

# Investigation and Study of the Use of Modern Chinese Characters

Yanmei Zhang<sup>1</sup> Yu Tian<sup>2</sup> Zhan Lv<sup>3</sup> Jiali Chi<sup>4</sup> Zixin Lin<sup>5</sup> Mengyuan He<sup>6</sup> Shuang Ye<sup>7</sup> Chengya Zhang<sup>8</sup> Nuo Zhang<sup>9</sup> Xin Ruan<sup>10</sup>

<sup>1,2,4,5,6,7,8,9,10</sup> Wuhan Institute of Technology, Wuhan, Hubei 430205, China

<sup>3</sup> Jinan University, Guangzhou, Guangdong 510610, China

<sup>3</sup>Corresponding author.

## ABSTRACT

This paper adopts the research methods of corpus linguistics and metrological linguistics, obtains a massive corpus, and builds a corpus by using Python crawler technology to carry out a systematic and comprehensive, large-scale statistical study on the frequency of modern Chinese characters. On this basis, it will reveal the frequency gradient of the current Chinese characters, the number and specific types of commonly used characters, and the number, type, and order of the basic Chinese teaching material and international Chinese education materials, providing a reference. A survey on the irregular use of characters was also conducted, which mainly investigated the irregular use of traditional characters and characters in different forms. The survey reveals: 1. the number of commonly used modern Chinese characters needs to be simplified; 2. the specific character types in the list of commonly used modern Chinese characters need to be updated; 3. the awareness of the standardization of traditional and variant characters needs to be strengthened. On this basis, this study argues that the current direction of Chinese character reform is not pinyinization, but rather simplifying Chinese characters (while giving better play to the role of Pinyin) and standardizing their use.

**Keywords:** *Character frequency, Character types, Sight characters, Traditional characters, Variant characters.*

## 1. INTRODUCTION

Chinese characters are the world's oldest existing scripts, running through the ages and bridging the north and south, and have made significant contributions to preserving and passing on history and culture, maintaining national unity, and promoting ethnic and cultural exchanges. However, Chinese characters also have many shortcomings: for example, the shape is complicated; the strokes are many; and the structure is complicated. There are too many characters, including ancient characters, common characters, traditional and simplified characters, variant characters, dialectal characters, transliteration characters, and so on. Several generations of patriotic scholars have worked hard on this issue, exploring reform proposals and launching a vigorous campaign to reform characters throughout the country. After the founding of the People's

Republic of China, of several Chinese characters were simplified and the "Scheme for the Chinese phonetic Alphabet" was formulated, which generally kept pace with the modernization of Chinese information processing and led to the healthy development of language life in the new era. However, in the current social life, problems such as confusing and uncontrolled use of characters abound and need urgent intervention and research.

Specifically: a. indiscriminate character creation in the media and on the Internet, especially names of people and places, uncommon words, etc.; b. changes in the use of high-frequency Chinese characters, which are still complicated, and the number of characters is still too much; c. high-frequency characters should be initially graded, and decreasing utility rate of Chinese characters once proposed by Mr. Zhou Youguang has not attracted sufficient attention; d. the characters used in basic Chinese teaching material for primary and

secondary schools and those used in foreign Chinese materials arbitrarily exceed the outline; e. in current language life, whether in the media and the Internet or in the social use of characters in streets, stores, dishes, billboards, building slogans, etc. in urban and rural areas, there is still unstandardized social use of characters in traditional characters, foreign languages, misspelled characters, variant characters, and uncommon characters; f. Statistical analysis techniques for language and characters have gained great progress over the years, but language life itself is intricate, and the construction of a dynamic corpus itself is difficult. There are still many unsatisfactory aspects in conducting field surveys and statistical analyses of the social use of characters on a large scale; etc.

There are already some studies on the use of modern Chinese characters. This is mainly reflected in some relevant journal articles and annual reports on the state of language life released by the national language committee for 16 consecutive years since 2006. a. Journal articles. The main ones are: "A Survey and Research on the Social Use of Characters in Beijing City" (Zhao Xue, Lu Jinfang, Liu Yifan, 2018), "A Survey on the Use of Characters in Newspapers, Radio and Television, and the Internet" (Wang Tiekun, Hou Min, Yang Erhong, 2007), "A Survey and Consideration on the Use of Characters in Four Sets of Basic Education Language Teaching Materials - Based on the Textbooks of Renjiao, Sujiao, Beishida" (Zhou Meiling, Su Xinchun, 2009), "A Survey on the Use of Characters in Chinese Language Teaching Materials for Foreigners since the 1950s: Five Sets of Chinese Language Teaching Materials for Foreigners" (Wang Yanjun, 2009), "A Survey on the Use of Characters in Major Chinese Language Media in Southeast Asia" (Liu Hua, 2010), etc. b. Annual report on the state of language life. Since the first annual report on language life ("Green Paper") was released to the society in the "Report on the State of Language Life in China (2005)", the national language committee has released an annual report on the state of language life to the society for 16 consecutive years. It has made positive contributions to the development of national conditions and the direction of language life in a healthy and harmonious direction, including some related surveys on the use of characters in newspapers, media, the Internet, society, and textbooks.

