

Construction of a Corpus of Character Names and Their Linguistic and Cultural Connotations in Online Games and Online Novels

LinYE Li

*Hebei University, Baoding, China
17692362981@163.com*

Abstract. In the midst of the rapid development of digital culture, online games and novels have become crucial means of presenting unique aspects of Chinese culture, attracting an ever - growing number of readers and gaining more and more international recognition; the aim of this study was to build a special corpus for analyzing character names in online gaming and literary settings divided into two main areas: the language features and cultural implications contained in these names; the corpus included complete information about character names, related literary works, types of works, gender of characters, their personality descriptions, and the cultural patterns they represented; through improving the data resources and analytical tools in the field of digital culture, this research tried to push forward academic linguistic research and boost the development of the cultural industry.

Keywords: online games, online novels, character naming, corpus, linguistic and cultural connotations

1. Introduction

As emerging cultural forms, online games and online novels attract a vast audience, with character naming serving as the soul of a game or novel. In *On the Naming and Significance of Characters in Lu Yao's Novels*, Likuai Dang and Renli Liu emphasize that character naming is a crucial component of literary creation and explore the naming characteristics in Lu Yao's novels [1]. Similarly, Hua Cheng pointed out in *The Art of Naming Characters in Thunderstorm** that character names often reflect the spiritual realm, cultural literacy, and aesthetic preferences of the namer [2]. Research on the cultural connotations of Chinese character naming has mainly focused on classical literary works. Scholars have thoroughly analyzed the naming methods and underlying cultural meanings in classics such as *Dream of the Red Chamber*, *Water Margin*, and *Romance of the Three Kingdoms*. For example, in *Dream of the Red Chamber*, character names frequently contain phonetic allusions, such as "Zhen Shiyin" (甄士隐), which is a homophone for "true events concealed" (真事隐). This unique naming technique demonstrates the author's approach to expressing thematic depth and conveying cultural connotations. However, in the realm of online culture, character naming often exhibits homogenization and formulaic patterns. For instance, in fantasy novels, male characters' names commonly include words such as "Dragon" (龙) and "Xiao" (霄). Currently, research on "character naming" remains limited and primarily focuses on character design rather than the naming itself. Pei Sun and other scholars have explored the integration of traditional cultural elements in character

design in games such as *Honor of Kings* and *League of Legends*, but have not specifically addressed character naming [3]. Thus, existing studies lack systematic and comprehensive analyses. There is no complete and well-structured dataset for collecting and organizing character names, with most research focusing on individual works or specific genres, making it difficult to present a holistic view and general patterns of character naming. To address this gap, this study extensively collects character names from online games and novels, drawing upon methodologies from studies such as Construction and Analytical Model of a Chinese Multi-party Dialogue Discourse Structure Corpus [4], Construction and Application of a Sentiment Recognition Corpus for Chinese MOOC Comments [5], and Construction of a Sentiment Corpus for Citations in Chinese Literature [6]. Using data mining and natural language processing techniques, supplemented by manual verification and annotation, this research aims to build a corpus of character names from online games and novels. It will analyze the linguistic features and cultural connotations of these names and explore their role in cultural transmission, providing valuable references for both creative and academic research in the online cultural field.

2. Practice and feedback on the corpus questionnaire

2.1. Explanation of the questionnaire survey

To thoroughly assess how linguistic corpora are utilized in naming characters within online gaming and virtual literature, this study developed a multi - part questionnaire which had four sections: one about demographic information, another regarding attitudes towards character names, a third concerning perceptions of cultural subtleties, and the last about the influence of cultural spread; data were collected using both digital and traditional methods to get a diverse and representative group of participants and eventually 300 questionnaires were completed but 12 were discarded because they were incomplete so there were 288 valid responses with a response rate of 96% and this set of data forms a solid basis for future analysis.

The survey tool consists of four different parts; first part collects demographic information such as age and gender from participants which allows people to look at differences in attitudes regarding character naming among various demographic groups; second part focuses on finding out how much online game players are involved with and like the names of characters in games; third part is aimed at readers of online novels and tries to figure out their views and evaluations of character names in stories including picking out their favorite character names from novels; last part explores the cultural meanings behind character names by assessing whether participants understand the cultural details in character names and how they contribute to spreading cultural stories; besides these fixed questions, the questionnaire also has open - ended questions to get qualitative opinions from both gamers and readers.

