Retraction

Retraction: Teaching Development and Application of Japanese Corpus Based on Computer Big Data (J. Phys.: Conf. Ser. 1744 042056)

Published 16 September 2022

This article has been retracted by IOP Publishing following an allegation that raises concerns this article may have been created, manipulated, and/or sold by a commercial entity. In addition, IOP Publishing has seen no evidence that reliable peer review was conducted on this article, despite the clear standards expected of and communicated to conference organisers.

The authors of the article have been given opportunity to present evidence that they were the original and genuine creators of the work, however at the time of publication of this notice, IOP Publishing has not received any response. IOP Publishing has analysed the article and agrees there are enough indicators to cause serious doubts over the legitimacy of the work and agree this article should be retracted. The authors are encouraged to contact IOP Publishing Limited if they have any comments on this retraction.

Retraction published: 16 September 2022
Teaching Development and Application of Japanese Corpus Based on Computer Big Data

Ya Li¹,*

¹Japanese Teaching and Research Section, Xi’an Fanyi University, Xi’an, China, 710105

*Corresponding author e-mail: 2013034@xafy.edu.cn

Abstract. In order to ensure the rationality, diversity and scientifi city of Japanese corpus, the process of its collection is random and the samples collected are extensive and rich, which provides conditions for the application of computer big data technology. Based on this, this paper first analyses the role of Japanese corpus in teaching development, and then studies the difficulties in the construction of Japanese corpus based on computer big data. Finally, the key technology and application of Japanese corpus teaching development and application based on computer big data are given.

Keywords: Japanese Corpus, Computer Big Data, Teaching Development and Application

1. Introduction

With the development of global trade and the increasingly frequent economic, social and cultural exchanges under the background of economic globalization, the demand and requirements for talents with transnational communication ability are deepening¹. At present, with the continuous growth of trade and exchanges between China and Japan, the requirement for talents with Japanese communication ability is not only to have the basic abilities of listening, speaking, reading, writing and translation, but also to actively comply with the development of information technology era and make full use of the conditions of computer big data technology. Based on the collection and collation of the existing Japanese corpus based on computer big data, making full use of the functions and characteristics of the Japanese corpus, will help to develop a Japanese corpus that meets the needs of the times and society.

Through the collection and collation of Japanese language materials in practical application, in order to ensure that the corpus and the real context application fit together as much as possible, the collection process should be random, and the collected samples should be extensive and rich, so that the construction of Japanese corpus can truly reflect the actual situation of Japanese application. In addition, corpus refers to the collection of representative real language materials collected on the basis of random sampling, which is the sample of language use. If the sample is representative, random and large enough, it can be considered that the sample is the real representative of the population; the
sample has the statistical characteristics of the whole, so the study of the real language materials can better reflect the Japanese culture and understand the standard Japanese language.

The application of computer big data technology can significantly improve the efficiency and accuracy of the collection and collation process of Japanese language materials, and can store the collected Japanese language materials based on linguistic principles. Japanese corpus based on computer big data technology can be divided into several types as shown in Table 1 below. It can be seen that the Japanese corpus is based on the expected collection of linguistic principles, so as to achieve its goal of building a database.

**Table 1.** Types of Japanese corpus based on computer big data technology.

| Levels               | Types          | Sub-types          |
|----------------------|----------------|--------------------|
| Application level    | General corpus | Purpose of use     |
| Large corpora        | Special corpus |                    |
| Medium corpora       | General purpose| Network based      |
| Small corpus         | Industrialization | Individualization |
|                      |                | Specialization     |

Japanese corpus based on computer big data is based on big data technology to realize the scientific adoption and large-scale construction of Japanese. The Japanese corpus based on computer big data has several typical characteristics as shown in Figure 1.

At present, corpus based on computer big data technology has not only been widely used in translation research, but also become a tool for teaching development and application, and provides research ideas. Corpus based on computer big data technology has successfully played a huge role and influence in theory and application[2]. In addition, the researches of Japanese corpus based on computer big data technology helps to code Japanese corpus concisely and effectively, and obtain a large number of useful data, so it has important practical research value.

