A Corpus Analysis of Rural Representations in English-language News Media in China

Chen Fangfang[*]

The Hong Kong Polytechnic University, Yuk Choi Road, Hung Hom, Kowloon, Hong Kong

Abstract. China, the world's most populous country, has a sizable rural population. From this point of view, rural China's focus can serve as a lens into global rural issues and poverty. This article employs corpus analysis to examine media representations of 'rural' in China Daily newspapers from 2001 to 2021, focusing on rural areas. Findings show that the rural social system in the twenty-first century is more diverse and complex than in the previous century. Economic development is the most frequently discussed factor in rural transformation development (RTD). Additionally, unlike the idyllic life described in previous studies, the Chinese news media are relatively objective and forthright about negative development issues.

1 Introduction

Regarding social representations of the countryside, most research has concentrated on the United Kingdom, the United States, and Central Europe. While some scholars have argued that it is necessary to investigate the peculiarities of regions other than the Western world [1], there is still a significant gap in the literature on this subject in countries other than Anglo-American and Central European countries, with a particular emphasis on rural China.

This paper aims to fill the current research gap by conducting a corpus-driven analysis of Rural representations in English-language news media in China, supplemented by critical discourse analysis. As the world's most populous country, China has a sizable rural population [2], and numerous historic villages in China are undergoing a rapid transformation [3; 4]. Rural issues are a critical component of Chinese nation-building, as stated in the *State Council Opinions on Comprehensively Promoting Rural Revitalization and Accelerating Agricultural and Rural Modernization* issued by the Communist Party of China's Central Committee [5]. Simultaneously, the fact that China is home to 56 officially recognised ethnic groups contributes to our interest in the social representations of the Chinese countryside [6]. China's regional disparities in socio-economic development and geographical and biophysical conditions distinguish China's rural problems from other countries and regions. From this point of view, rural China's focus can serve as a lens into global rural issues and poverty.

The following is how this paper will be organised. Section 2 summarises pertinent prior literature and exposes the research questions. This section establishes the theoretical

---

[*] Corresponding author: fangfang.chen@connect.polyu.hk

framework and situates our study within existing research. The research methodology is described in Section 3, along with why we chose this method. We then describe how we collected and analysed the data. In Section 4, we summarise the qualitative and quantitative data analysis findings. Section 5 is our conclusion, perception of our research findings and procedure. Section 6 discusses the paper's potential limitations.

2 Literature review

2.1 Examine rural issues from a language and cultural perspective

Rural representations and discourses vary according to their history and culture. Rural areas are diverse, and individuals have their characteristics and life experiences shaping their rurality discourses, resulting in multiple and sometimes contradictory discourses [7]. Certain studies paint a picture of a contemporary rural idyll [8]. This highly idealised view of the countryside resulted in the beautiful land that academics are familiar with [9]. In contrast, several studies have examined whether the general rural population identifies with the idyllic rural life described above [10]. A survey of films set in rural Portugal [11] reveals the rural idyll's harsh reality, with the countryside portrayed as a space with multiple characteristics. Along with idyllic rural life, the countryside is also a place of doom and structural problems, including abandonment, depopulation, ageing, and forest fires. Additionally, a study examines the relationship between rural idyll and rural poverty [12]. Rural idylls are considered contributors to rural poverty[13; 14].

The Rural's social representation is a composite of personal experiences and beliefs promulgated through literature, communications, the nation, household, peers, and organisations [15]. The approach to defining the rural area is based on observing that the rural area and its synonyms are daily terms and concepts that people understand and use. For instance, interviews with UK residents mentioned six distinct interpretations of the term 'rural' [10]. Additionally, five different representations of rural life exist in Portuguese society, and the distinctions between the representations have their correlates [7]. According to previous research, the discourse has aided in redefining rural areas' responsibilities and activities [16]. As a result, we believe that analysing rural representations and associated discourses provides a valuable lens for considering these images' context and the factors that shape them.

