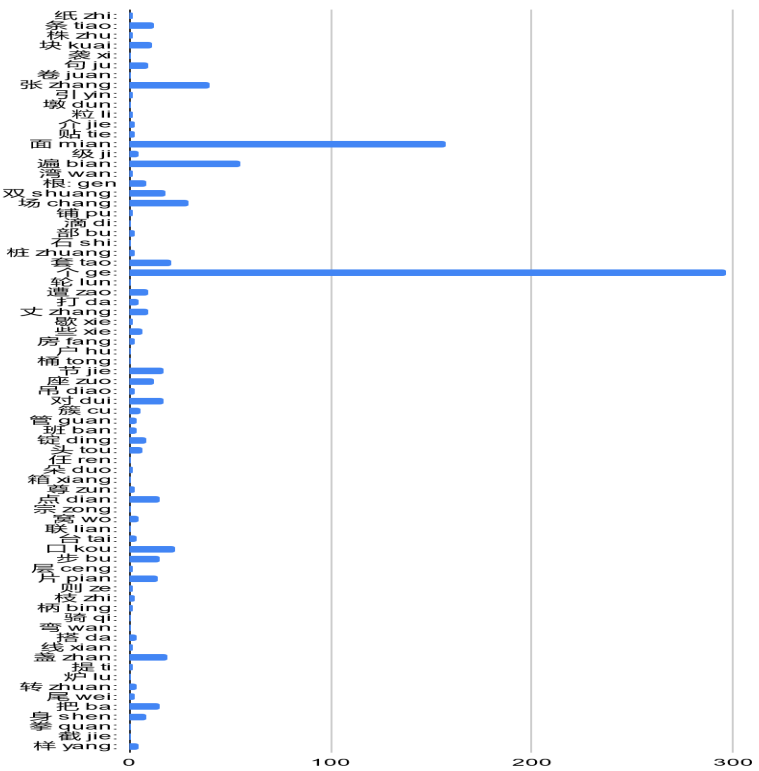


4.4c Journey to the West

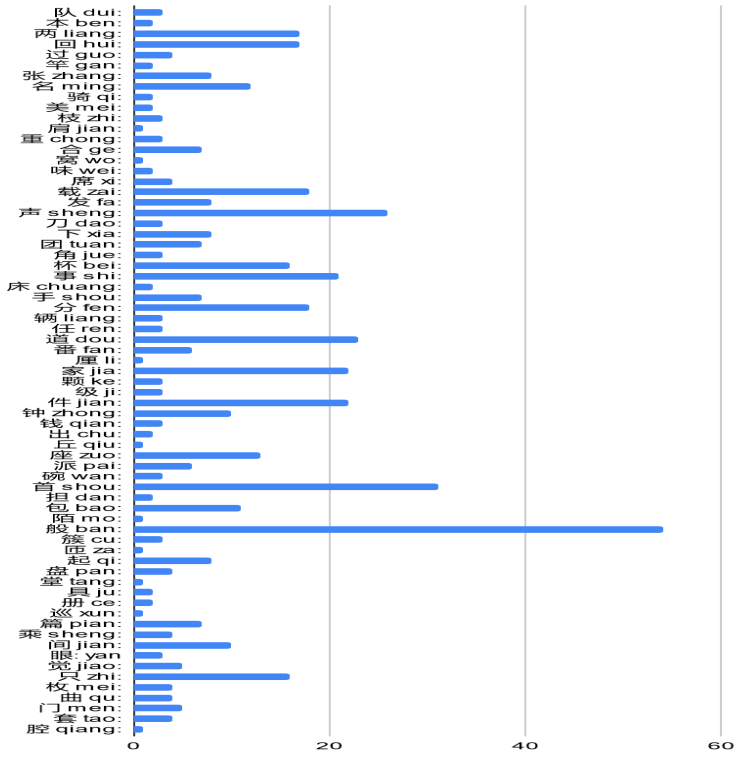


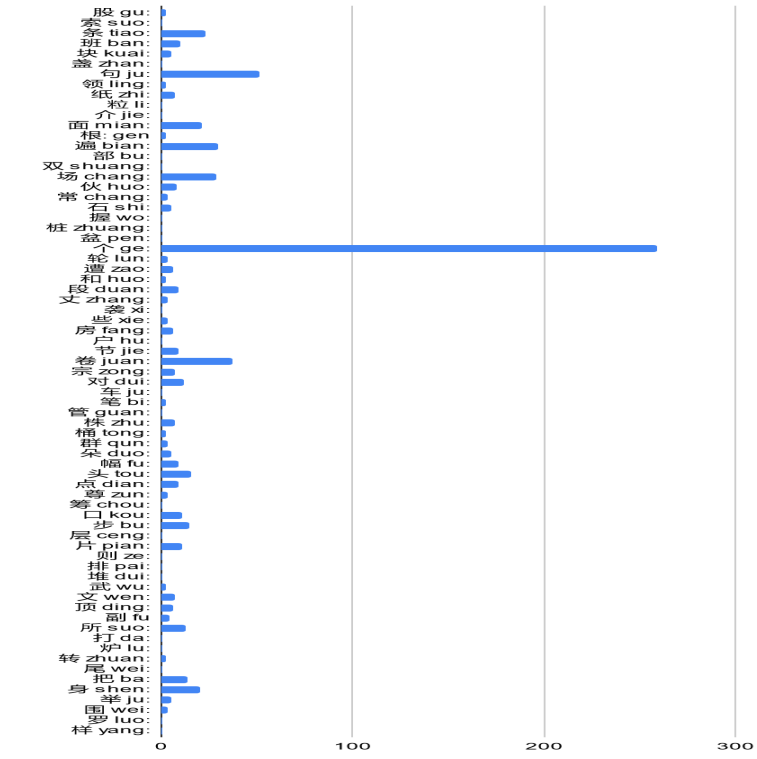
4.4d The Plum in the Golden Vase CL counts



