

Where Have All the Paragraphs Gone?

Detecting and Exposing Censorship in Chinese Translation

Mizhang Streisand
GFW Report

Eric Wustrow
University of Colorado Boulder

Amir Houmansadr
University of Massachusetts, Amherst

Abstract

Translated literature often gets cut or altered before being published in China. Worse yet, both readers and the original literature authors are not informed of such removal or changes most of the time. Such erasures can change the fundamental meaning of texts, and readers are left unaware of vital context.

In this work, we propose an NLP-based approach to detecting translation censorship automatically and effectively. We further conduct a case study to analyze censorship in translated literature. We build a demonstration website and argue that an effective way to combat this form of censorship is to actively trigger the Streisand effect: by highlighting what text has been censored and making it easy to access, we hope the censored text reaches a broader audience, including those who would not have read the censored version in the first place.

1 Introduction

*“The past was erased, the erasure was forgotten,
the lie became the truth.”*

George Orwell, Nineteen Eighty-Four [14, chap. 7]

Like other media, China censors translated books and literature before they are published in China. This censorship is often invisible to readers as a deliberate choice by the censor to avoid triggering Streisand effect [4, 18]. That is, telling readers that a book is (partially) censored provokes their curiosity and interest, and the censored content is thus more likely to be sought out by more people [7–9]. While this type of quiet censorship could be caught by painstakingly translating the original text for comparisons, it is often too expensive to do so in practice [12, §NO ONE IS CHECKING].

Surprisingly, it is often the publisher (and not the government) who self-censors translated books, as failure to do so could result in various forms of punishment, including harsh scrutiny of future publications [12, §CENSORSHIP ON THE MAINLAND], confiscating the published books [20, §Books published and then banned], and immediate suspension of the

publishing rights [5, 21]. This type of self-censorship based on fear is effective, because without clear and exact rules, a publisher errs on the side of conservatively censoring anything that might be interpreted as sensitive, resulting in more restrictive censorship than if it were done by the censoring government themselves [4, 18]. In many cases, authors were not aware that the translation of their work had been cut until well after its publication [12, §IN THE DARK]. In other cases, authors were aware of the censorship, but had to make a choice between publishing a censored version [2, 11], or not publishing at all [13].

In this work, we propose a low-cost and effective way to *automatically detect* when a translated book or text has omitted or substantially changed from the original work. We then *present* the original text and its translation side-by-side on a dedicated website, with censored contents highlighted. By highlighting these erasures and making them easy to access for the general public, we frustrate censors’ attempts to suppress information and make its censorship backfire.

Our tool operates by performing an automatic translation of the translated text (e.g. in Chinese) back to the original language (e.g. English), and comparing the paragraphs using state-of-the-art natural language processing (NLP) tools. If a paragraph is either missing or is substantially different in meaning, our tool can alert the reader to the deletion or change. By using NLP, our tool can gracefully handle the subtle differences between automatic and manual translation that do not change the semantic meaning of paragraphs. Our hope is to ultimately apply this tool to many translated works; we report on our prototype tool and preliminary results here.

2 Methodology

Program to detect censorship in translation. The way we detect censorship in translation can be framed as a typical bitext word alignment task in the NLP domain, where we *align* two texts and look for discrepancies. In particular, our program first uses the Google translation API [6] to machine-translate the Chinese book back into English. We

then break the bixtext (the machine translation and the English version) into paragraphs to form a bipartite graph, where each paragraph is a vertex, and the weight of each edge is the similarity between the two paragraphs. To calculate the weight of each edge, we first embed each paragraph using the NLP model *distiluse-base-multilingual-cased-v1* [15] and assign the negation of the cosine similarity of two paragraphs as the weight [16]. We set a similarity threshold, and when two paragraphs’ similarity is below the threshold, we do not consider them as similar at all by assigning a positive number to the weight. Finally, we calculate the linear sum assignment of the bipartite [19]. The result tells which machine-translated paragraph matches which original paragraph.

Once the program computes which paragraph in the machine translation matches which paragraph in the original version, it applies the same approach recursively to determine, for each pair of matched paragraphs, which translated sentence matches which original sentence. Finally, the program marks unmatched sentences as possibly being cut, and sentences with a low similarity as possibly being altered.