2.2 Rural Transformation Development (RTD) as an emerging phenomenon

Rural restructuring occurred in the late twentieth and early twenty-first centuries in Western Europe, North America, and Israel in the Middle East [17]. Simultaneously, this transformation occurred in rural areas of developing countries such as China [18; 19]. According to the literature, 56% of China's population lived in rural areas in 2011 [2]. Regional differences in socioeconomic development, geography, and biophysical conditions are significant in China. This is why, compared to other countries, regions' synergistic and balanced development is a problem unique to China. Fortunately, once dominated by agriculture, the traditional centralised economy has increasingly become urbanised and industrialised. As illustrated in Figure 1, the number of rural-related news in the China Daily has increased almost annually over the last two decades.

China's new rural development strategy aims to achieve advanced production, improved livelihoods, cultural and moral advancement, clean and tidy villages, and efficient management [2]. This is consistent with the theoretical concept of rural development level (RDL), which serves as a proxy for changes in rural areas' social, economic, cultural,

natural resource, and environmental conditions. Considering the literature mentioned above, we will use categories such as economy, livelihoods, culture and education, rural environment, and governance and rules to frame rural representation in this study.
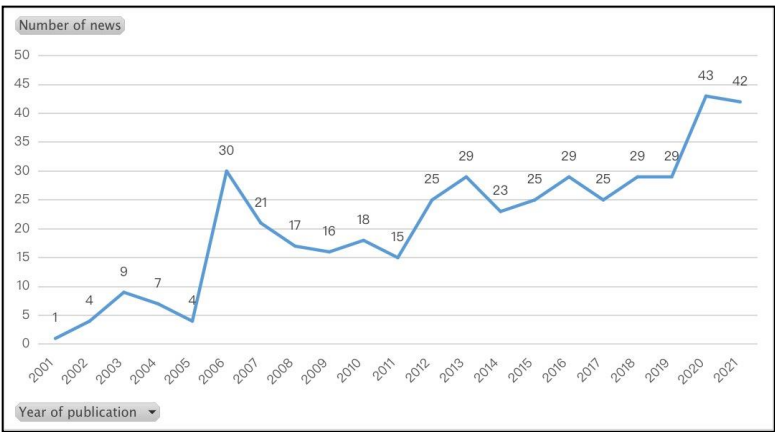

Fig. 1. The number of rural-related news in China Daily 2001-2021

The following research questions are posed in light of the literature above:

RQ: What kind of rural character does media representations of 'rural' reflect, and how does this further reflect the Chinese social environment?

3 Methodology

In general, this study will combine corpus linguistics and critical discourse analysis to examine rural representations in English-language news media in China. In other words, the quantitative analysis will be complemented by the qualitative analysis.

3.1 Data selection and corpus construction

China Daily's website is used as the data source for this study. As an English-language newspaper sponsored by the Publicity Department of the CPC Central Committee, China Daily serves as a window on China's national image. Collecting data from the China Daily website to create the corpus ensures the data's representativeness.

Data is collected automatically by computer. We analyse data using *AntConc*, and some of our statistical work is performed in Excel.

We use a data crawler on the China Daily website to look for news headlines containing "rural." The computer-generated data has been manually verified to ensure that each news item is rural-related. As presented on the China Daily website, China Daily is a daily English-language newspaper backed by China's state propaganda department. It serves as a critical window through which China projects its national image to the rest of the world. Collecting data from the China Daily website to create the corpus ensures the representativeness of the data. To examine rural life from a macro viewpoint, the news contained in the corpus we created spans 21 years (from 2001 to 2021).

The corpus for this study comprises headlines containing the word 'rural' extracted from the China Daily website. According to previous literature [20], the viewer's attention span is short in the digital age, and thus the headline serves as the news's hook for engaging the reader. Because the headline is frequently the only glance the reader gets of the article, hooking the reader begins before the reader reaches the introduction [21]. With this in

mind, internet news writers must understand how to craft the most effective headlines. Thus, the corpus is constructed using Internet news headlines rather than the full text of the news, primarily to account for the characteristics of Internet news and improve the corpus's conciseness.

