Text Analysis of Cross-border E-commerce Policy Based on Co-word Clustering Method: A Case Study of Gansu Province

Yuzhen Wang 1* and Jiaru Li 1

1 School of Information Engineering, Lanzhou University of Finance and Economics, Lanzhou, Gansu, 730020, China

*Corresponding author’s e-mail: wyz70214@163.com

Abstract. Taking the cross-border e-commerce policy texts introduced in the past five years as the research object, the co-word clustering method is used to divide the policy texts into four categories, and the content is analyzed based on the theory of the ecosystem. The study found that Gansu Province currently has unbalanced cross-border e-commerce policy structures, insufficient specific policies issued by prefecture-level city governments, and weak linkages among government departments in issuing policies. Therefore, the relevant departments need to further improve the cross-border e-commerce policy in response to the above issues, so as to promote the high-quality development of cross-border e-commerce in Gansu Province.

1. Introduction

Cross-border e-commerce is the product of the evolution of e-commerce to market segmentation. Under the background of rapid development of Internet economy and globalization of trade, the development of cross-border e-commerce, on the one hand, it provides new impetus for the transformation and upgrading of logistics, payment and other industries; on the other hand, it promotes the global circulation of domestic commodities and accelerates the development of open economy. Hence, the central and local governments have successively introduced cross-border e-commerce policies to promote rapid, standardized and orderly development in this field. As an important province along the "Belt and Road" and an important gateway for China to open to the west, Gansu Province has made the development of cross-border e-commerce an important part of regional economic upgrading, and has successively issued a series of policies to boost the development of cross-border e-commerce. However, cross-border e-commerce is a new thing and its update and development are fast. Therefore, cross-border e-commerce policies need to be continuously optimized and improved. By analyzing the text of the cross-border e-commerce policy, we can find out the deficiencies of the cross-border e-commerce policy that have been issued, so as to provide a basis for formulating a more comprehensive, systematic and perfect policy system.

Due to the importance of cross-border e-commerce policies, there have been certain research results in this field, mainly related to the construction of cross-border e-commerce policy systems and text analysis of cross-border e-commerce policies. In terms of building a cross-border e-commerce policy system, Zhang H believed that the establishment of a fair market environment was the direction of policy improvement, and put forward the formulation of more functional industrial policies and the coordination of multiple policies to form a policy synergy to ensure the effectiveness of policy implementation[1]; Wang X.H et al found that the policies of taxation, logistics, customs and legal supervision in the development of cross-border e-commerce were incomplete, and designed the system for the above four aspects from the perspective of business innovation[2]. In terms of policy text
analysis, Hou Z.X et al carried out a quantitative analysis of the agricultural product e-commerce policy text from the two-dimensional perspective of policy tools and ecosystem theory, and made suggestions for the development of regional rural e-commerce[3]; Wu A.P et al took innovation and entrepreneurship policy as the research object and used co-word clustering to quantify the text content, and then intuitively analyzed the overall structure and characteristics of the policy[4]; Zhang B.J et al used the K-means method to perform a cluster analysis on the keywords of national science and technology innovation policy content, and made targeted suggestions for policy improvement[5], and so on.

It can be seen that relevant scholars have made some research on the construction of policy system and policy text analysis, which provides an important theoretical basis for this paper. However, the existing research results are mainly studied from the perspective of the whole country or developed provinces, and there are few research results specifically targeted at the underdeveloped Gansu province. Therefore, by collecting the relevant policies of cross-border e-commerce issued by the government departments of Gansu Province in the past five years, this paper finds out the deficiencies of the cross-border e-commerce policies that have been issued, so as to put forward optimization suggestions in line with the actual situation of Gansu Province.

2. Collection and pre-processing of cross-border e-commerce policy texts

2.1. Data Sources
This article takes the cross-border e-commerce policy text as the research object and collects data through the following two channels: First, the government website search. On some official websites such as the Gansu Provincial Government and the Provincial Department of Commerce, the policy text is retrieved with "cross border e-commerce" as the search term; The second is the search of the policy database. Through the use of the third-party policy database, the subject words related to the development of cross-border e-commerce, such as "cross-border e-commerce", "foreign trade", are searched. A total of 97 cross-border e-commerce related policies in Gansu Province were retrieved through the above two channels, and relevant policies such as notices, announcements, and implementation opinions issued in the past five years were selected, and duplicate policy documents with low relevance were eliminated. Finally, after screening and sorting out, 41 policies were selected as research samples, including 33 policies issued by the Gansu Provincial Government. The Development and Reform Commission, the Department of Finance, the Department of Commerce, the Lanzhou Customs, the Foreign Exchange Department, and the Taxation Department (including joint issuance document) were respectively 3, 4, 4, 5, 1 and 2, accounting for 5.77%, 7.69%, 7.69%, 9.62%, 1.92% and 3.85% of the total number of policies.

