## 1. Introduction

This study is conducted to answer the research question: "What are the differences or similarities in the depiction of ghosts in two Chinese fictional novels from different eras?"

The aim is to find high-frequency words that appear near (either before or after) the word "ghost" and compare the collocations between novels. The concordances of the collocations should be also read through to get a better idea of the relation of the words.

The second section of the report introduces the data used in this study and goes through the methods that were used to extract the concordances and collocations. The third section analyses the results of the extracted data, and the fourth section discusses the conclusions made from the results. The list of tools and R-code can be found in the section five. Section six shows the references and section seven includes the attachments of the results.

### 2. Data and method

The text data I chose for this study comes from the Project Gutenberg collection. The data consists of two novels by two Chinese authors, written in traditional Chinese. The older text is authored by Qian Tao who lived between 372?-427 (birth year uncertain) (Tao, 2005). The other text is authored by Mei Yuan who lived between 1716-1798 (Yuan, 2008). The time difference between the creation of these two novels is therefore around 1400 years.

I wanted to try modifying the text data with R-code using R-studio. Before entering the text data into R, I removed the info-text from Project Gutenberg from the beginning and end of the text file.

Inside R-studio, I first downloaded recommended packages. Then I created a variable for stop words by downloading a Baidu stop word list online, removing whitespace and extracting

comma-separated words. This was done with the help of a manual found in R's own website (Chinese text handling 2022). The codes used in R-studio along with explanations can be found in the last section of this report (section 5.2.).

Secondly, I created variables for the two text files. I used regular expression "\\W" to remove any non-alphanumeric character (spaces, commas etc.).

After cleaning the text, it should be tokenized. Tokenization of Chinese language is challenging because there are no spaces between words. I used multiple guides and loaded countless of packages to tokenize the Chinese text. I managed to create a list of word tokens, but I was not able modify the text variable with it. Secondly, whatever I did, I couldn't get the commands to use stop words list to work.

With another guide (Chinese Text Processing 2022), I was able to create concordances from Qian Tao's novel (examples below). The concordance informs that there are 33 matches for the search word "鬼" (ghost), which is quite small number for an analysis.

[text1, 3640] 來今已四年為|鬼|所枉殺案生錄

[text1, 7050] 眾客悉人此|鬼|亦入既入戶

[text1, 7056] 亦入既入戶 | 鬼 | 便持斧行棺

[text1, 7072] 欲與亡人訣|鬼|便以斧正打其

[text1, 7084] 即倒地左右扶出 | 鬼 | 於棺上視戎

[text1, 7095] 而 笑 眾 悉 見 | 鬼 | 斧 而出 李子 豫 少

[text1, 7133] 後有參謂腹中 | 鬼 | 曰何不速殺之

I did the same for Mei Yuan's novel (examples below). The concordance shows 812 matches, which is clearly more than in Qian Tao's novel.

[text1, 11] 語 也 然 龍 血 | 鬼 | 車 繋 詞語 之 玄鳥

[text1, 737] 聲 宛然 家主 也 二 | 鬼 | 啟 棺 之前 和 扶

[text1, 753]	若不勝病者二 鬼 手摩其腹口
[text1, 811]	然注目下視二 鬼 扶家主自棺中
[text1, 834]	曰法敗矣二 鬼 猙獰繞屋尋覓卒
[text1, 852]	牀帳被褥碎裂之一 鬼 仰頭見老僕在
[text1, 881]	地 棺 合 如故 二   鬼   亦 不復 見 矣 群

I was also able to change the text variable into corpus variable, but further commands to draw information from the variable did not work. Sadly, all the guides I used either had some codes that were missing from my R, even though I had all the recommended packages running, or the codes returned error-messages that I was unable to fix.

As it seems that my skills are too limited to conduct this research fully on R, I moved on to use Sketch Engine. I uploaded both novels into Sketch Engine.

