ISSN 2616-7433 Vol. 6, Issue 10: 75-81, DOI: 10.25236/FSST.2024.061011

Research on Language Data and Its Functions in the Digital Era

Jingtao Liang[1,2,a], Jiayin Huo[3,b], Ziyuan Li[1,c], Tingting Su[4,d,*]

[1]*Institute Department of Humanities and Management and Hebei Key Laboratory of Health Care with Traditional Chinese Medicine, Hebei University of Chinese Medicine, Shijiazhuang, Hebei, China*
[2]*Institute for Language Policies and Standards, Beijing Language and Culture University, Beijing, China*
[3]*HBU-UCLan School of Media, Communication and Creative Industries, Hebei University, Baoding, Hebei, China*
[4]*School of International Education of Chinese Language, Beijing International Studies University, Beijing, China*
[a]*ljtblcu@foxmail.com,* [b]*19932158178@163.com,* [c]*lifuchun6972@163.com,* [d]*pistacie@163.com*
[*]*Corresponding author*

Abstract: *With the advent of digital economy era, data has become one of important production factors for national economy. Language data, as an essential part of data, consists of language structural data and language informative data. Language structural data refers to language structure units and rules such as phonetics, words, vocabulary, and grammar, while language informative data encompasses text discourse, conversations, and parallel corpora which are discourse data formed through human language use. Roughly speaking, the functions of language data consists of the function of living factor and that of production factor, which concerns the different functions of language data in daily life and social production. The function of production factor entails utilizing machine-readable data as a foundation to data mining, knowledge extracting, insights forming and their use in economic activity. It extends its function from essential living factor to production factor, driven by the national big data strategy, high-quality economic development and digital technology iteration of the People's Republic of China.*

Keywords: *Digital Era; Language Data; Functions; Shit to Production Factor*

1. Introduction

In March, 2021, *the Outline of the 14th Five-Year Plan(2021-2025) for National Economic and Social Development of the People's Republic of China and the Long-Range Objectives for 2035*(hereinafter referred to as the 14th Five-Year Plan) was officially released, proposing to embrace the digital era, activate the potential of data factors, enhance China's strength in cyberspace, and accelerate the development of digital economy, digital society, and digital government. According to the statistics of the China Academy of Information and Communications Technology, China's digital economy reached a scale of USD 17.2 trillion in 2022, playing a crucial supporting role in the growth of the national economy. The rapid growth of the digital economy is based on data, with the core lying in the information, knowledge mined and digital technologies developed through machine-readable data. In light of the significant role of data in economic activities, the state has formulated policies to include data as a kind of production factor. Among data, language data is the most crucial one and should also be categorized as production factor (Li Yuming, 2020)[11]. Language data is both a living factor and a production factor. In the past, it primarily played a role in human social life, serving human daily communication, cultural inheritance, information transmission, etc. As humanity moves towards the Information Age and the Era of Economic Globalization, it has begun to function as production factor. In the future, with the implementation of Big Data strategies, Artificial Intelligence strategies, digital economy and the 14th Five-Year Plan, language data will play an even more significant role in social production. However, its functions and functional shit in economic development deserve particular attention in the current times.