The above-related surveys and studies reflect the new forms and achievements of current linguistic research in terms of corpus acquisition

and technological breakthroughs. They record language life and draw some relevant conclusions and suggestions, and the survey methods have reference value for subsequent corpus construction and related word usage surveys. However, the above surveys and statistical analysis on language and characters focus more on reflecting language life and rarely consider the current use of characters from the perspectives of simplifying and standardizing Chinese characters.

It has been 34 years since the release of the "Modern Chinese Commonly Used Character List" in 1988. After its release in 2013, the first-level list of 3,500 commonly used characters in the "General Standardized Chinese Character List", which functionally replaced the "Modern Chinese Commonly Used Character List", has also been born for nearly 10 years. They have played a greater positive role in Chinese language life, the development of primary and secondary school Chinese textbooks and international Chinese education materials, the compilation of dictionaries, and the processing of Chinese character information. But after 34 years, do the 3500 commonly used characters still match the actual characters used in social life in today's time? Has the use of high-frequency characters changed? Are the usage of Chinese characters and high-frequency characters reasonable in Chinese textbooks for primary and secondary schools and international Chinese education materials compiled accordingly? In 2013, the "General Standardized Chinese Character List" was released, which focuses on "standardization". It sorts out the relationship between traditional and simple characters and between positive and negative characters and presents the corresponding relationship between standard characters, traditional characters, and variant characters in the form of an attached table, in an attempt to promote the standardization of contemporary Chinese characters at the application level. Nearly 10 years after its release in 2013, have the standardization criteria in the relationship between traditional and simplified Chinese characters and their orthography been well implemented?

This study conducts a systematic and comprehensive, large-scale statistical study on the frequency of modern Chinese characters through corpus self-construction and corpus measurement statistics. On this basis, it will reveal the frequency gradient of the current Chinese characters, the number and specific types of commonly used characters, and the number, type, and order of the

basic language materials and international Chinese education materials, providing a reference; A survey on irregular use of characters was also conducted, which mainly investigated the irregular use of traditional characters and characters in different forms. Considering the need for a large-scale diachronic corpus for word frequency statistics, this study selected the news corpus of Guangming Daily for the past 15 years from 2008 to 2022 (the first half of the year), the data on the use of characters in the "Annual Media Word Master List" for more than 10 years from 2011 to 2021 in the "Report on the State of Language Life in China" (Green Paper), the corpus of the basic language textbook of the unified edition for primary and secondary schools, the corpus of "Mandarin" in the Taiwan Hanlin edition of the elementary school, and the corpus of the intermediate and advanced reading textbooks in "Developing Chinese". As an authoritative newspaper in China, Guangming Daily not only covers important news at home and abroad but also contains quintessential articles by a wide range of intellectuals in the news sections of national studies and literature. Therefore, the corpus selected for this study covers the use of characters in news and press, literary works, media, basic language textbooks for primary and secondary schools on both sides of the Taiwan Strait, and international Chinese language education textbooks. The corpus has a large scale and a long duration, covering almost all aspects of the social use of characters, which can provide a scientific corpus basis for word frequency statistics. The survey on the use of irregular characters is based on the corpus of Guangming Daily for the past 15 years from 2008 to 2022 (the first half of the year). Because the news and articles published are normative and exemplary. The survey can provide reference data and ideas for the current work of Chinese character standardization, and is conducive to further promoting the better implementation of the standards for the relationship between traditional and simplified, positive and negative in the "General Standardized Chinese Character Table".

Through the research, the following three conclusions are drawn: 1. the number of commonly used modern Chinese characters needs to be simplified; 2. the specific character types in the list of commonly used modern Chinese characters need to be updated; 3. the awareness of the standardization of traditional and variant characters needs to be strengthened.

## 2. THE NUMBER OF COMMONLY USED MODERN CHINESE CHARACTERS NEEDS TO BE SIMPLIFIED

Regarding the simplifying of Chinese characters, there were relevant views in the 1930s, and China once made considerable achievements in simplified characters, which have been partially investigated by previous generations. However, the current understanding of simplifying characters is not enough, and the past views have not been well inherited and carried forward. There is no consensus on the current number of commonly used characters in modern Chinese.

