

A new type of multimedia content with Chinese characters as the core—— Barrage

Yang Pan, KiHong Kim, JiHui Yan

Doctor, Department of Visual Contents, Dongseo University, China.

Professor, Department of Visual Contents, Dongseo University, Korea.

Doctor, Department of Visual Contents, Dongseo University, China.

Professor, Department of VR Contents, Qingdao Huanghai University, China.

pyy8976341@naver.com, khkim@g.dongseo.ac.kr, ikkity@gmail.com

Abstract

Barrage language is a new media language based on Internet video. It is a representative expression of the Internet environment regardless of its format and content. As a unique movie viewing characteristic provided by the barrage function, the timeliness of feedback, entertainment, and interactivity of the content are excellent. Characteristic. Barrage language itself is closely related not only to the value of linguistic research, but also to the spread of the Internet in the Internet environment.

Based on the current situation of Chinese video sites, this thesis explores how Barrage is an organic cycle of culture and consumption in a specific platform and group, and analyzes its propagation methods and effects.

By analyzing the characteristics of content production patterns unique to the barrage culture, implications and reference values for production activities of other cultures, the effect of popularization on viewers and the production and consumption of the 'barrage' culture of the industry is studied, and furthermore, the 'barrage' culture is It was designed to be a reference for the development of the platform, internet culture, and animation industry culture.

Keywords: *Digital Contents, Barrage, Barrage Culture, Digital Language Culture.*

1. Introduction

According to the 47th China Internet Development Statistics Report released by the China Internet Information Center (CNNIC), the number of Chinese netizens was 989 million as of December 2020, up 85.4 million from March 2020, and the Internet penetration rate was 70.4%, up 5.9% from March 2020. The growing number of netizens has led to the development of the Internet in professional and sophisticated areas such as instant messaging, audio-visual entertainment, and other fields. With the advent of the video site,

Manuscript Received: April. 27, 2022 / Revised: May. 3, 2022 / Accepted: May. 6, 2022

Corresponding Author: khkim@g.dongseo.ac.kr

Tel: +82- 51-320-1704

Professor, Department of Visual Contents, Dongseo University, Korea.

Internet users can watch movies and dramas on their computers at home, and video sites such as Bilibili, Aichi, and Tencent, which have been able to purchase a large number of copyrights of popular videos, attracted a large number of Internet users.

However, traditional video sites had no choice but for viewers to passively absorb the video. Unlike ordinary video-sharing sites, the anthrax site, which is one of the pillars of the video-sharing site, allows viewers to review their videos in real time, and when all viewers watch the video at that point, the sliding subtitles are displayed in real time, adding to interaction between different viewers. On the ammunition site, animation series, animation-related music or personal performances, secondary creative animation, dongin animation, and parody videos are the main releases, and there are also works that have been moved from other video-sharing sites.

Japanese scholars believe that in the Internet age, people tend to build their identities in a mimetic environment, and barrage provides a virtual place for audiences to co-exist. Audiences speak and interact in the place, and confirm whether they are related to this collective through the speeches of others. content, to meet the needs of their own identity building. [1]

The longer the proportion of time people live on the Internet, the stronger the need for identity in this mimetic environment. Therefore, the emergence of a barrage culture in Generation Z is inevitable in the information age.

But if the Barrage is a common psychological need of contemporary people, then why is there no mature Barrage culture in other countries except China and Japan?

[Generation Z is a popular term in the United States and Europe, referring to people born between 1995 and 2009, also known as the Internet generation, Internet generation, collectively refers to the Internet, instant messaging, SMS, MP3, smart phones and tablet computers, etc. Technological products have a great impact on a generation.]

2. The fast-growing barrage culture only in the Chinese and Japanese markets

Japan is the world's first country to launch a video website with a bombshell. In January 2007, just a month after its release, it surpassed 100 million views on the Niconico page. In May 2008, the total number of video playback exceeded 5 billion, and in 2016, the total number of paid members exceeded 2.56 million. Japanese-style ammunition culture also emerged quickly. Soon, all UP states had to consider the attributes of their fans by displaying tags such as banning ammunition, cautioning ammunition, recommending ammunition, and welcoming ammunition in their videos.