2.2. Explanation of survey results

2.2.1. Basic information and usage habits: group characteristics and the necessity of corpus construction

The survey data showed that the main consumer group for online gaming and digital literature was young people aged between 18 and 30 who made up 80% of all respondents; in the area of digital literature, romantic stories were the favorite type among female readers as they appealed to 56% of women while male readers liked fantasy - themed novels which interested 20% of the participants; when analyzing the character names in these genres, it could be seen that the heroes in romance novels usually had names emphasizing grace and emotion whereas the characters in fantasy novels generally bore names showing their connections to elements or unique backgrounds.

To make certain that the corpus mirrors the varied preferences of different demographic groups, a wide range of genres and stylistic features were incorporated, which ensured the comprehensiveness and representativeness of the sample population.

2.2.2. Attention to character naming and preferences: naming characteristics and the standardization of the corpus

More than sixty percent of those who took part in the survey said they deliberately focused on the naming of characters in games or literary works and a large majority of seventy - six percent agreed that a name should match the character's personality, also around sixty - six percent of the respondents liked names filled with poetic and cultural charm while about thirty percent preferred names that were simple and easy to remember, this shows how consumers are engaged with character names which emphasizes the great significance of naming in both literature and gaming stories so it stresses the need for systematic standardization when compiling language databases, such databases should carefully note and sort out the special features of these names like their cultural origin, style details and emotional tendencies, for instance, the character "Li Bai" (李白) in the game "Honor of Kings" is not just a name but also reminds people of the cultural heritage of the well - known Tang dynasty poet, by strictly following annotation rules, the database could be a creative tool for content makers and a research place for scholars to look into the relationships between character naming, character identity and cultural inheritance.

2.2.3. Awareness of cultural connotations: cultural elements and the comprehensiveness of the corpus

Around 52% of the participants stated that character naming mirrors a great deal of cultural heritages and approximately 34% specifically mentioned the influence of traditional cultural patterns like poetry, historical allusions, and mythology; it was remarkable that the video game "Honor of Kings" included the name "Li Bai", while "Battle Through the Heavens" (斗破苍穹) had "Xiao Yan"(萧炎), which showed the combination of traditional culture and modern creative story - telling; these findings emphasized the significance of taking an inclusive method when compiling the corpus so the corpus to be proposed should contain various naming rules from different cultural sources covering traditional, present - day, and cross - cultural impacts; the compilation and analysis of this data would make the corpus serve as a resource for creators looking for names of cultural importance and provide crucial material for cultural research; scholars could use this corpus to study how traditional cultural elements were applied in modern creative works and evaluate their role in passing on culture.

2.2.4. Cultural communication effects: naming influence and the dynamic nature of the corpus

Forty - four percent of the participants stated that the naming of characters has a great influence on their tendency to recommend games or novels to others and another 43% agreed that character names with unique Chinese cultural features have strong appeal in the international market, these findings emphasize the crucial role of character naming in cultural spread and strengthen the need for a flexible and dynamic corpus, as the fields of online games and novels develop new types and cultural themes keep emerging so it was essential for the corpus to be able to be updated continuously to include new character names promptly to keep pace with industrial progress and thus the corpus remained a relevant and up - to - date resource for studying cultural diffusion.

2.3. Principles and framework of corpus design

2.3.1. Principles of corpus design

To make it easier for research and practical application regarding the naming of characters in online games and virtual stories, this study sticks to four main design principles: being comprehensive, standardized, dynamic, and useful, which are crucial for the corpus to cover the unique features of character names and at the same time maintain strong data - handling and updating systems.

2.3.1.1. Principle of comprehensiveness

The corpus was compiled to cover a wide range of types and style differences existing in online games and literature so as to accurately reflect the variety in character names across different cultural settings and creative methods; it included some character names from traditional genres like martial arts, fantasy, and science - fiction as well as those from emerging genres such as urban supernatural stories and post - apocalyptic survival fictions which had become very popular in modern literature and gaming; this approach helped cut down on sample bias and made sure of a complete presentation of the norms and tendencies in character naming.

