教育教学类论文

基于语料库的中国太空计划白皮书翻译中中国形象研究

A Corpus-Based Study of China's Image in the Translation of the White Paper *China's Space Program*

[redacted]

xhtongxue@163.com

Abstract. The white paper *China's Space Program* is published to offer a brief introduction to China's achievements in this field and its main tasks in the future, in order to help the international community better understand China's space industry. It is not only an official document of the government, but also an important way to constructing China's image. This paper conducts a corpus-based study of translation and China's image by analyzing the translation of original text and keywords of English text. Findings show that the white paper presents a multi-dimensional image of China, including technological leader, international cooperation partner, innovation driver, global contributor, advocate for sustainable development, and space exploration pioneer.

Keywords: China's Space Program · corpus · China's image · translation · ICASSE

1 Introduction

In recent years, with the continuous acceleration of globalization in our country and the significant improvement of China's international influence, the study of China's image construction has received great attention from the government and academia. Kaibao Hu and Xin Li (2017) believe that the existing research on China's image is mainly qualitative research, and corpus-based research on China's image is quite rare. As a special cross-cultural exchange activity, the role of translation in the shaping and dissemination of China's image cannot be ignored. Political text translation can directly disseminate the China's image it builds to foreign audiences. In view of this, we should vigorously carry out research on translation and China's image, especially corpus-based translation and China's image research.

General Secretary Xi Jinping pointed out that "exploring the vast universe, developing the space industry, and building a space power are our space dreams that we have pursued unremittingly." China has always regarded the development of the space industry as an important part of the country's overall development strategy, and has always adhered to the exploration and use of outer space for peaceful purposes. Our country's space industry has created brilliant achievements represented by "two bombs and one

Z. Jing et al. (Eds.): ICASSE 2024, LNEE 1360, pp. 262–270, 2025.
https://doi.org/10.1007/978-981-96-2440-9_25

satellite", manned spaceflight, lunar exploration, and Mars exploration. It has embarked on a development path of self-reliance and independent innovation, and has accumulated a profound and broad space spirit. Relevant white papers record China's development and achievements in the space field. So far, the State Council Information Office of China has issued five white papers on China's Space in 2000, 2006, 2011, 2016 and 2021. These white papers are not only official government documents, but also important tools for shaping a country's image.

This research will start from the theory of imagery, use corpus data as an analysis method, use the original Chinese text and official English translation of the 2011, 2016, 2021 *China's Space Program* white paper as the research corpus, and use the British National Corpus (BNC) as the reference database to explore the theme words and translation effects of the *China's Space Program*, and explain how the Chinese government shapes its own international image through space texts. This research helps to supplement the corpus-based space political texts and the relevant literature on the English translation of the white paper, in order to expand the research field of corpus linguistics and translation studies. Secondly, doing a good job in the translation of political texts can enhance the international community's understanding of China's policies and policies, and ultimately achieve the goal of building a good national image of China, enhancing the international influence of Chinese discourse and promoting the self-shaping of the national image (Li Xie and Yinquan Wang 2018).

This research mainly answers the following questions:

First, compared with the British National Corpus, what are the subject words in the English translation of the *China's Space Program*?

Second, what kind of China's image is constructed by the *China's Space Program*?

2 Imageology in Translation

Imageology is one of the research categories of comparative literature. The constructed nature of image or identity is the core of image research (Leerssen J 2007). Traditional imageology research mainly focuses on the attention of the image itself, especially the description of the "hetero image" constructed in literary works. In contrast, modern imageology not only examines the process of the generation and dissemination of "hetero image", but also pays special attention to how these images promote the construction and evolution of the national cultural identity and image. In the field of translation studies, the discussion of "hetero image" is not the ultimate goal of researchers, but more focused on analyzing how these images have an impact on the shaping of their own national image and identity through the production and dissemination of translation (Xiaoyun Zhang 2011).

Since the 1980s, many scholars of translation studies have discussed the issue of image building. *Cultural Encounters in Translated Children's Literature: Images of Australia in French Translation*, published by St. Jerome's Publishing House (2007), and *Interconnecting Translation Studies and Imagology*, published by John Benjamin Publishing House (2015), are representative research results of the combination of translation studies and iconography abroad. In China, Yilin Publishing House published Zhang Xiaoyun's doctoral thesis "Imagery Perspectives in Translation Studies - A Case

Study of Kerouac's on the Road" in 2011. Foreign Language Teaching and Research Press published Dr. Lu Xiaojun's "National Image and Foreign Propaganda Translation Strategies" in 2015. In 2016, Shanghai Jiao Tong University held the first high-end forum on China's image research. Many scholars emphasized that the role of translation in the construction of China's image must be given enough attention (Hongyun Wang 2018). With the wide application of imagery in translation studies, the construction of self-image and the interaction between hetero image and self-image have become important research contents of modern imagery.

The research methods of modern iconography include text internal research and text external research. Text internal research mainly discusses what kind of image is constructed in the text and how to construct the image, while text external research focuses on external historical and cultural factors that affect the construction of the image. This paper mainly starts from the internal research of the text, and makes a case analysis of the national image constructed in *China's Space Program*.

To sum up, the research field of China's image needs to be expanded and innovated. First, the application of corpus technology, giving full play to the technical advantages of corpus in text analysis, using both qualitative and quantitative methods, combining description and interpretation, starting from the regularity of the application of specific language structures, reveals the China's image constructed by specific texts; second, based on the relevant theories of iconography, explore the internal reasons for the formation of China's image from the text level and the ideological level, and analyze how to shape the China's image at the symbolic level; third, the translation of white papers in the aerospace field can supplement the relevant literature on the translation of scientific and technological and political texts. In the context of the rapid development of our country's space industry, it is of great significance for the times to enhance the international community's understanding of China's space industry and to carry out foreign translation research on space white papers.

3 Research Design

3.1 Introduction to the Corpus

White Paper is an internationally recognized official document, an important document or report officially issued by the government on an important policy or issue, and an important carrier and means of external publicity (Meisong Chen and Hui Wang 2020). So far, the Chinese government has issued five *China's Space Program* in 2000, 2006, 2011, 2016 and 2021, mainly introducing the practical achievements of China's aerospace to achieve innovative leapfrog development, promote the modernization of space governance, and actively carry out international space exchanges and cooperation. It expounds China's policy concepts and proposals for in-depth international exchanges and cooperation in space and promoting the construction of a community with a shared future for mankind in outer space. Looking forward to the next five years, China's aerospace will embark on a new journey of building a space power in an all-round way and build a new pattern of space international cooperation. Vision is not only an official document of the government, but also an important tool for shaping the country's image.

In order to ensure the authority of the corpus and the accuracy of the research, this study collected the Chinese and English bilingual texts of the three *China's Space Program* issued by the State Council Information Office (scio.gov.cn) of the People's Republic of China in 2011, 2016 and 2021 as the target corpus, including 33,719 characters in Chinese corpus and 20,617 characters in English corpus.

3.2 Research Methods and Procedures

In the process of text research, this study uses a combination of quantitative and qualitative analysis methods. First, the sentence-level alignment of the corpus is carried out using the Tmxmall platform software; second, the corpus retrieval and analysis software AntConc is used to count the subject words of the corpus; then, the corpus retrieval tool BFSU Paraconc1.2.1 is used to conduct text retrieval to label and classify representative Chinese vocabulary and translation statistics; finally, on the basis of the previous steps, in-depth analysis, research and discussion are carried out to draw research conclusions. The authoritative interpretation platforms for Chinese and English vocabulary involved in the research process include *Oxford Learner's Dictionaries* online dictionary, terminology online platform, etc.

First, we use the AntConc 3.5.8 to extract the top 20 keywords from the corpus of *China's Space Program*. We will sort the keywords according to their keyness to generate a thesaurus. Keywords refer to words that appear far more frequently than the normal when compared with the reference text (Scott M and Tribble C 2006). Since different keywords represent different work themes and their focuses, through the study of keywords, we can draw the focus of the text, and thus get the Chinese image it builds. In addition, we will also select representative Chinese vocabulary and its English counterwords in the original text for research to explore the translation effect of the white paper.

3.3 Research Findings

We use the Antconc 3.5.8 software to generate the BNC glossary of the English translation of *China's Space Program*, and sort it by keyness. The top 20 keywords are shown in Table 1:

It can be seen from the table that with BNC as the reference, the two words with the highest theme in the English translation of *China's Space Program* are space and China, which shows that the content of the white paper of *China's Space Program* is closely related to the development of the country's aerospace industry, which coincides with the title of the text. Since space is a word with a relatively broad concept, we study its collocation words in order to obtain more detailed information. Taking space as the node word and three words on the left and right as the span, we use Antconc to count its collocation words in the English translation of *China's Space Program*. We use the mutual information value (MI) of the two as the collocation intensity standard, select high-frequency real words with a mutual trust value of 3 or more with space, and arrange them by frequency to explore the specific fields that China's aerospace industry is involved in (Table 2).

