

# Diachronic Lexical Features of Chinese Government Work Reports: A Text Mining Approach

Yan Jiang

*Shandong Technology and Business University*  
Email: yuqingcool423@163.com

## ABSTRACT

As important political discourses, Chinese Government Work Reports are of great significance in conveying the government's policies and hence building national images. However, few studies have investigated the diachronic changes of lexical features in Chinese Government Work Reports. This study employed a text mining method to investigate high frequency words and changes of word use in terms of frequency in Chinese Government Work Reports across 20 years, which can be divided into four Five-Year-Plan periods. It is found that the continuity and changes of word use in terms of frequency reflect the governance philosophy of the Chinese government, such as governing the country according to law, administering the party strictly, taking the people as the foundation, keeping pace with the times and developing peacefully, which contributes to the diachronic construction of a responsible country. The development concept and diplomatic concept of our government are important factors for the success of these images.

**Keywords:** *Chinese Government Work Report, text mining approach, lexical features*

## 1. INTRODUCTION

Chinese government work report is an important political discourse, and its text structure mainly reviews the work of the previous year and arranges the tasks of the current year. The government work report conveys the governance notion of the Chinese government to the outside world, which is an important source for the international community to understand China's major policies and social and economic development [1], and also an important medium for building a national image [2]. It is of great significance to analyze the Chinese Government Work Reports.

In view of the significance of the Chinese Government Work Reports, the current literature has analyzed the Chinese government work report from different perspectives. Some scholars took The State of the Union Address of the same period as the reference corpus, and used corpus technology to compare the lexical and syntactic features of Chinese Government Work Reports [1][2]. Some scholars have pointed out that the application of translation techniques based on readers' acceptance will narrow the distance between the reports and overseas readers, and enhance the acceptability of texts[3]. Other scholars have pointed out that the application of translators' techniques is helpful to build a down-to-earth and enterprising national image of China

by examining the use of high-frequency notional words, first-person plural pronouns and modal verbs in the Chinese Government Work Reports from 2000 to 2016 [2].

The existing literature provides various perspectives for further investigation of the Chinese Government Work Reports, and shows that the approach with word use as a starting point is feasible and reliable. However, the current studies on Chinese Government Work Reports mainly focus on the use of translation strategies, while the diachronic studies on word use in Chinese Government Work Reports are rare. A diachronic study of word use in Chinese Government Work Reports can reflect the development and changes of Chinese society and reflect the subtle changes of Chinese government in governance philosophy. Therefore, this study adopts a text mining method to analyze word use in Chinese Government Work Reports from 2001 to 2020. The application of text mining technology can form objective and accurate criteria and quantitative basis, and help to overcome the defects of subjectivity of researchers in the traditional text analysis process. By examining the high frequency notional words and changes of word use, this paper analyzed the diachronic lexical features of Chinese Government Work Reports and the historical evolution of China's national image constructed in the reports in the past 20 years.

## 2. METHODS

### 2.1. Research questions

Our research questions are as follows:

- (1) What are the high frequency words in Chinese Government Work Reports in the past 20 years?
- (2) Are there variations in Chinese Government Work Reports in terms of words' frequencies?
- (3) What kind of continuity and changes in China's national image are reflected from the above questions?

### 2.2. Corpus

In order to answer the above research questions, a total of 20 Chinese Government Work Reports from 2001 to 2020 were downloaded from the official websites of China and were saved as plain text documents. Then a corpus of these reports was compiled, totalling 292,091 words. A text mining method was used to quantify and statistically analyze the corpus of Chinese Government Work Reports in the past 20 years.

### 2.3. Data analysis

The R programming language (version 4.0.3) was used for data analysis and ggplot2 for data visualization [4]. First, the corpus of 20 government work reports was preprocessed, including removing stop words and cleaning text. The steps of text mining in this paper are as follows: Then, the frequency of each notional word in the English version of the annual government work report and the total frequency of each notional word in 20 English versions are counted. Then the glm () function was used to statistically model the counting data [5], in which the year was the independent variable and two dependent variables, namely the frequency of a word in the annual government work report and the total vocabulary of each year, so as to statistically model every word with a total frequency exceeding 200 in 20 government work reports. According to the statistical modeling analysis results of corpus data, we used the software AntConc 3.5.9 [6] to examine the context of words whose frequency of use has changed significantly, and then investigate the evolution of the national image reflected by the significantly changed real words and their collocation in the reports in the past 20 years.