ISSN 2616-7433 Vol. 6, Issue 10: 75-81, DOI: 10.25236/FSST.2024.061011

2. Literature Review

Data serves as the object of computation and research. Originating primarily from computer science, data is a general term for symbolic media that possess certain meanings and can be processed by computer programs. *The Data Security Law of the People's Republic of China* defines data as any record of information in electronic or other forms. The International Organization for Standardization considers data as collected, raw and unprocessed facts about an object (Sun Yi, 2021)[23]. Yang Tao (2020) argues that data is the physical symbols or collections of these symbols used by human to describe and record the nature, state and interrelationships of the natural world and human society, encompassing both numerical (number) and non-numerical data, with the latter being capable of conversion into numerical data through technological means for digital presentation[32]. Data can be categorized as either digital or non-digital in form, with carriers including numerical values, texts, videos, images, etc. (Tu Zipei, 2014)[25]. Language data constitutes an important part of data. When introducing the developmental characteristics of computational linguistics, Feng Zhiwei (2013) translated Franz Josef Och's statement on the future of machine translation as, give me enough parallel data, and you can have translation system for any two languages in a matter of hours. Here, parallel data refers to parallel language data in digital format[6]. Overall, the academic exploration of language data is still in its infancy. From the perspective of the definition of language data, Li Yuming and Wang Chunhui (2022) define language data as data constituted based on the linguistic symbol system.[12] In terms of sources, Jiang Guoquan and Li Yifei (2021) believe that language data is constructed upon elements such as phonetics, semantics, pragmatics, and context, originating from human activities of observing and exploring the world[10]. From a classification perspective, Li Yuming (2020), considering the era's backdrop of data becoming production factor, classifies language data into linguistic symbol systems, information with language as its carrier, and translanguaging symbols[11]. Li Yuming and Wang Chunhui (2022) further categorize it into linguistic data, discourse data, language-derived data, man-made language data, and language code data[12]. Wang Chunhui (2022) explores the types of language data from the angles of form, content, subject and processing[26].

Diachronically speaking, the academic exploration of data functions spans three dimensions: past, present, and future. From a historical perspective, data serves as a means of knowledge inheritance and accumulation (Che Pinjue, 2017)[3]. In the present context, data acts as a tool for discovering new knowledge, optimizing economic activities, and regulating production relations. For instance, Schönberger and Cukier (2013) argue that analyzing data can lead to unexpected discoveries[29]; Yang Tao (2020) further elaborates that data can reduce transaction costs associated with trial and error, matching, and other activities in the real world; it decreases subjective uncertainty in human behavior and enhances processing efficiency[32]. The Big Data Strategy Key Laboratory (2017) posits that the integration of blockchain technology with data is innovating modern governance models[4]. Looking into the future, data serves as the foundmental material for Artificial Intelligence development and represents the overarching trend of human development. Scholars such as Che Pinjue (2017)[3], Tu Zipei (2014)[25], and Wu Jun (2016) [30]contend that data is a core element driving the advancement of Artificial Intelligence.

As an important part of data, language data also fulfills the functions of data. Moreover, relevant academic explorations on one hand reiterate the production factor function of language data (Li Yuming, 2020; Jiang Guoquan & Li Yifei, 2021; Huang Shao'an, 2022; Wang Hailan, 2022; Wang Zhanglong, 2022; Liang Jingtao, 2023, etc.)[9][10][11][16][27][28]. Wang Hailan (2022) discusses the economic attributes of language data from three aspects: language data and digital technology, language data and production decision-making, language data and industrial structure adjustment[27]. Liang Jingtao (2022)[15], Li Yuming and Liang Jingtao (2024) [13]have studied the ways in which language data participates in production link. Liang Jingtao and Zhang Hongjie (2024)[18], Liang Jingtao and Tang Peilan (2024) [17]have explored practical paths for the production factor function of language resource as data in the digital economy. On the other hand, the academic explorations primarily focuse on the domains where language data functions, such as language policy research (Dai Hongliang, 2013)[5], contemporary language studies (Qian Xiaofei, 2018)[20], translation studies (Huang Haiying, 2018; Hu Kaibao, 2022)[7][8], Artificial Intelligence (Li Yuming & Wang Chunhui, 2022)[], language teaching (Li Yuming & Wang Chunhui, 2022; Hu Kaibao, 2022)[7][12], language products development (Li Yuming & Wang Chunhui, 2022; Rao Gaoqi, 2022)[12][21], language preservation and language information processing (Li Yuming & Wang Chunhui, 2022)[12], technology research and development of Automatic Speech Recognition and Text To Speech (Wang Zhanglong, 2022)[28], and self-media operations (Song Hui & Gao Ning, 2022)[22]. Since language data can manifest as spoken language data, it can also be used in human social interactions. For example, Jiang Guoquan and Li Yifei (2021)

ISSN 2616-7433 Vol. 6, Issue 10: 75-81, DOI: 10.25236/FSST.2024.061011

believe that language data can help eliminate information barriers in implementing the Belt and Road Strategy[10].

