

Application of Digital Humanities and Corpus Linguistics in Language Teaching

Jialin Ren*, Le Zhang

University of Shanghai for Science and Technology, College of Foreign Languages, Shanghai 200093, China
Email: 1397445721@qq.com

How to cite this paper: Ren, J. L., & Zhang, L. (2025). Application of Digital Humanities and Corpus Linguistics in Language Teaching. *Social Sciences and Humanities*, 2(3), 44-51. ISSN Print: 3104-4328; ISSN Online: 3104-4336.

<https://doi.org/10.63313/SSH.9053>

Published: 2025-12-25

Copyright © 2025 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Abstract

This is a review article on the application of digital humanities and corpus linguistics in language teaching and learning. Corpus linguistics studies linguistic phenomena by constructing and analyzing large-scale text corpora, while digital humanities is an emerging cross-research field that applies computer and digital technologies to traditional humanities research and teaching. Combining the two to analyze their application in language teaching and learning can enrich the research methodology of language teaching and learning, and also provide digital humanities with powerful tools for language analysis.

The purpose of this paper is to comprehensively analyze the current status, achievements, challenges, and future development trends of the application of digital humanities and corpus linguistics in language teaching and learning, so as to provide systematic references for language teaching and learning research and practice.

Keywords

Digital Humanities; Corpus Linguistics; Language Teaching and Learning

1. Introduction

1.1. Research background

Nowadays, we are living in a new age of rapid digitalization and information explosion. At the same time, the education sector is also undergoing a significant transition phase. The classical ideas and practices of education are no longer accepted by the new generation of students, who are more amenable to educational practices related to teaching incorporating multimodal forms of language, like digital text, audio, and video, and so on, which are very interactive, engaging, and provide instant feedback.

In this context, digital humanities and corpus linguistics have rapidly developed. Digital humanities, being cross-disciplinary in nature, integrates technology with

humanities studies in-depth, shattering the conventional constraints of research in humanities studies in data processing, analysis, and display, thus imparting a new element of approach in language teaching. Digital humanities can enable teachers and learners to comprehend language from a macro level, culture, history, society, and multi-dimensions, besides affording a wide scope for innovative development in language teaching resources.

Corpus linguistics, on the other hand, revolves around large-scale empirical research on language data for purposes of organizing and analyzing it. Corpus linguistics helps highlight data on language actual use, structure, and semantics, based on constructed corpora types. In language teaching, corpus linguistics helps inform a scientific approach for determining language teaching content, teaching goals, language teaching methods, and student learning results, bringing language teaching closer to actual language use, making it more relevant and efficient. There appears a link between corpus linguistics and language teaching for better language use, based on corpus studies research.

In summary, the combination of digital humanities, corpus linguistics, and language teaching has great practical value and development prospects because of the dual driving factors of the digitization wave and the educational reform, which deserves to be explored and practiced.

1.2. Research Objectives and Significance

The purpose of the study is to systemically and deeply collate the multidimensional applications of digital humanities and corpus linguistics in language teaching. Theoretically, this study is very important for the enrichment of a theoretical system in language teaching. The union of digital humanities and corpus linguistics shatters the constraint of conventional language teaching theories, where attention was placed on the structure and function of language per se, and a different dimension, namely a cultural, social, and historical surrounding, and also a digital environment, is introduced for consideration. This aims at a deeper exploration of the nature of language, language acquisition law, and relationship between language and culture as well. In consideration of the macroscopic level of educational developments, it benefits for innovation and change for language teaching with a trend for educational modernization under the digital age.

2. Literature Review

2.1. Overview of Digital Humanities

As an interdisciplinary and burgeoning field with abundant and diversifying connotations, the application and target of digital humanities lie in conducting a deep exploration and widespread dissemination of materials and phenomena within humanities through the powerful driving force generated by digital technology. In a

narrow sense, it refers to building and studying digital humanities based on some specific areas requiring a universally standardized technological training process[1]. The development and application of digital humanities are not a simple overlaying process involving the application of digital technology over humanities research in traditional disciplines but also can profoundly transform the way and thinking mode of humanities research. In fact, based on abundant and advanced database technology and algorithms, on one hand, it focuses on applying various techniques, such as data mining and text analysis algorithms, and on analyzing and excavating the hidden information and relationship within humanities data through various data analyses and excavations by taking full advantage of computer algorithms and data analytical software.