**Figure 1.** Characteristics of Japanese corpus based on computer big data technology.

2. The role of Japanese corpus in teaching development
At present, the application of Japanese corpus in educational development is not only at the level of words and phrases, but also has been integrated into the various processes of Japanese teaching. Japanese anticipation driven teaching requires Japanese classroom teaching process to have its real language environment as far as possible, so as to promote teaching activities can effectively fit the teaching application[3]. In addition, the establishment of Japanese corpus helps teachers to choose textbooks and auxiliary materials that are more suitable for practical application. It can be seen that Japanese corpus can effectively promote the diversification and aggregation of teaching contents, teaching views and teaching genres in teaching development.

2.1. Promote autonomous learning with the help of human-computer interaction
In the process of Japanese teaching, the Japanese corpus platform can effectively use its problem driver and human-computer interaction mode to promote the interest of teaching and the realization of human-computer interaction process, and can effectively enhance students' interest in learning. For example, with the help of corpora, the Japanese collaborative platform module can automatically translate many expressions from the platform to enhance students' motivation and interest in autonomous learning, and promote their thinking and in-depth understanding of learning problems.[4]

In addition, the construction of Japanese corpus can build a student-centered teaching mode and promote the realization of students' subjectivity. With the help of the retrieval function of the corpus, students can realize autonomous cognition and judgment at several levels as shown in Figure 2 below, so as to self-analyze the deficiencies and problems, and make statistics and analysis of language errors. It can be seen that the use of Japanese corpus can significantly promote the improvement of students' language awareness, as well as the establishment of language sense, and promote the realization of Japanese internalization.

![Figure 2. The elements of the retrieval function of corpus.](image)

2.2. Promoting the contextualization and localization of Japanese communication

In the process of teaching development and application based on Japanese Corpus, the application of computer retrieval in Japanese corpus can help students better realize the synchronous development of Japanese language and structure. Secondly, based on the comparison with Japanese parallel corpus, it is helpful for students to better grasp the relevant expressions and concepts of Japanese, and can quickly find the corresponding content.

In addition, the application of Japanese Corpus can help the students to better grasp the practical application of Japanese, and promote the continuous improvement of their ability of contextualization and localization of Japanese communication.

3. Difficulties in the construction of Japanese Corpus based on computer big data

3.1. The status quo of Japanese corpus construction

At present, most of the Japanese corpora in teaching development and application are small in scale. They are mainly based on the principles of simple teaching, convenient use and quick retrieval to realize the construction of Japanese expectation. At present, most of the Japanese corpora have the functions of word retrieval, generating thesaurus and subject words, and software carriers are given to realize the expected coding and compression of Japanese, so as to avoid the garbled code in the process of Japanese corpus appearing in the software.[5] Secondly, most of the corpus selected in the current Japanese prediction database is relatively old, mainly to avoid copyright issues, so as to reduce the construction cost of the corpus, at the same time, it can be used for teachers and students free of charge in the process of application and teaching development.

In addition, the Japanese corpus covers a wide range of subjects, covering a variety of types, thus ensuring the richness and diversity of Japanese expectations. In order to ensure the scientificity of the Japanese corpus, the selection of the corpus will also check the quality of the language materials. In
addition to selecting more authoritative and professional materials, it will also select more well-known monographs, so as to ensure the artistry and culture of Japanese language teaching and application.

3.2. Difficulties in the construction of Japanese corpus
At present, the difficulties in the construction of Japanese corpus are mainly reflected in the scale, material selection and timeliness, as shown in Table 2 below. Among them, in the scale level of Japanese corpus construction, if it wants to build a universal, network-based super large Japanese corpus, it will put forward higher requirements for the performance of the computer, and its construction cost is high. However, limited by the current school financial and other practical factors, the Japanese corpus constructed by it has a large space for improvement in both scale and performance\(^6\). This leads to the fact that most of the current school Japanese corpus can only provide simple retrieval function. Secondly, in terms of the selection of Japanese corpus, most of the current construction of Japanese corpus is limited by funds and human resources, which leads to the imbalance of corpus construction, and the selection of materials is difficult to meet the needs of literature and art, and it is difficult to realize the rapid improvement of students' Japanese level and literary literacy through the corpus.

In addition, in terms of the expected timeliness of Japanese, the current choice of Japanese corpus is mainly based on the form of literary composition, rather than fixed on the basis of the current Japanese literary form. This is mainly due to the fact that language is not very sensitive to timeliness, for example, people can better read and understand the words of a long time ago. However, timeliness is still one of the key issues in the development of Japanese corpus.