Human analysis. As the first step, the program analysis provides a similarity score for each sentence. It allows us to quickly and effectively detect and identify potential censorship in translation. When the similarity is lower than a certain threshold, we manually confirm the missing text and identify potential meaning changes in context.

Limitations. The current approach does not consider the cases where two machine-translated sentences correspond to one original English sentence, or vice versa. In [Appendix B](#), we introduce this issue and a solution to it in detail. To reduce the potential bias introduced in human analysis, we will minimize the involvement of humans and measure the inter-rater reliability when human analysis is necessary.

3 Case Study

In this section, we conduct a case study to help readers better understand this form of censorship. In particular, we use Chapter 5 of *Chinese Literature: A Very Short Introduction* by Sabina Knight as a demonstration [10], comparing its original English version with its Chinese version translated by Li Yongyi side-by-side. We strongly encourage readers to visit the demonstration website for an interactive experience: <https://gfw.report/projects/translation-censorship>.

In [Table 1](#), we categorize the censored topics and count the censorship actions in the Chinese translation of Chapter 5 of *Chinese Literature: A Very Short Introduction*. In total, we identified seven censored topics, including 2000 Nobel Prize in Literature laureate Gao Xingjian and his work, the plasma economy in Henan province, China, the Tiananmen Square Massacre in 1989, and Mao Zedong and the political

Topics	# Paragraphs cut (altered)	# Sentences cut (altered)	# Words cut (altered)
Gao Xingjian	2	14	0
Plasma Economy	0	3	0
Mao Zedong	1	2	0(1)
Great Leap Forward	0	1	0
Cultural Revolution	1	6	1
Tiananmen Massacre	1	4	0
Book Censorship	0	1	0
Total Count	5	31	1(1)

Table 1: Censored topics and the count of censorship actions — We count the number of paragraphs, sentences, and words removed or altered in the Chinese translation of the Chapter 5 of *Chinese Literature: A Very Short Introduction*. When a paragraph is removed, we count it as both paragraph cuts and sentence cuts. When a sentence is removed, we do not count it towards word cuts though. When a sentence covers multiple topics, we count it toward all of them.

movements he led: the Great Leap Forward and the Cultural Revolution. Interestingly, as discussed previously, a discussion of the book censorship itself is cut, showing another effort by the censor to avoid triggering the Streisand effect.

4 Discussion

Alternative approaches. While our approach to detecting censorship in translation is based on state-of-art NLP models and is proved to be efficient and effective, it is by no mean the only approach. For example, an alternative approach we tried is bixtext mining. However, possibly due to our limited understanding and skills, we have not been able to use bixtext mining to yield a detection result that is as good as the current approach [3, 17]. [Appendix C](#) logs that ChatGPT identified a missing sentence in translation correctly and reasoned on it [1]. We note that having multiple independent approaches to detecting censorship in translation will help to improve the overall robustness of detection, as one can compare and evaluate each method’s output against the others’.

5 Conclusion

In this work, we explore censorship in translation. In particular, we propose an NLP-based approach to automatically and effectively detecting translation censorship. We further conduct a case study to analyze censorship in translated literature. In addition, we build a demonstration website and argue that an effective way to combat this form of censorship is to actively trigger the Streisand effect. That is, by highlighting the censored text and making it easy to access, we hope the censored text reach a broader audience, even including the curious who would not have read the censored version of the books in the first place.

Acknowledgments

The authors would like to express their sincere gratitude to Jeffrey Knockel and the anonymous reviewers for their insightful feedback and constructive comments. We also appreciate Xiao Qiang and Michael Carl Tschantz for their initial discussions on this work. We thank David Fifield for his guidance, comments, feedback, and support throughout the work.

This study was funded by the NSF grant 1953786, and by the Young Faculty Award program of the Defense Advanced Research Projects Agency (DARPA) under the grant DARPA-RA-21-03-09-YFA9-FP-003. The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

Availability

To maintain reproducibility and stimulate future work, we make our code and data public to the maximum extent that does not harm our anonymity. The code, data, and the project webpage are available at: <https://gfw.report/publications/foci23/en/>