Table 1. *China's Space Program* in the BNC glossary

Order	Keyword	Keyness	Order	Keyword	Keyness
1	space	6916.46	11	beidou	670.16
2	china	3808.59	12	science	589.49
3	satellite	2913.93	13	international	586.28
4	satellites	2071.87	14	tt	562.42
5	cooperation	1494.95	15	earth	559.93
6	and	1229.15	16	development	553.64
7	navigation	812.85	17	spacecraft	538.73
8	launch	800.11	18	observation	527.64
9	exploration	796.2	19	communications	525.71
10	lunar	721.75	20	spaceflight	471.02

Table 2. *China's Space Program* in the BNC glossary

Collocation	MI	Frequency
science	4.06	99
technology	3.69	60
outer	4.14	60
debris	4.34	46
deep	4.59	41
environment	3.78	28

By studying the collocation words of space, we can have a clearer understanding of China's space industry: first, China's space industry is based on science and technology, second, China's space industry mainly involves outer space and deep space exploration, and finally China's space industry is also committed to solving space environmental problems. According to the above research methods, we can clearly understand the key points of the text by studying the BNC keyword vocabulary in the English translation of *China's Space Program*, and thus get a clear understanding of what kind of China's image is constructed through the text:

Technology Leader

China's development and achievements in the aerospace field demonstrate its strong strength in the high-tech field. From the "science" in the keyword list, it can be seen that the Chinese government has invested heavily in aerospace science research, and has carried out research in many fields such as space physics, astronomy, and microgravity science, and the research level is at the forefront of the world. From the "beidou, tt (TT & C: Tracking, Telemetry and Command)" in the keyword list, it can be seen that the construction of the Beidou satellite system marks China's achievements in the field

of satellite navigation that have attracted worldwide attention. The successful research and development of this system has won China the voice over in the field of global navigation satellite systems, and has shown the world China's outstanding strength in scientific and technological innovation and engineering technology.

International Partner

From the words "cooperation, international, communications" in the glossary, it can be seen that China actively advocates international cooperation in the field of space, and conducts joint projects with other countries and international organizations to jointly promote space exploration and scientific research, which is reflected in the joint development of space exploration missions, international satellite cooperation, and international space station cooperation. At the opening ceremony of the United Nations/China Global Partnership for Space Exploration and Innovation Symposium, the National Space Administration issued the "China Aerospace Initiative to Promote the Construction of a New Global Partnership for Space Exploration and Innovation"; the National Space Administration also took the lead in launching multilateral cooperation projects such as the International Lunar Research Station, the BRICS Remote Sensing Satellite Constellation, and the "Belt and Road Initiative" Space Information Corridor, and continued to implement the Asia-Pacific Space Cooperation Organization Multi-mission Small Satellite Project. These measures reflect China's image as an open and cooperative country, willing to share knowledge and resources with global partners.

Innovation Driver

China has demonstrated its innovative capabilities in the space field through a series of exploration missions, satellite technology applications, etc. From the "exploration" in the subject thesaurus, it can be known that China has successfully carried out a series of lunar exploration and deep space exploration missions, such as the Chang'e lunar exploration project and the Tianwen Mars exploration mission. The Chinese government will continue to promote deep space exploration of Mars and asteroids, and make contributions to human exploration of the universe. From the "development" in the subject thesaurus, it can be known that the Chinese government regards the space industry as an important part of high-tech industries and innovation-driven, and supports the research and development and innovation of space enterprises through national policies to promote the continuous development of space technology. The Chinese government has encouraged space enterprises to carry out technological innovation through policy support and investment, demonstrating the country's determination and ability in innovation.

Global Contributor

From the words "satellite, beidou, navigation" in the thesaurus, it can be seen that the Chinese government actively promotes the development of satellite technology, covering communications, navigation, remote sensing and other fields. For example, China's Beidou Navigation Satellite System is one of the largest satellite navigation systems in the world, providing high-precision positioning and navigation services for global users. The Chinese government promotes navigation applications and cooperation at home and abroad by promoting the Beidou system. China's Beidou Navigation Satellite System has achieved global coverage, providing positioning, navigation, timing and other services

for global users. The Chinese government will continue to improve the Beidou system and provide more accurate and global navigation services. China's Beidou Navigation System provides services for global users, and China's lunar exploration missions and deep space exploration have also contributed to the space exploration of all mankind. This has created an image of China actively participating in global affairs and contributing to the common well-being of mankind.

Space Exploration Pioneer

From the words "launch, lunar, spaceflight" in the title list, it can be known that China has demonstrated its enthusiasm and strength for space exploration through deep space exploration missions such as lunar exploration and Mars exploration. The Chinese government supports satellite launch and deep space exploration missions by developing reliable rocket launch technologies. China's Long March series rockets have extensive commercial launch services at home and abroad, which has promoted the development of the space industry. The Chinese government has formulated a long-term plan for lunar exploration, including the implementation of lunar roaming and sampling return missions. The Chang'e series probes have successfully achieved missions such as soft landing, patrol and sampling on the lunar surface, providing valuable data for future lunar scientific research. The Chinese government is committed to achieving the long-term goal of manned spaceflight in space flight, developing astronaut training, space station construction, etc. China's spaceflight also demonstrates the country's leading position in space exploration, establishing an image of a pioneer in space exploration.

To sum up, through its development and policies in the space sector, China has built a multi-dimensional image of a scientific and technological leader, an international partner, an innovation driver, a global contributor, a supporter of sustainable development, and a pioneer in space exploration. These images reflect China's achievements and influence in the space sector, while also reflecting the country's values and long-term development goals.

4 Discussion

Through the study of the English translations of the three recent white papers *China's Space Program*, we obtain the BNC glossary of the English translation of the white paper *China's Space Program* and the modern Chinese corpus glossary of the original white paper *China's Space Program*. By studying the English translation of the *China's Space Program* white paper on the BNC subject word vocabulary, we can clearly understand the key points of the English translation, and thus get a multi-dimensional China's image of a scientific and technological leader, international partner, innovation driver, global contributor, sustainable development supporter, space exploration pioneer, etc.; by comparing the original white paper *China's Space Program* to the subject word vocabulary of the modern Chinese corpus and the English translation of the *China's Space Program* white paper to the BNC subject word vocabulary. According to the above research results, we can summarize the significance of this research into the following three points:

From "Aerospace White Paper" to "National Image"

As the official document of the Chinese government on space development, the white paper *China's Space Program* has important national significance. First of all, the white paper introduces the purpose, principles, current situation, goals and international cooperation of the Chinese government in space development at home and abroad, providing transparent and authoritative information for the international community and demonstrating China's ambitions in the space field. By studying the white paper, it is possible to deeply understand the Chinese government's strategic planning in the space industry and provide a more comprehensive and in-depth understanding for the international community. Secondly, the white paper is not only an official document of the government, but also an important tool for shaping the national image. By applying the theory of imagery, it is possible to deeply explore the China's image constructed in the white paper, such as multi-dimensional images of scientific and technological leaders, international partners, and innovation drivers. This helps to understand the strategic positioning of the Chinese government in the space field and its self-presentation on the international stage. By studying these constructed images, the position and role of the Chinese government in the global space field can be better understood.

From "Aerospace White Paper" to "Aerospace Industry"

At the level of the aerospace industry, the study of the white paper *China's Space Program* has a profound impact. First of all, the white paper reveals China's scientific and technological strength and innovation ability in the aerospace field. Through the analysis of keywords and keywords in the white paper, we can understand China's advantages and characteristics in aerospace technology, R&D, innovation, etc., providing important reference for domestic and foreign enterprises and scientific research institutions. This is of positive significance for promoting international cooperation in aviation science and technology and promoting industrial innovation. Secondly, the international partnerships involved in the white paper also have an important impact on the development of the aviation industry. Through the introduction and emphasis of international cooperation in the white paper, we can understand the partnerships in the aerospace field that China is actively expanding globally. This will contribute to the international development of the aviation industry, promote resource sharing and cooperation on a global scale, and promote technological innovation and exchange of experience.

From "Translating the World" to "Translating China"

In recent years, there have been some obvious new trends in the development of the translation industry. First, the shift from the "translating world" in the past to "translating China". At the translation level, studying the white paper *China's Space Program* has unique value for language dissemination and international exchanges. First, by comparing the original Chinese and official English translations of the white papers in 2011, 2016, and 2021, we can deeply explore the cultural nuances in the translation process, the adjustment of language expression, and the differences in information transmission. This is helpful for understanding the Chinese government's translation strategy in international communication, and provides a reference for improving the international recognition of China's aerospace image. Secondly, studying the translation effect of the

white paper can help improve the international dissemination level of Chinese scientific and technological documents. By analyzing the possible problems and challenges in translation, we can provide lessons for the translation of similar documents in the future, and improve the international readability and acceptance of Chinese scientific and technological documents.

Overall, the study of the white paper *China's Space Program* involves not only policy formulation and image construction at the national level, but also scientific and technological development and international cooperation in the aerospace industry. At the same time, it also pays attention to language dissemination and cultural expression at the translation level. This multi-faceted research helps to comprehensively understand China's development strategy in the aerospace field and provides a clearer understanding for the international community. At the same time, for the aerospace industry and the translation field, the research results also have positive guiding significance, promoting the healthy development of related fields.

5 Conclusion

This research is helpful to supplement the corpus-based aerospace political texts and the relevant literature on the English translation of the white paper, so as to expand the research field of corpus linguistics and translation studies. Secondly, it can enhance the international community's understanding of China's policies and policies, and ultimately achieve the purpose of building a good national image of China, enhancing the international influence of Chinese discourse and promoting the self-shaping of the national image. The limitation of this study is that only one representative real word was selected when conducting translation effect research, so the conclusion is not comprehensive enough. The next step of research can start with the English translation of traditional cultural words, and explore the wonderful relationship between China's aerospace and Chinese traditional culture from the Chinese and English texts of the white paper *China's Space Program*, so as to provide more materials for telling Chinese stories and Chinese culture going global.