It should be pointed out that all statistical operations in this study were based on the preprocessed texts. The data processing and visualization of this study are completed by R, an open source software.

## 3. RESULTS AND DISCUSSION

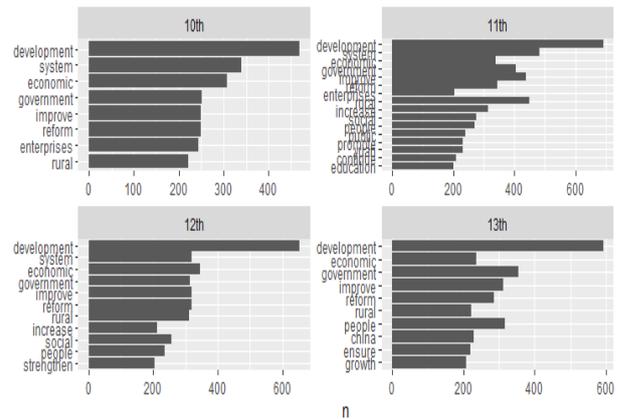


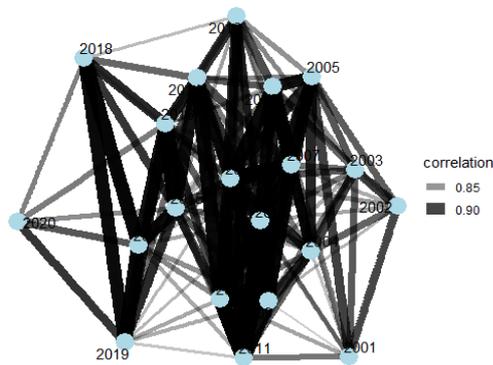
Figure 1 The most common words in Chinese Government Work Reports

As can be seen in Figure 1, among the most common words that appeared more than 200 times in the Chinese Government Work Reports of each Five-Year-Plan period, development, system, economic, government, improve, reform, enterprises, rural are the recurrently high frequency words. It clearly shows that China has consistently attached top priority to economic development and persisted in the policy of reform and opening up. Besides, the repeated high frequency of rural shows that promoting development in the rural area has been top priority in the past 20 years for the Chinese government. The Chinese government attaches great importance to the coordinated development of urban areas and the planning and implementation of rural policies. On the whole, these highly frequent words shared by Chinese Government Work Reports in the four Five-Year-Plan periods has constructed China's image as attaching great importance to economic development, reform and opening up and coordinated development of urban areas.

In addition to the consistency of adhering to basic policies illustrated above, it is also evident that the coverage of words whose frequency exceeds more than 200 times expanded in the reports during the latter periods. Relevant issues include people, public, education in the 11<sup>th</sup> and 12<sup>th</sup> Five-Year-Plan period respectively. It means that under the leadership of the government, the people's status has steadily improved and that more attention has been shifted to education. Similarly, the high frequency of China in the reports of the 13<sup>th</sup> Five-Year-Plan period may be related to the gradual enhancement of self-awareness of China's national status and an emerging national image of China as more confident in an all-round way.

It can be inferred from Figure 1 that the philosophy of the Chinese government in the past 20 years is stable and continuous, and the policies implemented are consistent and unified, always focusing on issues such as economy,

people’s livelihood and development. This assertion has been verified by the results of correlation analysis of these sets of word frequencies presented in Figure 2.



**Figure 2** Correlation of Chinese Government Work Reports from 2001 to 2020

For further analysis, we calculated the correlation coefficients between the reports of different years. Figure

2 shows the correlation coefficients between the reports in the past 20 years. Note that only coefficients which exceed 0.8 were filtered for visualization. The threshold 0.8 was chosen because statistical results revealed that all correlation coefficients between each report was highly correlated. This figure, once again, shows that the steps and policies of the Chinese government in the past 20 years have been stable and continuous, which has always focused on economic and social affairs.

The section above reported the results of high frequency words. Now let us turn to answer questions 2 and 3. In order to investigate the specific frequency changes of the words used in the reports in the past 20 years, we only kept words used at least 200 times in each report. For this count data, we used glm () with family = “binomial” for modelling. As many slopes were compared, and some of them were not statistically significant, we applied an adjustment to the p-values for multiple comparisons. The adjusted significance level was set at 0.05.