References

- [1] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020. <https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf>. (Cited on pp. 2 and 5)
- [2] Liz Bury. Author bows to Chinese censorship of his Deng Xiaoping biography. *The Guardian*, October 2013. <https://www.theguardian.com/books/2013/oct/22/author-chinese-censorship-den-xiaoping-biography>. (Cited on p. 1)
- [3] Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. Language-agnostic bert sentence embedding, 2020. <https://aclanthology.org/2022.acl-long.62.pdf>. (Cited on p. 2)
- [4] David Fifield. Reading group: Censored: Distraction and Diversion Inside China’s Great Firewall, August 2022. <https://github.com/net4people/bbs/issues/122>. (Cited on p. 1)
- [5] Nam fung Chang. Censorship in translation and translation studies in present-day China. In *Translation and censorship in different times and landscapes*, pages 229–240. Cambridge Scholars, 2008. <https://www.semanticscholar.org/paper/Censorship-in-translation-and-translation-studies-Chang/ac8ea1d3e29b61141141fd45e9b308fe0a3ec64f>. (Cited on p. 1)
- [6] Suhun Han. Googletrans. <https://pypi.org/project/googletrans/>. (Cited on p. 1)
- [7] William R. Hobbs and Margaret E. Roberts. How sudden censorship can increase access to information. *American Political Science Review*, 112(3):621–636, 2018. <http://www.margaretroberts.net/wp-content/uploads/2016/08/selfiecensorship.pdf>. (Cited on p. 1)
- [8] Sue Curry Jansen and Brian Martin. Making censorship backfire. 2003. <https://ro.uow.edu.au/cgi/viewcontent.cgi?article=1054&context=artspapers>. (Cited on p. 1)
- [9] Sue Curry Jansen and Brian Martin. The Streisand effect and censorship backfire. 2015. <https://ro.uow.edu.au/cgi/viewcontent.cgi?article=2890&context=lhapers>. (Cited on p. 1)
- [10] Sabina Knight. *Chinese Literature: A Very Short Introduction*. Oxford University Press, February 2012. <https://doi.org/10.1093/actrade/9780195392067.001.0001>. (Cited on p. 2)
- [11] Lily Kuo. Edward Snowden says autobiography has been censored in China. *The Guardian*, November 2019. <https://www.theguardian.com/us-news/2019/nov/12/edward-snowden-says-autobiography-permanent-record-censored-in-china>. (Cited on p. 1)
- [12] Alexa Olesen, Katy Glenn Bass, Suzanne Pettypiece, et al. Censorship and conscience: Foreign authors and the challenge of Chinese censorship, May 2015. <https://pen.org/research-resources/censorship-and-conscience>. (Cited on p. 1)
- [13] Natalie Ornell. Evan Osnos on not publishing his book in China. *China Digital Times*, May 2014. <https://chinadigitaltimes.net/2014/05/evan-osnos-publishing-book-china/>. (Cited on p. 1)
- [14] George Orwell. *Nineteen Eighty-Four*. Project Gutenberg Australia. <https://gutenberg.net.au/ebooks01/0100021.txt>. (Cited on p. 1)

- [15] Nils Reimers. Pretrained models. https://www.sbert.net/docs/pretrained_models.html. (Cited on p. 2)
- [16] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 11 2019. <https://arxiv.org/pdf/1908.10084.pdf>. (Cited on p. 2)
- [17] Nils Reimers and Iryna Gurevych. Making monolingual sentence embeddings multilingual using knowledge distillation. *arXiv preprint arXiv:2004.09813*, 2020. <https://arxiv.org/pdf/2004.09813.pdf>. (Cited on p. 2)
- [18] Margaret E. Roberts. *Censored*. Princeton University Press, 2018. <https://press.princeton.edu/books/hardcover/9780691178868/censored>. (Cited on p. 1)
- [19] scipy. Scipy: linear sum assignment function. https://docs.scipy.org/doc/scipy-0.18.1/reference/generated/scipy.optimize.linear_sum_assignment.html. (Cited on p. 2)
- [20] Yuwu Song. Book censorship in Post-Tiananmen China (1989-2019). *Journal of East Asian Libraries*, 2022(175):4, 2022. <https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=2800&context=jeal>. (Cited on p. 1)
- [21] Zaixi Tan. Censorship in translation: The case of the People’s Republic of China. *Neohelicon*, 42(1):313–339, 2015. <https://link.springer.com/article/10.1007/s11059-013-0231-8>. (Cited on p. 1)

both traditional Chinese theater (with its interplay of sound and drama) and French theater of the absurd, Gao's characters sometimes talk for the sake of talking. But their dialogues also express deep longings and trenchant social criticisms. For Glasses, who misses his final chance to take the College Entrance Exam, the waiting becomes unbearable: "Life has left us behind. The world has forgotten us. A lifespan passes by in vain right before your eyes." "Waiting's not so bad," the master carpenter counters later, when the actors break character to deliver overlapping lines. "People wait because they have hope."