2.2. Overview of Corpus Linguistics

Corpus linguistics first emerged in the mid-20th century. Most of the corpora in the early days were built in English-speaking countries. For example, the Brown Corpus, a large-scale English corpus containing texts of various genres which helped English research, was built in the US in the 1960s. Later, the LOB corpus was built in the UK, and both of them played an important role in promoting English grammar, vocabulary and discourse research at that time. Corpus linguistics not only opened up a new era of corpus language research, but is itself also a discipline: probabilistic induction and generalization of linguistic behavior[7].

The growth of computer technology and the expansion of storage capacity have also expanded the size and diversity of corpora, and a very significant event was the completion of the British National Corpus, or BNC, towards the end of the 1990s, comprising 100 million words of spoken and written material mostly gathered across a variety of sites within the boundaries of the United Kingdom, and comprising a representative and authoritative source for all language studies. Meanwhile, other language corpora such as the German Text Archive for German, or others such as the French National Corpus for French, have also been flourishing, and have been providing data for research and teaching purposes for their language.

The speed of the development of corpus linguistics in the 21st century is even more accelerated. There are multimodal corpora that not only involve text, but also incorporated audio and video information that can reflect the actual communicative context of language more thoroughly. For example, there are corpora of English speech that involve audio and video information that are beneficial for the study of the combined effect of linguistic and non-linguistic factors.

3. Research Status

The field of digital humanities and corpus linguistics in language teaching and learning has seen a vigorous and diverse exploration in recent years.

3.1. In the Aspect of Digital Humanities

Speaking of digital humanities, there are many scholars and educators who actively explore how to incorporate the valuable resources and innovative methods that digital technologies offer into the teaching practice of languages. On the one hand, there are many digital teaching resources and learning platforms that have arisen in the development of teaching resources, and the digital transformation is both technological progress and an upgrade of concepts in education[6].

Such online educational resources are not only confined to the conventional pattern of text, thereby including pictorial, auditory, as well as video dimensions. For instance, certain resources used for language instruction support the comprehension of the cultural implications of language on the part of the student by incorporating documentary videos related to history and culture, sound recordings of the recitation of literature, and corresponding pictorials. For example, some studies have pointed out that under the development of education informatization, the integration of information technology and early childhood language teaching is an inevitable trend, and that it is necessary to set up educational contexts based on the physical and mental development characteristics of young children and make reasonable use of information technology to optimize the teaching content and links, stimulate young children's interest in language learning, and enhance the ability to express themselves[5], and put forward the strategies for the application of information technology in early childhood language teaching[6], and at the same time, the application strategy of information technology based on the Digital humanities-based project-based learning is also being gradually promoted, with teachers guiding students to utilize digital tools to carry out research on specific language and cultural topics, such as by creating digital stories and constructing cultural knowledge maps, in order to cultivate students' comprehensive language use skills, critical thinking, and teamwork. As well as in the international academic community, more attention has been paid to the help and application of artificial intelligence to language teaching[4], and there are also studies exploring digital humanities tools that can be used in different stages of process genre writing and their application methods[8]. In terms of teaching evaluation, with the data collection and analysis function of digital humanities technology, it is possible to collect and analyze multi-dimensional data such as students' learning trajectories, interactive behaviors, and creation of works in digital learning environments, so as to build a more comprehensive, precise and personalized learning evaluation system, and to provide students with targeted feedback and guidance.