After completing automatic computer crawling and manual checking, a preliminary version of the corpus is constructed. The crawler's raw data are recorded in an Excel file. Each news headline contains the URL and publication date to ensure that the original story can be retrieved. This type of raw data storage can provide context and analytical clues for critical discourse analysis. Additionally, if the study includes multimodal discourse analysis, traceable news sources can help the researcher better understand the page layout. Then, de-duplication becomes critical. We removed duplicate article titles from the initial version of the corpus using Excel's de-duplication function. Finally, the corpus used in this study contains 441 titles of news articles totalling 3130 words. News headlines are stored separately in plain text files to facilitate analysis using the *AntConc* corpus tool.

3.2 Data analysis methods

First, we will examine the linguistic patterns associated with the word 'rural.' The *AntConc* collocate function is used to generate the 'rural' collocation. The search term is surrounded by a four-word range (2L+2L) to the left and right. Then we will use *AntConc's* n-gram function to generate clusters containing the term 'rural' and their co-occurring patterns. Once the collocations have been corrected, they are classified into function and content words. The classification is based on their most frequent association with rural, although some terms are polysemous and can be classified in two or more possible categories. This study will then count and categorise the most frequent noun collocates of the adjective rural. The reason for using a frequency-based metric to obtain collocations is that linguistic data, according to the literature, are subject to clustering, combination, and other factors rather than occurring randomly [22]. In addition to the data-driven inductive approach, this study will incorporate a deductive part. Representative excerpts will be used to demonstrate critical discourse analysis to summarise the natural characteristics conveyed by linguistic patterns.

Finally, the corpus-driven linguistic representations will be placed in a socio-cultural context to address how the linguistic patterns further reflect the Chinese social environment. The frequency distribution of collocate categories in tokens and types will be analysed statistically, and the correlation between the frequency and timeline of news headline topics will be investigated. We will examine how the media shape and observe the transformation of rural identities in China, the country with the world's largest rural population, through media representations of rural.

4 Findings and discussion

4.1 Rural transformation development (RTD) involves multiple development factors

As shown in Table 1, the following nouns are associated with the adjective Rural in this study's corpus. A first noteworthy observation is that the top 15 nouns in the frequency ranking pertain to the economy (e.g., tourism, revitalisation), livelihoods (e.g., life), culture and education (e.g., schools, teachers, education), rural environment (e.g., land), and governance and rules (e.g., reform). This subset of noun collocations serves as a primer for the key themes index using the adjective rural.

Table 1. Most frequent noun collocates of adjective Rural

Rank	Most frequent noun collocates	Freq.
1	Areas	58
2	Tourism	30
3	Revitalization	21
4	China	18
5	Development	15
6	Reform	10
7	Residents	9
8	Vitalization	9
9	Schools	8
10	Teachers	7
11	Education	6
12	Gap	6
13	Land	6
14	Life	6
15	Doctors	5

Following that, the collocations in Table 1 are manually reviewed and classified into thematic categories. The classification is based on the indexing topic of the corpus article rather than on its lexical meaning [23]. In Table 2, we classify noun collocations with the adjective rural.

Table 2. Categorized noun collocates of adjective Rural

Classification	Vocabulary
Economy	tourism, revitalization, development, vitalization, gap, e-commerce, urban, economy, income, revival, *taobao*, youth, banking, consumption, divide, economies, market, pain, poor, poverty, progress, road, roads, tax
Livelihood	residents, life, doctors, people, women, children, elderly, living, families, health, infrastructure, kids, medical, medicare, water, wife
Rural Environment	land, pollution
Culture and Education	schools, teachers, education, students, cultural, buzzwords, libraries, teacher
Governance and Rule	reform, democracy
Kind of community	areas, China, area, villages, issues, problems, communities, societies,