[0003]

A political allegory of China's passage from the countryside to the city, Gao's play suggests five themes central to the modernization and globalization of Chinese literature.

Though many critics interpret the play as signaling the government's failure to deliver the means of progress, others hear in it ongoing commitments to national pride, humanism, progress, memory, and pleasure. Although such positive aspirations have often been downplayed by critics of Gao's 2000 Nobel Prize in Literature, their resonance for the Nobel Committee and for his other admirers suggests these themes' centrality in contemporary world literature.

的特点，他们有时仅仅是为了说话而说话。但是他们的对话也表达了深沉的渴望和对社会的尖锐批评。“戴眼镜的”错过了最后一次高考机会，对他来说这样的等待已经无法忍受：“我们被生活甩了，世界把我们都忘了，生命就从你面前白白流走了。”后来，当演员跳出虚构的角色，同时说着各自的台词时，“扮师傅的演员丙”反驳道，“等不要紧。人等是因为人总有个盼头。”

高行健的剧作是一则寓言，解读了中国从乡村进入城市的变化，隐含着对中国文学的现代化和全球化至为关键的五个主题：对民族自豪感、人文主义、进步、记忆和快乐的追寻。尽管许多评论家将这部剧解释为政府未能提供进步手段的信号，但其他人则从剧中听到了对民族自豪感、人文主义、进步、记忆和快乐的持续承诺。尽管高 2000 年诺贝尔文学奖的批评者经常淡化这种积极的愿望，但他们对诺贝尔委员会和他的其他仰慕者的共鸣表明这些主题在当代世界文学中的中心地位。

Figure 1: **A screenshot of the demonstration website** — We use Chapter 5 of Chinese Literature: A Very Short Introduction by Sabina Knight as a demonstration, comparing its English version and its Chinese version translated by Li Yongyi side-by-side. In particular, we highlighted the English text whose Chinese translation got cut in yellow. We also added our Chinese translation of the cut text and highlighted it in yellow. Text whose meaning changed in the Chinese version is highlighted in purple. We also provide a button, that once clicked, shows a translation of the deleted text. The website is available at: <https://gfw.report/projects/translation-censorship>.

A The demonstration website

Figure 1 is a screenshot of the demonstration website. We encourage readers to visit the website for an interactive experience: <https://gfw.report/projects/translation-censorship>.

B An example of the one-on-one mapping issue

The current detection program does not consider cases where two machine-translated sentences correspond to one original English sentence, or vice versa. For example, when translating the sentence “[w]ear whatever you want, don’t wait until you are my age”, the translator may break it into two sentences “[w]ear whatever you want” and “[d]on’t wait until you are my age”. While no sentence is cut or altered, the current program will only match “[w]ear whatever you want” with the original text, and incorrectly mark “[d]on’t wait until you are my age” as missing.

To address this problem, one can improve the program by heuristically merging a missing sentence with its context and check if the similarity score goes up. If so, it may consider the sentence and its context together correspond to one sentence. In this example, the program would have first merged “[d]on’t wait until you are my age” with “[w]ear whatever you want”, and then check if the similarity score goes up.

C A conversation with ChatGPT

In this section, we log a conversation with ChatGPT [1], which suggests a promising alternative approach to detecting censorship in translation in future work.

As shown in Table 2, in the first session, we provided the chatbot with an original text and its partially censored translation. The chatbot identified the missing sentence correctly and reasoned on it.

Authors: Compare the following text and its Chinese translation. Is there any sentence missing in the translation? If so, make a 500 word case study on this.

Origin: “Though many critics interpret the play as signaling the government’s failure to deliver the means of progress, others hear in it ongoing commitments to national pride, humanism, progress, memory, and pleasure. Although such positive aspirations have often been downplayed by critics of Gao’s 2000 Nobel Prize in Literature, their resonance for the Nobel Committee and for his other admirers suggests these themes’ centrality in contemporary world literature.”