Data serves as the carrier of information and knowledge, while information and knowledge represent the content or the rules of the content which data conveys. This study defines language data as data that takes the form of language and text. In terms of form, language data can be classified into digital and non-digital, as well as structured and unstructured types. Non-digital formats can be converted into digital formats through technological means for computer use, and unstructured formats can also be transformed into structured formats through conscious human processing to serve specific purposes. Language data encompasses both language structural data and language informative data. Language structural data refers to language structure units and rules such as phonetics, words, vocabulary, and grammar, while language informative data encompasses text discourse, conversations, and parallel corpora which are discourse data formed through human language use. It is important to clarify that the language structural data mentioned here originates from the linguistic term language structure, which differs from the structure in structured and unstructured data; the former refers to units and rules of language structure, while the latter pertains to the form of induction and organization of language data. This paper, starting from the perspective of language data's functional roles as both living factor and production factor in human society, discusses the reasons why the functional focus of language data in the digital era has expanded from primarily being a living factor to production factor. The aim is to lay a theoretical and methodological foundation for the development of language data, contribute to the growth of the digital economy, and provide support of language for building build a modern strong socialist country.

3. Functions of Language Data

In the context of the digital era, data has emerged as a crucial production factor. The development of the digital economy necessitates leveraging the production function of data. Specifically, this entails utilizing machine-readable data as a foundation to data mining, extract knowledge, and form insights. Throughout this process, digital and intelligent technologies are developed to drive economic growth. From the perspective of mined information, it serves as a barometer of market trends and a source for discovering the market demands. In terms of extracted knowledge from data mining, it forms the basis for studying market patterns and user interests or preferences, addressing the shortcomings of the Rationality Assumption of Economics. As for the insights gained, they guide decision-making in economic activities such as production and marketing. As for the developed technologies, they aim to achieve automated and real-time intelligent analysis, enhancing the efficiency of these processes and, in turn, contributing to the improvement of economic activity efficiency. For instance, the industry leverages user behaviors data, combined with contextualized information, to mine his or her interests and preferences, and accurately reflect user needs, providing decision-making guidance, informing corporate investment, production, and marketing activities, thereby driving high-quality business development (Che Pinjue, 2014)[2].

As an important part of data, language data serves as the fundamental raw material for human society's production and is a crucial labor object and production factor. The researchers urgently need to consider how to conduct research and development centered on promoting the development of the digital economy. Since language data is still a relatively new concept, there is limited discussion on which types of language data can fulfill the role of production factor and how they can do so. Addressing the first question broadly, solely in terms of raw materials, language structural data rarely generates direct economic value. Theoretically, language structural data belongs to public goods, meaning their use by any individual, any nation, or even a country does not confer exclusivity. However, learning resources such as grammar books, dictionaries developed based on language structural data are classified as club goods, requiring payment for access and possessing significant economic potential. These learning resources are derivatives, not raw materials. In contrast, language informative data is relatively more valuable. Production factors correspond to living factors, both of which align respectively with human society's production and living aspects (Xu Shoubo, 2007)[31]. The development and utilization of language data by human must inevitably fall into two broad categories: social living and social production. Broadly speaking, the functions of language data can be categorized into living factor function and production factor function.

ISSN 2616-7433 Vol. 6, Issue 10: 75-81, DOI: 10.25236/FSST.2024.061011

3.1 Function of Living Factor

Language data, originating from human society, plays a pivotal role in social life. It serves as the most crucial information carrier, providing essential foundational support for human daily communication, knowledge accumulation, cultural heritage and transmission. It is an essential living factor. Since the founding of the People's Republic of China, the Party Central Committee has formulated language policies for the promotion and popularization of the national common language based on China's basic national conditions, language situations, and economic and social development needs. After decades of efforts, the national common language has been basically popularized nationwide, and language communication barriers have been largely eliminated (Liu Pengjian, 2020)[19]. People utilize language informative data carried by the national common language for cross-regional and cross-ethnic communication, significantly facilitating their daily life.