References

Wang, H.: Imagology and translation studies. J. Foreign Lang. **41**(04), 86–93 (2018)

Kaibao, H., Li, X.: Corpus-based translation and chinese image research: connotation and significance. Foreign Lang. Res. **34**(04), 70–75 (2017)

Leerssen, J.: Imagology: history and method. In: Beller, M., Leerssen, J. (eds.) Imagology-The Cultural Construction and Literary Representation of National Character, Rodopi, Amsterdam & New York (2007)

Xie, L., Wang, Y.: Research on political discourse translation from the perspective of China's International Image Construction. Foreign Lang. Educ. **39**(05), 7–11 (2018)

Chen, M., Wang, H.: Identity construction in the shaping of china's national image: with evidence from the white paper china and the world in the New Era. J. Zhejiang Inter. Stud. Univ. **01**, 13–22 (2020)

Scott, M., Tribble, C.: Textual Patterns: Keywords and Corpus Analysis in Language Education. John Benjamins, Amsterdam (2006)

Zhang, X.: The Image Perspective of Translation Studies - A Case Study of Kerouac's Chinese Translation of On the Road. Yilin Press, Shanghai (2011)


CSSCI 来源集刊
语言研究集刊
第三十五辑
复旦大学汉语言文字学科
《语言研究集刊》编委会 编
1905
上海辞书出版社

目　录

语言研究集刊 . 2025 (01) : 109-125+341-342 查看该刊数据库...

汉语范畴量词“种”和“类”的构式搭配及构式化对比

1 [redacted]

摘要： 文章基于多语料库和R语言平台，采用共时与历时相结合、定性与定量相结合的研究思路，以汉语近义范畴类量词“种”和“类”为研究对象，考察它们的名词搭配范畴及构式化轨迹。研究发现：1)“X种N”构式中名词搭配多属上位概念范畴，主要表达具有相似特征事物的范畴集合，以区别于其他范畴；“X类N”构式中名词搭配则多属下位概念范畴，主要表达具有相似特征事物的范畴集合，强调范畴内部的统一性。2)“种”和“类”构式化的建构机制主要涉及隐喻和转喻。转喻在其构式的形成阶段起主导作用，负责“具象关联”;隐喻则在其构式演化阶段起主导作用，负责“抽象类推”。这一发现修补了先前Goldberg有关构式承接关系的研究结论，反映出汉语量词构建的特色。3)“种”和“类”的量词用法大同小异，但各有侧重，搭配范畴的演化整体上呈现出“基本范畴→扩展的具体范畴→扩展的抽象范畴”及“紧邻范畴→远距范畴”的轨迹，并反映出汉语建构的精细化认知和体认性特征。

关键词： 范畴近义量词; 构式搭配; 构式化;

汉语范畴量词“种”和“类”的构式搭配及构式化对比*

[redacted]

提要 文章基于多语料库和R语言平台，采用共时与历时相结合、定性与定量相结合的研究思路，以汉语近义范畴类量词“种”和“类”为研究对象，考察它们的名词搭配范畴及构式化轨迹。研究发现：1）“X种N”构式中名词搭配多属上位概念范畴，主要表达具有相似特征事物的范畴集合，以区别于其他范畴；“X类N”构式中名词搭配则多属下位概念范畴，主要表达具有相似特征事物的范畴集合，强调范畴内部的统一性。2）“种”和“类”构式化的建构机制主要涉及隐喻和转喻。转喻在其构式的形成阶段起主导作用，负责“具象关联”；隐喻则在其构式演化阶段起主导作用，负责“抽象类推”。这一发现修补了先前Goldberg有关构式承接关系的研究结论，反映出汉语量词构建的特色。3）“种”和“类”的量词用法大同小异，但各有侧重，搭配范畴的演化整体上呈现出“基本范畴→扩展的具体范畴→扩展的抽象范畴”及“紧邻范畴→远距范畴”的轨迹，并反映出汉语建构的精细化认知和体认性特征。

关键词 范畴近义量词 构式搭配 构式化

* 本文受[redacted]目编号23ZHTD02007）资助，是河南省兴[redacted]字文化研究”（项目编号2023XWH037）、河南省教育科学规划项目（项目编号2024YB0152）、河南省高等教育教学改革研究与实践项目（项目编号2024SJGLX0406）及河南省研究生教育改革与质量提升工程项目（项目编号YJS2023JD50）的阶段性成果之一。《语言研究集刊》匿名审稿专家和编辑部为本文提出宝贵意见，谨深表谢意。不当之处概由笔者负责。

一、问题的提出

相较于印欧语系,量词是以汉语为代表的汉藏语系所独有的特色词类,是汉语使用者对事物量化或属性标记的重要手段。(吕叔湘 1979;郭锐 2021 等)对外部世界中不同范畴进行分类是人们认识世界和改造世界的基础,"种"和"类"是汉语中常用来表范畴的一组近义量词。《现代汉语词典》第 7 版将"种"定义为"表示种类,用于人和任何事物",将"类"定义为"许多相似或相同的事物的综合"。可见二者的语义十分相似,例如:

(1) 以下五种方法提高孩子记忆力,值得家长收藏。

(2) 请家长务必要防范这类拐卖方式。

例(1)中的"种"和例(2)的"类"分别与"方法"和"方式"搭配,表示一种类别集合,差异难辨,但根据语言的经济性原则(le principe d´économie),人们使用语言时倾向于以最低的消耗实现最高效的信息传递。因此,任何不同的语言形式都会在语义或功能等方面存在一定差异,不存在绝对的同义词。(Martinet 1975: 228;陈满华 2013)那么,两者有何区别? 对此,学界研究十分有限,郭先珍(2002: 181)曾提出"种"侧重事物内在相似性,语义范围较小,"类"侧重事物外在相似性,语义范围更大;缑瑞隆(2009)则认为虽然"种"的搭配范围比"类"小,但搭配数量却比"类"多。可见,学界现有少数研究主要关注这两类量词的搭配范围,考察视角较单一。此外,汉语量词常用于 XCN 或 NXC 结构。① 从形式上看,该结构属于 Goldberg(2006: 8)分类中典型的半固定构式(partially filled construction),即在内容性和图式性上都呈现半固定形式。但据中国知网和 Springer 检索显示,国内外从构式角度对汉语量词开展的研究不足 20 项,从构式角度考察汉语量词的共时性和历时性的研究则迄无开展。

有鉴于此,本文将基于语料库,辅以现代量化分析工具,从构式视角对范畴近义量词"种"和"类"在现代汉语中名词搭配分布和古代汉语中的构式化进程展开描述性和解释性考察,尝试填补以上研究缺憾。

① X 代表非定指词性(动词、形容词、数词及代词等),C(Classifier)代表量词,N(Noun)代表名词。

二、现代汉语中“种”和“类”的名词范畴分布对比

Goldberg(2006：69)认为,构式中的论元只有与构式在语义上具有一致性才能进入该构式,且语义一致性越高,与该构式的搭配强度就越高。因此,对量词与搭配名词的搭配强度考察就成了揭示量词名词范畴分布和构式义的重点。本节将基于北京语言大学 BCC 现代汉语语料库(以下简称“BCC 语料库”)和 R 语言统计平台,对此展开考察。

2.1 研究设计

2.1.1 研究问题

本节主要考察以下两个问题：

1)“X 种 N”和“X 类 N”量词构式的名词搭配范畴有何异同?

2)“X 种 N”和“X 类 N”量词构式的原型构式义有何异同?

2.1.2 语料来源和研究对象

本研究语料主要选自 BCC 语料库,原因如下：1) 该语料库为平衡和通用语料库,能较全面、真实地反映研究对象的实际用法;2) 该语料库提供词性标注,可通过限定词性获取相关量词构式的总现频数和与某名词的具体搭配频数,便于开展后期研究。

本研究以“X 种 N”和“X 类 N”构式的名词搭配为研究对象,对 X 分别以数词、动词、不定代词及指示代词进行限定,以“[a v m r]种 n”和“[a v m r]类 n”①为具体检索项,辅以人工筛选,建立封闭语料库。

2.1.3 研究方法

构式搭配分析法(Collostructional Analysis, CA)是一种基于编程语言 R 的构式量化研究方法,它以语料库中的语言实例为基础,通过考察构式与构式成分间的搭配强度,既可从微观层面纵向考察构式与词项间的搭配强度和范畴化关系,又可从宏观上横向对比不同近义构式间的异同。(Gries 2017：505;张立英 2021)该研究方法主要包括以下三类：1) 共现词位分析法(Collexeme analysis);2) 多项差异共现词位分析法(Multiple distinctive

① 在 BCC 语料库检索体例中,a 代表形容词,v 代表动词,m 代表数词,r 代表代词,n 代表名词。

collexeme analysis);3) 互为变化的共现词位分析法(Covarying collexeme analysis)。出于研究目的,本节将采用共现词位分析法,从语料库中的语言事实出发,考察"X 种 N"和"X 类 N"构式中 N 词项与该构式的搭配强度,以此来确定它们各自的名词搭配倾向及范畴分布。