What is the current number of commonly used characters in modern Chinese? This study has conducted a series of investigations and analyses.

### 2.1 This Study Discovered That the High-frequency Characters Used During These 15 Years Were Stable and Concentrated, with A Large Number of Shared Characters in Each Frequency Ladder

The high-frequency characters used in Guangming Daily in the past 15 years are very stable and concentrated. On average, 523 characters cover 80% of the corpus, 861 characters cover 90% of the corpus, and 2315 characters (2315.87) cover 99% of the corpus each year. And the number of high-frequency characters used is some distance from the 3500 characters in the first level of the "General Standardized Chinese Character List". The relationship between the average coverage rate of Chinese characters and the number of character types in Guangming Daily in the past 15 years is shown in "Table 1".

Table 1. Statistics of the average number of character types required for the 10%-99% coverage rate segment of the general character list of Guangming Daily in the past 15 years

Coverage (%)	Average Number of Character Types Required by Year
10	8.20
20	27.47
30	57.20
40	99.87
50	160.00
60	240.47
70	353.13

Coverage (%)	Average Number of Character Types Required by Year
80	523.40
90	861.80
91	918.87
92	984.40
93	1060.00
94	1149.00
95	1256.40
96	1391.67
97	1571.87
98	1836.53
99	2315.87

Secondly, in the last 15 years, there have been a large number of shared characters in each frequency gradient in the characters used in Guangming Daily. Even among the high-frequency characters, there is an obvious frequency gradient. As shown in "Table 1", when the coverage rate segment gradually increases, the number of required character types gradually increases, the first 10% coverage rate only needs about 8 characters, while the 10% coverage rate segment

Table 2. Statistics on the number of shared characters in the 10%-90% coverage segment of Guangming Daily in the past 15 years

Coverage	10%	20%	30%	40%	50%	60%	70%	80%	90%
Number of shared character types by year	8	20	46	78	124	197	294	449	756

## 2.2 Study of the "Annual Summary Table of Media Terms", as published in the "Report on the State of Language Life in China"

First, the top 4,000 characters used in each year of the "Annual Media Word Master List" for more than 10 years were sorted and obtained a total of 4,362 characters and their total frequency of use and total number of texts in the "Annual Media Word Master List" for more than 10 years. On this basis, all the characters were arranged in descending order of frequency to form the 10-plus-year word list. It is found that when the "Modern Chinese Commonly Used Character List" is set at 3500 commonly used characters, it can cover 99.48% of 2 million characters of text after testing, while only 2858 characters in the 10-plus-year media word list can cover 99.48% of the 10.9 billion corpus (the total frequency of the 10-plus-year "Annual Media Word Master List" is

from 80% to 90% needs about 340 characters, the utility of these 8 characters is much greater than the latter 340 characters. The coverage of each frequency gradient has a large number of shared characters during these 15 years, indicating that the utility of these high-frequency characters is very stable and the frequency gradient is stable. The statistics of shared characters in each coverage segment of Guangming Daily in the past 15 years are shown in "Table 2: Statistics on the Number of Shared Characters in the 10%-90% Coverage Segment of Guangming Daily in the Past 15 Years". In other words, the 8 characters "的(de)、国(guó)、和(hé)、人(rén)、是(shì)、一(yī)、在(zài)、中(zhōng)" have been in the top 10% coverage segments of Guangming Daily in each of the 15 years; the 20 characters "不(bù)、大(dà)、的(de)、发(fā)、国(guó)、和(hé)、会(huì)、了(le)、年(nián)、人(rén)、生(shēng)、是(shì)、为(wéi)、学(xué)、要(yào)、一(yī)、有(yǒu)、在(zài)、中(zhōng)、作(zuò)" have been ranked in the 20%-30% coverage segment in Guangming Daily for each of the 15 years. Due to the limitation of space, not all of them are listed.

10940610401). Therefore, since the role of high-frequency characters is much greater than that of low-frequency characters. As the number of corpora gradually increases, the number of characters needed to reach the rated coverage will be reduced accordingly, and the number of commonly used modern Chinese characters will not reach 3500 characters.