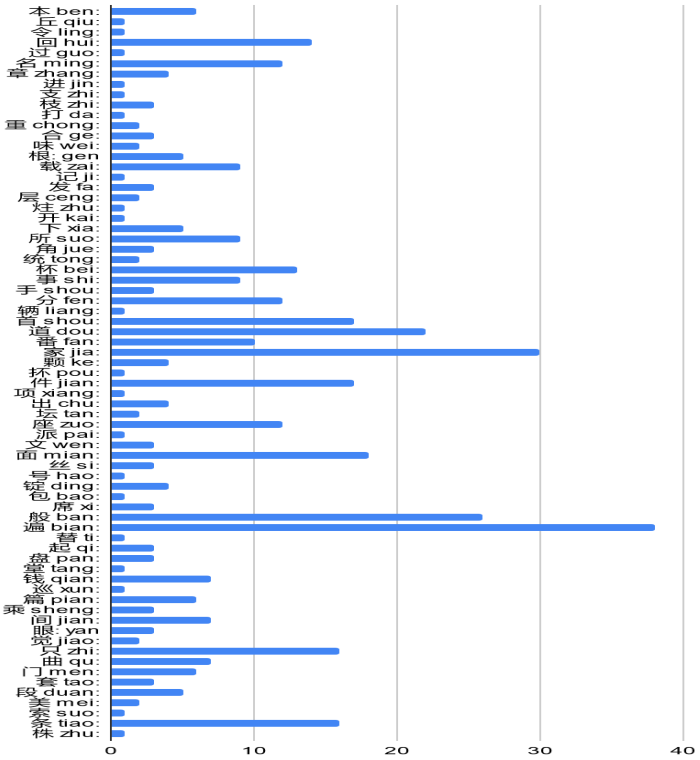


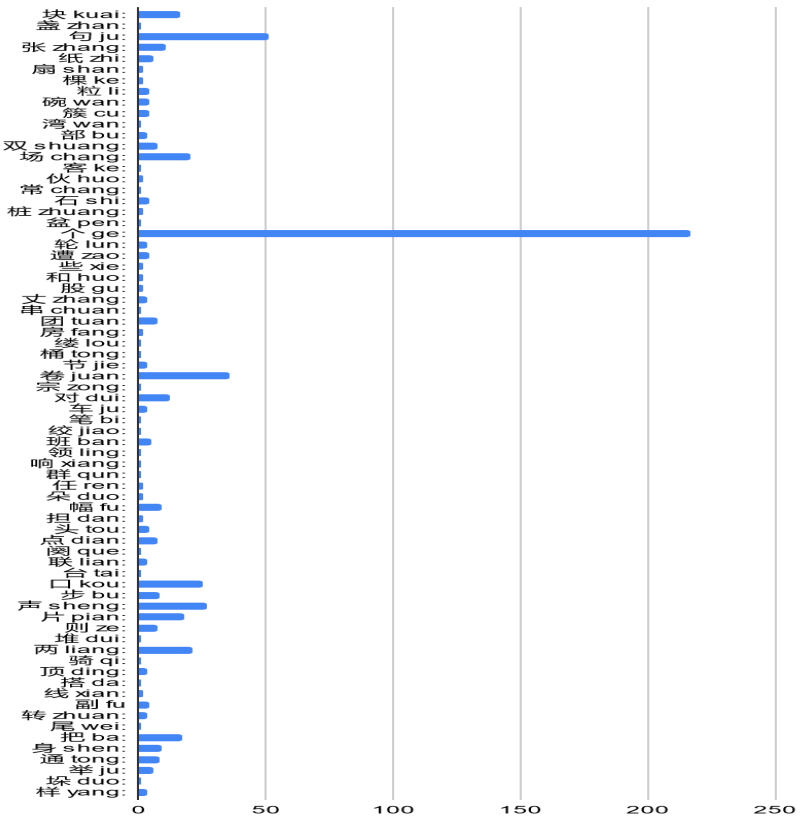
4.4e Stories to Instruct the World CL counts