As shown in Figure 2, economic development is a primary focus of the news media's coverage of rural areas. Economic terms are mentioned 130 times, accounting for 39% of all noun collocations.
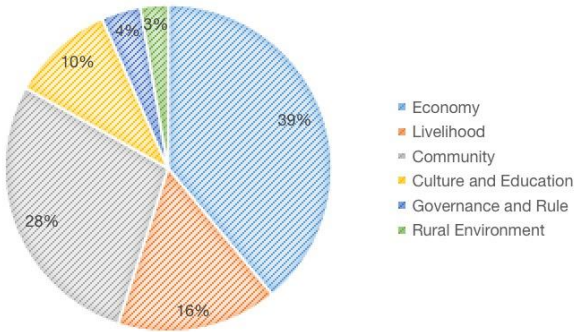

Fig. 2. Frequency distribution of collocate categories in terms of tokens

Simultaneously, the economic category contains 24 distinct words (singular and plural counted separately), accounting for 40% of all nouns collocated with the adjective Rural (see Figure 3).
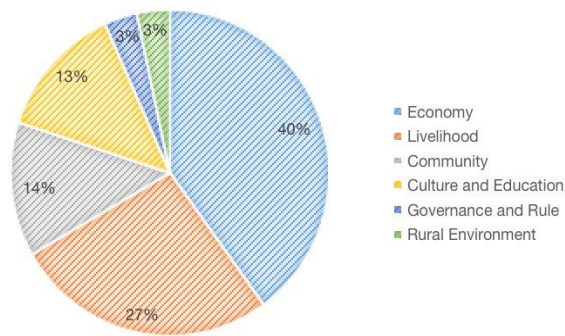

Fig. 3. Frequency distribution of collocate categories in terms of different words

There are several abstract nouns in the economic lexicon, including revitalisation, development, and capitalisation. These nouns imply positive semantic prosody, as they are associated with concepts such as the decent, the vibrant, and the better life. These words encapsulate or explain rural regeneration's values, visions, and ideals. In addition, among the economic terms associated with the adjective Rural, there are also nouns referring to specific financial industries, such as e-commerce and tourism. These words reflect the practical economic development efforts, which concretise the goal of rural economic development. The concrete economic terms indicate that rural revitalisation is a genuine effort by the Chinese in the modern era.

A noteworthy observation is that, as rural e-commerce has grown, Taobao, a central Chinese e-commerce platform, has seen its name appear in rural-related news. Premier Li Keqiang of China first mentioned the term 'Internet+' in his 2015 government work report, referring to the integration of the Internet with traditional industries. The excerpts below demonstrate that the 'Internet+ rural economy' revives China's rural development.

Excerpts 1	Taobao boosts rural e-commerce; Rural Taobao brings e-commerce to the countryside; Live e-commerce plays an active role in helping rural revitalization; E-commerce plays key role in battling poverty in rural areas;

As longitudinal data demonstrate, economic development is the most frequently discussed aspect of rural development. The subjects of news headlines have been coded and counted according to their publication date. As illustrated in Figure 4, news media coverage of economic issues increased from 0 articles in 2001 to 24 articles in 2021. Over the last two decades, there has been a significant rise in media coverage of rural economic issues. In comparison, media coverage of livelihoods, culture and education, rural environment, and governance and rules have remained relatively stable, with minor fluctuations.

According to the data, rural transformation development (RTD) involves multiple development factors, with economic topics gaining prominence in media coverage. In other words, rural economic issues are gaining increasing media attention. This reflects the fact that, over the past two decades, there have been an increasing number of new phenomena in the rural economy that merit media coverage and discussion.
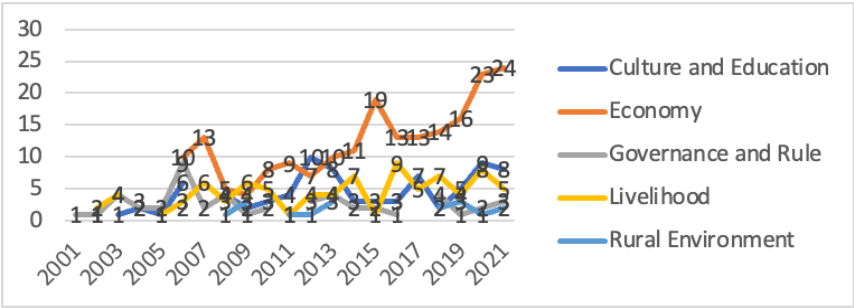
Fig. 4. Correlation between the frequency and timeline of rural-related news topics

4.2 Rural transformation development (RTD) involves both positive and negative factors

Table 4 classifies nouns associated with the adjective Rural according to semantic prosody. As evidenced by the article titles about rural areas, rural development has positive and negative aspects.