Second, this study also conducted a statistical analysis on the number and coverage of Chinese characters in the "Annual Media Word Master List" for more than 10 years, as shown in "Table 3: Coverage and Number Types of Characters in the Annual Media Word Master List from 2010 to 2020". It is found that the coverage rate and number of characters used in media from 2010 to 2020 are relatively stable. When the coverage rate is 80%, the number of characters used in each year fluctuates between 557 and 601, with a standard deviation of about 15.10; the proportion of characters used fluctuates between 4.86% and

5.66%, with a small proportion and a standard deviation of about 0.25. When the coverage rate is 90%, the number of characters used fluctuates between 877 and 972, with a standard deviation of about 30.81; the proportion of characters used fluctuates between 7.82% and 9.15%. When the

coverage rate is 99%, the number of character species used fluctuates between 2377 and 2449 with a standard deviation of about 62.17; the proportion of character species used fluctuates between 19.75% and 22.90% with a small proportion and a standard deviation of about 0.97%.

Table 3. Coverage and number of types of characters in the annual media word master list from 2010 to 2020

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Standard deviation
Coverage												
Number of character types with 80% coverage	601	593	594	597	597	600	585	583	576	558	557	15.10
Percentage of number of character types with 80% coverage	5.66%	5.60%	5.50%	5.23%	5.06%	5.10%	4.86%	5.03%	5.04%	5.00%	5.08%	0.25
Number of character types with 90% coverage	972	957	959	965	963	967	941	938	925	888	877	30.81
Percentage of number of character types with 90% coverage	9.15%	9.04%	8.87%	8.46%	8.17%	8.22%	7.82%	8.10%	8.09%	7.95%	7.99%	0.44
Number of character types with 99% coverage	2431	2396	2405	2428	2443	2449	2378	2377	2362	2280	2247	62.17
Percentage of number of character types with 99% coverage	22.90%	22.63%	22.25%	21.28%	20.71%	20.82%	19.75%	20.52%	20.66%	20.41%	20.48%	0.97

The standard deviation is the most common form of quantification to reflect the degree of dispersion of a set of data and is an important indicator of accuracy. The smaller the standard deviation, the less discrete the data is, and the closer this set of data is to the mean. When the coverage rate reaches 80%, 90%, and 99% the standard deviation of the proportion of word species is about 0.25, 0.44, 0.97, reflecting the differences between the data groups are small and the proportion of the number of characters is particularly stable. That is to say, it shows that the number of characters used each year to cover how many media texts is very stable, and the proportion to the total number of characters is very fixed. It is further demonstrated that the number of commonly used characters is small; the frequentness of use is high; the frequency of use is stable, and the coverage utility is large.

### 2.3 Survey and Analysis of the Use of Characters in Teaching Materials for Taiwan's Primary and Secondary Schools and "Mandarin" of the Hanlin Edition

This study investigated and compared the character coverage and the number of character types in the textbook of the unified edition, and "Mandarin" in the Taiwan Hanlin edition of the elementary school. It found that the number of high-frequency characters used in both textbooks was small, with only about 2,700 medium- and high-frequency characters (2710) among the character types used in the textbook of the unified edition for the three grades, and only about 2,000 (2,061) medium- and high-frequency characters in the three grades of the textbook "Mandarin" in the Taiwan Hanlin edition. Meanwhile, this study conducted a detailed analysis of the number of character types used in the basic language material of the unified edition for primary and secondary schools (18 volumes for grades 1-9). The specific situation of the number of characters used in each

textbook is found in the research. The number of characters used in the basic Chinese textbooks for grades 1 to 9 is between 879 and 3,133, and the number of characters used in the fourth grade (grades 7-9) also includes a large part of classical Chinese characters. It can be found that the number of characters used in primary and secondary school textbooks is still a long way from the 3,500 commonly used characters set in the modern Chinese commonly used character list.

#### ***2.4 Survey on the Character Usage of the Intermediate and Advanced Reading Textbooks of "Developing Chinese"***

In this study, the number of characters used in the intermediate and advanced reading materials of the classic international Chinese education textbook "Developing Chinese" was analyzed as follows: the number of characters used in the two intermediate reading materials ranged from 52,672 to 57,157, and the number of character types used did not exceed 2,000; the number of characters used in the two advanced reading materials ranged from 73,944 to 74,420, and the number of character types used did not exceed 2,500. The number of character types used in the intermediate and advanced reading of "Developing Chinese" is similar to the number of character types used in the second and third sections of the unified edition of the basic language textbooks surveyed above, and far from the number of 3500 in the table of commonly used modern Chinese characters. The number of commonly used character lists is large, which also has a certain negative impact on the choice of characters in textbooks: the range of commonly used characters is wide because it includes some characters that are currently not frequently used. Therefore, even if the textbooks choose characters according to the commonly used character list, some "commonly used characters" that do not conform to the current language life and are not considered high frequency will be selected.

