

Theme and Motif Extraction of Bilingual Writers Using Natural Language Processing

Ming Qian, Charles River Analytics, Cambridge, MA, USA (mqian@cra.com)

Research questions

Do bilingual writers frequently focus on similar themes/motifs for works in different languages?

Can natural language processing (NLP), which is a branch of artificial intelligence (AI), be used to automatically detect and extract themes and motifs across different languages?

Methods

Human evaluators read and analyzed prose and novels [1, 2, 3, 4] written by two bilingual writers in both English and Chinese: Lin Yutang (1895-1976) and Ye Geling (1958-). Themes and motifs were then identified and verified by human evaluators.

Infranodus text network analysis and visualization tool performed unsupervised clustering and topical keyword extraction. GPT-3 and Chat-GPT generative language models were used to generate short summarizations (2-3 sentences) on the writing samples based on topical keywords provided by Infranodus tool. The results were then used to confirm the human evaluation.

Sample Results

Lin YuTang's English and Chinese works were evaluated [Table 1, 2 and Figure 1]. While the Chinese prose tried to explain the British cultural character to a Chinese audience, and the English book tried to explain the Chinese cultural characters to an English audience, the author adopted the same theme/motif: balanced abstract/logical thought and 'common sense'. Table 1 shows the human highlighted keywords. Figure 1 shows the topical keywords clusters identified by Infranodus text network analysis and visualization tool [5]. Table 2 shows the summarization results returned by GPT-3 [6] based on original content and topical keywords identified.

Conclusion

Bilingual writers frequently focus on similar themes/motifs for works in different languages. They sometimes finetune themes/motifs to appeal to audiences with different cultural backgrounds. A new generation of NLP can be leveraged to effectively automate the tasks of detecting themes/motifs across different languages

《英国人与中国人》 ("British and Chinese People")	"THE DOCTRINE OF THE GOLDEN MEAN" in the book "MY COUNTRY AND MY PEOPLE"
英国人曾被人非难为虚伪、矛盾、有“糊涂混过”的天才，却显然缺乏逻辑的民族……所有一切对英国人的误解，是由于对思想的真正功能的谬误见解所致……我们要把抽象的思想认为人类心性的最高功能，认为它的价值超过了简单的常识……这种帮助一个人去获得食物和生活下去的思想是一种较高的，而不是较低的思想，因为这一类思想常常比较健全。这一类的思想通常便叫做常识……一个具有健全常识的民族并不是一个不会思想的民族，而是一个把他的思想归纳到生活的本能那里，使它们和谐相处的民族。 The British have been denounced as hypocritical, contradictory, a "foolish" genius, but apparently has been a nation without logic... All misconceptions about the English are due to a false view of the true function of the mind... We regard abstract thought as the highest	The religion of common sense or the spirit of reasonableness is part and parcel of Confucian humanism . It is this spirit of reasonableness which has given birth to the Doctrine of the Golden Mean, the central doctrine of Confucianism. Reference has already been made in the preceding chapter to the spirit of reasonableness , as contrasted with logic or reason itself... the spirit of reasonableness is largely intuitive and practically identical with English common sense . It has been further

function of human mind, and consider its value beyond simple **common sense**... The kind of thought which helps a man to get food and to live is a higher thought, not a lower one, for such thoughts are often more healthy. Thoughts of this kind are usually called **common sense**... **A nation with sound common sense** is not a nation without thinking, but a nation that **reduces its thoughts to the instincts of life and makes them live in harmony**.

shown that for a Chinese it is not enough that a proposition be **"logically correct"**; it is much more important that it be **"in accord with human nature"**... A typically Chinese judgment is: **"A is right, and B is not wrong either."** ... **The reasonable mind keeps a balance** when the logical mind has lost it.

Table 1: The same theme/motif (the doctrine of the golden mean) was adopted by Lin YuTang to analyze characteristics of British and Chinese cultural characters in his bilingual pieces.