As illustrated in excerpts 2, the positive aspects include the emergence of numerous new phenomena in rural areas, such as e-commerce enhancing people's lives, bloggers identifying new directions for rural economies, and rural toilet construction projects.

Excerpts 2	Rural bloggers cultivate loyal followers on social media and boost income; Rural toilet revolution flush of vitalization; Rural e-commerce helps farmers save during spring planting

Table 3. Classification of nouns according to semantic prosody

Semantic Prosody	Categories	Vocabulary	Number of nouns
Neutral	Economy	tourism, e-commerce, urban, economy, income, taobao, youth, banking, consumption, economies, market, progress, road, roads, tax	15
	Livelihood	residents, life, doctors, people, women, children, elderly, living, families, health, infrastructure, kids, medical, medicare, water, wife	16
	Community	areas, china, area, villages, issues, communities, societies	7
	Culture and Education	schools, teachers, education, students, cultural, buzzwords, libraries, teacher	8
	Governance and Rule	democracy	1
	Rural Environment	land	1
Positive	Economy	revitalization, development, vitalization, revival	4
	Governance and Rule	reform	1
Negative	Economy	gap, divide, pain, poor, poverty	5
	Community	problems	1
	Rural Environment	pollution	1

Meanwhile, excerpts 3 demonstrates that the media is concerned about potential issues with rural development, such as the disparity between urban and rural areas. Additionally,

the negative vocabulary includes poor, pain, and poverty. These words imply that rural development issues are confronted, not glorified, to motivate people to act to achieve greater levels of achievement and development.

	Beijing moves to close the urban-rural gap.
Excerpts 3	A welcome move to bridge the urban-rural education gap
	Urban-rural income gap gets bigger: report

5 Conclusion

Although rural development has historically been regarded as an imprecise term that is difficult to define, quantify, and assess, it is still possible to discern from the titles of the last 21 years the characteristics of rural China and the Chinese social environment reflected in the media representations of 'rural'.

Consistent with previous research, the twenty-first century's rural economy and social systems are more diverse, complex, and globalised than those of the last century [24].To begin, China's rural transformation development (RTD) encompasses a range of development factors, including the economy, livelihoods, culture and education, rural environment, and governance and rules, with economic development receiving the most attention, both horizontally and vertically. Second, rural development involves both positive and negative outcomes, such as poverty and the urban-rural divide. In contrast to the idyllic rural life mentioned by previous literature, the English-language news media in China observe reviews and comment on positive and negative developments objectively in China's rural development.

6 Potential Limitation

Due to time constraints, this study concentrated exclusively on the keyword 'rural' and excluded synonyms and nouns with similar meanings (e.g. village, countryside, rurality). We propose that future research incorporate as many rural terms as possible into its focus to creating a larger corpus. This would provide a more comprehensive view of Rural representation and its reflection on society, thereby increasing the study's reliability and validity.