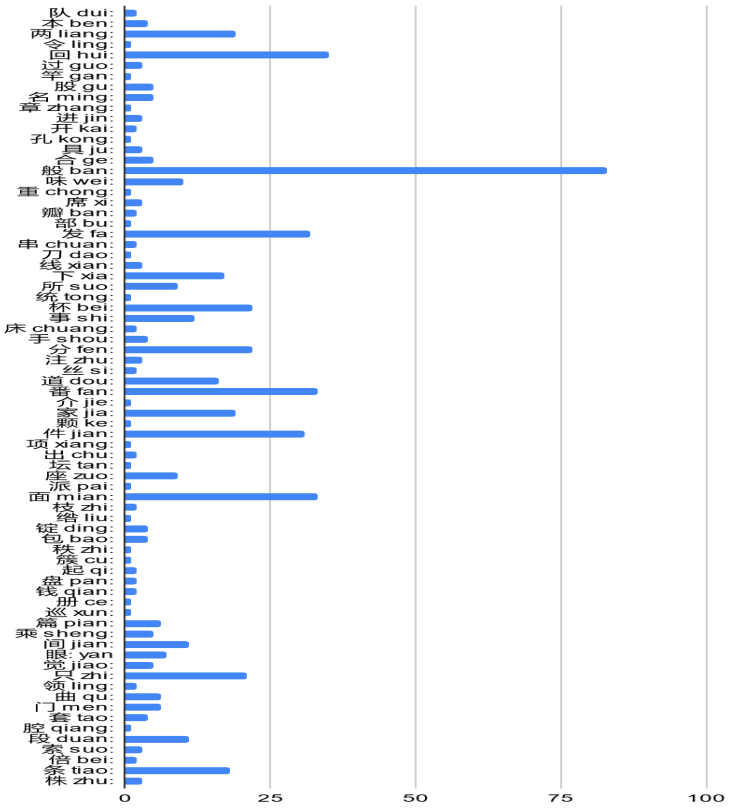


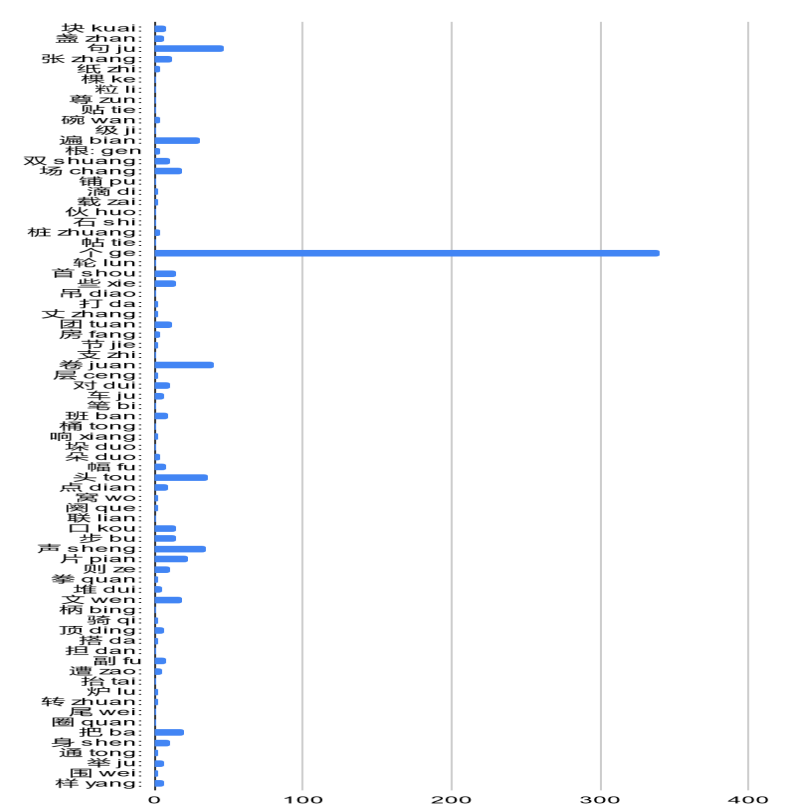
4.4f Stories to Caution the World CL counts