2.3.1.2. Principle of standardization

To maintain consistency and accuracy in the makeup of the dataset, every collected item undergoes a strict and standardized annotation process which carefully records categorical information like the literary work related to the character, genre categorization, gender, personality traits, and the cultural subtleties contained in names; for example, the character Wuqing (无情) from the story Justice Online (Nishuihan, 逆水寒) was included in the corpus and was assigned personality descriptions such as being outwardly cold but inwardly kind - hearted and connected to the Taoist philosophical idea of transcending emotions (Taishang Wangqing, 太上忘情); this systematic annotation approach ensures a uniform data structure, making the data more useful and easier to analyze and study later on.

2.3.1.3. Principle of dynamism

The corpus was designed to handle the continuous development of online games and literary works through the integration of dynamic updating functions, systematically collecting character names from new and popular titles to keep up with emerging genres and new trends, and notably, in light of the rapidly growing metaverse concept over the past few years, quickly incorporating character names from relevant genres like that of Hongxian (红线) from the novel Ballad of the Boundless (燕云十六声), so as to maintain its alignment with the concurrent progress in the industry.

2.3.1.4. Principle of usability

The database was designed to include an easy - to - use search and retrieval system which made it easier for researchers to get information quickly, specifically speaking, the system allowed users to look up particular character names like Weifang He (何惟芳) and then immediately get relevant information from National Beauty and Fragrance (Guose Fanghua, 国色芳华), such as the character's personality traits and the cultural subtleties associated with the name, this design characteristic significantly increased the usefulness and effectiveness of the research process.

2.3.2. Process and framework of corpus construction

The compilation of the character name corpus was established through a rigorous methodology encompassing four principal phases: the aggregation of raw data, systematic organization of the information, analytical processing of the dataset, and the subsequent construction and management of the corpus. This resultant corpus encompasses a diverse array of genres and stylistic variations.

2.3.2.1. Data collection

The compilation of the corpus depends on the collection of a large amount of data mainly from official gaming websites, digital novel platforms, and special forums for online literature and gaming communities and this combination of different sources was essential to form a comprehensive and representative dataset.

2.3.2.2. Information organization

After collecting the data, the dataset was systematically arranged by categorizing it and the compilation made clear distinctions between online literary works and games based on their genres and stylistic features; for instance, online novels were divided into different subgenres like historical, contemporary, science - fiction, romance, etc., while at the same time online games were classified into types such as role - playing games (RPGs), open - world adventure games, and other genres, which enabled the corpus to better reflect naming norms in various cultural contexts.

2.3.2.3. Data analysis

Based on the organized data, the study conducts two major types of analyses:

Linguistic Investigation: Carrying out a thorough exploration of the phonetic, semantic, and pragmatic features of character names within the context of linguistic research so as to clarify inherent naming patterns.

Cultural Connotation Analysis: This study involves looking into the origin, present - day meaning, and symbolic implications of character names and includes a detailed exploration of traditional cultural patterns - such as poetic allusions, historical references, and mythological impacts - contained in character names and evaluates how these patterns mirror and exemplify cultural spirit.

2.3.2.4. Corpus construction and operation

In the initial stage of corpus construction, data collection was followed by meticulous linguistic and cultural analysis, after which the polished data was combined into a special web - based platform or database having an easy - to - use search feature allowing users to carry out searches by character names, types of works, cultural themes, and various relevant keywords, and to keep the corpus updated, it followed a dynamic updating procedure, methodically including new character names from official gaming websites, literary platforms, and online discussion forums so as to ensure that the corpus stayed a contemporary storehouse reflecting the present - day progress in the industry and naming tendencies in this field.

