Derick Olson
18 December 2015
V. H. Mair

Bypassing Censorship: The Use of Term Morphs on Weibo

A study on the censorship of "Grass Mud Horse Lexicon" terms on China's largest microblog, Sina Weibo. This paper surveys methods used by studies of Chinese internet censorship and frames its question in terms of "morphs", or alternate representations of controversial terms and ideas.

Background

Sina Weibo is the primary microblogging service in China with 500 million registered users, and over 46 million daily active users[1]. In the past decade, microblogging services such as Twitter have risen in prominence in the realm of social media, as well as that of politics, news, and finance. In the United States, companies such as Dataminr and Datasift "mine" twitter for live-streaming data about events around the world[2]. Short messages and additional contextual information associated with a post, or "tweet", have made microblogs like Twitter important corpus of structured text in academic research[3]. As a similar service, Sina Weibo is another useful resource for the above applications. However, Sina's cooperation with China's government censorship policies make the Weibo corpus a uniquely positioned proxy to understand the censorship policies of China.

As with all websites and public platforms in China, social media platforms are subject to government censorship. The Chinese government sometimes involves itself in pursuing particular posters, such as the recent case of Pu Zhiqiang's trial for his posts on Weibo[4], and the

[1]How many people really use Sina Weibo?
http://blogs.wsj.com/chinarealtime/2013/03/12/how-many-people-really-use-sina-weibo/
[2]Twitter is Selling Access to Your Tweets for Millions
http://business.time.com/2013/10/08/twitter-is-selling-access-to-your-tweets-for-millions/
[3] Twitter Data Analytics. http://tweettracker.fulton.asu.edu/tda/
[4]Chinese rights lawyer Pu Zhiqiang stands trial for Weibo posts criticizing government.
http://mashable.com/2015/12/14/pu-zhiqiang-trial/?utm_campaign=Mash-Prod-RSS-Feedburner-All-Partial&utm_cid=Mash-Prod-RSS-Feedburner-All-Partial#F7CZK27wyEqV

controversial imprisonment of Ai Wei Wei in 2011[5]. In the case of Weibo, censorship consists of entire posts being removed from the timeline of searchable posts. Post and account blocking are handled by Sina, prior to government intervention. Estimates of censored Weibo posts range from under 1% to 16% of all posts.

The motivations for censorship can be grouped into two primary categories: (1) to suppress criticism of the state, and (2) to reduce the probability of collective action. According to an empirical study[6] by King et. al, China's censorship policies are almost entirely focussed on suppressing the latter case of collective action potential, rather than state critique.

There are multiple methods of internet censorship in China. These can be grouped into three categories: site-wide censorship (the Great Firewall), automatic blocking (based on topics and keywords), and manual blocking (human censors). Examples of the first case are the ban on sites like Google, Facebook, and Twitter from the Chinese internet due to these companies' refusal to comply with China's censorship policies. The latter two cases exist in most companies that do comply with China's censorship policies. For example, Sina has blacklisted terms which are blocked automatically. The advantages of such an approach is speed and scalability, because this form of censorship can be performed by computers. Manual blocking, on the other hand, requires a human to read and respond to a particular post, or to add a term to an automatically-blocked blacklist. Jason Ng's post on Citizenlab.org goes into further detail on the actual flow human-computer post filtering[7]. Keyword-based censorship has led to the introduction of alternative words and phrases known as *morphs*, which have entered the Chinese internet language.[8]

Morphs, defined as an alternative form of a preexisting, original word or phrase, can be categorized as one of the following: homophone, homograph, acronym, pun, and pinyin. Homophonic *morphs* include 河蟹 (he xie) "river crab" as a *morph* for the sarcastic use of 和谐 the word "harmony" in reference to the campaign for China as a "harmonious society". A

[5] https://en.wikipedia.org/wiki/Ai_Weiwei
[6] King et. al. How Censorship in China Allows Government Criticism but Silences Collective Expression.
[7] Tracing the Path of a Censored Weibo Post https://citizenlab.org/2014/11/tracing-path-censored-weibo-post-compiling-keywords-trigger-automatic-review/
[8] Chen et. al. Tweeting under Pressure.

homographic *morph* is 目田 (mu tian), which looks like 自由 (zi you), meaning "freedom". Acronyms generally use the pinyin romanization to represent a Chinese phrase, such as TMD for "ta ma de", meaning "fuck your mother". A example of pun or metaphor is 西朝鲜 (xi chao xian) which means "West Korea", and is an alternative way of saying China[9]. Finally, the pinyin representation of a Chinese word acts as an easily-accessible *morph*, as the use or absence of accent tones change the underlying representation, allowing such terms to bypass automated censorship methods.[10]

Methodology

This study addresses pinyin as a *morph* that can be used to bypass censorship. It uses the the China Digital Times' "Grass Mud Horse Lexicon"[11] as a source of potentially sensitive terms and phrases, and predicts the term's censorship status.