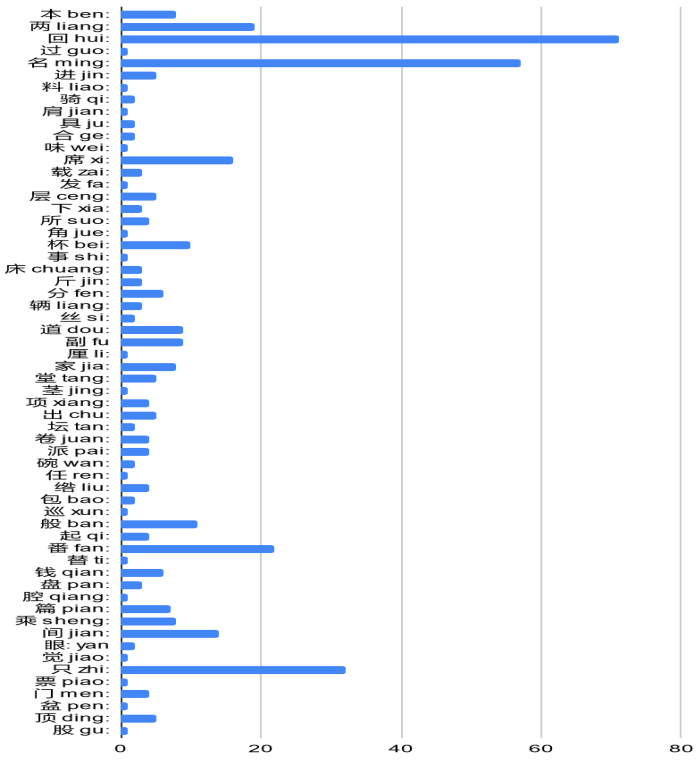


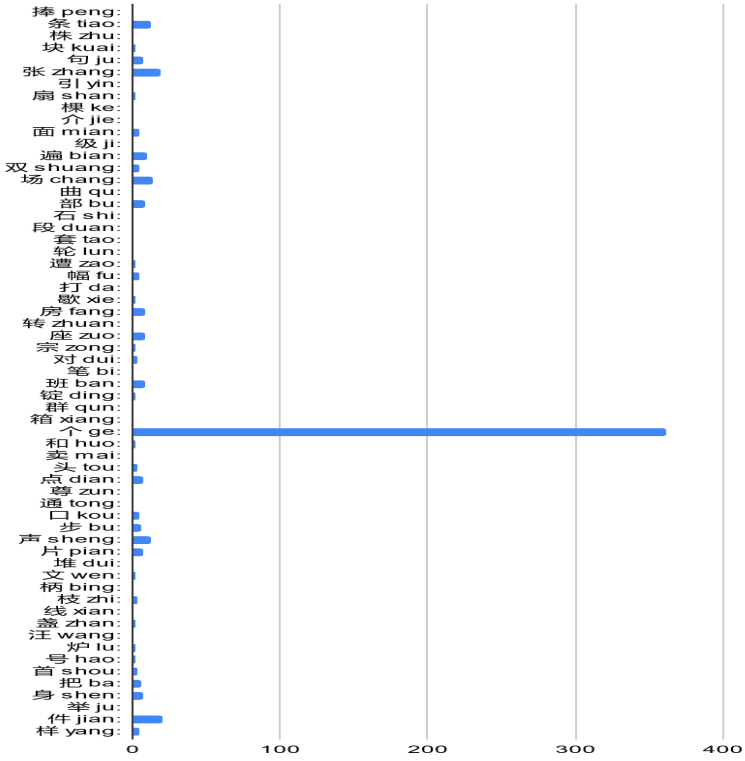
4.4g Stories to Awaken the World CL counts