The concordances in R showed that the matches for "鬼" (ghost) in Qian Tao's novel were 33, but when searching for the word in Sketch Engine, it only gave 8 concordances. Then I proceeded to use regular expressions and searched "\*鬼\*" and indeed got 33 results. The reason behind this difference is with tokenization: In addition to word "鬼", the Sketch Engine has created words such as "鬼神" (supernatural beings), "惡鬼" (evil spirit), "鬼怪" (ghosts and monsters), in total 22 different word types. I read through all 33 concordances to see that all the words referred to ghosts and decided to keep all the tokenized word types. In addition, I tried searching other words for ghosts in Chinese (\*幽灵\*, \*崇\*, \*亡灵\*, \*阴魂\*, \*幽魂\*, \*魅影\*), but none of these amounted any results.

Then I moved on to Sketch Engine's other function, Word Sketch. Comes out that the search with regular expressions (\*鬼\*) regrettably does not work, because the search only supports lemmas. I resulted in searching all word types separately in this smaller corpus and

combining them into Excel file. Some of the search words do not give any results because the corpus in this case is very small (Qian Tao's novel corpus only has 13.289 tokens).

	Word	Grammatical relation	Count	Score ↓	
1	<b>長三丈</b> 鬼,長三丈	and/or	1	11.8	
2	<b>縛得</b> 縛得鬼	Object_of	1	11.8	•••
3	暨令 鬼。 廬江 杜謙為諸 暨令	and/or	1	11.8	•••
4	<b>依諸叢</b> 鬼 依諸叢	Modifies	1	11.8	
5	<b>來住劉</b> 鬼來住劉	Modifies	1	11.8	•••

Picture 1. A part of Word Sketch search for a search word "鬼".

To familiarize myself with another corpus (Mei Yuan's novel) as well, I ran concordance search with a search word "鬼" (337 matches) and ("\*鬼\*") (809 matches). This corpus is considerably larger with 182.389 word tokens. I decided to make the Word Sketch search with the same search words as Qian Tao's novel and save them in Excel file as well (see the last section 7 for attachments of these files).

### 3. Results

I will first look at the resulted collocations in smaller and older Qian Tao's corpus. The reliability of the collocations is questionable as all the results show a count of one, so each of them appear with the search word only once. I sorted the list in order of the typicality score, which tells how much more typically the word in question appears with the search word compared to other words in the corpus. The words with highest scores are 外來 ("come from outside"), 還祠 ("return to the shrine"), 得以 ("so that"), 眾悉見 ("everyone saw") and 與人 ("compared to people") all of them with typicality score of 14. To get an idea of scoring system in Sketch Engine, I searched the word "ghost" in British National Corpus (BNC) and

concluded that the scores in large corpora likes this does not often go above 10 (for example a collocate "haunt" got 15 matches and a score of 9.7 with a search word ghost). The high scores in Qian Tao's corpus can then be explained with the small size of the corpus.

After going through all the collocations in Qian Tao corpus search, the words (all with typicality score over 11) that could be seen describing ghosts are 還祠 ("return to the shrine"), 即絕 ("to cut off/kill"), 群 ("group" of ghosts), 役使 ("to use as servant"), 腹中 ("inside the stomach"), 打 ("hit"), 刀 ("knife"), 長三丈 ("3 (Chinese) feet long" =9.9m), 來 住劉 ("came to live"), 縛得 ("to bind") and 偷食 ("to steal food").

In the longer and newer Mei Yuan novel's corpus, unlike in Qian Tao's corpus, there are collocates with several hits. The words with most hits are 日 ("speak/say", 36 hits), 耶 (yesound that is added to the end of sentence to express a questioning, enthusiasm, impatience etc., 10 hits), 群 ("group" of ghosts, 7 hits) and 卒 ("soldier", 6 hits). The words with the highest typicality scores are 逡巡 ("to hesitate", score 12,42), 勿懼 ("do not fear", score 12), 豈 (asking a rhetorical question, score 12), 乘機而至 ("take advantage of the opportunity", score 11,68). Other collocations (with typicality score over 10) that can be used to describe ghosts are 驅除 ("exorcise"), 昏聵 ("decrepit /muddle-headed") and 欺人 ("deceive people").

## 4. Conclusion and discussion

After close reading the concordances with the focus on the most meaningful collocations, I find Qian Tao's novel to depict ghosts as violent and eager to kill. The ghost can inhabit a person's body (there is twice a mention of a ghost inside the person's stomach), use a knife or axe to hit people or steal food. They seem often to appear in groups and move to inhabit shrines or houses. They may look a lot like humans and use clothes, but they also may be 10

meters tall. Buddhist monks can have an ability to recite matras to bind ghosts and to make ghosts or spirits obey them.