**Table 1.** Words with significant frequency changes

accelerate	conditions	health	people	services
administrative	create	improving	poverty	social
agricultural	economic	industries	production	socialist
agriculture	education	innovation	projects	stable
billion	employment	insurance	reforms	strong
business	ensure	job	region	structure
businesses	enterprises	management	resources	supply
China	farmers	military	science	system
China’s	grain	million	scientific	technological
Chinese	growth	party	security	technology

The results of statistical modeling show that the frequency of 50 words that were used more than 200 times in the reports have changed significantly in Chinese Government Work Reports in the past 20 years ( $p < .05$ ) as presented in Table 1. It can be seen that these words with significant changes in terms of frequency cover many areas, including economic, political, educational, social, and international areas. The rise and fall of the frequency of these different words and the specific context each word is situated in can reflect the subtle changes in the government’s governing notions and describe the goals of our country in respective fields in the 21<sup>st</sup> century. Limited by space, this paper cannot discuss the changes in great detail. We just focus on the high frequency nouns, because nouns can more clearly highlight the issues of concern to our government and reflect the concerns or tendencies in the Chinese government work report over the years.

Combined with contextual analysis, we found that these nouns appeared not only in the review of government work in the previous year, but also in the

arrangement of major tasks of government work in that year. In some circumstances they appeared together with modal verbs such as will, must and/or should, indicating that the annual Chinese government work report not only states the facts, but also expresses the firm determination of the Chinese government to realize the aims of that year. The diachronic change of these words not only reflects the diachronic change of China’s outlook, but also indicates that the government has always updating its governance concept, and that a diachronic national image is constructed.

In terms of political aspects, our government has always attached great importance to self-reform and self-improvement. Administering the party strictly in an all-round way has been pushed forward in depth, with remarkable anti-corruption achievements, comprehensive and in-depth implementation of the rule of law, deepening reform of the administrative system, and making positive progress in government capacity building. These aspects show that the Chinese government attaches great importance to boldly

displaying political ideas with Chinese characteristics in the government work reports.

Regarding economic aspects, promoting development has always been the top priority of our government in governing and rejuvenating the country. By establishing an effective corporate governance structure, the reform of state-owned enterprises has been continuously promoted. Measures have also been taken to encourage the development of small and medium-sized enterprises. The word changes in Chinese Government Work Reports from 2001 to 2020 reflect China's confidence in developing its own economy and promoting the development of the world economy as a big developing country.

In terms of people's livelihood, the Chinese government has always practiced the idea of people-centered and constantly solved all kinds of problems related to the vital interests of ordinary people, such as education, employment, social security and medical care. Such issues have received more and more attention, and the measures taken have become more and more specific, which reflects the people-oriented governance concept of our government.

#### **4. CONCLUSION**

This study investigated the lexical features and their diachronic changes in Chinese Government Work Reports from 2001 to 2020 through a text mining approach. High frequency notional words were found constant in the four Five-Year-Plan periods, which shows the consistency and continuity of Chinese government's policies. Besides, the frequency of some notional words used was found to vary in the examined government reports. The top highly-frequent notional words as well as the words that changed in frequency in the reports over

the years both reflect the national image of the Chinese government, solving problems in a down-to-earth way and adhering to the road of win-win development through cooperation. We thus argue that the change of Chinese government's governance concept is an important cause of the diachronic changes of the words identified in this study, and a decisive factor for the successful construction of China's national image.

#### **AUTHORS' CONTRIBUTIONS**

Yan Jiang: Conceptualization, Methodology, Writing – original draft, Visualization, Writing – review & editing.

#### **REFERENCES**

- [1] G. Wu, A corpus based analysis of the stylistic features of Chinese government work report [J]. *Journal of Xi'an International Studies University*, 2014, (4): 118-121.
- [2] K. Hu, X. Tian, A corpus-based study of the linguistic features and textual effect of the English translation of the government work report [J]. *Foreign Language and Literature*, 2018, (5): 1-11.
- [3] Y. Wang, A study of tone weakening in the English translation of government work report [J]. *Chinese Translators Journal*, 2016, (5): 110-113.
- [4] H. Wickham, *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag, 2009.
- [5] J. Silge, D. Robinson, *Text Mining with R: A Tidy Approach*. Boston: O'Reilly Media, 2017.
- [6] L. Anthony, *AntConc (Version 3.5.9)* [Computer Software]. Tokyo, Japan: Waseda University. From <https://www.laurenceanthony.net/software>, 2020.