There have been several approaches to researching censorship on the Chinese internet.

Chen's research on pinyin acronyms used a questionnaire-based approach to determine which acronyms were most popular among a sample population of Mainland China Mandarin speakers. This approach is yields high quality data in small amounts, and is particularly useful for seeding terms that are as yet unknown. Since I am using an existing corpus of terms for my study, I do not take this approach.

A second approach to Chinese internet censorship studies is long-term monitoring of websites to detect when certain terms are censored. For example, in the King et. al. study on the effects of censorship on collective expression, researchers crawled the Chinese internet to find a myriad of blogs and websites, sampling for terms at varying levels of controversy over time. In the Chen et. al. study on censorship on Weibo, researchers determined a set of the most influential posters on Weibo, and followed the lifecycle of their posts and comments over time. Both of these approaches are very effective for determining when terms are censored, and have

[9] Ways to say China. http://languagelog.ldc.upenn.edu/nll/?p=22520
[10] Pinyin spam text message. http://languagelog.ldc.upenn.edu/nll/?p=21744
[11] http://chinadigitaltimes.net/space/The_Grass-Mud_Horse_Lexicon

shown the dynamic nature of Chinese keyword-based censorship. Due to time constraints, I elected not to take this approach, as these studies were conducted over many months.

Lastly, we come to Jason Ng's query-based approach used to select terms for his book *Blocked on Weibo*. His particular approach crawled Wikipedia for Chinese article titles, and tested whether or not they were blocked on Weibo. This approach is targeted in that it pre-selects queries, rather than ingest the live firehose of data that is Weibo. It's primary advantage is that it does not require many months of monitoring in order to return interesting results. It's primary disadvantage is that queries can become "stale", or no longer relevant, and this approach does not account for changes due to current events.

I chose a similar query-based approach, as it balances the advantage of computer-aided research to ingest large amounts of data, but stays within the resources and timeline available to my study. One disadvantage of this approach is that it is limited to pre-selected terms. However, my particular study focusses on known *morphs*, and the extent to which their pinyin representations are censored.

There were three main stages to my implementation: (1) Collection, (2) Querying, (3) Response handling.

In the collection stage, I parsed the entire "Grass Mud Horse Lexicon" (found at http://chinadigitaltimes.net/space/Grass-Mud_Horse_Lexicon:_Browse_by_Pinyin). For each entry, I stored the script-based term and pinyin representation (along with the English meaning for reference). After filtering for certain duplicates undocumented terms, I was left with 336 (script, pinyin) pairs. Each entry is tagged with its *morph*-type:

Chinese script:	'og'
Pinyin script:	'py'

Now that we have a pairing, I can state my hypothesis:

Given a blocked Chinese character/word, the pinyin representation will sometimes be used as a morph in order to bypass keyword-based censorship methods.