Mei Yuan's novel depicts more human-like behaviour of ghosts compared to Qian Tao's novel. The ghosts in Mei Yuan's novel can speak to humans, which did not happen even once in Qian Tao's novel. The ghosts also show human-like feelings like being hesitant or decrepit, and they show ability of thinking as they try to deceive humans. Both novels had ghosts appearing in groups, but only in Mei Yuan's novel there are ghost soldiers. Skilled people can exorcise ghosts. This made with Buddhist mantras as was done in Qian Tao's novel.

In summary, it seems that in the newer Mei Yuan's novel the ghosts are more human-like in many ways: they can feel human emotions, act as a soldier, scheme, and even talk. The ghosts in Qian Tao's novel on the other hand are simply drawn to haunt empty houses or shrines and merely act based on a killing instinct.

Whether these depictions of ghosts truly represent the general depiction of ghosts in the given era cannot truly be based on this study alone, as these two novels are only one example of one author of the era. Moreover, it is very possible that the story or the text has been modified over the centuries. It can still maybe show a rough change in the culture of mythologies of over thousand years. To make this find more eligible, more novels and other texts from the given eras should be studied to gain a bigger picture.

#### 5. List of tools and codes used

#### 5.1.Tools

- R-code, R studio
- Sketch Engine

```
5.2. Code
> library("corpus")
> set.seed(100)
> library(stringi)
> library(tidyverse)
> library(tidytext)
> library(quanteda)
> library(stringr)
> library(jiebaR)
> library(readtext)
## Download a stop word list suitable for Chinese (the Baidu stop words), remove
whitespace and extract the comma-separated words
> cstops <-
"https://raw.githubusercontent.com/ropensci/textworkshop17/master/demos/chineseDemo/Ch
ineseStopWords.txt"
> csw <- paste(readLines(cstops, encoding = "UTF-8"), collapse = "\n")
> csw <- gsub("\\s", "", csw)
> stop_words <- strsplit(csw, ",")[[1]]
## Download the text file and remove whitespace, commas etc.
> qian_tao1 <-"C:/Users/crazy/OneDrive - University of Helsinki/UUDET
KURSSIT/Computational literacy/Final Project/original texts/Qian Tao (older) – cleaned.txt"
> qian_tao <- paste(readLines(qian_tao1, encoding = "UTF-8"), collapse = "\n")
> qian_tao <- gsub("\\W", "", qian_tao)
##Create a corpus variable
> corpus_qian_tao <- corpus(qian_tao)
##Creating a list of tokenized words
> text_tokens <- tokens(qian_tao)
##Creating a concordance
```

> kwic(text\_tokens, pattern = "鬼")

## 6. References

M. Yuan. (2008). 子不語. Urbana, Illinois: Project Gutenberg. Retrieved November 29, 2022, from <a href="https://www.gutenberg.org/ebooks/25245">https://www.gutenberg.org/ebooks/25245</a>.

Q. Tao. (2005). 搜神後記. Urbana, Illinois: Project Gutenberg. Retrieved November 29, 2022, from <a href="https://www.gutenberg.org/ebooks/7266">https://www.gutenberg.org/ebooks/7266</a>.

Chinese text handling (2022). Retrieved November 29, 2022, from <a href="https://cran.r-project.org/web/packages/corpus/vignettes/chinese.html">https://cran.r-project.org/web/packages/corpus/vignettes/chinese.html</a>.

Chinese Text Processing (2022). Retrieved December 16, 2022, from https://alvinntnu.github.io/NTNU\_ENC2036\_LECTURES/chinese-text-processing.html.

