# Appendix 10
# Notes on Topic Models

## 10.1 Stylistic Topics in Which *Xin* Appears

| Topic | Weight | Name | Top 10 Characters |
|---|---|---|---|
| 64 | 0.11826 | *Shijing* 詩經 stylistic | 彼 兮 爾 心 人 君 思 匪 予 莫 |
| 43 | 0.04009 | *Chuci* 楚辭 stylistic | 兮 余 些 心 思 只 離 魂 悲 歸 |
| 92 | 0.06683 | *Wenshi Zhenjing* 文始真經 stylistic | 人 物 聖 尹 關 道 地 神 生 天 |

Topic #64 immediately strikes one as representing distinctively *Book of Odes* (s*hijing* 詩經) language, with its archaic personal pronouns and poetic exclamation terms, and indeed it loads 30% on the *Shijing* and very little on any other text except the lexicon *Erya* (7%) and the *Explanations of the Odes* (*shisuo* 詩說 (5%), the latter making great sense since it is a commentary on the *Shijing*. Topic #64 appears to be a stylistic rather than conceptual topic, reflecting a distinctive cluster of punctuation, grammar terms, or repeated phrases linked to a particular text, and therefore does not tell us much about *xin* conceptually, even though *xin* loads rather highly. It is also worth noting that, while a verb for cognitive action appears (*si* 思, "to think, reflect, long for"), no emotions words are present.

Topics #43 and #92 also seem to be stylistic marker topics. In topic #43, the poetic exclamation *xi* 兮 ("oh") loads most highly (.277), but the topic then rapidly falls off to the first-person pronoun *yu* 余 (.019), with the rest of the components having very little weight in the topic. This seems to be a *Songs of Chi* (*Chuci* 楚辭) stylistic marker topic, since these two bits of punctuation are pervasive in this Late Warring States / Early Han poetic text. This would also explain terms related to spiritual travel or physical journeys, such as 離魂歸, as well as 'sorrow' (悲). Indeed, a check of topics loading per

text shows Topic #43 loading 40% in the *Chuci* and into almost no other text at all, and no more than 1-2% in the few where it appears.
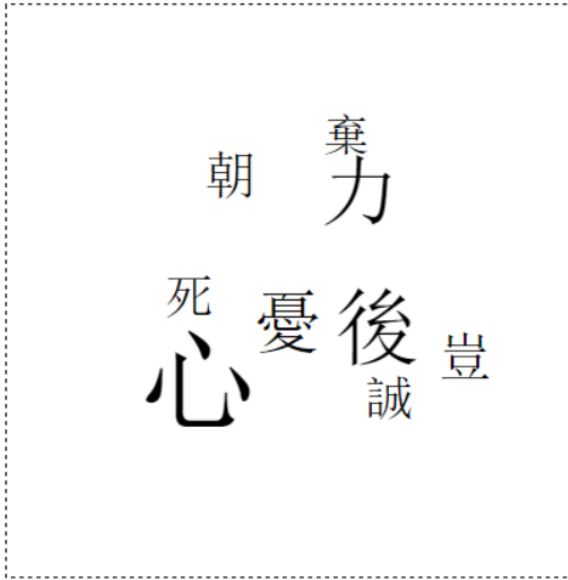
Topic #92 is another purely stylistic topic, loading 50% onto the *Wenshi Zhenjing* 文始真經, a loosely "Daoist" text of unclear origin also known as the *Yiguanzi* 關尹子, or *[Book of] Master Yiguan*. This immediately explains four of the top five heaviest loading characters, since every passage in the text begins "Master Yiguan said," (*yiguanzi yue* 關尹子曰), and many concern speculations about the "sagely person" (*shengren* 聖人).

**10.2 Word Cloud Representations of Topics #97 and #10**



Topic 97

**Topic 10**



Word clouds represent the "weight" within the topic by size.


**10.3  Traditional Chinese Medicine Topics Involving *Mai* 脈**

        The only other organ that appears in any of our 100 topics is *mai* 脈
("vein/artery/pulse"), which shows up, as one might expect, in two wonderfully coherent
"Traditional Chinese Medicine" topics (#73 and #84), that consist almost entirely of
technical, medical terminology and load almost exclusively in the Medical text portion of
the corpus.


| Topic | Weight | Name | Top 14 characters |
|-------|--------|------|-------------------|
| 73 | 0.0876 | Traditional Chinese Medicine | 氣 病 陽 脈 陰 下 刺 大 歧 熱 藏 血 痛 治 |
| 84 | 0.06703 | Traditional Chinese Medicine | 兩 湯 升 脈 下 服 病 大 寒 味 熱 汗 水 黃 |


        Topic #73 loads at high levels into two of our four medical texts, 60% in the
*Huangdi Neijing* (黃帝經, *The Yellow Emperor's Inner Canon*) and 70% in the *Huangdi
Bashiyi Nanjing* (黃帝八十一難經, *The Yellow Emperor's Canon of Eighty-one Difficult*

*Issues*), with some minor loading in the other two texts *Shang Han Lun* (傷寒論, *Treatise on Cold Injury*) (9%) and *Jinkui Yaolue* (金匱要略, *Essential Medical Treasures of the Golden Chamber*) (6%). Topic #84, on the other hand, loads at 80% into the *Shang Han Lun* and the *Jinkui Yaolue*, although only 4% into the *Nanjing* and 1% in the *Shang Han Lun*. The former texts are generally attributed to the late Warring States or earlier portion of the Han Dynasty (206 BCE – 220 CE) and the latter texts to the second half, so our topic model might here be corroborating these attributions by picking up an interesting chronological shift in medical terminology, one that could then be used to date less well-attested texts.

**10.4 Topic #11, Containing *Wei* 胃**

| Topic | Weight | Name | Top 14 characters |
|-------|--------|------|-------------------|
| 11 | 0.01866 | Archaeological *Daodejing* Stylistic | 善 大 胃 弗 恆 呵 器 居 果 久 俞 邦 單 驚 |

Topic #11 should be mentioned because, at first glance, "stomach" (*wei* 胃) seems to appear in it. Topic #11 loads almost exclusively into the two archaeological versions of the *Daodejing*—the Mawangdui (20%) and Guodian (6%)—in the CTP corpus. Digging into the passages themselves, however, it is clear that *wei* 胃 here is simply the graphic variant for *wei* 謂 ("to be called") that is employed in the Mawangdui *Daodejing*. Topic #11 thus seems to be another stylistic topic, picking out terms distinctive to the archaeological versions of the text. These include the use of *bang* 邦 in place of *guo* 國 to convey "state," a feature of texts written before the tabooing of the personal name of the first Han Emperor, Liu Bang 劉邦, upon his accession in 206 BCE. This topic also loads in the two lexicons in the corpus (the *Shuowen* and *Erya*, both at 2%), which no doubt reflects their inclusion of archaic word forms. It also shows up in the poetry collection, the *Songs of Chu* (*chuci* 楚辭), at 1%, which might reflect stylistic features common to texts from the ancient state of Chu, where the Mawangdui and Guodian tombs were located. Topic #11 thus gives us another good example of how topic modeling can pick

up clusters of distinctive grammar or terminology that could be useful in dating texts or adjudicating debates about origins or transmission of received texts.