Since the 21st century, the country has successively formulated language policies such as the National Language Resources Monitoring and Research Program (2004), the Construction of the Chinese Language Resources Audio Database (2008), and the Chinese Language Resources Protection Project (2012). The National Language Resources Monitoring and Research Program constructs language data from various perspectives, including print media (Beijing Language and Culture University), audio media (Communication University of China), online media (Central China Normal University), ethnic languages (Minzu University of China), educational materials (Xiamen University), and overseas Chinese languages (Jinan University), to monitor and study language situation. Each year, it uses detailed language data to reflect the current state of language situation in China from different perspectives, releasing significant events, hot topics, and actual data from annual language situation. By offering language data as a kind of public goods, it facilitates direct access and encourages society to pay more attention to the language situation of the whole country. Simultaneously, these data serve policy adjustments and formulation, guiding language situation towards continuous, harmonious, and healthy development. The Construction of the Chinese Language Resources Audio Database and the Chinese Language Resources Protection Project collect and construct language data in a structured way to record and preserve the dialects and ethnic minority languages in China, safeguarding the cultural heritage they carry, and drawing public's attention to the spiritual homeland constructed by language. This raises social awareness and participation (Tian Lixin, 2015)[24], contributing to the preservation of world cultural diversity. Both the National Language Resources Monitoring and Research Program and the Chinese Language Resources Audio Database and Protection Project are formulated with the starting point of the role of language data in social life, which can be called the function of living factor.

3.2 Function of Production Factor

Language data can break down language communication barriers both domestically and internationally, enhancing the coordination efficiency of economic activities and playing an important role in social production, , which can be called the function of production factor. Domestically, the promotion and popularization of the national common language have reduced interpersonal communication costs, coordinated production and cooperation, and facilitated the cross-regional flow of capital, goods, and human resources, making it an indispensable element in economic activities. In particular, the quality of a region's barrier-free information communication environment is a direct influencing factor of its investment environment. Internationally, with the deepening of economic globalization and world economic integration, China has embarked on a new stage of opening up and integration into the world. This surge in foreign economic exchanges and cooperation has generated a vast demand for translation (including translation talents and software), leading to the emergence of diverse translation professions (such as oral and written translation, translation project management, and translation technology support), which in turn form a translation industry chain. The cultivation of translation talents relies heavily on language informative data. People need to acquire linguistic knowledge and improve their language proficiency through studying language informative data to achieve cross-lingual information transmission and meet market demands. Furthermore, with the accumulation of large-scale authentic bilingual parallel data, data-driven machine translation technology has achieved higher accuracy, making machine translation tools a valuable assistant for translators and capable of fulfilling basic cross-lingual communication needs. Moreover, in the march towards the Information Age, human have built an information world that depends on the nature and social world by processing natural language based on machine-readable language data, enabling machines to read and understand natural language (Liang Jingtao, 2020)[14]. Throughout this process, various social divisions of labor have emerged, encompassing design and development aspects such as encoding, character sets, font libraries, glyphs, input methods, typesetting, storage, transmission,

ISSN 2616-7433 Vol. 6, Issue 10: 75-81, DOI: 10.25236/FSST.2024.061011

editing, output, word segmentation, tagging, knowledge bases, and dictionaries, all centered around different aspects of natural language processing (Durkheim, 2013)[1]. These divisions, in turn, form an industrial chain centered on natural language processing. This industrial chain is guided by linguistic knowledge and takes language data as its labor object and dominant production factor.