# 7. Attachments

## 1. Collocates for 鬼 Qian Tao's novel

corpus	subcor	keyword 💌	pos 🔻	freq 💌	coll	hits 💌	score 💌	query	<b>v</b>	gramrel 💌
user/ravanka/qian_tao	-	鬼	-n	8	長三丈	2	11,83	[ws(2,	35125761)]	and/or
user/ravanka/qian_tao	-	鬼	-n	8	暨令	:	11,83	[ws(2,	35125760)]	and/or
user/ravanka/qian_tao	-	鬼	-n	8	來住劉	<i>'</i>	11,83	[ws(2,	35093506)]	Modifies
user/ravanka/qian_tao	-	鬼	-n	8	依諸叢	2	11,83	[ws(2,	35093505)]	Modifies
user/ravanka/qian_tao	-	鬼	-n	8	縛得	<i>-</i>	11,83	[ws(2,	35066304)]	Object_of
user/ravanka/qian_tao	-	鬼	-n	8	見一		11,68	[ws(2,	, 35112768)]	N_Modifier
user/ravanka/qian_tao	-	鬼	-n	8	腹中	<i>-</i>	11,54	[ws(2,	35112769)]	N_Modifier
user/ravanka/qian_tao	-	鬼	-n	8	偷食	1	11,42	[ws(2,	, 35073538)]	Subject_of
user/ravanka/qian_tao	-	鬼	-n	8	病		10,75	[ws(2,	35093504)]	Modifies
user/ravanka/qian_tao	-	鬼	-n	8	入	2	10,48	[ws(2,	35073537)]	Subject_of
user/ravanka/qian_tao	-	鬼	-n	8	日	<i>'</i>	8,12	[ws(2,	35073536)]	Subject_of
user/ravanka/qian_tao	-	卿家鬼	-n	1	何在	2	13,42	[ws(2,	37601472)]	Subject_of
user/ravanka/qian_tao	-	呼諸鬼	-n	1	王	<i>'</i>	11,83	[ws(2,	15147904)]	Modifies
user/ravanka/qian_tao	-	惡鬼	-n	1	忽有	2	11,42	[ws(2,	15074944)]	Object_of
user/ravanka/qian_tao	-	聞鬼從	-n		外來	<i>'</i>	14	[ws(2,	37939648)]	Subject_of
user/ravanka/qian_tao	-	見鬼	-v	2	能	1	10,48	[ws(2,	, 35853440)]	SentObject_
user/ravanka/qian_tao	-	見鬼	-v		人	<i>'</i>	7,63	[ws(2,	35860928)]	Subject
user/ravanka/qian_tao	-	諸鬼	-n	1	還祠	1	14	[ws(2,	36153408)]	Subject_of
user/ravanka/qian_tao	-	諸鬼兩兩三	n	1	抱持		13,42	[ws(2,	36108352)]	Subject_of
user/ravanka/qian_tao	-	飯遺鬼	-n	1	得以	1	14	[ws(2,	36801600)]	Object_of
user/ravanka/qian_tao	-	飯遺鬼	-n	1	酒	1	13	[ws(2,	36809792)]	N_Modifier
user/ravanka/qian_tao	-	鬼便	-n	1	打	:	12,19	[ws(2,	35178688)]	Subject_of
user/ravanka/qian_tao	-	鬼對	-n	1	腹中	1	13	[ws(2,	35441024)]	N_Modifier
user/ravanka/qian_tao	-	鬼對	-n	1	日	:	8,21	[ws(2,	35434304)]	Subject_of
user/ravanka/qian_tao	-	鬼怪	-n	2	即絕	1	13,42	[ws(2,	. 15644544)]	Modifies
user/ravanka/qian_tao	-	鬼怪	-n	2	刀	:	12	[ws(2,	. 15651328)]	N_Modifier
user/ravanka/qian_tao	-	鬼斧而出	-v	1	眾悉見	:	14	[ws(2,	35246016)]	Subject
user/ravanka/qian_tao	-	鬼為	-v	1	何在	1	13,42	[ws(2,	40970112)]	SentObject
user/ravanka/qian_tao	-	鬼相叱	-n	1	群	- 1	13,42	[ws(2,	36003392)]	N_Modifier
user/ravanka/qian_tao	-	鬼神	-n	2	役使	1	13,42	[ws(2,	. 14814656)]	Object_of
user/ravanka/qian_tao	-	鬼神	-n	2	事	- 1	11,3	[ws(2,	, 14821376)]	Modifies
user/ravanka/qian_tao	-	鬼騎馬滿道	-n	1	與人		14	[ws(2,	, 36641984)]	and/or