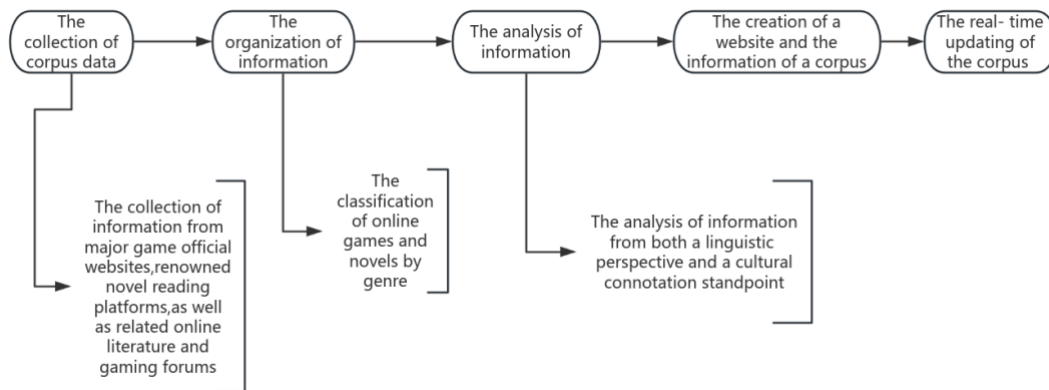


Figure 1. Corpus construction workflow (illustration reference)

2.4. Case study of corpus content

In the part about presenting corpus content, we conducted a thorough exploration of the language features and cultural significance in character naming through targeted case studies but because of space limitations, only a portion of the material was shown and in order to illustrate the abundance and diversity in the corpus, we chose to emphasize two typical character names: Wuqing (meaning "Without Feeling") from the online game Justice Online (Nishuihan, 逆水寒) and Weifang He (何惟芳) from the online novel National Beauty and Fragrance (Guose Fanghua, 国色芳华).

Table 1. Character name analysis from justice online

Game	Character Name (Aliases)	Linguistic Analysis	Cultural Connotation Analysis
Justice Online (Nishuihan, 逆水寒)	Wuqing (无情) (Aliases: Yuyu Sheng 盛崖余 Yueya'er 月牙儿)	Phonetics Wuqing (wú qíng) has a level-tone pronunciation with a steady and crisp articulation. The concise and sharp sound reflects the character's cold and emotionless exterior.	Origin The name Wuqing may be influenced by the traditional Chinese concept of "detachment from emotions" (Taishang Wangqing, 太上忘情) in Daoism, which emphasizes transcending emotions to observe the world objectively and calmly. In wuxia culture, such figures are often known for their exceptional martial arts skills and ruthless demeanor.
		Semantics Wuqing means "without emotion" or "cold-hearted." In the game, this character is often portrayed with a stern and distant image, showing no personal bias when solving cases—his name aptly reflects his personality.	Reflection of Modern Society In real life, people often need to maintain rationality and composure. The qualities represented by Wuqing—calmness and logic—resonate with individuals who adopt a detached or reserved attitude when dealing with complex interpersonal relationships and societal challenges as a form of self-protection.

Table 1. (continued)

Function	Symbolic Expectation
The name Wuqing serves as the character's primary identifier, reinforcing his image as an iron-faced detective in the wuxia world. His name instills a sense of respect and awe, strengthening his authority and impartiality in executing justice and solving crimes.	Through future story developments, the character is expected to maintain his sense of justice and integrity. As he navigates the darker aspects of the martial world, he embodies the steadfast force of righteousness. Players are also encouraged to learn from his adherence to principles.

Table 2. Character name analysis from national beauty and fragrance

Novel	Character Name	Linguistic Analysis	Cultural Connotation Analysis
National Beauty and Fragrance (Guose Fanghua, 国色芳华)	Weifang He (何惟芳)	Phonetics The three characters in Weifang He (何惟芳) follow the tone pattern of rising tone (阳平), rising tone (阳平), and level tone (阴平), creating a natural tonal variation. The first character He (何) is pronounced as "hé," an open vowel with a clear and resonant sound. The second character Wei (惟) is pronounced as "wéi," featuring the "ei" vowel, which gives it a soft and gentle tone. The third character Fang (芳) is pronounced as "fāng," with a nasal ending, making it crisp and melodious. The combination of these three characters creates a harmonious and rhythmic name that is pleasant to pronounce.	Origin The character Fang (芳) frequently appears in classical Chinese literature and is often used to describe women. For example, in Li Sao (离骚), the line "I gathered autumn orchids as adornments" (纫秋兰以为佩) uses fragrant plants as metaphors for noble personalities. Fang inherits this cultural tradition, symbolizing beauty and virtue. Meanwhile, the character Wei (惟) carries deep cultural significance, as seen in The Book of Documents (尚书), where the phrase "唯精惟一, 允执厥中" emphasizes a spirit of focus and purity. The name Weifang He embodies both elements, reflecting a heritage of traditional virtues and the pursuit of moral integrity.
		Semantics The character Wei (惟) means "only" or "to think," representing focus and purity, suggesting the character's determined and unwavering nature. The character Fang (芳) originally refers to the fragrance of flowers and is commonly used in literature to describe beauty and virtue. In this context, it implies that Weifang He possesses both exceptional beauty and a noble character. Overall, the semantic combination of Weifang He conveys the image of a devoted and virtuous woman.	Reflection of Modern Society In the novel's setting, Weifang He lives in a complex social environment where women face numerous restrictions and challenges, such as the expectation to be virtuous wives and adhere to Confucian moral codes. Her name signifies that despite these hardships, she remains steadfast in her integrity and refuses to conform to societal corruption. This reflects the broader reality of women striving for self-worth and truth under the constraints of feudal ethics. From a contemporary perspective, the name embodies people's admiration for purity and virtue—qualities that remain cherished regardless of changing times. Sincerity, kindness, and self-commitment continue to be valued and celebrated.