Since 2006, AcFun and bilibili have introduced this method to China, which has quickly created a Chinese charcoal culture. Not only did it prosper on the streaming media site, but it also made a big hit on the offline screen.

In 2014, the "New York Times" published an article that surprisingly introduced readers to the mysterious oriental phenomenon of "coalition." According to the article, young Chinese are no longer satisfied with just sitting in the theater and watching the movie quietly, but they text their cell phones and leave comments on the huge screen.

What the news says is an anthrax film attempted by "Sausage 3" at various movie theaters in Beijing in 2014. In this attempt, the audience can send comments to the movie theater through a mobile phone, and after processing the comments in the movie theater, the audience's comments are scrolled in the form of bombing on the movie screen in real time.

The New York Times commented that it did not know whether it was the scariest nightmare of moviegoers or the coolest of the future.

In Europe and the United States, Twitch is the only form that is close to a coal mine. Founded in San Francisco in 2011, Twitch displayed a side window on the right side of the video that allows you to scroll through reviews in real time, reminiscent of scrolling comments during current live broadcasts.

It provides users with the possibility of exchanging comments online, but it does not appear in the form of ammunition in the video. The same goes for YouTube. Not only is there no ammunition, but there are fewer subtitles. In the eyes of some European and American video casters, the only acceptable ammunition is the first scene of the movie Star Wars.



Figure 1. Star Wars first scene

Chinese and Japanese netizens made the most of their subjective initiatives to enjoy the fun of the barrage on YouTube, created impossible possibilities, and forced the development of various Chrome browser plugins and added a barrage to YouTube.

However, due to most of my personal work, these plugins fall out quickly. Available so far are TUBETUBE developed by Japanese netizens, DanMage supporting Crunchyroll & Hulu & Funimation, and Dmooji弹幕君 developed by Chinese netizens.

In Japan, the reason for the rapid growth of the barrage with China goes back to the essence of the text. The Chinese characters used in China and Japan are more suitable for barrage than English. If you look at the countries that have escaped the barrage, both countries use Chinese characters. Chinese characters representing meanings are far superior to the phonetic alphabets such as English in information entropy.

Linguistically, English, French, German, Russian, Spanish, etc. belong to phonetic scripts, and the language handwritten symbols of phonetic characters obey direct pronunciation, and shapes obey consonants and do not directly indicate meaning.

Chinese belongs to ideograms that have a certain phonetic function, and the language handwriting signs representing ideograms express meanings, and although there are phonetic components in the structure of the text, consonants obey the shape, and letter signs do not directly indicate the voice. Chinese characters were introduced to Japan and used by Japanese ancestors.

When first using Chinese characters to express concepts without Chinese characters, the Japanese thought of a way to take on the sound and shape of a word and discard its meaning. In this way, the usage of Chinese characters was invented, and manyogana appeared in literature from <Kojiki> to <Manyoshu>. Then, pure phonetic characters, hiragana, appeared. It is used in combination with Chinese characters.

2.1 Difficulty of scanning

When reading Chinese, instead of scanning character by character, I scan through a section. When a person stares intently at a Chinese character, if you add 1-2 Chinese characters to the left and 2-3 Chinese characters to the right of a Chinese character, you can see up to 5-6 characters at a time.

Therefore, there is no problem in reading two letters that are adjacent to each other or are separated by one

letter. However, if there are more Chinese characters exchanged, it becomes difficult to read. Replacing only a pair of Chinese characters in a short passage makes it difficult to read. On the other hand, the alphabet of phonetic characters constitutes an affix, an affix constitutes a word, and a word constitutes a phrase. Even when Westerners read phonetic characters, skilled people read ten lines at a time. [2]

However, if you do a word substitution test for English sentences and read it again, the difficulty is higher. Do you understand? Changing only the alphabetical order of the words in a sentence is usually a quick way to understand the original meaning.

Next, reverse the order of several words. Find an example first. You make millions of decisions that mean nothing. Reorder words alphabetically: You make millions of decisions that mean nothing. There is no problem even if you change the word alphabet, and you can recognize it with the inertia of reading like a quick scan.