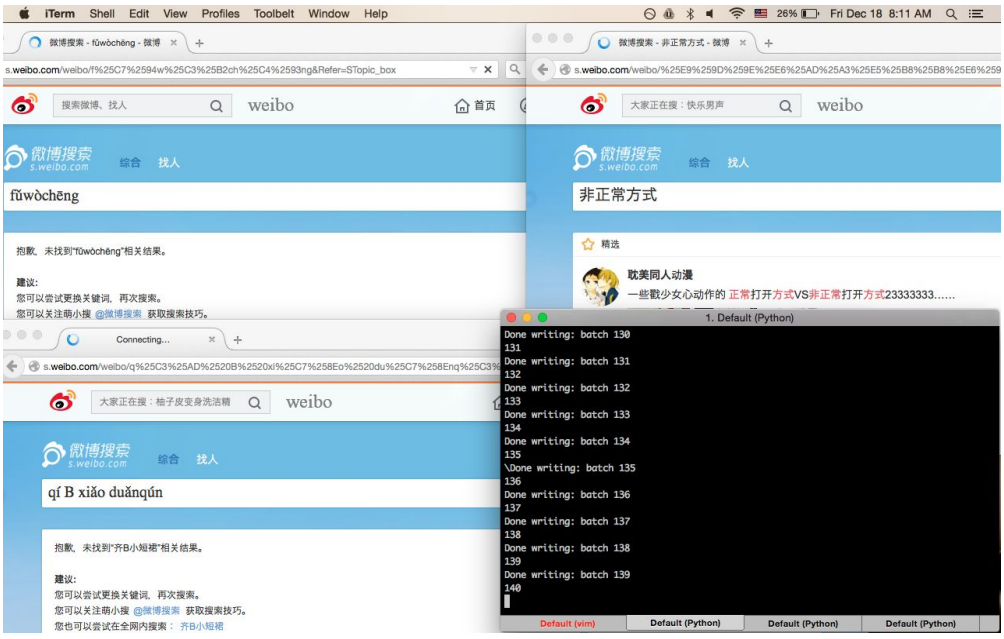
The reasoning behind this hypothesis is that in the case that the automated censors have discovered a new *morph* (the script term) for some controversial idea, the netizen will have continued using the same conceptual *morph*, but represented in an alternative way such as pinyin.

In the querying stage, I used a Selenium[12] browser driver, along with my own scripts to query Sina Weibo's search engine (found at http://s.weibo.com/weibo) for each term of each pair.

In the response handling stage, I parsed the html responses to my queries to determine which the response type to assign to each *morph*. This study designates four response types:

BLOCKED:	message stating query was censored
NO_RESULTS:	message stating query returned no results
FEW_RESULTS:	less than 5 results
MANY_RESULTS:	more than 5 results

In addition, there is a fifth ERROR state in which timeouts and other web-crawling-related issues prevented the script from determining the appropriate response. After response-handling, I saved the results to a .CSV file (see appendix).



Screen shot of graphical interfaces for automated browser querying

[12] http://docs.seleniumhq.org/

Analysis

My hypothesis that certain blocked terms would be presented with pinyin was not supported by the data. In fact, only one term containing pinyin was blocked across the entire dataset: *GFW zhī fù,* which refers to the "Father of the Great Fire Wall", or China's nationwide internet censorship policy. It is possible that some of the errors could have resolved in terms that were either blocked or unblocked in such a way that pinyin *morph* was used in place of a blocked character term, but this is not particularly likely[13]. Below is a sample of the overall results (see the link in Appendix for full results).

Base Term	Meaning	Morph	Type	Result
不差钱	no shortage of money	不差钱	og	MANY_RESULTS
不差钱	no shortage of money	bù chā qián	py	NO_RESULTS
不折腾	free from turmoil	不折腾	og	MANY_RESULTS
不折腾	free from turmoil	bù zhēténg	py	NO_RESULTS
不明真相	don't understand the actual situation	不明真相	og	MANY_RESULTS
不明真相	don't understand the actual situation	bù míng zhēnxiàng	py	NO_RESULTS
不要乱说话	do not make irresponsible remarks	bú yào luàn shuōhuà	py	ERROR
不要乱说话	do not make irresponsible remarks	不要乱说话	og	MANY_RESULTS
专业孙子	professional grandchild	专业孙子	og	ERROR
专业孙子	professional grandchild	zhuānyè sūnzi	py	NO_RESULTS
中国人是需要管的	Chinese people need to be controlled	中国人是需要管的	og	MANY_RESULTS
中国人是需要管的	Chinese people need to be controlled	Zhōngguórén shì xūyào gu ǎ n de	py	NO_RESULTS
中国人民的老朋友	old friends of the Chinese people	中国人民的老朋友	og	MANY_RESULTS
中国人民的老朋友	old friends of the Chinese people	Zhōngguó rénmín de l ǎ o péngy ǒ u	py	NO_RESULTS

[13] There were about 100 errors in total, but 0 pairings of blocked-unblocked *morphs* in the remaining 550+ entries

中国感恩节	Chinese Thanksgiving	中国感恩节	og	MANY_RESULTS
中国感恩节	Chinese Thanksgiving	Zhōngguó G ǎ n'ēnjié	py	NO_RESULTS
中国的互联网是开放的	China's Internet is open	中国的互联网是开放的	og	MANY_RESULTS
中国的互联网是开放的	China's Internet is open	Zhōngguó de hùliánw ǎ ng shì kāifàng de	py	NO_RESULTS

Typical Sample of Dataset. Observe that pinyin generally does not return any results, and Chinese script generally passes through censors.