Based on the opinions of various schools and the specific survey conducted in this study, it found that the number of commonly used modern Chinese characters is about 2500, which can cover 99.18% of the corpus of Guangming Daily for the past 15 years from 2008 to 2022 (the first half of the year) and 99.36% of the corpus of the "2020 Annual Media Word Master List" (the Green Paper of 2021), which can already meet people's needs.

### **3. THE SPECIFIC CHARACTER TYPES IN THE LIST OF COMMONLY USED MODERN CHINESE CHARACTERS NEED TO BE UPDATED**

In this study, a series of investigations and analyses were conducted on the specific character types of the "Modern Chinese Commonly Used Character List" and the "General Standardized Chinese Character List", and it was concluded that the specific character types of the commonly used character list need to be updated.

#### ***3.1 Building a Corpus based on News and Texts in Guangming Daily to Statistically Analyze Text Usage***

The Chinese characters used in Guangming Daily from 2008 to 2022 (the first half of the year) were ranked in descending order of frequency, resulting in a total of 345,451 documents, 43,244,258 characters, 360,240,272 words, and 10,556 Chinese character types in the "General Table". The corpus of characters used by newspapers and periodicals involved in the "General Table" is large in scale and long in duration. It objectively reflects the general situation of characters used by newspapers and periodicals in the past 15 years, and can provide a reference for the formulation and updating of commonly used character lists for modern Chinese characters. By comparing the characters ranked in the top 3500 frequencies of the "General Table" with the 3500 words of the "General Standardized Chinese Character List" (Level 1 character list), it can be found that the two lists share 3137 characters, while 363 characters of the "General Standardized Chinese Character List" (Level 1 character list) fall outside the 3500 frequencies of the "General Table". These characters may not be considered high-frequency characters in current language and writing life, so the specific characters in the Level 1 character list need to be updated.

#### ***3.2 An In-depth Analysis of the "Ten Years of Summary Table of Media Terms" Based on the Limitations of the Corpus of the Guangming Daily***

The 10-plus-year "Annual Media Character Master List" in the "Report on the State of Language Life in China" from 2011 to 2021 can objectively show the current frequency of Chinese

character usage. Therefore, this study compares the "General Standard Chinese Characters Table" (Level 1 character list) with the top 3,500 characters used in the general list of media used in the Green Paper for more than 10 years and analyzes whether the "General Standardized Chinese Characters Table" (Level 1 character list) is not in line with the actual situation of the current language life. Statistics found that 413 of the 3,500 types of characters in the "General Standard Chinese Character List" (Level 1 character list) are not among the top 3,500 types of characters used in the general list of media used in the Green Paper for more than 10 years, indicating that these 413 characters no longer meet the standard of commonly used characters in current actual language life. That is, the frequency of use is not high.

"General Standard Chinese Character List" (2013) has been partially updated for the "Modern Chinese Commonly Used Character List" (1988), replacing 103 characters in it. Comparing the 103 characters updated in the "General Standard Chinese Character List" with the top 3500 characters in the frequency of use in the media list for more than 10 years, it is found that 18 of the 103 characters are still not among the top 3500 characters in the media list for more than 10 years: 畎(chǒu)、咄(duō)、癸(guǐ)、萑(huāng)、叩(kòu)、婪(lán)、矜(jīn)、卯(mǎo)、瑙(nǎo)、啪(pā)、壬(rén)、巳(sì)、迢(tiáo)、戍(wù)、戍(xū)、貽(yí)、酉(yǒu)、哉(zāi). This shows that there are still problems with the specific type of characters selected in the "General Standardized Chinese Character Table" (2013). A considerable number of Chinese characters that are not currently used frequently are selected.

#### **4. THE AWARENESS OF THE STANDARDIZATION OF TRADITIONAL AND VARIANT CHARACTERS NEEDS TO BE STRENGTHENED**

The implementation of language standardization has always been the focus of language standardization, with the press as the main forum. In 1956, the People's Daily published an editorial "To the Readers", pointing out that "newspapers are published every day, and they use tens of thousands of characters to influence millions of readers every day". This shows that the standardization of the use of characters in the press has always been an important issue of public concern.

Guangming Daily, the same age as the Republic, was founded on June 16, 1949. As a bridge and link between the Party the state and the intellectuals, it has played a unique and important leading role in the development of language and writing in China at all historical stages. Since its inception, it has always attached importance to the standardization of language and characters; one of the three special language and writing magazines created once, "Language and Characters", opened the classic column "Seeking Faults", which "analyzes the irregularities in the use of language and character with the force of a lion fighting a rabbit"; On May 8, 2005, Guangming Daily even published "Some Common Errors in Our Newspaper", and corrected itself with a broad-minded and realistic attitude. It can be said that Guangming Daily has been the vanguard, leader, practitioner, and promoter of language standardization work. But there is a flaw in the jade, and the flaw does not cover up the yogurt, and the same is true of Guangming Daily.