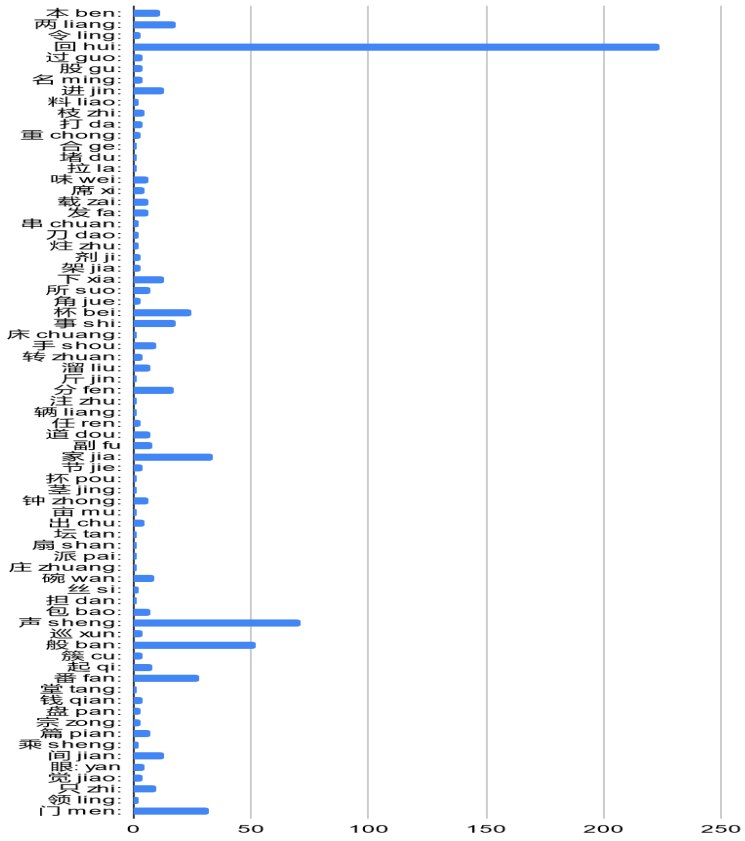


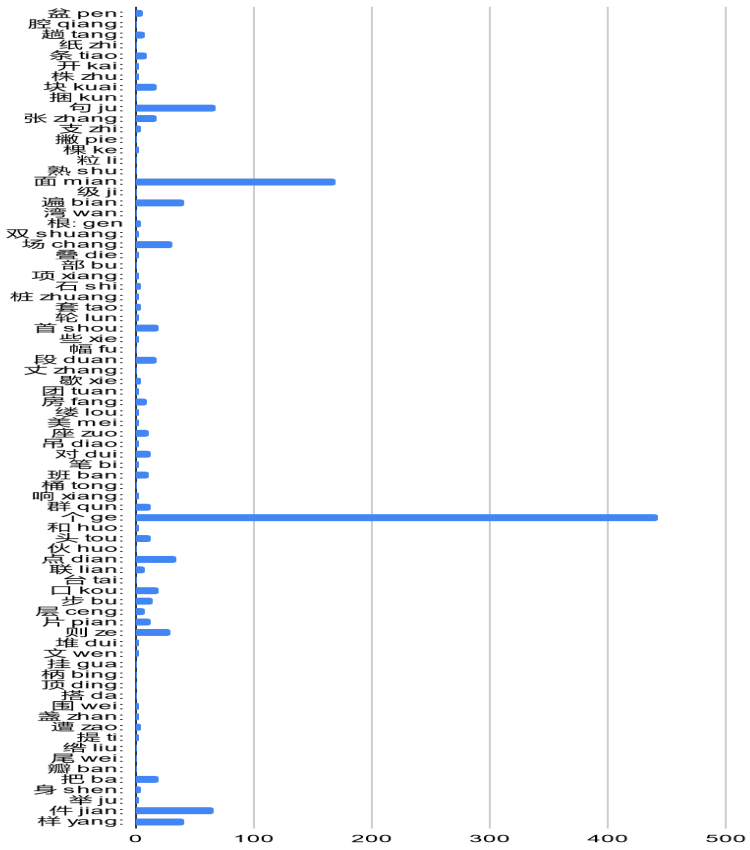
4.4h The Unofficial History of the Scholars CL counts