An interesting result of this is that the pinyin terms with accents were very rarely used on Weibo. The vast majority of pinyin morphs yielded no results, even when the base term *morph* yielded many. There are likely two reasons for this. The first is that computer-based entry systems are generally geared toward displaying characters, even though the user may be typing in pinyin. In the case of stroke-based entry, the pinyin is completely bypassed. The second reason for the lack of pinyin results is that the computer-based entry for pinyin with tones is particularly cumbersome, often requiring the user to hold down a key in order to select the accented version. Particularly with idiomatic terms such as these, it is likely that any pinyin input would have been entered without tone marks. We discuss the addition of toneless pinyin as an additional *morph*, along with other further steps, in the conclusion section.

Below we find all of the blocked-status *morphs* from the dataset. It is of interest that only 15 of over 300 terms (over 600 morph-pairs) are currently blocked by censors. On one hand, it could indicate that the internet slang found in the Grass Mud Horse Lexicon is new enough to pass undetected by censors.

Base Term	**Meaning**	**Morph**	**Type**	**Status**	**Topic**	**Connotation**
GFW	Great Firewall	GFW	og	BLOCKED	censorship	pejorative
GFW之父	father of the Great Firewall	GFW之父	og	BLOCKED	censorship	pejorative
GFW之父	father of the Great Firewall	GFW zhī fù	py	BLOCKED	censorship	pejorative
习包子	Steamed Bun Xi	习包子	og	BLOCKED	political	endearing

五月三十五日	Thirty-Fifth of May (June 4, 1989 - Tiananmen)	五月三十五日	og	BLOCKED	collective action	neutral
土共	TG	土共	og	BLOCKED	political	neutral
带鱼包子	cutlassfish bun	带鱼包子	og	BLOCKED	political	pejorative
当今皇上	reigning emperor	当今皇上	og	BLOCKED	political	pejorative
捅鸡局	Bureau of Dicking Around	捅鸡局	og	BLOCKED	political	pejorative
日人民报	Screwing People Post	日人民报	og	BLOCKED	political	pejorative
正腐	govern-rot	正腐	og	BLOCKED	political	pejorative
淋巴县长	Mayor Lymph	淋巴县长	og	BLOCKED	collective action	neutral
电婊	power whore	电婊	og	BLOCKED	political	pejorative
糊煮席	muddled boiled banquet	糊煮席	og	BLOCKED	political	pejorative
翻墙	scale the wall	翻墙	og	BLOCKED	collective action	neutral

All blocked entries from Grass Mud Horse lexicon

Yet, if these sorts of terms are widely available on the internet, it would seem that it would be very simple for censors to simply add the list to their own blacklist of terms. Rather, it seems that the majority of the lexicon's terms are not intended to be blocked. This is in line with the King et. al. hypothesis that criticism is not necessarily something that the government aims to suppress, as long as it is not backed by collective action.

Of the blocked terms below, I have classified three to refer to collective-action events: 35th May, Mayor lymph, Scale the wall. The 35th of May refers to June 4, 1989, or the Tiananmen square crackdown[14]. Given the controversy surrounding the topic, it is not surprising to see a reference to Tiananmen censored. Of particular note is "Mayor Lymph", which refers to a manifesto for reform and democratization in China signed by 350 scholars and activists in

[14] http://chinadigitaltimes.net/space/Thirty-Fifth_of_May

2008[15]. Scale the wall refers to methods of bypassing China's firewall via VPN[16], which also has collective action connotations.

The remaining terms have to do with censorship or political criticism. It is interesting that only these particular terms were blocked, as they do not seem particularly distinguishable from the majority of unblocked terms. One of particular note is "习包子", or "Steamed Bun Xi", which is an endearing reference to Xi Jinping's dining at a common steamed bun restaurant. This act was generally popular among netizens for its representation of the average citizen, although some criticized it as a "political show"[17]. Regardless, it appears that it is the potential for viral conversation around politics, neither decisively positive or negative, that has landed this term on the blocked list.