Table 2. (continued)

Function	Symbolic Expectation
From a narrative perspective, Weifang He's name encapsulates her defining traits. In the novel, she remains loyal to her love and values, resisting the temptations of fame and fortune, mirroring the Wei (惟) character's sense of focus and purity. Her beauty, talent, and noble character align with the refined meaning of Fang (芳). Additionally, her name plays a crucial role in advancing the plot—her exceptional reputation and striking appearance attract attention from various characters, drawing her into complex interpersonal conflicts and significant events.	From the character's personal fate, Weifang He's name symbolizes her ability to uphold her virtues and dignity despite hardships, ultimately leading to happiness and inner peace. In terms of the story's development, she is expected to break free from feudal constraints, pursuing freedom and true love, thereby driving the plot in a positive direction. From the reader's emotional perspective, her name embodies an aspiration for idealized human qualities and romantic love. Readers hope to see a character like Weifang He fulfill her life's purpose, bringing warmth and hope to the narrative.

3. Conclusion

This study abided by the principles of being comprehensive, standardized, dynamic, and useful with the aim of thoroughly probing into the linguistic features and the cultural implications hidden in character names, so through the use of advanced data - mining techniques and NLP tools, we assembled a large corpus of character names from online games and literary works which includes various types and styles and serves as an organized resource for academic research and practical use in the domain of digital culture, but the present progress is confronted with difficulties such as the standardization of corpus annotation procedures not being adequate and some cultural nuances not being fully incorporated, thus future research will concentrate on improving the annotation rules, increasing the coverage of cultural references, and using complex analytical methods to figure out the inner relationships among character naming, cultural background, and social psychology in order to achieve the final objective of gaining more profound understandings which can be helpful for academic debates and the growth of cultural industry applications.

References

- [1] Dang, Likuai, & Liu, Renli. (2023). On the Naming and Significance of Characters in Lu Yao's Novels. *Journal of Weinan Normal University*, 38(11), 54-62. DOI:10.15924/j.cnki.1009-5128.2023.11.008.
- [2] Cheng, Huazao. (2017). The Art of Character Naming in *Thunderstorm*. *Chinese Language Teaching and Research*, (29), 8-9.
- [3] Sun, Bei, Li, Yuanlin, Cao, Yong, et al. (2024). Application of Traditional Cultural Elements in Character Design of Competitive Online Games. *Toy World*, (02), 128-130.
- [4] He, Weikai, Jiang, Yuru, Chen, Jie, et al. (2023). Construction and Analytical Model of a Chinese Multi-party Dialogue Discourse Structure Corpus. *Journal of Beijing Information Science and Technology University (Natural Science Edition)*, 38(05), 10-18. DOI:10.16508/j.cnki.11-5866/n.2023.05.002.
- [5] Wei, Xiaocong, & Yu, Lan. (2023). Construction and Application of a Sentiment Recognition Corpus for Chinese MOOC Comments. *Journal of Chongqing University of Technology (Natural Science)*, 37(04), 174-181.
- [6] Xu, Linhong, Ding, Kun, Chen, Na, et al. (2020). Construction of a Sentiment Corpus for Citations in Chinese Literature. *Journal of Information Science*, 39(01), 25-37.