Change multiple word positions: You make millions of decisions that mean nothing. When you quickly scan these words, you can quickly extract keywords as well, You, make, decisions, nothing, mean, but it becomes cumbersome and you can't intuitively react to what word to put after make, make. It tells whether of is possible, make decisions is possible, or make nothing is possible. So dissection is hindered and it is not good.

In addition to verb inflections, which are easy to confuse, English has a set word order, such as profanity and simplified colloquial words. For example I have to hit the books. Among them, hit the books is a slang word that means to study hard. If you change the order a little, I have books to the hit. The meaning of the profanity disappears immediately, and the meaning of the entire sentence is completely different.

Therefore, basically, under similar sentence structure and complexity, there are more core parts of English phrases than Chinese, so it may be more difficult than reading English phrases in the character compatibility reading test. What could be explained? Instead of looking at the bullets on a bulletin board site one by one, you can intuitively know what the bulletin is saying by looking at each part and pulling out keywords. You can extract keywords when you scan the English barrage, but because the keywords are intricately entangled, you cannot instantly know what the English barrage is saying. [3]

Overseas media have been discussing why bullet sites are not popular for a long time. KrASIA has a long article on why barrage sites aren't popular in the US.

Despite the popularity of bullet comments in China, major video websites in the United States did not make any significant moves to change their video-viewing cultures. Neither Facebook nor Youtube adapted the bullet comment feature to their website. Language barriers may be an obvious obstacle. Although Youtube and Facebook services are offered in different languages, the bullet comments are hard to distinguish and put apart. Unlike Bilibili and AcFun, whose users are predominantly Chinese, Youtube and Facebook users may see their comments in various languages, making the potential bullet comments hard to understand and separate.

Figure 2. KrASIA has a long article on why barrage sites aren't popular in the US.

One of the main reasons is language barriers, the bullet comments are hard to distinguish and put apart. However, this argument points to the complexity of the user's language.

2.2 Visual reading efficiency

The process of human interpretation of phonetic characters is the process of interpreting meaning from text to pronunciation and again from pronunciation. This is because this language system was initially invented according to the phonological interpretation method. However, the ideogram interpretation process is the process of interpreting the meaning directly from the text. Chinese characters were the first to use images as a medium of interpretation. Under the same text length, the same reader's eyesight, and the same proficiency conditions, the reading efficiency of phonetic characters is much higher than that of phonetic characters. So for Westerners, it is much easier to hear the language (pronunciation → meaning) in the video than to read the text (letter → pronunciation → meaning) in the video. Because language is inherently phonetic.

Of course, it is not necessary to use the language to mark the bulletin board, but it can be displayed with symbols, facial emoticons, and emojis. Therefore, attack websites of Western spoken languages are inherently limited by the language itself and are not doing well.

2. Information entropy

Information is a very abstract concept. People say they have a lot of information or they say it is little, but it's hard to say exactly how much. For example, how much information is in a 500,000-character Chinese book?

It was not until 1948 that C. E. Shannon solved the quantitative problem of information by suggesting the concept of information entropy in an article titled *A Mathematical Theory of Communication*. The word information entropy was borrowed from thermodynamics by C. E. Shannon. In thermodynamics, thermal entropy is a physical quantity that describes the degree of confusion in molecular states. Shannon describes the uncertainty of information sources with the concept of information entropy. Claude Shannon, the founder of information theory, was the first to describe the tedious correlation between probability and information in the language of mathematics. [4]

In general, it is unclear which code a single source of information is sending, but it can be quantified according to the probability of its appearance. Probability is high, opportunity is high, so uncertainty is low, and conversely, uncertainty is high.

The uncertainty function f is a decreasing function of the probability P , and the uncertainty created by two independent symbols must be equal to $f(P_1, P_2) = f(P_1) + f(P_2)$, which is the sum of the respective uncertainties, and is called 'additivity'. A function f that satisfies these two conditions simultaneously is a logarithmic function, that is:

$$f(P) = \log \frac{1}{p} = -\log p \quad (1)$$

The Information Source should consider the average uncertainty of all situations in which this identity may arise, rather than the uncertainty arising from a single code. If the source symbol has n values: $U_1 \dots U_i \dots U_n$, the corresponding probability is: $P_1 \dots P_i \dots P_n$, and the appearance of each code is independent of each other. In this case, the average uncertainty of the information source is the uncertainty for each code $-\log P_i$, and the statistical average value (E) is called information entropy. that is:

$$H(U) = E[-\log p_i] = -\sum_{i=1}^n p_i \log p_i \quad (2)$$

In formulas, logarithms are usually base 2 and units are bits. However, other logarithmic bases may be used, and other corresponding units may be applied to convert between them into a base conversion formula.