Conclusion

It would be worth revisiting the crawling stage in order to re-evaluate the ERROR-results. The primary limitations in this regard were not having accesses to the Weibo API for search queries (which would essentially give a clean backdoor to search, where no web crawling would be necessary), and having to monitor queries due to CAPTCHAs. The API issues would require getting Weibo developer application approval, which may turn out to be unfeasible for academic purposes. The CAPTCHA issue may be solvable by using proxies.

One interesting direction to take this study would be comparing the result types with term classification. For example, tagging sensitive with topics such as: government policies, criticism, pornography, profanity, news would be useful supporting or refuting the collective action potential hypothesis (set forth by King et. al) in the microblogging space.

The code for this project was designed to allow for more *morphs*. To begin, two additional *morphs* can be derived programmatically from each entry: no tones (pīnyīn→pinyin), and acronym (pīnyīn→PY). There many more romanizations (Wade-Giles, dialect-based), syllabary-based such as bopomofo, homophones lookup up via rhyming dictionary, and

[15] http://chinadigitaltimes.net/space/Mayor_Lymph
[16] http://chinadigitaltimes.net/space/Scale_the_wall
[17] http://chinadigitaltimes.net/space/Steamed_Bun_Xi

potentially even homographs from a graphically-organized Chinese dictionary. All of these methods are certainly doable by hand, but it would be interesting to explore programmatic techniques to accurately predict the possible "up and coming" *morphs* by finding all potential morphs, and checking to see if any are being used in to represent the initial term.

With this in mind, a programmatic approach to exploring potential word *morphs* applies to more than patterns of internet censorship. The homophonic and multi-topolectical nature of Sinitic languages in China may be particularly suitable to programmatic approach to the problem of *morph* generation because of the ease of generating novels ways of presenting ideas based on similar sound. An exploration of *morph* usage and proliferation could offer insight into the development of internet language in China, and lend itself to similar studies in other languages.

Bibliography

Le Chen, Chi Zhang, Christo Wilson. "Tweeting Under Pressure: Analyzing Trending Topics and Evolving Word Choice on Sina Weibo". Proceedings of the first ACM conference on Online social networks. 2013. Source: http://www.ccs.neu.edu/home/cbw/pdf/weibo-cosn13.pdf.

Gary King, Jennifer Pan, Margaret Roberts. "How Censorship in China Allows Government Criticism but Silences Collective Expression". American Political Science Review. 2013. Source: http://gking.harvard.edu/files/gking/files/censored.pdf?m=1448321711.

K. Fu, C. Chan, M. Chao. "Assessing Censorship on Microblogs in China: Discriminatory Keyword Analysis and Impact Evaluation of the 'Real Name Registration' Policy". IEEE Internet Computing. 2013.

Victor Mair. "Ways to say 'China' that can circumvent the censors". Language Log. 2015. Source:http://languagelog.ldc.upenn.edu/nll/?p=22520.

ibid. "Pinyin spam text message". Language Log. 2015. Source: http://languagelog.ldc.upenn.edu/nll/?p=21744.

Ben Zimmer. “Censoring ‘Occupy’ in China”. Language Log. 2011. Source: http://languagelog.ldc.upenn.edu/nll/?p=3523.

Sherry Chen. “From OMG to TMD -- Internet and Pinyin Acronyms in Mandarin Chinese”. Language@Internet Volume 11. 2014. Source: http://www.languageatinternet.org/articles/2014/chen.

Mark Liberman. “Franco-Croation Squid in pepper sauce”. Language Log. 2009. Source: http://languagelog.ldc.upenn.edu/nll/?p=1225.

Jason Ng. *Blocked on Weibo: What Gets Suppressed on China’s Version of Twitter (and Why)*. New York and London: The New Press. 2013.

Jason Ng. Email Correspondence. 2015.

Shengyun Sun, Hongyan Liu, Jun He, Xiaoyong Du. “Detecting Event Rumors on Sina Weibo Automatically”. Computer Science pp 120-131.

Jason Ng. “Tracing the Path of a Censored Weibo Post”. Citizenlab. 2014. Source: https://citizenlab.org/2014/11/tracing-path-censored-weibo-post-compiling-keywords-trigger-automatic-review/

Appendix

Full results can be found here: https://docs.google.com/spreadsheets/d/18OUejuJGU_TWgKGDDCPpwUYDzh87ymPwN6BFCZkFcYg/edit?usp=sharing

Complete project code can be found here: https://github.com/dericko/ealc231-proj