2.1.4 数据收集及构式搭配强度统计

本节基于构式搭配分析法,首先以"[a v m r]种 n"和"[a v m r]类 n"为检索项,在 BCC 语料库中进行检索,通过人工筛选剔除无效项(如"买种子""这类活动量"等),并将重复搭配项合并(如"一类方法""这类方法"等),得到统计对象如下:"X 种 N"构式语例有效总数量为 16 565 个,"X 类 N"构式语例有效总数量为 4 841 个。可见,从频数上看,"X 种 N"构式语例要显著多于"X 类 N"构式语例。考虑到数据统计的可操作性和有效性,我们选取出现频度最高的前 20 个名词搭配为深入研究对象(含"X 种 N"构式语例 12 053 个,"X 类 N"构式语例 2 928 个),[①]并在 R 语言软件平台运行 Coll. analysis 3.5 统计分析程序,对以上数据展开构式搭配强度分析。具体步骤如下:1) 首先获取"X 种 N"和"X 类 N"构式中出现频率前 20 词项的交叉数据(以名词"问题"与构式"X 种 N"的搭配强度为例,详见表 1);2) 接着将所得数据导入 R 语言程序进行费舍尔精确检验(Fisher Exact Test);3) 最后选择共现词位分析(Collexeme analysis)测算出名词项与构式的搭配强度值(coll. strength)。具体名词搭配和搭配强度排名详见表 2[②]:

表 1 名词"问题"与构式"X 种 N"搭配强度测算的交叉数据统计

	问题	¬ 问题	总计
"X 种 N"构式	1 482	15 083	16 565
¬ "X 种 N"构式	1 145 461	28 521 401	29 666 862
总计	1 146 943	28 536 484	29 683 427

① 原因在于出现频数排名在前 20 以外的名词搭配数量均低于 10,不具有显著统计价值。

② 根据统计学上的界定,当搭配强度值(coll. strength)大于 1.301 03 时,反映该词位与构式呈现显著搭配关联,搭配强度值越大,搭配关联越显著;搭配强度值为 inf 时,表示搭配无限高度关联。(Gries 2017)

表 2　X 种/类 N 构式中搭配强度值位列前 20 的名词

X 种 N				X 类 N			
words	col. sth	words	col. sth	words	col. sth	words	col. sth
事情	Inf	语言	183. 22	作品	Inf	观念	126. 32
问题	Inf	态度	171. 85	事情	Inf	游戏	115. 74
人	Inf	情绪	162. 74	问题	Inf	消息	108. 53
方式	Inf	习惯	147. 53	故事	Inf	题材	96. 63
思想	Inf	精神	136. 85	情况	Inf	国家	94. 86
方法	Inf	情况	115. 32	文章	209. 54	字眼	87. 89
类型	Inf	耻辱	91. 83	事物	184. 74	琐事	78. 12
游戏	Inf	感觉	85. 22	运动	179. 51	错误	66. 74
手段	Inf	关系	81. 41	动物	163. 39	案件	53. 85
乐趣	189. 32	行为	69. 52	话题	148. 77	现象	45. 93

2.1.5　统计结果分析

表 2 显示,“种”和“类”在“X 种 N”和“X 类 N”构式中的高关联名词搭配存在交叉,但交叉项并不多,这说明它们各自原型搭配的语义特征和范畴分布均存在显著差异,例如:

(3) a. 每天清晨早饭之后,小轩都会帮嘉欣梳辫子,这是她的一种乐趣。

b. 在这类游戏中,读者成了发起人,能充分利用自己的原有知识。

(4) a. 这个剧本深刻地揭示出生活的荒诞和徒劳,被译成三十多种语言。

b. 澳大利亚有很多休眠火山群,但迄今未遇喷发,因此国民全然漠视这类话题。

可以看出,“X 种 N”与“X 类 N”在例(3)、例(4)中搭配的名词虽不相同,但又存在一定关联:“一种乐趣”与“一类游戏”相比,前者中的搭配对象是概念上更宽泛的上位范畴,后者中的搭配对象则是前者范畴中的一个次范畴,即“游戏”是“乐趣”的一种(实现形式);“三十多种语言”与“这类话

题”也呈现类似关系,即“语言”是“话题”的上位范畴,“话题”虽然不直接隶属于某种“语言”范畴,但在概念层次上仍然属于“语言”的下位层次。此外,例(3)、例(4)中的“种”和“类”一般不能双向互换,前者可替换后者,但后者一般不能替换前者,即可以说“这种游戏”“这种话题”,但通常不能说“*一类乐趣”“*三十多类语言”。为了进一步考察以上现象是否呈现一定倾向,我们对表中二者的高关联名词搭配进行分析,将它们的概念层次领属关系呈现如图1。①

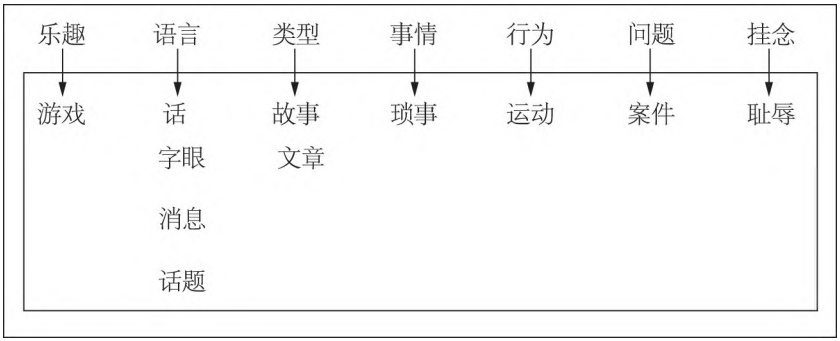


图1 “X种N”和“X类N”构式中高关联名词的领属关系分布

由图1可知,“X种N”构式中反映上位概念的高关联名词明显多于“X类N”,倾向性显著,即“种”搭配概念宽泛对象的倾向要高于“类”。同时,这也解释了为什么有些使用量词“类”的情况可用“种”替代,而有些使用量词“种”的情况却不能用“类”替代。因为从构式的语义一致性原则来看,若“种”搭配对象的概念更宽泛,那么“X种N”的概念就更宽泛,限制性条件就更少,适用性则更强,能进入该构式的搭配也更多。前文语料统计时“X种N”语例显著多于“X类N”语例的结果也印证了这一特征。

2.2 研究结论

范畴具有统一相似对象和区别差异对象的双重作用。范畴内部各成员

① 该统计结果不代表绝对性的语言事实或规则,而是二者在用法相似性基础上的用法倾向差异;大方框内名词为“X种N”构式的高关联名词,小方格内名词为“X类N”构式的高关联名词;实线箭头代表“X种N”中高关联名词之于“X类N”中相应高关联名词的领属关系,虚线箭头则代表“X类N”中高关联名词之于“X种N”中相应高关联名词的领属关系。

在认知中的地位并不相同,而是具有一定层级性的。范畴量词“种”和“类”的用法也反映出这一特征。两者用法虽十分相似,多可互用,但仍呈现出一定的用法倾向差异。这些差异主要源于二者对事物范畴化时的立足点不同:“种”既强调范畴内成员的同一性,又侧重与其他范畴的差异性;“类”则主要强调范畴事物的同一性,并不侧重与其他范畴的差异性。此外,“种”更侧重与语义宽泛、表示上位概念的对象搭配,范畴容量更大,而“类”侧重与语义具体、表示下位概念的对象搭配,范畴容量更小。因此,总体上看,“种”的适用范围更广,总使用频率也更高。

最后,我们可将现代汉语中“X 种 N”和“X 类 N”的原型构式义界定如下:“X 种 N”多与表上位概念的范畴搭配,主要表达具有相似特征事物的范畴集合,以区别于其他范畴;“X 类 N”多与表下位概念的范畴搭配,主要表达具有相似特征事物的范畴集合,强调范畴内部的统一性。我们将在后文对以上共时研究结论的源流及理据展开进一步考察。

三、汉语范畴类近义量词构式化的古代汉语历时考察

在考察了现代汉语中“X 种 N”和“X 类 N”构式的语义范畴分布后,我们已对两者在“是什么”“怎么样”的问题有了较清晰的把握,但在“怎么来”“为什么”的解释维度上仍不清楚。量词在早期汉语中并不存在,而是基于语用需求由其他实词(主要为名词、动词及形容词)演化而来,发挥着量化、分类及标记等功能。(Aikhenvald 2000:353)基于特定的认知机制,词源词的原词义或词性逐渐改变,量化、分类及标记等特征逐渐凸显,继而形成较为固定的使用结构(构式形成),并选择性地搭配相应的名词范畴(构式演化),这一完整过程便是量词的构式化(Traugott & Trousdale 2013;施春宏 2022)。

本节将基于双语料库(北京大学中国语言学研究中心 CCL 语料库和国家语委古代汉语语料库),[①]从两者的字源出发,考察它们在古代汉语中构式化的整体轨迹,探究其背后的认知动因及机制。构式化主要涉及两个方面

① 鉴于古代汉语历时研究的周期跨度较大,语料的全面性对提升研究结论的可靠性至关重要,故本研究采取双语料库检索的方法对相关古代汉语语料进行搜集和分析。

变化:功能性调整(语义和语用)和结构性调整(词法和句法)。(Traugott & Trousdale 2013;文旭,杨旭 2016)本研究主要关注量词的功能性调整,对于量词的结构性调整不做重点考察。

3.1 范畴量词构式化的历时考察:"种"