3. Information entropy of Chinese characters

Strictly speaking, Chinese is the largest mainstream language of information entropy in the world.

The information entropy of one Chinese character (Chinese) is 9.56. On the other hand, in general, the information entropy of one alphabet is 3.9. However, Chinese characters generally occupy two bytes on the display screen, and the amount of information per unit area is larger than that of the alphabet. (The information entropy of two alphabets is not directly multiplied by two. Shannon originally calculated the information entropy of a single English word to be 2.62.) [5]

Informational scientists can only estimate the information entropy of each language in several ways. For example, Shannon thinks that the information entropy of English is between 0.6 and 1.3 bits/character, Cover and King think that it is 1.25 bits/character. The difference is because the sample and the experimental method are different. [6]

English and other phonetic texts only have 24 characters, but in the case of Chinese, statistics become difficult. Fortunately, not long after the development of information theory at that time, scientists in each field paid great attention to exploring various languages. Although the Chinese language has great specificity, the predecessors of information theory overcame all difficulties and calculated the Chinese information entropy by statistical sampling. Chinese has much more statistical base than English, from letters to radicals to syllables. [7]

Table 1. Chinese information entropy

	Entropy (BITS)	Max Entropy (BITS)
Word	9.63	12.20
Radicals	6.41	7.71
Syllable	8.68	10.17

In 2002, Frederi, et al. of Harvard University conducted a comparative experiment again. They believe that natural language has many common statistical characteristics and similar patterns in past natural language research. They hypothesized that, for different kinds of languages, Markov-based compression algorithms such as PPM would ignore language characteristics and compress text to an information compression lower bound.[8]

In other words, if the compression algorithm used is not specifically optimized for a particular language, different languages can approximate the information entropy by comparing the compression efficiencies of the algorithms. So they designed an experiment using the PPM algorithm to compress different versions of the Bible. [9]

Language	Original Size (bytes)	Ratio (Original)	Compressed Size (bytes)	Ratio (Compressed)
English	1936473	1	390846	1
Spanish	1804756	0.932	384681	0.962
French	1896459	0.979	393805	1.01
Chinese	884860	0.457	337505	0.864
Korean	1259920	0.651	346478	0.886
Arabic	1875204	0.968	418443	1.071
Japanese	1519224	0.785	452337	1.157
Russian	1506920	0.778	376162	0.962

Figure 3. PPM Algorithm Experiment

As shown in the figure above, the researchers compared the file sizes of various versions of the Bible in English, Spanish, French, Chinese, Korean, Arabic, Japanese, and Russian before compression, and the ratio

of file sizes to English before compression. Properties such as the file size after compression and the ratio of the compressed file size to English files. Under ideal conditions, if there is no loss of information in the process of translation, compression, etc., the ratio of the file size of English files to English files after compression should be 1.

It is clear that the compression efficiency of Chinese is lower than that of other texts. Could this compression efficiency be caused by the text and the compression algorithm? They completed two experiments as follows.

Language	Original Size (bytes)	Ratio (Original)	Compressed Size (bytes)	Ratio (Compressed)
English	1936473	1	411732	1
Spanish	1804756	0.932	414206	1.006
French	1896459	0.979	422532	1.026
Chinese	884860	0.457	370168	0.899
Korean	1259920	0.651	387532	0.941
Arabic	1875204	0.968	448962	1.090
Japanese	1519224	0.785	481649	1.170
Russian	1506920	0.778	412354	1.002

Figure 4. BZIP2 Algorithm Experiment

In the second experiment, another compression algorithm (BZIP2) was adopted, and the results were similar, showing that the compression efficiency was not reduced by the compression algorithm.