Data serves as the engine driving the development of the digital age, necessitating language data to function as production factor more in digital format. While the function of non-digital production factor in the past primarily aimed to eliminate language barriers in interpersonal communication and enhance efficiency, digital-format language data in the digital age takes on a technical form to propel social production and development. It is worth noting that when transitioning towards the Information Age, human developed technological tools to enhance the efficiency of natural language processing. From this perspective, it may seem that the function of language data as production factor in the digital age has not changed in nature. However, upon deep research, there are both closely connected and different from one another. Specifically: Informatization is a transitional product of humanity's pursuit of intelligence, having constructed an information world that allows human activities to shift from offline to online in the forms of complete, partial, or temporary cross-border transitions (Liang Jingtao, 2020)[14]. The technological tools developed during this period enable human to publish and access information better and faster. In contrast, the goal of the digital age is to build a data world and an intelligent world upon the foundation of the information world. By leveraging the vast amounts of data generated through circulation and aggregation, digital, data, and intelligent technologies are developed with the objectives of automating, intelligentizing, and accelerating information mining, knowledge extraction, and wisdom formation. This aims to more precisely and rapidly identify and match user needs (from community needs to individual needs), thereby enhancing economic efficiency and optimizing industrial structures. The two represent distinct stages of development, with informatization providing the premise and foundation for datafication and intelligentization, which further deepen the development level of informatization, showcasing qualitative differences.

4. Shift from Focusing Primarily on Living Factor to Production Factor

In April 2020, the CPC Central Committee and the State Council issued the *Notice by the General Office of the State Council Regarding Issuing the Master Plan for the Pilot Program of the comprehensive Reform of the Market-oriented Allocation of Factors* (hereinafter referred to as the *Notice*), which listed data as a production factor alongside land, labor, capital, and technology, and called for accelerating the cultivation of a data factor market. With the development of the digital economy, data factor has become a fundamental resource for social production. At present, the bottleneck hindering China's high-quality economic development lies in structural contradictions, and accurately targeting market and user demands is crucial to supply-side structural reform. Language informative data, based on which user's interests and preferences can be mined and analyzed, provides direction for production and marketing, thereby enhancing the quality of economic development. As a significant component of data, language data has also participated in social production, functioning as production factor. In light of national policy requirements, language data should conform to the needs of the current times and actively expand from primarily serving as living factor to production factor. It is important to study the reasons behind this functional shit or functional expansion.

Firstly, the functional shit of language data is an inevitable requirement for implementing the Big Data Strategy. In 2017, China proposed to build a digital economy with data as key factor, marking the first time that data was regarded as a production factor. Subsequently, the *Notice* issued by the CPC Central Committee and the State Council in 2020 formally incorporated data into production factors. The 14th Five-Year Plan (2021-2025) in 2021, further emphasized the fundamental role of data in China's national development strategy over the next five years by giving full play to the advantages of massive data and rich application scenarios, promote the deep integration of digital technology with the real economy, empower the transformation and upgrading of traditional industries, foster new industries, new business forms, and new models, and strengthen new engines for economic development. In November 2021, the Ministry of Industry and Information Technology issued the *14th Five-Year Plan for the Development of the Big Data Industry*, which included cultivating the data factor market, consolidating the foundation for industrial development, and promoting economic and social development. In summary, as data increasingly becomes an important strategic resource and a new type of production factor in the implementation of the Big Data Strategy, language data, as a crucial component of data, will also serve as labor object and a fundamental raw material for intelligent technology research and development, entering the social production links as an indispensable new type of production factor. The functional shit or expansion of language data is an inevitable requirement for adapting to the Big Data Strategy. In the future, the integration of intelligent analysis technologies developed based on language data with traditional industries will not only spawn new industries and sectors, but also facilitate the upgrading and transformation of traditional industries, thus

ISSN 2616-7433 Vol. 6, Issue 10: 75-81, DOI: 10.25236/FSST.2024.061011

holding significant strategic meanings.