A comprehensive survey and analysis of the use of traditional characters and variant characters in Guangming Daily from June 2008 to 2022 have been conducted. In addition to the six scenarios prescribed by the 'Law of the People's Republic of China on the Common State Language and Character', a comprehensive investigation and analysis of the usage of traditional characters and variant forms has been undertaken. (2014, Article 17), there are also some irregularities in the use of traditional and variant characters in Guangming Daily, mainly including three cases of irregular cases, such as the mixed-use of traditional and simple characters/proper and variant characters when quoting, the mixed-use of traditional and simple characters/proper and variant characters when writing, and the mixed-use of traditional/variant characters.

#### **4.1 A Survey of the Use of Traditional Characters in Guangming Daily in the Past Fifteen Years**

In this study, the word usage statistics of Guangming Daily in the past fifteen years from January 1, 2008, to June 1, 2022, were conducted, involving 345,451 news articles, with a total corpus of 43,244,2658 characters, 360,240,272 total characters, and a total of 10,556 character types. The number of traditional Chinese characters used in Guangming Daily in the past fifteen years amounted to 498 characters, reaching 2833 total number of characters, involving 1307 articles.

Among them, a total of 124 traditional Chinese characters are irregular in use, accounting for 24.89% of the number of traditional Chinese

characters. The number of traditional characters used in Guangming Daily by year is shown in the following “Figure 1”.

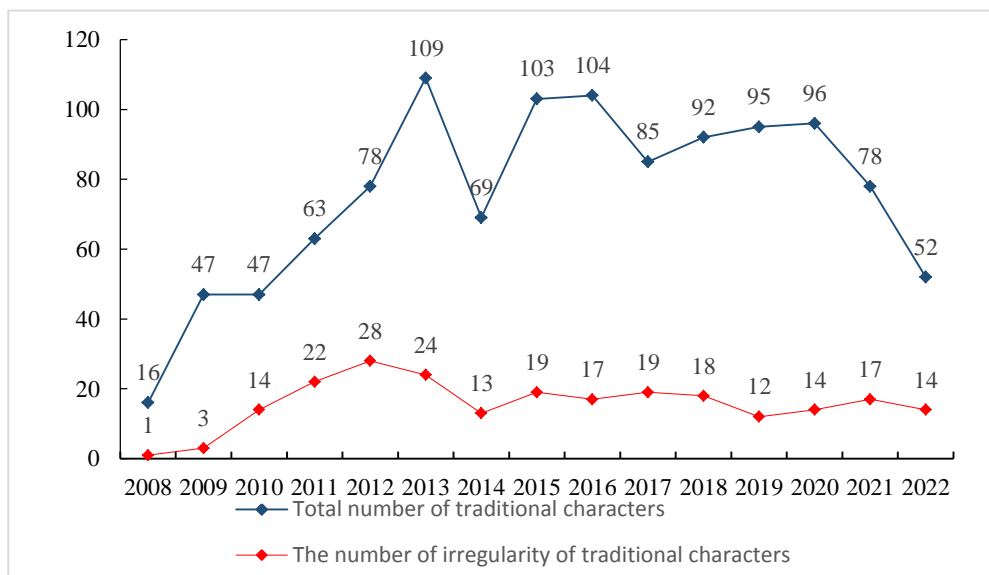


Figure 1 Statistics on the Number of traditional characters used in Guangming Daily from 2008 to 2022 (first half of the year).

As shown in “Figure 1”, the number of traditional characters used in Guangming Daily from 2008 to 2022 (the first half of the year) fluctuates greatly in general: the number of traditional characters used in the six years from 2008 to 2013 shows a rapid growth trend, increasing from 16 to 109 with a total of 93 characters; after 2013, the number of traditional characters used dropped sharply, with a difference of 40 characters from 2014; the number of traditional characters used in the seven years from 2015 to 2021 shows a trend of first decline and then rise, stabilizing at about 90 traditional characters; only the first half of 2022 is counted, so the number of traditional Chinese characters used is relatively small.

At the same time, “Figure 1” shows that the traditional characters used in Guangming Daily are partially out of standard every year. In 2012, the number reached 28, accounting for 35.90% of the total number of traditional characters. In the past 15 years, only in 2008 and 2009, there were fewer irregularities in the traditional characters used in Guangming Daily, while in the rest of the years, a certain number of traditional characters involved irregularities.

