Title
Sifting texts: What can a digital tool tell us about late imperial literacy?

Permalink
https://escholarship.org/uc/item/3jr3z5rt

Author
SCHNEEWIND, Sarah

Publication Date
2020-03-23

Peer reviewed
Sifting texts: What can a digital tool tell us about late imperial literacy?


This talk was prepared for the AAS 2020 in Boston. Please cite it properly if you use it.

Introduction

The Late Imperial Primer Literacy Sieve was designed, by Joshua Day with assistance from historian Jenny Huangfu Day, to answer one research question that I had as I worked on shrines to living men in Ming times. But as it turned out, it showed some other things, too. Other scholars have used it for other research questions. This paper will suggest some more possible uses, although I myself do not plan to pursue them.

I was studying a particular genre of text: essays commemorating shrines to living men. I mostly found the commemorative records in local gazetteers and in Ming men’s collected works, but they were also engraved on stones and set up at the shrines. I thought that they sent a particular message. The question I wanted the Late Imperial Primer Literacy Sieve to answer was whether this message would come across to people who could not read the full record, because they did not have a full classical education. So the research question was, “What would a partially-literate Ming person understand from a stone inscription commemorating a premortem shrine?”

Most people in Ming times probably started their education, and many ended it, with one of a few standard primers. So, the idea of the Sieve was simply to run one or more of the standard primers against the texts I was interested in. It would sift out any characters in the stele text that did not occur in the primer, and I would read the result. That was the basic idea. The designer, Joshua Day, designed it with three elements: primers, target texts, and depleted texts.

In the lead-up to the panel, Monica Klasing Chen asked whether I had read Wu Huifang 吳蕙芳’s work and thought about using 雜字 booklets as primers. Wu considers them to be more 'popular,' and the 千字文, baijiaxing, etc. to have been mainly used by the elite, or at a later stage in education. This would be something that someone could pursue. New primers can be built into the Sieve. The scholar would have to choose which Zazishu was most widespread or applicable to her research.
It’s actually quite hard to prove what books were used in elementary education; I know because I tried. But even the teachers Li Yu discusses in Qing who were innovating in pedagogy still relied mainly on these standard primers.\(^1\) I noticed that Wu’s claim that these books were widely used rests in part on the many editions she found, which is good evidence, but in part on local gazetteers from the 1990s.\(^2\) When you search the Ming gazetteers in the Erudite data base only one of them mentions one of the texts that Wu talked about.

2. The second element of the Primer Literacy Sieve is the “target texts” -- those I was interested in, or those another research wants to study.

3. The third element is the output, called “depleted texts”: target texts with the characters not found in primers removed entirely or marked in various ways.

For a demonstration or to try out the Sieve, please see


\(^2\) 吳蕙芳，明清以來民間生活知識的建構與傳遞 *Taipei Shi* : *Taiwan xue sheng shu ju*, 2007, pp. 213 ff for list of gazetteer mentions.

Schneewind, “Sifting Texts,” 3
But this is one way depleted text can appear: the red characters on a black background are not found in the selected primer.

There are all kinds of problems with “reading” the depleted text, but the Journal of Chinese History has just published my discussion of that so I’m not going to talk about that today. Today I’m going to discuss some things that others might want to do with the Sieve.

In designing the Sieve, Josh had three priorities. He followed the guidelines developed by Peter H. Salus.

Schneewind, “Sifting Texts,” 4
The principles are that programs should each do one thing well, but be able to work with other programs, and be able to handle text. In the case of the Sieve, what this meant to Josh was that: First, the software had to be fast enough to handle a lot of texts. Second, the results had to be something that a historian could interpret – not just statistics but something texty. Third, the results had to be explorable. That is, since it was not clear at the start what results would be most useful, the historian had to be able to use the software to make discoveries.

I got my question answered, but Josh’s design meant that other things turned up, too. For instance, it turned out that the amount of one of the target texts that was legible was pretty consistent across the 17 target texts we ran.

How much of a premortem stele could one read?
Roughly, if one had memorized
Heart Sutra: about one-tenth
Hundred Surnames: about one-fifth.
Guanyin Sutra: just under one-third.
Classic of Filial Piety: just over one-third.
Trimestrical Classic: about 40%.
Thousand Character Essay: a bit over half.
Elementary Learning: 80% or more.

This flexibility means that we can ask other questions. When I went to a workshop at the University of Chicago run by Haun Saussy and Jeff Tharsen recently, the audience did raise other, interesting questions.

Schneewind, “Sifting Texts,” 5
Other Questions the Sieve Might Answer

Are the levels of legibility, and the consistency of those levels, a matter of genre? Aside from the two sutras, the other primers have a fairly Confucian outlook, and so do the premortem steles. If the target texts were short stories or Daoist prayers, or legal briefs, or a chunk of the Ming Code, would the levels of legibility be different, and would they vary more? You could run the Guanyin sutra against another sutra, or for instance, the biography of a Buddhist nun, and see if you get better than the 30% it makes legible of premortem steles.