从字源上看,有学者(如谷衍奎 2008)认为"种"通"穜",本义为动词"种植",如《周礼》"以物地相其宜而为之穜"等。亦有学者(如李学勤 2012)认为其通"種",本义指"早种晚熟的禾类",后引申为名词"种子",如《周礼》"生種稑之種,而献之于王"等,近代后期简化作"种"。本研究更倾向于后者,一来据文献显示,"种"做动词和名词的出现时间十分相近,且使用频率相当;二来从逻辑上看,实体和行为相较,前者应是后者产生的施事或受事基础。"种"的量词用法早见于汉代,本节将从以下两个阶段考察其量词的构式化进程。

3.1.1 量词"种"的构式形成

总体上,"种"主要经历了"名词(一种谷物)→名词(谷物种子)→名词(种族)→量词"的演化轨迹:

(5) 其谷宜五種,谓黍、稷、菽、麦、稻也。(周《周礼》)

(6) 诞降嘉種,维秬维秠,维糜维芑。(春秋《诗经》)

(7) 今吴稻蟹不遗種,子将助天为虐,不忌其不祥乎!(战国《国语》)

(8) 王侯将相,宁有种乎?(西汉《史记》)

(9) 家人尚不欲绝種祀,况于国之神宝旧畤。(东汉《汉书》)

(10) 索隐果菜千种。千种者,言其多也。(东汉《汉书》)

在上古早期,[①]"种"主要用作名词,起初泛指"谷物"(共 19 例),如"谷宜五種"等,后可转称"谷物的种子",如"诞降嘉種"等。"种子"义的产生是通过部分(种子)—整体(谷物)的转喻思维实现的,并带有明显的[+源头]概念特征。后来,"种"又可泛指"生物的种子"(共 15 例),如"稻蟹不遗種"等。到了上古后期,"种"表"种族"(共 16 例)的用法开始出现,如"宁有种乎"等,对[+源头]概念的隐喻映射是实现这一语义转变的关键。"种"的量词用法早见于东汉,表示"相同范畴事物或概念的集合",并多搭配生物对象

① 本研究采纳学界较为普遍接受的古代汉语分期方法:上古汉语—先秦至两汉时期、中古汉语—魏晋至隋唐五代时期、近古汉语—宋元至五四运动前。(王力 1989:35)

(共58例),如“果菜千种”等。通过转喻将[+源头]概念从实体中抽离并凸显是量词“种”构式形成的认知机制。

3.1.2 量词“种”的构式演化

在上古汉语中,量词“种”的构式搭配便开始了一系列演化,总体上呈现出“生物类搭配→非生物搭配→抽象类搭配”的演化轨迹:

(11) 纵广正等二十五由旬,其叶千色,有百种画,如天璎珞。(东汉《论衡》)

(12) 云何行,得八种声入万亿音?(东汉《佛说般舟》)

(13) 故灾变万种兴起,不可胜纪,此所由来者积久复久。(东汉《太平经》)

(14) 夷狄自伏法万种,其类不同,俱得老寿。(东汉《太平经》)

如前所述,量词“种”诞生之初主要用于生物类对象,随后扩展至非生物类的具体事物(共45例),如“百种画”等,并逐渐成为这一时期的主流用法。“种”也偶与抽象事物搭配(共17例),但多见于宗教典籍,如《佛说般舟》《太平经》等,在世俗作品中鲜有出现。这或许是因为宗教是一种探求人与世界关系的哲学,义理深邃,只有将抽象难解的宗教教义相对具体化,才能为世俗人群所理解。例如,佛教的逻辑基础因明学由宗、因、喻、合、结等“五支作法”构成,其中隐喻是重要的论证手段之一。(刘文英2012:421-425)在形式上,该时期“种”的量词构式主要以“数+量+名”结构为主,且数词通常数额庞大,如“千”“万”等;从认知机制上看,量词“种”从生物类事物到非生物类具体事物,再到抽象事物的搭配扩展,均是通过隐喻机制对[+源头]概念的投射扩展实现的。其名词范畴及数量分布如表3所示:

表3 上古汉语中量词“种”的名词搭配范畴

范畴类别	高频名词搭配	事物集合特征
生物类(58)	果蔬、鸟等	[+源头][+生物范畴集合]
非生物类(45)	画、石等	[+源头][+非生物范畴集合]
抽象类(17)	音、灾等	[+源头][+抽象事物范畴集合]

鉴于量词“种”的搭配在中古和近古时期变化不大,故在此将两段时期

合并讨论。中古汉语的用例如:

(15) 凡服药千种,三牲之养,而不知房中之术。(魏晋南北朝《抱朴子》)

(16) 年年长是阻佳期,万种恩情只自知。(《全唐诗》)

(17) 如人得到宝山中,百种珠珍遍寻觅。(五代《敦煌变文》)

近古汉语的用例如:

(18) 百种凄凉,几般烦恼,没个人怜。(宋《柳梢青》)

(19) 你撇下半天风韵,我拾得万种思量。(元《西厢记》)

(20) 一别家山音信杳,百种相思,肠断何时了!(明《水浒传》)

(21) 及到筵散花谢,虽有万种悲伤,也就没奈何了。(清《红楼梦》)

量词"种"的用法在中古时期较为多样,除延续上古时期的用法外,其与抽象事物的搭配不再仅限于宗教典籍,开始在世俗文本中大量出现,基本可用于任何"同范畴事物或概念的集合"。这说明量词"种"的用法在这一时期已趋近成熟。到了近古时期,其用法较中古时期未出现明显变化。在形式上,依旧以大数值的"数+量+名"结构为主,这从一个侧面说明,量词"种"除了表示事物的相同范畴,同时还侧重对不同范畴的区分。其名词范畴及数量分布如表4所示:

表4 中古汉语和近古汉语中量词"种"的名词搭配范畴

范畴类别	高频名词搭配	事物集合特征
具体类(153)	货物、草等	[+源头][+具体事物范畴集合]
抽象类(84)	思、音等	[+源头][+抽象事物范畴集合]

综上,"种"的量词构式早见于上古后期,由名词"種"(谷物)演化而来。在形式上,它在上古、中古、近古三个时期都以"数+量+名"的构式结构为主,并且数词通常较大,如"百""千""万"等,这反映出它在语义上倾向于凸显不同范畴间的差异;在搭配上,它主要继承了名词(种子)中的[+源头]概念,并且语义一直较为稳定,多表示"同范畴事物或概念的集合",基本呈现出"生物类事物→非生物类具体事物→抽象事物"的演化轨迹。相较"类"(见后文),其量词用法出现更早,因此也较早与上位概念范畴搭配并逐渐固化;在认知机制上,转喻和隐喻都在其量词的构式形成阶段发挥了关键作用,而其构式演化则主要是借助隐喻机制对[+源头]概念的投射实现的。其

构式化的认知路径如图2所示①：

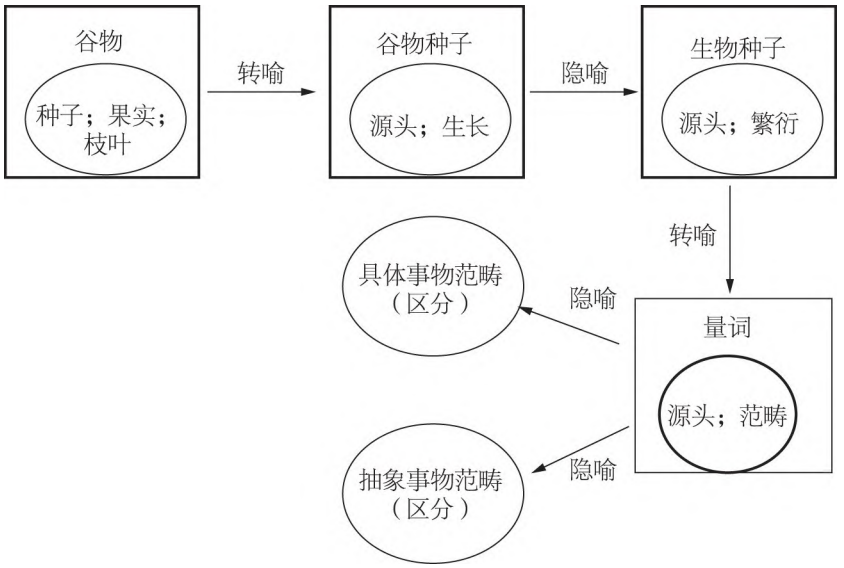


图2　量词“种”构式化的认知路径

3.2　范畴量词构式化的历时考察：“类”

从字源上看，“类”通“類”，早见于上古时期，据《说文》载“种類相似，唯犬为甚。从犬，頪声”，本义为“相似事物的集合”，如《荀子》“禽兽群焉，物各从其類也”。金代明刻本《改并四声篇海》中开始出现简体“类”字。量词“类”大致出现在中古汉语时期，本节将从以下两个阶段考察其量词的构式化进程。

3.2.1　量词“类”的构式形成

从字源来看，早期“类”主要包含[+范畴][+相似]这两大概念特征。其在形成量词前主要经历了“名词→动词→量词”的演化轨迹：

名词用例如：

(22) 本乎天者亲上，本乎地者亲下，则各从其类也。(周《易经·乾》)

(23) 子曰：“有教无类。”(春秋战国《论语》)

(24) 身不正，言不信，则义不壹，行无类也。(西汉《礼记》)