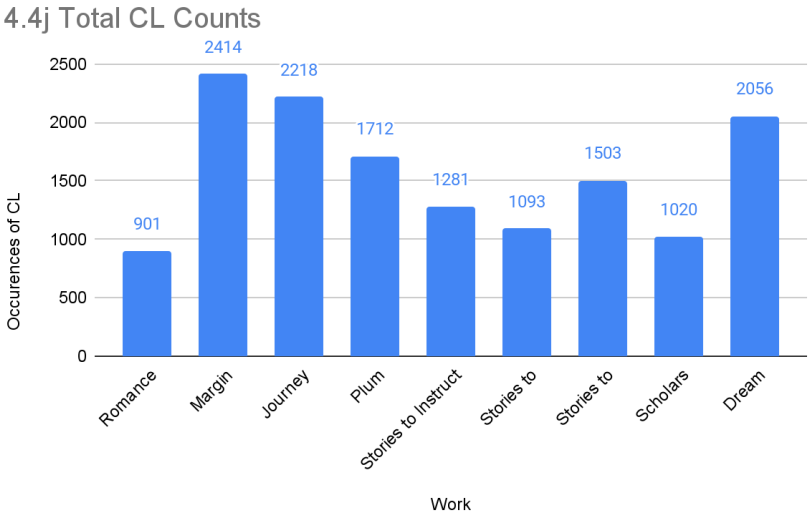


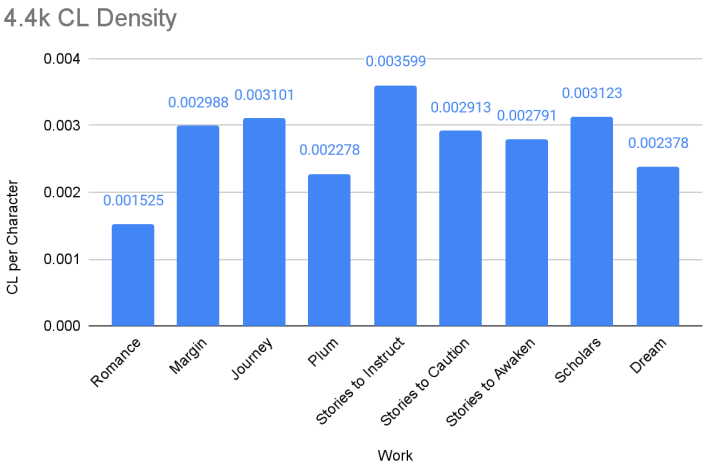


4.4i Dream of Red Chambers CL counts






4.4j Total CL Counts
Occurences of CL
2500
2000
1500
1000
500
0
901
2414
2218
1712
1281
1093
1503
1020
2056
Romance
Margin
Journey
Plum
Stories to Instruct
Stories to
Stories to
Scholars
Dream
Work


4.4k CL Density
CL per Character
0.004
0.003
0.002
0.001
0.000
0.001525
0.002988
0.003101
0.002278
0.003599
0.002913
0.002791
0.003123
0.002378
Romance
Margin
Journey
Plum
Stories to Instruct
Stories to Caution
Stories to Awaken
Scholars
Dream
Work

`

Bibliography

Baxter, William H., and Laurent Sagart. *Old Chinese: A new reconstruction*. Oxford University Press, 2014.

Brewitt-Taylor, C. H. (1925). *San kuo, or Romance of the three kingdoms*. Kelly & Walsh.

Borer, H. (2005). *Structuring sense: Volume 1: In name only* (Vol. 1). Oxford University Press.

Cheng, L. L. S., & Sybesma, R. (1998). Yi-wan tang, yi-ge tang: Classifiers and massifiers. *Tsing Hua journal of Chinese studies*, *28*(3), 385-412

Cheng, L. L. S., & Sybesma, R. (2005). Classifiers in four varieties of Chinese. *Handbook of comparative syntax*, 259–292

Croft, W. (1994). Semantic universals in classifier systems. *Word*, *45*(2), 145-171.

Denton, K. A. (1996). *Modern Chinese literary thought: writings on literature, 1893-1945*. Stanford University Press.

Evans, W.N. (2021). *From Roadside to Centerfold: A Study of the Development of Vernacular Literature in Late Medieval and Early Modern China*. Unpublished manuscript. Haverford College, East Asian Languages and Cultures department. Undergraduate thesis, Fall 2021.