#### 4.2 A Survey of the Use of Variant Characters in Guangming Daily in the Past Fifteen Years

In this study, the word usage statistics of Guangming Daily in the past fifteen years from January 1, 2008, to June 1, 2022, were conducted, involving 345,451 news articles, with a total corpus of 43,244,2658 characters, 360,240,272 total characters, and a total of 10,556 character types. The number of variant characters used in Guangming Daily in the past fifteen years amounted to 332 characters, reaching 7723 total number of characters, involving 4783 articles. Among them, a total of 192 variant characters are irregular in use, accounting for 57.83% of the number of traditional characters. The number of variant characters used in Guangming Daily each year is shown in the following “Figure 2”.

As shown in “Figure 2”, the number of variants used in Guangming Daily from 2008 to 2022 (the first half of the year) fluctuates greatly: the number of variants used in the six years from 2008 to 2013 shows a surge, increasing from 37 to 112 characters with a total increase of 75 characters; the number of variants used from 2013 to 2014 decreases more, with a difference of nearly 30 characters; the number of variants used after 2015 is relatively stable, fluctuating between 88 and 109 characters,



and most years the number of variants used is above 100; the number of variants used in 2022 is relatively small because it only covers the first half of the year. The number of variant characters used after 2015 is relatively stable, fluctuating between

88 and 109 characters, with most of the variant characters used in most years above 100; 2022 is only counted in the first half of the year, so the number of traditional Chinese characters used is relatively small.

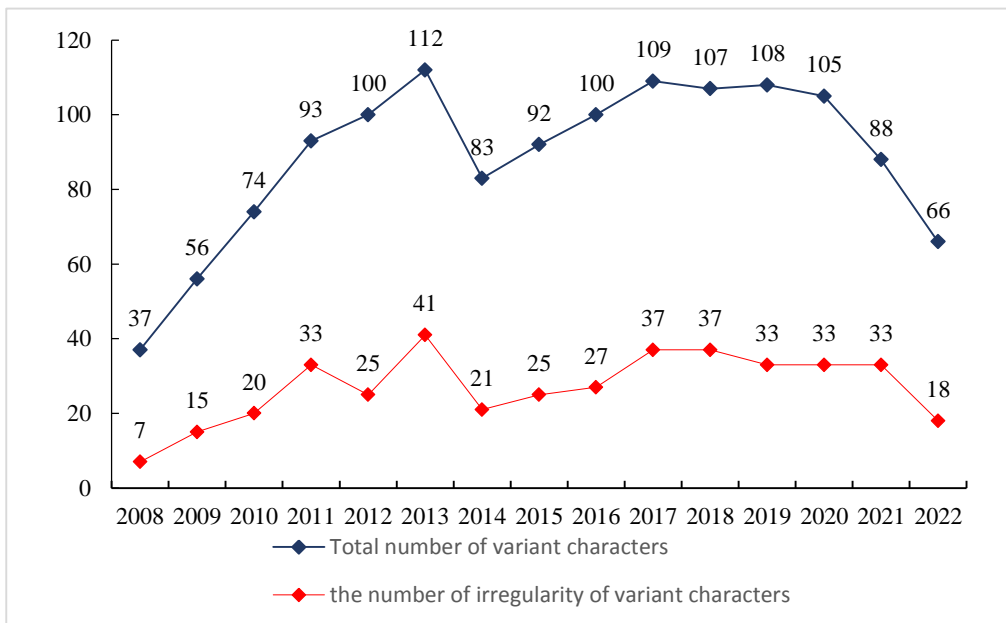


Figure 2 Statistics on the number of variant characters used in Guangming Daily from 2008 to 2022 (first half of the year).

At the same time, “Figure 2” reveals that each year, the variant characters used in Guangming Daily are partially out of standard, and the number of variant characters involved in the irregularity of variant characters in 2013 reached 41, accounting for 36.60% of the total number of variant characters in that year. In the past 15 years, a certain number of variant characters were involved in irregularities in each year of Guangming Daily.

The study conducted a comprehensive survey and analysis of the use of traditional and foreign characters in Guangming Daily for nearly 15 years from 2008 to June 2022. More extensive field surveys on the use of traditional characters, foreign characters, misspelled characters, variant characters, and other non-standard social characters, as well as analysis of the causes and countermeasures, will be further explored in depth. In an attempt to better promote the public’s common concern for language standardization work.