# 2. Collocates for 鬼 Mei Yuan's novel

_								
	subcorpus keyword	y pos y			hits 🔻 so			gramrel 💌
user/ravanka/mei_yuan	- 鬼	-n	337		10		[ws(2, 73962575)]	
user/ravanka/mei_yuan		-n	337		7		[ws(2, 74253448)]	
user/ravanka/mei_yuan		-n	337		6		[ws(2, 73962568)]	
user/ravanka/mei_yuan	- 鬼	-n	337		36		[ws(2, 73150849)]	
user/ravanka/mei_yuan		-n	337		4		[ws(2, 73150907)]	
user/ravanka/mei_yuan		-n	337		3	8,07	[ws(2, 73150902)]	Subject_of
user/ravanka/mei_yuan	- 鬼	-n	337	入	4	8,02	[ws(2, 73150862)]	Subject_of
user/ravanka/mei_yuan	- 鬼	-n	337	去	5	7,87	[ws(2, 73150851)]	Subject_of
user/ravanka/mei_yuan	- 鬼	-n	337	肯	3	7,87	[ws(2, 73150861)]	Subject_of
user/ravanka/mei_yuan	- 鬼	-n	337	持	3	7,8	[ws(2, 73150873)]	Subject_of
user/ravanka/mei yuan	- 鬼	-n	337	笑	3	7,71	[ws(2, 73150850)]	Subject of
user/ravanka/mei yuan	- 鬼	-n	337	言	3		[ws(2, 73150866)]	
user/ravanka/mei yuan	- 鬼	-n		呵欠	2		[ws(2, 74582700)]	
user/ravanka/mei yuan		-n		避人	2		[ws(2, 73562925)]	-
user/ravanka/mei yuan		-n	337		2		[ws(2, 73150881)]	
user/ravanka/mei_yuan		-n		拍手	2		[ws(2, 73562886)]	
user/ravanka/mei_yuan	- 鬼	-n		為群	2		[ws(2, 74253454)]	
· -			1	河水	2			
user/ravanka/mei_yuan		-n					[ws(2, 74253469)]	
user/ravanka/mei_yuan	- 鬼	-n	337		2		[ws(2, 74253452)]	
user/ravanka/mei_yuan		-n	337		3		[ws(2, 73962562)]	
user/ravanka/mei_yuan	- 鬼	-n	337		2		[ws(2, 72934792)]	
user/ravanka/mei_yuan		-n	337		2		[ws(2, 72934794)]	
user/ravanka/mei_yuan	- 鬼	-n	337		4	7,48	[ws(2, 72934788)]	Object_of
user/ravanka/mei_yuan	- 鬼	-n	337	散	2	7,46	[ws(2, 73150900)]	Subject_of
user/ravanka/mei_yuan	- 鬼	-n	337	喝	2	7,45	[ws(2, 73150883)]	Subject_of
user/ravanka/mei_yuan	- 鬼	-n	337	答	2		[ws(2, 73150878)]	
user/ravanka/mei_yuan	- 鬼	-n	337		2		[ws(2, 73150856)]	
user/ravanka/mei yuan		-n	337	杭州	2		[ws(2, 73962570)]	
user/ravanka/mei_yuan	- 鬼	-n	337		2		[ws(2, 74582660)]	
user/ravanka/mei_yuan		-n	337	_	3		[ws(2, 74362666)] [ws(2, 73150854)]	
	- 鬼		337		2		[ws(2, 73150854)] [ws(2, 73150860)]	
user/ravanka/mei_yuan		-n	1					
user/ravanka/mei_yuan		-n	337		2		[ws(2, 73150888)]	
user/ravanka/mei_yuan	- 鬼	-n	337		5	_	[ws(2, 74582656)]	
user/ravanka/mei_yuan		-n	337		2		[ws(2, 74253442)]	
user/ravanka/mei_yuan	- 鬼	-n	337		3		[ws(2, 73150848)]	
user/ravanka/mei_yuan	- 鬼	-n	337	得	2	6,59	[ws(2, 73150857)]	Subject_of
user/ravanka/mei_yuan	- 鬼	-n	337	人	2	5,53	[ws(2, 74253440)]	N_Modifier