Secondly, the functional shit or expansion of language data is an internal requirement for high-quality economic development. In the current times, the primary contradiction hindering China's high-quality economic development lies in structural issues, which determines that supply-side structural reform is the main thrust. From the perspective of Political Economics, production determines consumption, while consumption exerts a counteractive influence on production. Supply-side reform aims to eliminate excess, low-end, and homogenized products, and instead, provide personalized and differentiated high-end products based on consumer demands. In the context of the digital economy, language data containing information such as product evaluations and feedbacks has become a crucial basis for analyzing consumer demands. Therefore, the industrialization of the production, collection, storage, processing, analysis, and service contribute to the high-quality economic development by providing information support for the supply side. As production factor, vast amounts of language data needs to be processed, mined, and analyzed, in order to serve other industries with derivatives such as intelligent technologies and consulting reports. For example, iFLYTEK's speech recognition technology, developed based on speech data, can be applied to various scenarios such as telemarketing, customer service, meetings, interviews, subtitles, and classrooms, recognizing both Mandarin and Chinese dialects. The company also offers speech-to-text services on its website. When integrated with headphones, this speech recognition technology becomes an intelligent product that provides real-time transcription of phone calls, expanding the functionality of headphones while creating consumer demand and achieving product value enhancement, as well as the upgrading and transformation of the headphone industry.

Thirdly, the functional shift or expansion of language data serves as the fundamental material for the development of digital technologies. The scale of the digital economy's development is attributed to advancements in digital technologies, and the upgrade of these technologies is contingent upon the volume of language data. In 2004, with the help of increased language data input, the error rate of speech recognition technology decreased by 50% compared to a decade ago. In 2005, Google put more language data in the development of its machine translation system, achieving an accuracy rate of 51% in Chinese-to-English translation tests, 17 percentage points higher than the second-ranked competitor (Wu Jun, 2016)[30]. Both speech recognition technology and machine translation technology rely on language data to continuously iterate and update, ultimately integrating with other products to transform human life and work. The iterative upgrading of digital technologies is directly related to the quality of supply-side products and services, constraining the effectiveness of high-quality economic development. Therefore, language data, as the basic raw material for digital technologies, must break through the boundaries of social life and participate in social production as production factor, realizing a shift from primarily being living factor to production factor.

5. Conclusion

In the context of the digital era, the implementation of Big Data strategies and Artificial Intelligence strategies, the pursuit of high-quality economic development, and the iterative upgrading of data and intelligent technologies are external factors that necessitate the shift of the functions of language data. The development of the digital economy demands the knowledge mining of machine-readable language data, which serves as the foundation for the development of intelligent technologies, continuously enhancing the efficiency of economic activities and promoting the optimization of industrial structures. In particular, intelligent technologies enable faster access to information and knowledge, improving efficiency and driving the upgrading and transformation of existing industries, marking an epoch-making significance and bringing immeasurable changes to human society. This underscores the need for the functions of language data to expand from primarily being living factor to production factor. These changes profoundly impact the development and governance of language data in the digital era. The realization of the production factor function of language data, as well as issues such as how language data should flow and converge, are academic and practical topics that require consideration in the future research.

Acknowledgement

This work was supported by China National Social Science Fund (No. 23BYY056, Title: Research on Multimodal Language Landscape Investigation in Northeastern Border Cities in China) and China National Committee for Terminology in Science and Technology Fund (No. YB2023016, Title: A study of the Production Factor Function of Term Data).