## 5. CONCLUSION

From the above research, it shows that the current direction of Chinese character reform is not pinyinization, but rather the simplifying of

characters (while making better use of pinyin) and the standardization of their use. Modern Chinese characters have an important role to play in China’s efforts to achieve informatization and internationalization. And it is the mission of contemporary scholars to focus on the actual use of characters and their future. Investigating the current use of characters and regulating the application of language and character is also part of the research on the innovative development of the excellent Chinese character culture, which can provide a reference for decision-making on the setting of commonly used modern Chinese characters, and play a good role in promoting awareness of the use of characters in society and defending language and writing norms.

## ACKNOWLEDGMENTS

This study is funded by the following projects.

1. 2020 Project funded by the Hubei Provincial Philosophy and Social Science Research Major Project (Hubei Provincial Philosophy and Social Science Fund Preliminary Funding Project) named "Research on the Integration of Excellent Chinese

Character Culture into Primary and Secondary School Education" (20ZD049).

2. 2022 Joint Chinese and foreign special project on international Chinese language education research organized by the Centre for Sino-foreign Language Exchange and Cooperation of the Ministry of Education, "A graded Chinese character book 'Chinese Characters Can Talk' based on the *Standard for Chinese Proficiency in International Chinese Language Education*. (22YH29ZW).

3. 2021 Hubei Higher Education Teaching Research Project "Interesting Stories of Chinese Characters - Construction of First-class Curriculum for Social Practice of "Chinese Characters and Culture" (2021323).

4. "Research on the Integration of Excellent Chinese Character Culture into Hubei Higher Education", Intramural Education Reform Program of Wuhan Institute of Technology (X2018049)

5. Humanities and Social Sciences Foundation of Wuhan Institute of Technology, Research on Collaborative Innovation of Chinese and Foreign Languages and Cultures" (R202107).

6. The 14th Postgraduate Education Innovation Fund of Wuhan Engineering University: "Foreign translation of Chinese academic achievements: the example of the translation of "Chinese Character Study" by teachers and students" (CX2022496).

7. 2022 "National Student Innovation and Entrepreneurship Training Program" provincial project in Hubei Province, "Survey and Research on Chinese Character Usage in the Era of Big Data and Cloud Computing" (S202210490068)

8. 2022 "National Student Innovation and Entrepreneurship Training Program" provincial project in Hubei Province, "' Word in the World' - Fun Chinese Characters Situation drama" (S202210490064).

9. 2022 National Project of "National Students' Innovation and Entrepreneurship Training Program": "Volunteer services for children with language difficulties" (202210490022).

10. The 14th Postgraduate Education Innovation Fund Grant Project of Wuhan Engineering University "English Translation of Li Shoukui's One Hundred and Fifty Lectures on Explaining and Interpreting Chinese Characters" (CX2022497).

11. The 14th Postgraduate Education Innovation Fund Project of Wuhan Institute of Technology, "Chinese Scholarship in Foreign Translation: The

Case of English Translation of *Interesting Talks on Ancient Chinese Characters* written by Weizhan Chen" (CX2022500).

## REFERENCES

- [1] Zhou Youguang. *Chaowendao ji* [M]. Beijing: World Book Publishing Company, 2014.
- [2] Feng Zhiwei. *Modern Chinese characters and computers* [M]. Beijing: Peking University Press, 1989.
- [3] Qiu Xigui. *Outline of philology* [M]. Beijing: The Commercial Press, 2006.
- [4] Su Pei-Cheng. *Outline of Modern Chinese Characterology* [M]. Beijing: The Commercial Press, 2014.
- [5] Su Pei-Cheng. *Language reform and language regulation in contemporary China* [M]. Beijing: The Commercial Press, 2010.
- [6] Department of Language and Writing Information Management, Ministry of Education. *New Development of Chinese Character Specification in the Information Age: A Collection of Literature on the Universal Standardized Chinese Character List* [M]. Beijing: The Commercial Press, 2015.
- [7] Editor-in-Chief Wang Ning, Verified by Li Yuming and Wang Tiekun. *Interpretation of the General Standard Chinese Character Table* [M]. Beijing: Commercial Press, 2013.
- [8] Wang Tiekun, Hou Min, Yang Erhong. *A survey on the use of words in newspapers, radio and television, and the Internet* [J]. Language and writing applications, 2007(01):29-37.
- [9] Wang Tiekun, Hou Min. *Language life in China from the 2008 survey data*[J]. Language Applications, 2010(02):26-35.
- [10] National Center for Language Resources Monitoring and Research. *China's Language Life Report 2011-2021* [M]. Beijing: The Commercial Press, 2011-2021.