Reference

1. K. Halfacree., I. KovÁch, ., & R.Woodward, *Leadership and Local Power in European Rural Development*. Ashgate. (2002).
2. H. Long, ., J.Zou., J.Pykett,., & Y. Li.. Analysis of rural transformation development in China since the turn of the new millennium. *Applied Geography*, *31*(3), 1094-1105. (2011)
3. L. Cui., Y. Zhao., D. Li., Jing, R., & T. Zhou.. A typological approach to the transformation of cave dwellings in Baishe Village, Shaanxi, China. *URBAN DESIGN International*, *27*(2), 115-131. (2022).
4. L. Jiang., J. Luo., C. Zhang., L. Tian., Q. Liu., G. Chen., & Y. Tian,. Study on the level and type identification of rural development in Wuhan city's new urban districts. *Isprs international journal of geo-information*, *9*(3), 172. (2020).
5. Z. Q. Meng,. *Opinions of the Central Committee of the Communist Party of China State Council on Comprehensively Promoting the Revitalisation of the Countryside and Accelerating the Modernisation of Agriculture and Rural Areas*. Xinhua News Agency.

(2021).

6. Q. Shen., & X. Gao,. Multilingualism and policy making in Greater China: Ideological and implementational spaces. *Language Policy*, *18*(1), 1-16. (2019).

7. D. S. Silva,.E, Figueiredo., C, Eusébio., & M.J. Carneiro, The countryside is worth a thousand words–Portuguese representations on rural areas. *Journal of Rural Studies*, *44*, 77-88. (2016).

8. J. R. Short, *Imagined country: environment, culture, and society*. Syracuse University Press. (2005).

9. H. Newby, *Green and pleasant land? Social change in rural England*. Penguin Books Ltd. (1980).

10. K. H. Halfacree, Talking about rurality: social representations of the rural as expressed by residents of six English parishes. *Journal of Rural Studies*, *11*(1), 1-20. (1995).

11. E. Figueiredo., J. Fernandes., A.Duque., & D,.Soares da Silva. Rural Portugal in the Movies–between preservation and change. Re-imagining Rurality: Conference and Exhibition. 27–28 February 2015, Faculty of Architecture and the Built Environment, (2015).

12. R. Woodward, 'Deprivation'and 'the rural': an investigation into contradictory discourses. *Journal of Rural Studies*, *12*(1), 55-67. (1996).

13. P. Milbourne Poverty, place, and rurality: material and sociocultural disconnections. *Environment and Planning A*, *46*(3), 566-580. (2014).

14. P. Milbourne Poverty and welfare in rural places. In *Routledge international handbook of rural studies* (pp. 480-491). Routledge. (2016).

15. K. H, Halfacree. Locality and social representation: space, discourse and alternative definitions of the rural. In *The Rural* (pp. 245-260). Routledge. (2017).

16. E, Figueiredo. E, Kastenholz & J, Lima Recreating rurality through tourism: Visions of hosts and guests in two Portuguese villages. *Recreating Rurality through Tourism: Visions of Hosts and Guests in two Portuguese Villages*, 43-70. (2013).

17. E. Capo. Transformation and development in Italian rural society. *Sociologia Ruralis*, *35*(3-4), 297-308. (1995).

18. M. U, Ahmed, Development of Rural Industries and Transformation of China's Rural Economy. *Asia-Pacific Journal of Rural Development*, *3*(1), 1-19. (1993).

19. F. Zhu, F. Zhang, & X.Ke,. Rural industrial restructuring in China's metropolitan suburbs: Evidence from the land use transition of rural enterprises in suburban Beijing. *Land Use Policy*, *74*, 121-129. (2018)

20. A.M. Grant, & T. G.Pollock, Publishing in AMJ—Part 3: Setting the Hook. *Academy of Management Journal*, *54*(5), 873-879. (2011).

21. R., Russell-Bennett, & S. Baron, Editorial: the importance of the snappy title. *The Journal of services marketing*, *30*(5), 477-479. (2016).

22. M. Stubbs, Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, *2*(1), 23-55. (1995).

23. P., Baker, C.,Gabrielatos, & T. McEnery, Sketching Muslims: A Corpus Driven Analysis of Representations Around the Word 'Muslim' in the British Press 1998–2009. *Applied Linguistics*, *34*(3), 255-278. (2012).

24. J. J.,Kennedy, J. W., Thomas, & P. Glueck, Evolving forestry and rural development beliefs at midpoint and close of the 20th century. *Forest Policy and Economics*, *3*(1), 81-95. (2001).