3.2. In the Aspect of Corpus Linguistics

In the term of corpus linguistics, researches on its application in language teaching have achieved fruitful results too. In vocabulary teaching, corpora are widely used in statistical analysis of vocabulary frequency by researchers to determine the core

vocabulary and high-frequency vocabulary of different language levels and different fields of specialization and provide a scientific basis for the screening of vocabulary teaching content. For instance, in the vocabulary teaching in Business English, the high-frequency vocabulary “negotiation” and “contract” can be exactly positioned by means of the analysis of the corpus of Business English, and we can create the teaching activities so that the students can master the vocabulary with different language levels and professional fields in accordance with the collocation of vocabulary in the corpus. The corpus can provide the learners with a huge amount of neatly classified real corpora by the search of KWIC. Collocation teaching plays a crucial role in language study since the study of the foreign language actually represents the process of mastering the “explicit knowledge” consciously, which needs a great number of language exposure to materials[10].

In the field of grammar teaching, corpus linguistics provides a wealth of real-life contextualized examples for the teaching of grammatical rules. Instead of relying only on artificial examples in grammar books, teachers can extract a large number of examples of grammar use in natural contexts from the corpus, so that students can more intuitively feel the diversity and flexibility of grammar rules in actual language communication. For example, when teaching English tenses, the corpus selects examples of tenses in different genres of texts (such as news reports, novels, academic papers, etc.), so that students can experience the subtle differences in the choice of tenses in different contexts. For the teaching of discourse, the corpus tool can be used to analyze the structural features, articulation devices and genre styles of the discourse. Through the corpus analysis of a large number of discourse pieces of the same genre, the typical structural frameworks and common articulators of discourse pieces of specific genres are summarized, which help students better understand and create different types of discourse pieces, such as academic papers, narrative texts, expository texts, etc. There are also studies that propose the use of corpora to allow students to explore new modes of learning on their own, a training process that encourages students' curiosity about linguistic patterns and, in a broader sense, facilitates their continued development as independent language learners[3].

In addition, there have been a number of exploratory studies on the application of the integration of the two. Some studies have attempted to combine the visualization techniques of digital humanities with the data mining results of corpus linguistics, presenting corpus analysis data in the form of intuitive charts and graphs, etc., to help students and teachers understand linguistic patterns and features more conveniently. For example, using a self-constructed comparable fiction corpus, the syntactic usage features of translated Chinese relative to original Chinese are examined through the text classification task and feature analysis of the random forest algorithm[11]. The frequency distribution of vocabulary in a particular text is demonstrated through word cloud mapping, and semantic associations between

vocabulary are presented with network mapping, etc. There are also studies aiming to combine the interactive teaching platform for digital humanities with corpus resources in order to create an overall environment for language training by combining language learning, resource searching, and interactive communication, so that learners are able to apply the corpus resources for independent learning, investigation, as well as project-based learning activities about digital humanities, and communicate with their teachers and schoolmates. Huang (2018) explored the current application scenarios of online corpora in foreign language teaching and the challenges they face, emphasized the need for teachers to play a moderating role in the integration of corpora and teaching methods, and pointed out the under-application of corpora in classroom teaching as well as in teacher training programs[2].

However, in spite of all these improvements, there are still some limitations within the current research. As far as the application of technology is concerned, some digital humanities and corpus linguistics related software and tools are quite difficult to handle and demand high IT literacy among teaching and non-teaching staff. As far as the quality and ethical issues are concerned, the representativeness and accuracy of corpus information and the concerns related to copyright and privacy during the process of data utilization still have to be resolved. Meanwhile, in teaching practice, how to better balance the relationship between traditional teaching methods and the application of digital humanities and corpus linguistics, as well as how to personalize the design of teaching programs according to the individual differences and learning needs of different students, etc., are the directions that need to be deeply considered and explored in the current research.

4. Conclusion

Through a thorough and in-depth analysis of the application of digital humanities and corpus linguistics studies in language teaching, a number of conclusions have been reached, which have been of great importance.

In the first place, the integration of digital humanities into language teaching presents a unique appeal and achieves remarkable effects. On one hand, it succeeds in shattering the integrity and limitedness of traditional language teaching resources by adding numerous and diverse cultural resources and information from historical documents to the classroom using digital modes. On the other hand, the teaching patterns of project-based learning and inquiry-based learning presented by digital humanities make great efforts to train students' ability of independent learning and teamwork spirit.