① 方框表示实体；箭头表示行为；圆形内为语义属性；椭圆内为名词搭配范畴；粗体线代表凸显成分；细体线代表非凸显成分。

动词用例如:

(25) 五色清明,其状类怒。(春秋战国《相马经》)

(26) 陷为天下轻薄子,所谓画虎不成反类狗也。(东汉《后汉书》)

(27) 有神兽,似马,其声类牛,导引历年乃出。(南北朝《北史》)

量词用例如:

(28) 因此便有一类道士,儒生,闲僧泛参禅理者。(唐《禅源诠序》)

(29) 如是世界十二类生,不能自全,依四食住。(唐《首楞严经》)

在上古早期,名词"类"主要表示"相似事物的集合"(共 12 例),即"范畴",如"有教无类"等,后来又衍生出"法则,规则"(共 9 例)的含义,如"甚僻违而无类"等,这一转变是通过隐喻投射"相似事物的集合"中的[+相似]概念实现的。此外,中国古代名辩学的发展应也是这一转变的动因之一。探究事物间的类属联系是名辩学关注的基本问题。墨家最早系统论述了"类"的逻辑意义,并将"类"作为逻辑推论的基本法则之一。(方克立 1994)春秋战国时期,"类"的动词用法开始出现,主要表"相似,像"(共 25 例),如"其状类怒"等。通过转喻(整体—部分)用范畴整体去转指范畴内个体间的关系特征是实现这一转变的主要认知机制。

"类"直到中古晚期才出现量词用法,表"同范畴事物或概念的集合"(共 31 例)。在形式上,它常以"一+量+名"的构式结构出现,如唐代的《禅源诠序》中有"一类道士"的记载,这从形式上反映出"类"强调范畴统一性的特征。在搭配上,量词"类"产生初期搭配对象多是人(共 26 例)或与人相关的事物(共 5 例),反映出明显的体认性特征。中国自古也有"近取诸身,远取诸物"的观点,即人们总是由自身出发,再到外部事物及抽象概念去建构认知。(王寅 2007:10)这一认知特点不仅存在于我们的思维层面,还外化于语言层面。

本文认为,量词"类"的演化并非经动词"类"而来,而是从名词"类"直接演化而来。因为一方面名词"类"并未消亡,而是与动词并行使用;另一方面,从认知经济性上来看,名词与量词无论在语法还是语义上认知都更接近。借助转喻将名词范畴中的[+相似][+范畴]特征从实体中抽离并凸显是量词"类"产生的主要认知手段。

3.2.2 量词"类"的构式演化

在中古汉语中,量词"类"的构式搭配不断演化,总体上呈现出以下搭配

范畴及演化轨迹：

（30）佛道如斯一类人，生生大不易见如来面。（唐《敦煌变文》）

（31）今推寻分析，色有地水火风四类。（五代《禅源诠序》）

（32）又曰：“身心是道，四生六类皆有身心，悉是道不？若有见闻，请对圣说！”（五代《祖堂集》）

量词“类”在唐代仍多用于人（共69例），到了五代时期，搭配对象开始变得更加多样，大体呈现从人到具体事物（共32例），再到抽象事物（共19例）的轨迹，几乎无所不包。在认知机制上，通过隐喻将对象“人”中[+相似][+范畴]的概念向其他对象投射，使搭配范畴不断扩展泛化。在形式上，“名+数+量”或“量+数”构式最为多见。其名词范畴及数量分布如表5所示：

表5　中古汉语中量词“类”的名词搭配范畴

范畴类别	高频名词搭配	事物集合特征
人（69）	人、道士等	[+相似][+人集合]
具体事物（32）	阳乌、食等	[+相似][+具体事物范畴集合]
抽象事物（19）	色、变文等	[+相似][+抽象事物范畴集合]

在近古汉语中，量词“类”的搭配范畴无明显演化，但在具体用法上较中古汉语时期仍产生了一定差异：

（33）其国，有一类人如马裸露，王运神力分身为蚕。（宋《五灯会元》）

（34）有一类人家儿子，不行孝养，不会礼仪。（元《西厢记》）

（35）三藏听言，跳起来叫道：“汝等皆是一类邪物，这般诱我！”（明《西游记》）

（36）俩人已经各有十万好答赚了。诸如此类事，不胜枚举。（清《官场现形记》）

到了近古时期，量词“类”的搭配范畴较上一阶段无明显增列，语义上较为稳定，仍表“同范畴事物或概念的集合”。但在量词构式的形式上，转以“数+量+名”结构居多，与现代汉语十分贴近。此外，在数词搭配上，以“一”最为多见，其他数词较少出现，这反映出其更侧重事物相似性的特征，与另一范畴量词“种”的功能存在互补性。其名词范畴及数量分布如表6所示：

表 6　近古汉语中量词“类”的名词搭配范畴

范畴类别	高频名词搭配	事物集合特征
人(165)	人、人家等	[+相似][+人集合]
具体事物(86)	邪物、药等	[+相似][+具体事物范畴集合]
抽象事物(37)	事、灾祸等	[+相似][+抽象事物范畴集合]

综上,“类”的量词构式由名词“类”演化而来,最早出现于中古晚期。在形式上,“数+量+名”“名+数+量”及“量+数”的构式结构均有出现,但总体上以“数+量+名”结构居多,其中“一+量+名”结构使用倾向较为显著,这也反映出其凸显范畴内部统一性的特征;在语义搭配上,它主要继承了名词“类”中的[+范畴][+相似]概念,且语义一致较为稳定。在形成之初,其搭配尤以“人”居多,这主要与人类认知的体认性有关。随后,搭配对象大体沿着由具体对象到抽象对象的轨迹扩展,到了近古时期,已基本可用于任何表示“同范畴事物或概念的集合”;在认知机制上,总体来看,转喻在其构式形成阶段发挥了主要作用,隐喻则在其构式演化阶段发挥了主要作用,其构式化认知路径如图 3 所示:

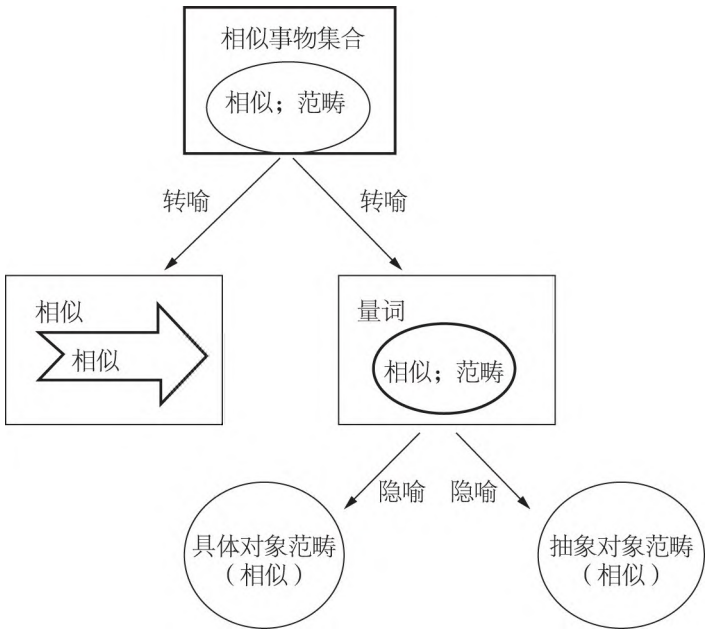


图 3　量词“类”构式化的认知路径

四、发现与启示

通过对范畴量词“种”和“类”的构式搭配和构式化路径考察,本文有如下发现和启示:

1)汉语近义范畴量词的构式搭配及构式化的古今异同。总体上看,量词“种”的用例比量词“类”的用例更多。在现代汉语中,“X 种 N”和“X 类 N”构式的名词搭配范畴分布十分相近,但前者多与表上位概念的范畴搭配,主要表达具有相似特征事物的范畴集合,以区别于其他范畴;后者则多与表下位概念的范畴搭配,虽亦多表达具有相似特征事物的范畴集合,但强调范畴内部的统一性。在古代汉语演变中,量词“种”和“类”均由名词演化而来,表示同范畴事物或概念的集合。但前者早见于上古汉语,整个构式化进程主要围绕[+源头]概念展开,多与上位概念范畴搭配,搭配数词数额一般较大,如“百”“千”“万”等,侧重不同范畴的多样性;后者直到中古汉语晚期才出现,整个构式化进程主要围绕[+相似]概念展开,但多用于“一+量+名”结构,侧重范畴内部的统一性。可见,“种”和“类”的量词用法特征在古代汉语中已逐步形成,并一直沿用至现代汉语。

2)汉语近义范畴量词构式化的认知机制。量词“种”和“类”构式化背后的认知机制主要涉及隐喻和转喻。其中,在两者的构式形成阶段,转喻的主导作用更为明显,为基于“相邻”关系的不同名词、动词或量词间的词性演化提供认知助推,即语法转喻。而在两者的构式演化阶段,隐喻则统一起着绝对主导作用,负责“相似”但不同类搭配对象间的抽象类推。这一发现补充了先前 Goldberg(2006)提出的构式承接关系(inheritance link)限于隐喻延伸连接(metaphorical extension link)、多义连接(polysemy link)、例示承接(instance link)及局部成分连接(subpart link)的研究结论,深化了相关理论,反映出汉语量词构建的特色。