Language	Original Size (bytes)	Ratio (Original)	Compressed Size (bytes)	Ratio (Compressed)
English	977885	1	110848	1
Spanish	1064225	1.088	123695	1.116
French	1050979	1.075	125214	1.130
Chinese	420178	0.430	109279	0.986
English	944778	1	106501	1
Russian	1009347	1.068	127750	1.200
English	478026	1	65310	1
Arabic	328606	0.687	69350	0.942

Figure 5. European Union Code Text Experiment

In the third experiment, different texts (regulations of the European Union) were applied, so all translations except English were expanded, and among them, Chinese was expanded the most. In response, the researchers explain that biblical texts are very common words, and European Union legislation contains many special words, which require long letter extensions to translate from other languages. This is because legal texts always want to be translated using special vocabulary, but these words do not appear very often in everyday life, so they appear to be informative. If you are translating legal texts into universal words, you will need some language extensions. However, Chinese is still the least-compressed language.

According to the results of three comparative experiments, it is concluded that Chinese is the language with the lowest compression efficiency or the closest to the limit of information entropy.

Although the design of this experiment is not perfect, according to various experimental results, Chinese was the language in which the most information is entropy among mainstream languages such as English, Spanish, French, Chinese, Arabic, Japanese, and Russian.

Another problem is visual capture and short-term memory. Although the information entropy of Chinese characters is very large, since the relative complexity of the characters is relatively large, it is considered that

the writing time and writing speed of various characters basically coincide. In addition, it is calculated that the amount of information transmitted per unit time and the bandwidth of someone's voice are all about 39 bits/s. (A single Chinese character is great at expressing meaning, but multiple characters share the same voice, and only a coined word can dilute this advantage).

However, when there is a dynamic visual effect, the area that can be captured by the human eye is limited, and tracking and concentration are usually required to read information. At the same time, the short-term memory capacity of the human cerebrum is very limited, about 7 ± 2 blocks, and the 'parallel processing' of the cerebrum is 'upper-parallel processing', so it is important to capture information and store the cache.



Figure 6. United Nations Triple Text Capture

You can see this by capturing a scene from the Bilibili video, which has nothing to do with the barrage, compare the length of the 'UN' trigram at the back. Chinese: 联合国, English: United Nations, Russian: Организация Объединенных Наций. Actually, this is not a big problem. Even if you get used to reading Russian, the speed is fast, but the problem is that when a video hits or a video with a large amount of playback is released, the number of bullets increases and letters and letters overlap, and Chinese characters are easily recognizable because they are square. If all the letters are mixed, it is difficult to tell them apart.

4. The inevitability of the rapid development of barrage culture in the Chinese market

In the language type, Chinese is more discriminating. Chinese and Japanese have a characteristic that they do not use participles when they use kanji in common or when kanji are mixed with kanji. (In Japanese, it is called '分かち書き') Even with the same syllable, Chinese has better discrimination power than Hangul. The anti-siege structure is a structure that cannot exist in Korean.

The habit of using subtitles is also an important factor in the development of the barrage culture. The rate of subtitles for movies, TV shows, and Internet programs distributed in China is significantly higher than that of English, American, Japanese, and Korean. One of the reasons is that the Chinese dialect is diverse. Chinese audiences are accustomed to seeing subtitles on video programs, so Chinese gofong songs, ghosts, vocaloids, etc. can be popular because they are difficult to understand without subtitles. Language and user habits are important conditions that support the rapid growth of barrage in the Chinese market, and the character of Chinese and the number of people using Chinese have laid the foundation for the development of Chinese barrage culture, but there are other aspects as well. These factors are neither quantitative nor sustainable.

The development and improvement of cultural copyright awareness and copyright law in the Chinese market is relatively late compared to other developed countries. In the case of YouTube, Americans are very interested

in copyright, and in particular, copyright protection of videos, music, and animations is large. As a result, netizens' re-creation is blocked and re-creation is not activated, hindering the development of the barrage.

In the early days of Niconico's founding, many YouTube external links were hijacked and blocked on YouTube in the second month. Web sites Acfun and Bilibili also started as piracy in the early days. YouTube also had copyright problems in the early days, but could not stop the surge of users and huge traffic after that, and distributors compromised with YouTube to form a strategic alliance to allow the release of video content on YouTube.