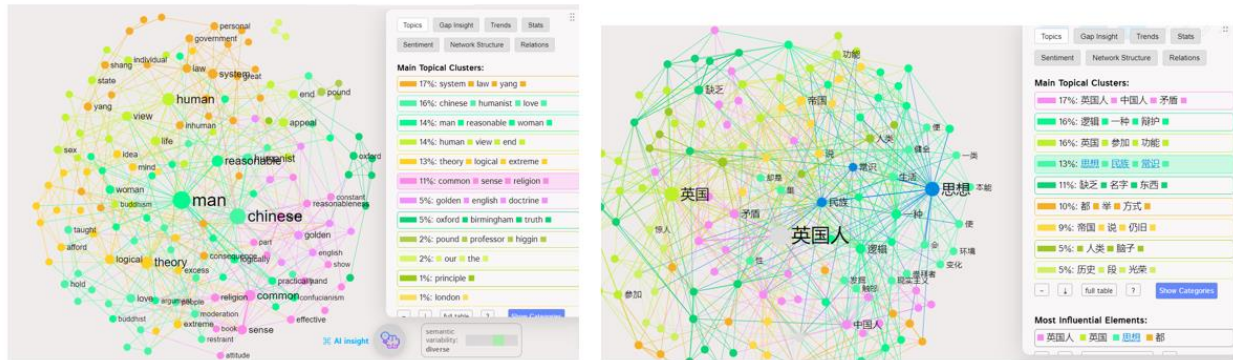


Figure 1: Topical modeling results based on Lin YuTang's Chinese prose 《英国人与中国人》 ("British and Chinese People") and the section titled as "THE DOCTRINE OF THE GOLDEN MEAN" from his English book "MY COUNTRY AND MY PEOPLE". Similar keyword combinations can be identified ("common-sense-religion" for the English writing and "思想-民族-常识" <"thinking-nation-common sense">).

<p>Chinese prose "英国人与中国人" ("British and Chinese People")</p>	<p>聚焦在“思想”、“民族”、“常识”，用 2 句话总结以上的论述如下： 英国人的性格体现了他们对逻辑的轻蔑，以及重视思想和常识的和谐统一。他们的常识被赋予了道德上的力量从而为他们带来了成功。 To provide a two-sentence summary for the above text using the highlighted topic words "thinking", "nationality" and "consciousness": The character of English people shows their contempt for logic and their emphasis on the harmony of thought and common sense. Their common sense is empowered with moral power, which brings them success.</p>
<p>The section titled as "THE DOCTRINE OF THE GOLDEN MEAN" in the English book "MY COUNTRY AND MY PEOPLE"</p>	<p>To provide a two-sentence summary for the above text with the highlighted topic words "common sense" and "religion", we have: The religion of common sense is a central doctrine of Confucianism, emphasizing moderation in thought and conduct, and aiming to cultivate a reasonable individual as the model of culture. It is the appeal to human nature, expressed in the Doctrine of the Golden Mean, that has enabled the Chinese to develop a capacity for compromise and a dislike of mechanistic systems.</p>

Table 2: GPT-3 (based on text-Davinci-003 model) summarization results based on the original writing contents and extracted topical keywords listed in Figure 1 (the bolded text are prompts and non-bolded text are summarization generated by GPT-3). The summarization results show that the author had similar theme/motif on his writing in English and Chinese for a similar topic. The finding confirmed human evaluators' conclusion.

Reference

[1] 《英国人与中国人》，《林语堂散文集》，"British and Chinese People", *Collection of proses* by Lin Yutang, <https://www.rain8.com/wenzhang/9075.html> (accessed on January 10th, 2023).
 [2] Yutang, L. (2021). *My country and my people*.
 [3] Geling, Y. (2006). Banquet Bug.
 [4] 《严歌苓作品集》，*Collection of works by Yan Geling*, kanunu8.com/files/writer/8675.html (accessed on January 10th, 2023).
 [5] Infranodus text network analysis and visualization tool, infranodus.com (accessed on January 10th, 2023).
 [6] OpenAI GPT3 and ChatGPT tool, openai.com/api (accessed on January 10th, 2023).