Feng, S. (1996). Prosodically constrained syntactic changes in early archaic Chinese. *Journal of East Asian Linguistics*, *5*(4), 323-371.

Feng, S. (2012). The syntax and prosody of classifiers in classical Chinese. *Plurality and classifiers across languages of China*, 67-99.

Jacques, G. (2017). Traditional Chinese Phonology. *webpage, http://www. academia. edu/2261629/Traditional_Chinese_Phonology*.

Mair, Victor H. *The Columbia History of Chinese Literature* New York: Columbia University Press, 2001.

Mather et al. (2002). *Shih-shuo Hsin-yu: A New Account of Tales of the World, Second Edition*. University of Michigan Press. https://doi.org/10.3998/mpub.19860

Peyraube, Alain. (1991) 'Some remarks on the history of Chinese classifiers.' *Santa Barbara Papers in Linguistics* 3 : 106–126.

Senner WM (1991). editor. The origins of writing. U of Nebraska Press;

Stordalen, T. (2007). The Canonization of Ancient Hebrew and Confucian Literature. *Journal for the Study of the Old Testament*, *32*(1), 3-22.

Sybesma, Rint (2017). 'Classifiers, Nominal.' Encyclopedia of Chinese language and linguistics, ed. Rint Sybesma, Wolfgang Behr, Yueguo Gu, Zev Handel, C.-T. James Huang and James Myers. Leiden: Brill, Vol I, 620-627

Wong, Tak-sum, and John SY Lee (2019). 'Vernacularization in Medieval Chinese: A quantitative study on classifiers, demonstratives, and copulae in the Chinese Buddhist Canon.' *Digital Scholarship in the Humanities* 34, no. 1, 64–81.

Zhang, H. (2007). Numeral classifiers in mandarin chinese. *Journal of East Asian Linguistics*, *16*(1), 43-59.