Table 2. Difficulties in the construction of Japanese corpus.

| Aspects       | Requirements                                      | Difficulties          |
|---------------|---------------------------------------------------|-----------------------|
| Scale         | Higher requirements for the performance of the computer | High construction cost |
| Material selection | Corpus should be scientific and representative | Balance of corpus |
| Timeliness    | Corpus be fixed in the form of literary composition | Relatively fixed      |

4. Teaching development and application of Japanese corpus based on computer big data

4.1. Function and technology of Japanese corpus based on computer big data
The functions of Japanese Corpus Based on computer big data mainly include the following aspects: firstly, collect all kinds of Japanese Corpus, and analyze and process the source and information of the corpus. Secondly, the collected Japanese Corpus is input into the database to lay the foundation for the realization of vocabulary and sentence retrieval. In addition, the functional modules of Japanese Corpus mainly include several modules and processes as shown in Figure 3.

Figure 3. Functional modules and process of Japanese corpus.
Among them, in the module level of acquiring corpus, the structure, content and characteristics of the collected corpus are analyzed, and the transcoding, parsing and content input of the documents are realized by coding. Secondly, at the level of corpus preprocessing module, the key lies in the preprocessing process of segmentation and alignment, and then it is imported into Japanese corpus. In addition, the preprocessing modules in Japanese corpus need to be sorted out and proofread manually. Finally, the functions of storing, searching and querying Japanese corpus database are realized.

4.2. Development and application of Japanese corpus based on computer big data
First of all, at the development level of Japanese corpus based on computer big data, we should realize the establishment of network corpus under the background of big data, ensure that the corpus is practiced and updated, optimize the scoring system, and improve the reliability and validity of the system by integrating human-computer scoring with teacher's marking. Secondly, at the application level of Japanese corpus, it should improve the ability of morphology, writing syntax, Japanese thinking and high discourse analysis ability.

In addition, the application of Japanese corpus improves students' mastery of sentence usage, collocation and context, and promotes students to understand and feel the differences of Chinese and Japanese language thinking based on Japanese anticipation database. The improvement of discourse analysis ability is to enable students to add corresponding sentences based on their own understanding, so as to make the text more logical and further improve the teaching effect of Japanese.

5. Conclusion
In summary, the application of computer big data technology can significantly improve the efficiency and accuracy of the collection and collation process of Japanese language materials, and can store the collected Japanese language materials based on linguistic principles. The research of Japanese corpus based on computer big data technology helps to code the Japanese corpus succinctly and effectively, and obtain a large number of useful data, so as to promote the development and application of Japanese teaching tools, as well as to promote the development and aggregation of teaching contents, viewpoints and genres.

Based on the analysis of the role of Japanese corpus in teaching development, this paper studies the role of Japanese corpus in promoting the contextualization of Japanese communication and students' dominant position in learning. Secondly, through the research on the difficulties in the construction of Japanese corpus based on computer big data, this paper analyzes the current situation and key points of Japanese corpus construction. Finally, through the research on the teaching development and application of Japanese corpus based on computer big data, the key functions and application key technologies of Japanese corpus based on computer big data are given, and its specific application is given.

References
[1] Du Zebing. A corpus based study on the structural and functional features of lexical chunks in Chinese Academic English [J]. Audio visual foreign language teaching, 2016 (05): 24-28.
[2] Geng Zhimeng, Zhong Chunlin, Liu Yuqin. Development and utilization of Japanese text corpus [J]. China education informatization, 2015 (1): 58-60.
[3] Hu Kaibao, Li Yi. A review of foreign corpus translation studies [J]. Contemporary linguistics, 2012, 14 (04): 380-395.
[4] Hu Kaibao, Mao Pengfei. Research on the creation and application of Contemporary English textbook corpus [J]. Foreign language audio visual teaching, 2016 (03): 8-10.
[5] Song Yajun. Applied research on corpus data driven English writing teaching mode in Higher Vocational Colleges [J]. Journal of Henan electromechanical college, 2014, 03:88-91.
[6] Yang Mei, Bai Nan. A survey of corpus translation research in China -- Based on data analysis of domestic academic journals [J]. Chinese translation, 2010, 31 (06): 46-50.