2.2. High-frequency word extraction
First of all, the word frequency statistics of the above 41 policies are used to obtain a total of 1175 policy words with different word frequencies, of which 573 are words with a frequency of 1. Then, these words need to be normalized. Finally, the high frequency words are found out according to the high and low frequency word boundary equation. The calculation equation is as follows [6]:

\[ T = \left[ \frac{(1+bI_1)^2-1}{2} \right] \]

In equation (1), \( I_1 \) represents the total number of words that appear once; \( T \) represents the boundary of high and low frequency words. Through calculation, \( T \) is equal to 33, so 33 is the boundary to determine the high frequency words studied in this paper. The selected high-frequency words are shown in Table 1.

| Num | Selected Word  | Frequency | Num | Selected Word | Frequency |
|-----|----------------|-----------|-----|---------------|-----------|
| 1   | Service Provider | 389       | 18  | Bonded       | 63        |
| 2   | Platform        | 272       | 19  | Financing     | 60        |
3. Text analysis of cross-border e-commerce policies

3.1. Construct co-word matrix
The co-word matrix is to count the number of times that two pairs of words appear in the same text, and then analyze the changes of the theme structure they represent. Therefore, based on the high-frequency words obtained in Table 1, the number of times that each word appears together in the policy text is counted, and a 33 × 33 co-word matrix is generated, limited to the length, and only partial results are listed, as shown in Table 2.

Table 2. Co-word matrix of high-frequency words (part)

| Service Provider | Platform | Trading Port | Manufacturer | Commodity | ... |
|------------------|----------|--------------|--------------|-----------|-----|
| Service Provider | 0        | 17           | 3            | 9         | 8   |
| Platform         | 17       | 0            | 2            | 3         | 7   |
| Trading Port     | 3        | 2            | 0            | 3         | 16  |

3.2. Cluster analysis
In order to facilitate the next clustering analysis, we first need to transform the co-word matrix into the similarity matrix. The similarity matrix is a matrix describing the correlation between word pairs in order to eliminate the impact caused by frequency discrepancy. In this paper, the cosine similarity is used to find the high-frequency word similarity matrix. According to the similarity matrix, the Ward hierarchical clustering algorithm is used to classify 33 high-frequency words into different groups to reflect cross-border e-commerce policy concerns and text content involved, as shown in Figure 1.
According to the clustering results in Figure 1 and the theory of business ecosystem [7], the high frequency words are divided into four categories by selecting the horizontal distance range from 2 to 6. The first category is the central group, which includes 4 high-frequency words such as "service provider" and "platform"; The second category is the critical group, including 4 high-frequency words such as "manufacturer" and "retailer"; The third category is the attached group, which includes 6 high-frequency words such as "talent" and "brand"; The fourth category is supporting group, including 19 high-frequency words such as "trading port" and "commodity". The specific results are shown in Table 3.

### Table 3. Cross-border e-commerce policy text high-frequency words ecosystem structure

| Central Group          | Critical Group          | Supporting Group        | Attached Group       |
|------------------------|-------------------------|-------------------------|----------------------|
| Service Provider       | Manufacturer            | Trading Port            | Talent               |
| Platform               | Retailer                | Commodity               | Brand                |
| Pattern                | Foreign Enterprise      | Government              | Marketing            |
| Supervision            | Free Trade Zone         | Logistics               | Culture              |
|                        |                         |                         | Tourism              |
|                        |                         |                         | Insurance            |
|                        |                         |                         | The Silk Road        |

### 3.3. Ecosystem perspective analysis

As shown in Table 3, we can find out that Gansu's cross-border e-commerce policy system basically involves various groups of the ecosystem. Next, the proportion of high-frequency words in each group is statistically analyzed, and the results are shown in Table 4. The following sections analyze the text of cross-border e-commerce policies from four aspects: central group, critical group, supporting group and attached group.
3.3.1. Central Group
The central group is a leader in the entire cross-border e-commerce ecosystem and plays a role in integrating and coordinating the entire system resources. From the clustering results of high-frequency words, it is concluded that service provider, platform, pattern, and supervision are classified as central group, accounting for 20.11%, indicating that relevant policies pay more attention to them. First of all, in the cultivation of central group, the government has taken distinctive measures to encourage various service providers such as information technology, logistics, and finance to participate in the construction of cross-border e-commerce platforms. At the same time, it supports well-known enterprises to build or jointly build cross-border e-commerce platforms with local characteristics and brands, so as to develop new patterns of platform operation according to local conditions. Moreover, the government has enhanced supervision measures such as product quality and corporate qualification audits to attract more enterprises to actively launch cross-border e-commerce business, and then continue to expand the size of the central group, providing professional platform support for the development of cross-border e-commerce.

3.3.2. Critical Group
The critical group is the subjects of cross-border e-commerce transactions, including individuals or organizations involved in commodity exchanges such as retailer, manufacturer, and supplier, serving other groups. From the clustering results of high-frequency words, the critical group is composed of manufacturer, retailer, foreign enterprise and free trade zone, accounting for 14.11%. In terms of cultivating critical group, first, for the product circulation links of deep processing and sales, actively cultivate cross-border e-commerce market players such as manufacturers and retailers. Second, take measures such as reducing operating costs, securing project land, and supporting specialty industries to attract foreign enterprises investment, and further connect with platform enterprises in cross-border e-commerce business. Third, increase the construction of international dry ports and international airports, replicate and promote the experience of other free trade zones reform pilots, and strive to establish a China (Lanzhou) free trade zone. It can be seen that relevant policies encourage and support manufacturers, retailers, foreign enterprises, and free trade zone, playing a key role in the scale development of cross-border e-commerce in Gansu Province.