Translation: “高行健的剧作是一则寓言，解读了中国从乡村进入城市的变化，隐含着对中国文学的现代化和全球化至为关键的五个主题：对民族自豪感、人文主义、进步、记忆和快乐的追寻。”

ChatGPT: It appears that the Chinese translation of the text is missing the sentences that refer to the interpretation of Gao’s play by different critics and the resonance of positive aspirations with the Nobel Committee and other admirers. These sentences are present in the original text but do not appear in the Chinese translation.

One potential explanation for the missing sentences in the Chinese translation is censorship. It is possible that the translators or publishers of the text decided to omit the sentences in order to avoid controversy or to comply with Chinese censorship laws and regulations. In China, the government and the ruling Communist Party have a long history of censoring books, media, and other forms of expression that they consider to be a threat to social stability, national security, or the party’s authority. This censorship can take many forms, including the outright banning of certain books and media, the removal of certain content from websites and social media platforms, and the suppression of certain topics or viewpoints in the media.

Another possible explanation for the missing sentences in the Chinese translation is the desire to make the text more concise or to better align with the intended audience or purpose of the translation. Translators often have to make decisions about what to include or exclude from a translation in order to convey the main points of the original text in a clear and effective way. In this case, it is possible that the translators felt that the missing sentences were not essential to the overall message of the text and therefore chose to omit them.

Regardless of the reasons for the missing sentences in the Chinese translation, their absence has significant implications for the way in which the text is understood and interpreted. The original text presents a nuanced and balanced view of Gao’s play, with some critics interpreting it as a criticism of the government and others seeing it as a celebration of national pride and other positive values. By omitting the sentences that mention these different interpretations, the Chinese translation presents a more one-dimensional view of the play and downplays the complexity and diversity of critical perspectives on it.

This case study highlights the complexities and challenges of translating texts in a censored environment, where translators must navigate competing demands of accuracy, clarity, and compliance with censorship laws and regulations. It also underscores the importance of freedom of expression and the free exchange of ideas in fostering a rich and diverse cultural landscape.

Table 2: **The first conversation session with ChatGPT** — We provided the chatbot with an original text and its partially censored translation. The chatbot identified the missing sentence correctly and reasoned on it.

Authors:	Compare the following text and its Chinese translation. Is there any sentence missing in the translation? If so, make a 500 word case study on this. Origin: “Though many critics interpret the play as signaling the government’s failure to deliver the means of progress, others hear in it ongoing commitments to national pride, humanism, progress, memory, and pleasure. Although such positive aspirations have often been downplayed by critics of Gao’s 2000 Nobel Prize in Literature, their resonance for the Nobel Committee and for his other admirers suggests these themes’ centrality in contemporary world literature.” Translation: “高行健的剧作是一则寓言，解读了中国从乡村进入城市的变化，隐含着对中国文学的现代化和全球化至为关键的五个主题：对民族自豪感、人文主义、进步、记忆和快乐的追寻。尽管许多评论家将这部剧解释为政府未能提供进步手段的信号，但其他人则从剧中听到了对民族自豪感、人文主义、进步、记忆和快乐的持续承诺。尽管高2000年诺贝尔文学奖的批评者经常淡化这种积极的愿望，但他们对诺贝尔委员会和他的其他仰慕者的共鸣表明这些主题在当代世界文学中的中心地位。”
ChatGPT:	There does not appear to be any sentence missing in the Chinese translation. The Chinese translation appears to be a faithful and complete rendering of the original text. It covers all of the main points and themes mentioned in the original text, including the various interpretations of the play, the themes of national pride, humanism, progress, memory, and pleasure, and the reception of Gao’s Nobel Prize in Literature by critics and admirers. In order to provide a more detailed analysis, it would be helpful to have more context about the play in question, as well as information about Gao’s Nobel Prize and the specific criticisms that have been leveled against it. However, based on the information provided, it seems that the Chinese translation is a thorough and accurate representation of the original text.

Table 3: **The second conversation session with ChatGPT** — We provided the chatbot with the same original text and completed the translation ourselves. The chatbot correctly stated that the translation was complete.

Table 3 shows the second conversation session with ChatGPT. As a control group, we provided the chatbot with the same original text and completed the translation ourselves. This time, the chatbot correctly stated that the translation was complete.

D Ethics

This work does not involve human subjects.