Another question: How much did the primers overlap? That is, if you had learned one, how many words in another would you already know?

This question was raised at the Chicago workshop by Arnd Helmut Hafner, and programmer Leonora Tindall answered it in the Sieve within an hour as the workshop proceeded:
If you knew that Xiaojing, you could read between 16 and 36% of the other primers, and the most useful one for reading the others would be the Qianziwen, which would let you read over half of each of the others.

Another way to ask this question would be to reverse it. How many more characters would you learn by learning a second primer? That would give us some sense of what families were investing in when they kept a kid in school.

Another kind of question:

**Apply this knowledge:**
- If a primer was more common in a certain area or time,
- estimate local legibility of texts
- Or estimate temporal legibility of texts

If you knew that a certain primer was more commonly used in a certain place or time – which I must say I think would be very difficult indeed to know – you could estimate how legible other texts would be in that particular social context.

A more interesting question is whether the elite authors of public texts purposely and effectively made them more legible by sticking to the vocabulary of the primers. For a given
author, you could run some elite-only texts, such as his letters to family members or friends, or social poetry, and get a legibility percentage for those. Then you could run something from another genre, such as a public placard or commemorative record, and see if the legibility percentage was higher. Suppose you wanted to know whether Ming officials hoped, when writing certain kinds of memorials, to reach a wider audience? This kind of comparison would be one way to gauge that.

Dan Knorr asked: did officials posted to border areas or poorer, less-educated areas write commemorative essays or other public texts using a smaller vocabulary?

More questions

- Did the elite write more simply when they were posted to “backward” areas?
- Did writers design some texts for both the primer literate and the elite?
- Did vocabulary change within one genre?

Another possible question: how did vocabulary change within one genre over time? For instance, there were premortem shrines in Tang. Was their vocabulary more high-falutin, were they less legible, than in the very different society of Ming?

Improving the Sieve

Hilde de Weerdt has already improved the Sieve by making an on-line version. Other developments are certainly possible. For instance:

The Future

- Use publicly available datasets to figure out “how readable” particular unknown characters are, on the basis of their constituent radicals.
- Take account of more research on what texts teachers used in particular place times.
- Run lots of Ming texts to see how many would have been readable with primer literacy.
Some people have asked about whether readers could use grammatical predictability to increase bottom-up comprehension. I did not consider this in my original analysis, but there are some tools that might be combined with the Sieve to make that possible to see:


http://godel.iis.sinica.edu.tw/CKIP/engversion/onlinesystem.htm KIP Chinese Knowledge and Information Processing

Even without this grammatical sophistication, others have used the tool for other things. Paola Zampierini uses it in her literature classes so students can see how legible vernacular fiction was in Ming times, and one other scholar’s use appears in my article. I’ll conclude with one more way in which I think the tool could be used.

One More Use for the Sieve

Angela Ki Che Leung has written about “Medical Instruction and Popularization in Ming-Qing China.” She studies a number of Ming-Qing medical primers. These medical primers refer to medical classics, but use simpler language than those classics and claimed to be for beginners, but adult beginners. The Sieve offers a way to gauge whether students who had studied a regular primer would do have an easier time learning the medical primers.

But Leung also argues that the language of medical primers became increasingly simple over time, so that those of the late Qing are easier to understand (136). The Sieve, if appropriately modified, could systematize that increasing easiness: one could run the medical primers against the more scholarly works they were based on, and test whether the vocabulary required diminished.

Third, Leung discusses the popularization of medical knowledge in encyclopedias and other sources; the Sieve, again, by running medical primers or classics against a popular medical manual, could pinpoint differences in vocabulary and thus offer a more systematic overview of whether knowledge for professionals differed from that designed for popular use. How many bodies of medical knowledge were there? The Sieve offers one way to begin answering that question, and this method could extend to other fields to suggest how elite and popular knowledge converged or diverged.

I hope that many of you will pick up the tool and adapt it to your own needs. If you find it useful, please do let me know, and please cite it in your work. Many thanks to Monica Klasing Chen for including me in the panel.

---

3 Chris Atwood commented on these: “My sense is that classical Chinese is so genre- and context-specific that if you wanted machine translation, you'd have to have several versions: Buddhist Classical Chinese, medical classical Chinese, philosophical classical Chinese, historical classical Chinese, and yes, exam essay classical Chinese.” Facebook, Sinologists Nov 7 2019. Paul Vierthaler’s work on genre may be relevant.

4 Angela Ki Che Leung, “Medical Instruction and Popularization in Ming-Qing China,” Late Imperial China 24.1 (2003): 130-152., 131, 145.
Where is the Sieve?

- knit.ucsd.edu/minghistoryinenglish/late-imperial-primer-literacy-sieve/
- Demonstration: paleoludic.com/sieve/ming-literacy-demo.html
- “The Sieve Online,” dh.chinese-empires.eu/markus/beta/sieveOnline.html