3.3.3. Supporting Group
The supporting group refers to the organizations that must rely on for cross-border e-commerce transactions, including government departments, financial institutions, information technology institutions, etc. to jointly support the operation of the cross-border e-commerce ecosystem. From the clustering results of high-frequency words, it is concluded that government, logistics, and network technology, etc. are classified as supporting group, accounting for 56.01%, indicating that the development of cross-border e-commerce is in the initial stage as a whole, and relevant policies attach great importance to the construction of supporting group. The high-frequency words of the supporting group are subdivided again, and the analysis results are as follows: First, the government, taxation department, customs, quality inspection department, and foreign exchange department are government agencies and are the main bodies of policy-making. The introduction of relevant supporting policies has played a powerful guiding role in regulating the order of cross-border e-commerce markets. Second, logistics, network technology, trading port, machining equipment, and business environment belong to infrastructure. The relevant policies improve the cross-border e-commerce software and

| Ecosystem Group | Total Frequency | Proportion |
|-----------------|----------------|------------|
| Central Group   | 765            | 20.11%     |
| Critical Group  | 537            | 14.11%     |
| Attached Group  | 372            | 9.78%      |
| Supporting Group| 2131           | 56.01%     |
hardware infrastructure from the aspects of cross-border transportation route subsidies, network technology innovation and foreign trading port investment and construction. Third, financing, bonded, drawback, investment, payment, financial institution, and insurance belong to financial institutions or their related business terms. Relevant policies improve the cross-border e-commerce financial service system in terms of regulating third-party payment behavior, adjusting cross-border payment handling fee settlement, and increasing the upper limit of tax-preferred commodities. Finally, from the perspective of commodity, relevant policies popularize local high-quality products and promote the circulation of commodities from the aspects of brand building, construction of quality traceability system, and introduction of modern agricultural technology.

3.3.4. Attached Group
The attached group is provider of value-added services for cross-border e-commerce transactions, including brand marketing companies, cultural media service providers, consulting service agencies, etc. From the clustering results of high-frequency words, it is concluded that the attached group is composed of talent, brand, marketing, culture, tourism, and the Silk Road, accounting for 9.78%, indicating that relevant policies have insufficient support for them. Currently, Gansu's cross-border e-commerce policy mentions funding for talents and the development of cultural and creative products, but it lacks support for third-party service providers such as talent training, brand promotion, and marketing planning, resulting in the enthusiasm of the merchants engaged in the third-party service is less high.

4. Conclusion
This article uses the co-word clustering method and ecosystem theory to analyze the cross-border e-commerce policy texts issued by Gansu Province in the past five years. The study discovers that the following problems exist in the cross-border e-commerce policy system in Gansu: First of all, the cross-border e-commerce policy structure is not reasonable. Compared with the central and supporting groups, the policy support for the cultivation of critical group and attached group is obviously insufficient. Furthermore, there are fewer specific policies to support the development of cross-border e-commerce. At present, the cross-border e-commerce policies formulated by different government departments in Gansu Province are within the scope of regional development planning, which belong to the universality policy. The detailed implementation rules of the policy are implemented in the municipal governments at all levels, while the introduction of specific policies by the municipal governments at all levels is less. The last but not the least, the number of joint document issuance between governments is not large. According to the investigation, there is a significant gap in the number of policy-making between the provincial government and the other six departments, and the number of joint participation in policy-making among departments is small. Therefore, in view of the above problems, the relevant government departments should make corresponding improvements to the cross-border e-commerce policies and formulate preferential policies embraced by enterprises, so as to promote the high-quality development of cross-border e-commerce in Gansu Province.

References
[1] Zhang H. (2018) The multiple attributes and synergies of cross-border e-commerce policy. J. China Business and Market., 5: 64-74.
[2] Wang X.H., Li L.L. (2018) Business model innovation and policy system design of cross-border e-commerce under the new normal. J. Contemporary Economic Management., 7: 22-27.
[3] Hou Z.X., Lu Y. (2017) A quantitative study of regional agricultural product e-commerce policy texts: A case study of Gansu Province. J. China Business and Market., 11: 45-53.
[4] Wu A.P., Dong M., Li H., (2018) Analysis of "Internet +" and "Popular Entrepreneurship and Innovation" policy structure: Based on grounded theory and co-word analysis. J. Science and Technology Management Research., 10: 44-52.
[5] Zhang B.J., Li F.L., Chen J. (2019) Theme analysis and evolution process of national science and
technology innovation policy: From the perspective of text mining. J. Science of Science and Management of S. & T., 5: 1-20.

[6] Donohue J C. (1973) Understanding Scientific Literature: a bibliographic approach. Cambridge: MIT Press, USA.

[7] Moore J F. (1993) Predators and prey: a new ecology of competition. J. Harvard business review., 3:75-83.