Since Barrage is an Asian culture, globalization will be difficult. The global video website YouTube has users all over the world, each with their own YouTube website in their own language. Each country has its own country, but a multinational video site such as YouTube cannot offer strong local features. Barrage is an Asian culture that can only be seen in Asia, and not everyone accepts it. On the other hand, Chinese video platforms distribute and produce content based on Chinese users. This growth method is speeding up the growth of the market in China and limiting the speed of internationalization of such content.

5. Conclusion

The emergence of the post-epidemic era has changed the needs and use of the Internet and digital media content by humans all over the world. Under the premise of the rapid development of digital media content, more and better media content has been developed and created. Not limited to traditional media content such as movies, games, animations, etc. Metaverse content, virtual digital content, virtual social networking, virtual network business and other content are also developing rapidly.

Barrage, born from digital media content, will also give birth to new forms of digital media content.

Smartphones, computers and other multimedia devices have influenced and changed the behavioral patterns of consumers watching and using digital media content from the hardware perspective. The direction of software influences and changes the behavioral patterns of consumers viewing and consuming digital media content.

Barrage is based on digital media content and performs well on streaming platforms in China and Japan. As the core of the barrage, that is, the language, which is different from English, Arabic and other languages. This paper analyzes the advantages of the Chinese language family as the core of the bullet screen. Chinese, Japanese, and Korean have inherent advantages that are very compatible with Barrage. Combined with the support of Barrage for digital media content, and the Asian culture with Chinese, Japanese and Korean characteristics, I think digital media content products that combine Barrage or similar functions will be a new characteristic development direction of the Asian digital media industry. While exporting Asian culture to the world, the barrage is also very interactive, which increases user feedback and experience.

In terms of practical value, the Barrage, as subtitle text, enriches the information of the video, and sometimes helps users understand the video. The impact of digital media content, from content planning, content production, user surveys, multi-perspective content experience, content complement and support, and even publicity and promotion, will play a positive role. At the same time, it will play a huge role in maintaining users' stickiness to video websites and the new profit model based on barrages derived from video websites.

As the key carrier of the Barrage, based on the culture and characters of China, Japan and South Korea, combined with the development technology of modern digital content and the national traditional culture, the author believes that the content of the Barrage will become an indispensable part of the future Asian digital media culture.

References

- [1] Deng XiangYang, "Media Economics", Hunan University Press, 2006
ISBN: 9787811130256
- [2] Yang Bin: Introduction to Media Convergence, Peking University Press, 2013
ISBN: 9787301233672
- [3] Kagawa, K., Itou, J., & Munemori, J. (2010). Effects of an intuitional pictograph comment function in a video sharing web system. *Int. J. Informat. Soc.*, 2, 94-99.
http://www.infsoc.org/journal/vol02/IJIS_02_3_094-099.pdf
- [4] Shannon C E. Prediction and entropy of printed English [J]. *Bell system technical journal*, pp. 50-64, 1951
DOI: <https://doi.org/10.1002/j.1538-7305.1951.tb01366.x>
- [5] Q Tian, Comparing information entropy between Chinese and English language, 2012 IEEE International Conference on Cyber Technology in Automation. 2012
DOI: <https://doi.org/10.1109/cyber.2012.6320050>
- [6] Cover T, King R. A convergent gambling estimate of the entropy of English[J]. *IEEE Transactions on Information Theory*, Vol 24(4), pp.413-421, 1978
DOI: <https://doi.org/10.1109/tit.1978.1055912>
- [7] Wong K, Poon R. A Comment on the Entropy of the Chinese Language[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24(6): pp. 583-585, 1976
DOI: <https://doi.org/10.1109/TASSP.1976.1162865>
- [8] Fromkin, V., Rodman, R., & Hyams, N. (2018). *An introduction to language*. Cengage Learning.
ISBN: 9781337559584
- [9] Behr Jr F H, Fossum V, Mitzenmacher M D, et al. Estimating and comparing entropy across written natural languages using PPM compression[J]. 2002
DOI: <https://doi.org/10.1109/dcc.2003.1194035>