3)汉语近义范畴量词建构的精细化认知和体认性特征。量词“种”和“类”均表示同范畴事物或概念的集合,但又在范畴层级、统一性及区别性等特征上存在不同搭配倾向,大同小异,又各有侧重,这既反映出普遍性的语言建构经济性原则,又反映出汉语言民族在语言建构中的精细化认知特征。体认性是认知语言学秉承的重要原则,即一方面人类的基本认知是通过身

体和外部世界的互动获取的,体验决定了主体认识世界的方式;另一方面人们常从身边的临近事物及其相关的基本概念出发,再到外部其他事物及抽象概念去不断丰富认知。(牛保义 2021)这一认知特点不仅存在于汉语民族的思维层面,也反映在汉语量词建构的语言层面。总体上,范畴量词名词搭配的扩展基本呈现出"基本范畴→扩展的具体范畴→扩展的抽象范畴"及"紧邻范畴→远距范畴"的轨迹,反映出较显著的体认性特征。

五、结　　语

语料库数据分析与语言学理论阐释相结合是认知语言学研究中理论创新和方法论完善的赓续之路。本文基于语料库,将现代统计分析方法和人工内省法结合,从语言事实出发比较了范畴近义量词"种"和"类"在现代汉语中名词项的范畴化分布异同,并对两者在古代汉语中的构式化轨迹进行历时考察,发现了它们建构背后的认知机制及认知精细化、体认性等特征,为进一步挖掘汉语量词特色,建构汉语特色话语体系带来了有益启示。

参考文献

陈满华(2013)惠特尼和叶斯柏森的语言经济思想——兼谈语言经济原则的产生及其发展,《中国人民大学学报》第 4 期:113-121。

方克立(1994)二十世纪中国哲学的宏观审视,《中国社会科学院研究生院学报》第 4 期:1-6。

缑瑞隆(2009)示形量词"股、束、缕、绺"的用法及认知基础——对外汉语量词教学个案研究,《云南师范大学学报》第 2 期:18-22。

谷衍奎(2008)《汉字源流字典》,北京:语文出版社。

郭　锐(2021)量词的功能扩张,《中国语文》第 6 期:691-712,767。

郭先珍(2002)《现代汉语量词用法词典》,北京:语文出版社。

李学勤(2012)《字源》,天津:天津古籍出版社。

刘文英(2012)《中国哲学史(上)》,天津:南开大学出版社。

吕叔湘(1979)《汉语语法分析问题》,北京:商务印书馆。

牛保义(2021)试析体认语言学的核心原则,《天津外国语大学报》第 1 期:26-35,158。

许　慎(1963)《说文解字》,北京:中华书局。

王　力(1989)《汉语语法史》,北京:商务印书馆。

王　寅(2007)《认知语言学》,上海:上海外语教育出版社。

文　旭,杨　旭(2016)构式化:历时构式语法研究的新路径,《现代外语》第 6 期:731-741,872。

张立英(2021)现在完成体中动词的语义特征——基于 BNC 语料库子库的构式搭配分析,《中国外语》第 6 期:36-44。

Givon T. (1990) *Syntax: A Functional - Typological Introduction*. Amsterdam/Philadephpia: John Benjamins Publishing Company.

Goldberg A. (2006) *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Gries S. (2017) Corpus Data in Usage-based Linguistic. //Brdar M, Gries S, Fuchs M. (eds.) *Cognitive Linguistics: Convergence and Expansion*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Martinet S. (1975) *Studies in Functional Syntax*. Cambridge: Cambridge University Press: 228.

Traugott E, Trousdale G. (2013) *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

cognitive characteristics of intricate categorization reflected in their constructions from both synchronic and diachronic perspectives. Using qualitative and quantitative methods, the study yields the following findings: 1) The construction *X zhong N* is primarily associated with higher-level conceptual categories, typically denoting an entire category of members that share similar characteristics, thereby distinguishing it from other categories. In contrast, *X lei N* tends to align with lower-level conceptual categories, emphasizing the internal unity of the category. 2) Metaphor and metonymy serve as the primary cognitive mechanisms in the formation of these constructions. However, while metonymy plays a dominant role in their initial formation, metaphor takes precedence in their evolution. This distinction deepens Goldberg's previous research on constructional inheritance and highlights key characteristics of Chinese classifier constructions. 3) The usage patterns of *X zhong N* and *X lei N* exhibit broad similarities but diverge in finer details, each with its own emphasis. The overall evolution of their collocational categories follows a trajectory from "basic category → extended concrete category → extended abstract category" and "immediate category → distant category", reflecting the cognitive processes and embodiment in the construction of Chinese category classifiers.

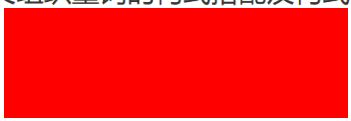
Key words: category classifier; collocation; constructionnalization

Abstract: This paper proposes a new classification system for Chinese progressive complex sentences, categorizing them into forward progressive sentences, backward progressive sentences, bidirectional progressive sentences, and inward progressive sentences. These sentence types exhibit distinct characteristics across multiple dimensions. At the cognitive level, they also display clear differences, each corresponding to a unique image schema that represents different cognitive structures and modes of information processing. The image schemas for bidirectional, forward, inward, and backward progressive sentences are represented as "A ⇙ B", "A→C", "A ⊕ D", and "E←A" respectively. The order of their emergence follows the pattern: bidirectional pro-

汉语组织近义量词的构式化及名词搭配范畴化对比

语言学研究，2024 (01)：133-149 查看该刊数据库收录来源

汉语近义组织量词的构式搭配及构式化路径分析



摘要： 本文基于语料库和R语言平台，首先从现代汉语共时角度考察组织量词构式“X家/户N”的原型构式搭配及构式义异同，继而从古代汉语历时角度考察该异同形成的源流路径及认知动因，扩展了汉语量词的研究视角，深化了对汉语量词特征的认识。研究发现：（1）“X家N”构式主要表达“行业组织集合”，偶表达“家庭和居所组织集合”；“X户N”构式主要表达“家庭组织的管理集合”，偶表达“行业和居所的组织管理集合”。（2）隐喻和转喻是二者构式化的主要认知机制，但转喻在它们的构式形成阶段起主导作用，隐喻则在它们的构式演化阶段起主导作用。（3）汉语近义组织量词的构式化除涉及Goldberg提出的隐喻连接、多义连接、例示承接及局部成分连接等常见构式连接外，还包括转喻连接，并反映出显著的精细化认知特征和社会认知属性。

关键词： 组织量词；构式搭配；“家”；“户”；构式化；

基金资助： 教育部人文社科基金项目(22YJC740075)；河南省哲学社会科学规划项目(2023XWH037)的阶段性研究成果；

专辑： 哲学与人文科学

专题： 中国语言文字

汉语近义组织量词的构式搭配及构式化路径分析

[redacted]

[提　要] 本文基于语料库和R语言平台，首先从现代汉语共时角度考察组织量词构式“X家/户N”的原型构式搭配及构式义异同，继而从古代汉语历时角度考察该异同形成的源流路径及认知动因，扩展了汉语量词的研究视角，深化了对汉语量词特征的认识。研究发现：（1）“X家N”构式主要表达“行业组织集合”，偶表达“家庭和居所组织集合”；“X户N”构式主要表达“家庭组织的管理集合”，偶表达“行业和居所的组织管理集合”。（2）隐喻和转喻是二者构式化的主要认知机制，但转喻在它们的构式形成阶段起主导作用，隐喻则在它们的构式演化阶段起主导作用。（3）汉语近义组织量词的构式化除涉及Goldberg提出的隐喻连接、多义连接、例示承接及局部成分连接等常见构式连接外，还包括转喻连接，并反映出显著的精细化认知特征和社会认知属性。

[关键词] 组织量词；构式搭配；“家”；“户”；构式化

1. 引言

量词是汉语使用者对事物量化或属性标记的重要手段，也是语言使用者思维方式的重要体现（吕叔湘，1978；郭锐，2021等）。其中，“家”和“户”是一组常用来搭配家庭或行业组织的近义量词，使用广泛。然而，两者虽有诸多相

* 作者简介：[redacted]

似，但也存在一定差异。对此，陈伟琳（2002）在考察汉语可修饰“多胞胎”的量词时指出“家”和“户”必须与“家庭”搭配才具有量化功能，计量功能并不显著；宗守云（2010）则指出“家”和“户”不能用于“成X”结构，只能修饰人，不是典型的集合量词；刘子平（2013）认为“家”可用于家庭，还有企事业单位，“户”用于门户、住户及房子。郑邵琳（2016）在考察魏晋南北朝石刻中的名量词时，指出“家”和“户”是不定量集体量词，侧重组织义。

以上学者虽业已关注两者异同，但十分有限，且多为年代久远的附带性论述，最新研究较少，缺乏专门性的对比研究。此外，“家”和“户”常用于“XCN”或“NXC”构式中。其中，现代汉语常用于“XCN”构式，古代汉语早期多用于“NCX”构式，后期则是两者兼用。X代表槽位词性的不确定性（动词、形容词、数词及代词等），C（Classifier）代表量词，N（Noun）代表名词。从结构上看，该构式属于Goldberg（2006）分类中的半固定构式（partially filled construction），即在内容性和图式性上都呈现半固定形式。但所查文献显示，相关构式视角研究尚未开展。有鉴于此，本文基于语料库和R语言平台，从构式视角分别从共时和历时两个维度对“家”和“户”的构式搭配及构式化路径展开考察。