Corpus linguistics also proves to be very fruitful in language teaching. In vocabulary teaching, corpus-based vocabulary frequency analysis can give accurate data support for determining the priority of teaching. Taking English teaching as an example, through the large-scale English corpus analysis, the core vocabulary to be mastered at different levels of students' English proficiency has been clarified; meanwhile,

with the help of corpus examples of vocabulary collocations, students can learn the usage of vocabulary in real contexts, which avoids the drawbacks of mechanical memorization and detachment from contexts in traditional vocabulary teaching.

From the perspective of synergistic use between digital humanities and corpus linguistics, it can be found that there is a strong level of complementarity between these two disciplines. The visualization technology and data expression form of digital humanities can visually represent the complex data that is extracted from corpus linguistics in a more intuitive form, for that matter, and in addition, based on large-scale real-world language data from corpus linguistics, a strong foundation can be constructed for digital humanities research at a language level, which can, in turn, allow the research in digital humanities to dig much deeper into the relationship between language, culture, and society, or other variables.

However, it must not be neglected that there exist some challenges in digital humanities and corpus linguistics in the process of language learning. On one hand, in terms of technology, because of the complexity of software and tools, there exist limitations to popularizing these applications in class. Additionally, a lack of IT training exists in most teachers. Furthermore, with regard to data quality, issues associated with copyright and privacy in data use exist.

In a word, digital humanities and corpus linguistics have made remarkable achievements in language teaching by introducing new ideas, abundant resources, and effective teaching methods. However, many technical, data, and practical problems have to be overcome to realize their wider and deeper application in language teaching. Therefore, future research and practice should simplify the technical operating process, improve the quality of teachers' IT literacy training; strengthen corpus data construction and management, perfect the system of quality management and control of corpus data, and elucidate the ethical norms of corpus data; explore further and design personalized teaching schemes according to different teaching contexts and learners' needs.

References

- [1] Battershill C, Ross S. *Using digital humanities in the classroom: a practical introduction for teachers, lecturers, and students*[M]. Bloomsbury Publishing, 2022.
- [2] Huang L S. Taking stock of corpus-based instruction in teaching English as an international language[J]. *RELC Journal*, 2018, 49(3): 381-401.
- [3] Kennedy C, Miceli T. Cultivating effective corpus use by language learners[J]. *Computer Assisted Language Learning*, 2017, 30(1-2): 91-114.
- [4] Liang J C, Hwang G J, Chen M R A, et al. Roles and research foci of artificial intelligence in language education: an integrated bibliographic analysis and systematic review approach[J]. *Interactive Learning Environments*, 2023, 31(7): 4270-4296.
- [5] Shisi Fu. Exploration of the Integration of Information Technology and Preschool Language Teaching in Practice [J]. *Journal of Huzhou Normal University*, 2024, 46 (06): 112-116
- [6] Xiaobing Fang. Electronic Education, Online Education, and Intelligent Education: The Digital Transformation of International Language Communication Institutions [J]. *Jour-*

- nal of Guizhou Normal University (Social Sciences Edition), 2024, (05):23-32.
- [7] Shichun Gui et al. Corpus Linguistics and Chinese Foreign Language Teaching [J]. Modern Foreign Languages, 2010, 33 (04): 419-426
 - [8] Haiying Liang. Teaching College English Writing from the Perspective of Digital Humanities: Ideas and Methods [J]. Foreign Language World, 2022, (06):50-57.
 - [9] Jinwei Ma. Research on the Application Strategies of Information Technology in Pre-school Language Teaching [J]. National General Language Teaching and Research, 2024, (06):179-181.
 - [10] Zhonghua Xiao et al. Corpus and Language Education [J]. Chinese Foreign Language Education, 2008,1 (02): 48-58+79
 - [11] Shuneng Zhong et al. Exploring the Path of Applying Machine Learning Text Classification to Translation Research under the Digital Humanities Paradigm: Taking the Study of Syntactic Features in Chinese Translation as an Example [J]. Journal of Shanghai Jiao Tong University (Philosophy and Social Sciences Edition), 2024, 32 (08): 1-17