ISSN 2616-7433 Vol. 6, Issue 10: 75-81, DOI: 10.25236/FSST.2024.061011

References

[1] Émile Durkheim, 2013, The Division of Labor in Society, translated by Qu Dong, Sanlian Bookstore.
[2] Che Pinjue, 2014, Fighting in the Age of Big Data: A Powerful Tool for Future Business, Zhejiang People's Publishing House.
[3] Che Pinjue, 2017, The Nature of Data, Beijing United Publishing Co.,Ltd.
[4] Key Laboratory of Big Data Strategy, 2017, Redefine the Big Data, China Machine Press.
[5] Dai Hongliang, 2013, The Analysis about the Linguistic Data of the Population Census of Hong Kong, Macao and Taiwan, Applied Linguistics, Vol.2: 19-25.
[6] Feng Zhiwei, 2013, Modern Language Schools (Revised Edition), The Commercial Press.
[7] Hu Kaibao, 2022, Language Data Science and Applications: Features, Research Areas and Approaches, Foreign Language World, Vol. 3:37-44.
[8] Huang Haiying, 2018, Copyright Risks of the Belt and Road Initiative Language Data in the Context of Cloud Computing, Library Tribune, Vol. 7: 40-46.
[9] Huang Shao'an, 2022, Digital Technology Will Substantially Increase the Economic Value of Language Data, Chinese Journal of Language Policy and Planning, Vol. 4: 61.
[10] Jiang Guoquan, Li Yifei, 2021, Language Data Embraces New Development Challenges of the Belt and Road, Language and Words Daily, October the thirteenth (002).
[11] Li Yuming, 2020, Language Industry in the Information Age, Journal of Shandong Normal University (Social Science Edition), Vol. 5: 87-98.
[12] Li Yuming, Wang Chunhui, 2022, Host Words: From Data to Language Data, Chinese Journal of Language Policy and Planning, Vol. 4: 13-14.
[13] Li Yuming, Liang Jingtao, 2024, On the Function as a production factor and Property Rights Systems Construction of Language Data, Language Teaching and Linguistic Studies, Vol. 2: 1-11.
[14] Liang Jingtao, 2020, On Functions of Language Resources, a doctoral thesis of Beijing Language and Culture University.
[15] Liang Jingtao, 2022, On How Language Data Performs as Production Factor, Journal of Institutional Economics Studies, Vol. 4: 222-233.
[16] Liang Jingtao, 2023, On Language Data, a Report on Postdoctoral Research of Ministry of Education Institute of Language Application and Beijing Normal University.
[17] Liang Jingtao, Tang Peilan, 2024, On why language is a resource, Journal of Tianjin Normal University (Social Science Edition), Vol. 4: 151-160.
[18] Liang Jingtao, Zhang Hongjie, 2024, On Practical Path of Language Resource from the Perspective of Digital Economy, China Language Strategies, Vol. 1: 189-198.
[19] Liu Pengjian, 2020, Language Affairs Development of the People's Republic of China: Accomplishments and Experiences, Applied Linguistics, Vol. 4: 2-6.
[20] Qian Xiaofei, 2021, Key issues and Countermeasures in the Construction of Language Resource, Corpus Linguistics, Vol. 2: 94-105.
[21] Rao Gaoqi, 2022, Contribute to the Construction of a Powerful Cultural Country Through the Data Mining of Ancient Text Resource, Chinese Journal of Language Policy and Planning, Vol.4: 64-65.
[22] Song Huizhou, Gao Ning, 2022, Optimization of Language Data and Improvement of Self-media Communication Efficiency, Chinese Journal of Language Policy and Planning, Vol. 4: 62-63.
[23] Sun Yi, 2021, Digital Economy Study, China Machine Press.
[24] Tian Lixin, 2015, On the Origins and Significances of the Project for Protecting Language Resources of China, Applied Linguistics, Vol. 4: 2-9.
[25] Tu Zipei, 2014, Top of the Data: The Big Data Revolution——History, Reality, and Future, CITIC Press.
[26] Wang Chunhui, 2022, On Language Data Security, Chinese Journal of Language Policy and Planning, Vol.4: 15-25.
[27] Wang Hailan, 2022, On the Economic Attributes of Language Data, Chinese Journal of Language Policy and Planning, Vol.4: 26-34.
[28] Wang Zhanglong, 2022, Intelligent Technology Boosts Digital Transformation in Language Training and Language Testing Industry, Chinese Journal of Language Policy and Planning, Vol. 4: 62.
[29] Viktor Mayer-Schönberge, Kenneth Cukier, 2013, Big Data: A Revolution That Will Transform How We Live, Work, and Think, Zhejiang People's Publishing House.
[30] Wu Jun, 2016, The Age of Intelligence: Big Data and Smart Revolution Redefine the Future, CITIC Press.
[31] Xu Shoubo, 2007, Theory of Six Forces on Essential Factors of Living, Journal of Beijing Jiaotong University(Social Sciences Edition), Vol. 6: 15-18.
[32] Yang Tao, 2020, The Factor of Data: Open Class for Leading Cadres, People's Daily Press.