user/ravanka/mei_yuan	- 惡鬼	-n	9	乘機而	1	11,68	[ws(2, 250794368]	Subject_of
user/ravanka/mei_yuan	- 惡鬼	-n	9	輕生	1	11,42	[ws(2, 250810176]	]and/or
user/ravanka/mei_yuan	- 惡鬼	-n	9	驅除	1	11,3	[ws(2, 250786240]	Object_of
user/ravanka/mei yuan	- 惡鬼	-n	9	耶	1	7,62	[ws(2, 250802176]	]Modifies
user/ravanka/mei yuan	- 見鬼	-v	7	勿懼	1	12	[ws(2, 422358337	Object
user/ravanka/mei yuan		-v		下午	1		[ws(2, 422381248	
user/ravanka/mei_yuan		-v		面目	1		[ws(2, 422358336]	
user/ravanka/mei_yuan		-v		常	1		[ws(2, 422392449]	
user/ravanka/mei_yuan		-v		能	4		[ws(2, 422372800]	
user/ravanka/mei_yuan		-v	_	不	1		[ws(2, 422392448]	
user/ravanka/mei_yuan		-n		逡巡	1		[ws(2, 471826944	
user/ravanka/mei_yuan		-v		豈	1		[ws(2, 450050816]	
user/ravanka/mei_yuan		-v	1	敢	1	7,52	[ws(2, 450042944]	SentObject
user/ravanka/mei_yuan		-n	20	果相認	1	10,61	[ws(2, 154009671]	]Modifies
user/ravanka/mei_yuan	- 鬼神	-n	20	幽明道	1	10,61	[ws(2, 154009665]	Modifies
user/ravanka/mei_yuan	- 鬼神	-n	20	為口實	1	10,61	[ws(2, 154009664]	]Modifies
user/ravanka/mei_yuan	- 鬼神	-n	20	昏聵	1		[ws(2, 153975236	
user/ravanka/mei_yuan		-n		情狀	1		[ws(2, 154085312]	
user/ravanka/mei yuan		-n		欺人	1		[ws(2, 154009669]	
user/ravanka/mei yuan	, ,	-n		唸經	1		[ws(2, 154009668]	
user/ravanka/mei_yuan		-n		以為	1		[ws(2, 154059971]	
user/ravanka/mei_yuan	, -, ,			力量			[ws(2, 154009667]	
		-n			1			
user/ravanka/mei_yuan		-n		言明	1		[ws(2, 153975234]	
user/ravanka/mei_yuan		-n		怪物	1		[ws(2, 154009670]	
user/ravanka/mei_yuan		-n		交接	1		[ws(2, 154009666	
user/ravanka/mei_yuan		-n		敬	1		[ws(2, 153929728]	
user/ravanka/mei_yuan		-n		何在	1		[ws(2, 153975235	
user/ravanka/mei_yuan	- 鬼神	-n		托	1	9,87	[ws(2, 153929731]	]Object_of
user/ravanka/mei_yuan	- 鬼神	-n	20	陰司	1	9,79	[ws(2, 154059969]	N_Modifier
user/ravanka/mei_yuan	- 鬼神	-n	20	天地	1	9,64	[ws(2, 154059968	]N_Modifier
user/ravanka/mei_yuan		-n		供	1		[ws(2, 153929733	_
user/ravanka/mei_yuan		-n		信	1		[ws(2, 153929729	
user/ravanka/mei yuan		-n		為	1		[ws(2, 154059970]	
user/ravanka/mei_yuan		-n		與	1		[ws(2, 153929732]	
user/ravanka/mei_yuan		-n		與	1		[ws(2, 153929732]	
user/ravanka/mei_yuan								
		-n		怪	1		[ws(2, 153975232]	
user/ravanka/mei_yuan		-n	20		1		[ws(2, 154093185]	
user/ravanka/mei_yuan user/ravanka/mei_yuan	- 鬼神	-n	20	事	1	6,6	[ws(2, 154093184	]and/or
user/ravanka/mei_yuan	- 鬼神 - 鬼神		20 20			6,6 6,3		)]and/or ) Object_of