

Final Project Report 1

Why the Project? (why these files & why relevant/helpful for my own research)

My corpus-building project takes a range of Japanese *ero guro* literature (literally, erotic grotesque, an artistic genre mainly emerged in 1920-30 which focuses on depicting eroticism, the supernatural, sexual corruption, and body decay) to study how they share adaptations from medieval *monogatari-shu* (Japanese term for tales) and how they are in common in creating this sense of simultaneous erotic and grotesque. In light of this, my chosen files consist of twelve pieces of 19th-and-20th-century fictions and tales, with two famous compilations of medieval tales which have been widely adapted and borrowed.¹ Some of the borrowings, for example those in Lafcadio Hearn's *Kaidan*, are fairly clear-cut, presenting explicitly as rendition of the original stories, whereas others could be implicit and intangible, and can only be noticed by paying attention to the use of certain words. I expect my project to shed some light on this.

The focus of the project is part of my broader research interest in the “strange” in the context of Japanese Buddhism. Being an umbrella term here, the strange concerns with wonder, marvel, miracle, and whatever anomalies departing from ordinary life. In this sense, *ero guro* is apparently closely related. Despite being a relatively modern concept that only appeared in popular literary and visual cultures, which is miles away from the “traditional” Buddhism, it is important to note that many of its prototypes, namely the medieval tales, uphold a Buddhist overtone, or have been Buddhist moral tales *per se*. Making use of Buddhist elements, I assume, is such an adaptation in modern Japanese stories of the strange. Studying *euro guro* as a segment of my broader research interest, as I noted above, is helpful to explore how these works perform this genre respectively, their intertextuality to one other, and shared adaptation from medieval texts, all of which will contribute to a more subtle understanding of its thematic evolutions and socio-religious implications². Together with the nature of the project leveraging digital tools, this research will be able to be conducted in a more efficient, manageable, and intuitive manner, thereby revealing something that might otherwise remain obscure.

How was TEI Made?

The conversion of TEI actually started when choosing the original files to process. By this I mean the quality of files selected will largely impact the experience of processing files thereafter and, in this case, low-quality files would add complexities and obstacles, which I indeed encountered. Starting with EPUB files, I originally chose seventeen files in total, with four of them later being eliminated due to the obscurity in their respective TXT generated. For example, chapters of one eliminated file consist of four lines, including one line with four asterisks, one empty line, one chapter number line, and one chapter name line. Since I still had a choice, I thought it would be more efficient to eliminate or replace it rather than sticking here for a long

¹ Note that this disparity in their respective amounts stems from the nature of the medieval tales, that is, they are mostly fairly short, and two compilations can contain over one hundred tales.

² The latter part, to understand its socio-religious implications, is outside the scope of the project, as it requires literary analyses.

time. The rest fourteen files were, lamentably, processed separately to TEI, for they were all from different sources and none of them are in identical pattern. Yet, they do share a degree of similarities, sometimes in identifying chapter titles while others in identifying subsections, which liberated me from repetitive tasks and just tweaking them a little based on one or two prepared scripts. In other words, this process is more of a matter of permutation and combination. One difficulty I've encountered was matching those text-only titles, which sometimes included non-English characters. I ended up writing a list of title strings and to match the list. Another obstacle lies in the file *Seven Japanese Tales*, where headlines are not only text-only but also have appeared multiple times. My first attempt was to write a line counter, which didn't work well (as mismatching still happened, or the first headline being eliminated), but was finally resolved by `re.compile()`.

However, there are also advantages in processing them separately. For instance, each of them are independent and you can run the script freely without worrying about previously processed TEI being overwritten. I have also left the processed TEI unorganized (i.e., not collecting them in a sub-directory) intentionally, as it allowed me to switch from different files (e.g., to compare one's TXT and TEI) and to quickly navigate if there are leftovers (for my processing wasn't completed at a time). [Update: I have organized them at the end.](#)

The work was overall smooth, yet sometimes might not be done in the “smartest” way and, I assume, might be only applicable to this kind of small workload. Further deliberations in optimizing the code will be essential if performing larger-scale conversion.