2. 现代汉语构式搭配的共时考察

2.1 构式搭配强度分析

量词“家”和“户”在现代汉语中十分常见。《现代汉语词典》将量词“家”描述为“用于住户或企业等”；将量词“户”描述为“用以计户数”。可见，二者在用法上存在一定交叉[①]：

（1）a. 他随后从自己口袋掏出100元钱，委托这家农民的亲人转交。
b. 经过三年的快速发展，公司业务已辐射到鄂、豫、陕三省六县5万多户农民。

在上述例句中，“家”和“户”都与“农民”搭配在“XCN”构式中表示“家庭组织集合”，语义十分相似，但a例中“家”可用“户”替换，但b例中“户”却不宜用“家”替换。显然，二者用法存在一定差异。本节将基于语料库和R语言平台，尝试通过对“家”和“户”构式中高关联名词的构式搭配强度测算，并结合语境分析，考察二者名词搭配的原型语义聚类和范畴分布异同。

鉴于量词性词缀构式“N家”和“N户”在汉语中使用亦十分广泛，能辅助

① 本文中现代汉语引证例句均取自BCC语料库。

考察量词构式“X家N”和“X户N”中高关联名词的范畴分布，本节将在BCC语料库中对它们均进行检索，并辅以人工筛选，剔除无效项（如“我家厨房”“濑户”等）后，得到预选统计对象如下：“X家N”搭配构式语例有效总数量为6825个，“X户N”搭配构式语例有效总数量为3292个；“N家”搭配构式语例有效总数量为2472个，“N户”搭配构式语例有效总数量为953个。随后，我们进一步合并重复搭配项（如“一户人家”“这户人家”等），选取出现频度最高的前30类名词作为进一步考察对象，其中含“X家N”构式词例4083个，“X户N”构式词例1326个，“N家”构式词例928个，“N户”构式词例505个。最后，在R语言软件平台运行Coll.analysis 3.5统计分析程序（Gries，2011），对以上数据做构式搭配强度分析，并将每组构式中搭配强度值位列前20的名词列表如下[①]：

表1　X家/户N构式中搭配强度值位列前20的名词

X家N				X户N			
搭配词	搭配强度	搭配词	搭配强度	搭配词	搭配强度	搭配词	搭配强度
企业	Inf	书店	173.32	人家	Inf	公寓	136.55
公司	Inf	杂货店	166.85	民宅	Inf	邻居	117.74
老小	Inf	药店	152.44	农舍	Inf	房客	99.42
酒店	Inf	宾馆	142.37	村民	Inf	农奴	91.96
人	Inf	工厂	136.42	平民	Inf	农家	84.74
商场	Inf	旅馆	118.63	家庭	184.49	苗寨	79.24
饭店	205.71	客栈	93.82	贫民	181.37	企业	68.79
银行	189.64	餐厅	81.55	居民	176.46	地主	56.24
医院	186.22	商店	73.41	佃农	158.31	人	49.85
酒吧	184.94	农民	65.21	社员	145.74	租客	41.73

① 根据 R语言平台的统计标准界定，当搭配强度值大于1.301时，此时反映该词位与构式呈现显著搭配关联，搭配强度值越大，搭配关联度越显著；搭配强度值为inf,即无限高度关联；当搭配强度值小于1.301时，表明两者搭配关联度较低，不具有统计意义。

表2 N家/户构式中搭配强度值位列前20的名词

N家				N户			
搭配词	搭配强度	搭配词	搭配强度	搭配词	搭配强度	搭配词	搭配强度
专	Inf	赢	183.53	商	Inf	租	148.63
画	Inf	军事	186.27	账	Inf	个体	137.85
企业	Inf	散文	172.63	农	Inf	养牛	127.94
作	Inf	买	166.59	住	Inf	养羊	111.86
翻译	Inf	仇	158.74	养殖	Inf	集体	94.71
科学	Inf	商	142.98	客	Inf	马	79.24
音乐	Inf	卖	133.43	猎	Inf	电	78.83
书法	Inf	医学	99.53	税	186.95	网	66.96
艺术	206.83	玩	83.74	粮	178.52	养猪	59.25
儒	189.76	收藏	76.92	奶牛	158.29	气	51.52

2.2 “X家N”和“X户N”的名词搭配原型范畴

如上表所示，“X家N”/“X户N”和“N家”/“N户”近义构式组中高关联名词搭配的语义聚类虽都有一定交叉，但各自聚类倾向显著，这说明它们各自的名词搭配原型范畴存在显著差异：

（2）a. 大鹏要养活一家人，生存对他来说没有舒适，忍受才是他的唯一感受。
b. 得知农民子弟学校老师们没钱购买棉衣，她联系两家企业捐助了羽绒服。

以上两例中的“X家N”构式所搭配名词都表示“组织集合”，但又明显不同。其中，“一家人”指所有家庭成员集合，与“家”的基本义十分接近；“一家企业”既脱离了家庭组织概念，也不再单纯指人的集合，而是指营利性的社会组织集合，具有类似特征的搭配对象在表1中还有广泛分布，值得进一步考察。

（3）a. 郑州市公安局家属院居住着近200户居民。
b. 县里决定改造破旧危房200户，使1138户农舍环境得到改善。

与例（2）不同，以上两例中的“X户N”构式中名词并不都表示组织集合概念。其中，“200户居民”表示家庭的组织集合；而“1138户农舍”搭配对象不再是家庭组织，而是表单数称谓的处所名词。但它实际是“1138户（人家的）农

舍”的缩合结构，依然带有鲜明的家庭组织特征。可见，搭配家庭组织在该构式中具有一定的聚类倾向，值得进一步关注。

综上，我们发现“X家N”和“X户N”构式中高关联名词在人称、居所及行业组织等对象上有潜在搭配倾向。为进一步验证，我们基于BCC语料库，对以上特征及其分布差异进行深入考察，并将统计结果呈现如下：

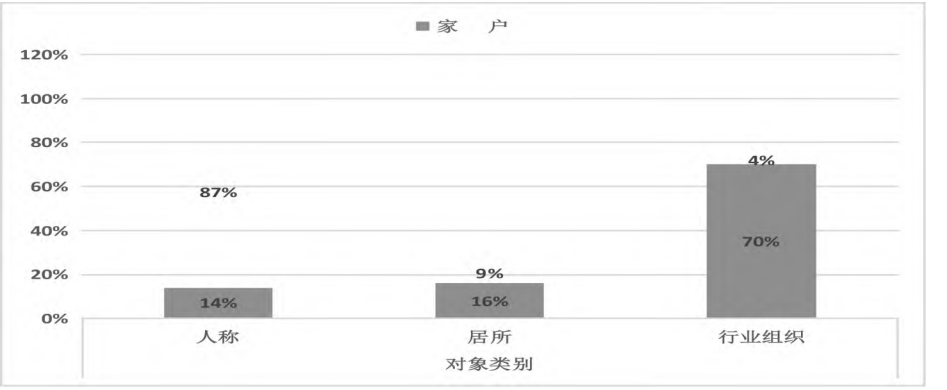


图1 “X家N”和“X户N”构式中名词语义聚类及范畴分布

如图1所示，这两类构式中高关联名词在人称、居所及行业组织等类别上存在显著聚类倾向和差异。具体来看，“X家N”构式中高关联名词搭配主要以行业组织为主（70%），以人称（14%）和居所（16%）搭配为辅；而“X户N”构式中高关联名词搭配则以人称为主（87%），极少与行业组织（4%）和居所（6%）搭配。更多用例如下所示：

（4）a.一家人/全家老小/一家农民等（人称）
b.一家客栈/一家农舍/一家宾馆等（居所）
c.这家公司/一家餐厅/一家商场等（行业组织）
（5）a.十户家庭/一户平民/两户农民等（人称）
b.一户农舍/一户住宅/千户苗寨等（居所）
c.一百户企业/一户农场/一户商铺等（行业组织）

至此，我们已对“X家N”和“X户N”构式中高关联名词的语义特征进行了考察，但这似乎还不够充分，因为量词构式中的名词除了本身所固有的语义特征外，有时还有来自量词及其他构式要素赋予的语义特征。通过上文对“X家N”和“X户N”构式在BCC语料库中的检索，我们发现数词(m)修饰语在这两类构式的X槽位中占据最高份额（其中“X家N”中有约3071例，占45%，在“X户N”中有约2535例，占77%）。显然，作为以量化功能为主的量词构式，这不足为奇。然而，我们进一步考察后发现，二者中的数词类型却差异显著，具体语料查

询截图及比重统计结果如下所示：

一户人家	154	几户人家	50
两户人家	22	三户人家	20
一户农家	17	四户人家	11
十几户人家	9	几十户人家	8
好几户人家	7	一户姓	6
十来户人家	6	五六户人家	6
二十户人家	5	十户人家	5
很多户人家	5	一户农民	5
一户民宅	5	一户好人家	4
七户农民	4	百十户人家	4

一家旅馆	214	一家咖啡	164
一家饭店	116	一家商店	113
一家公司	112	一家银行	111
一家酒店	95	一家医院	90
一家餐馆	70	一家报纸	66
一家饭馆	64	一家酒吧	63
一家老小	63	一家酒馆	58
万家楼	55	一家工厂	55
一家大小	54	一家店铺	51
一家铺子	45	一家书店	45

图2 “X家N”和“X户N”量化特征语料统计截图

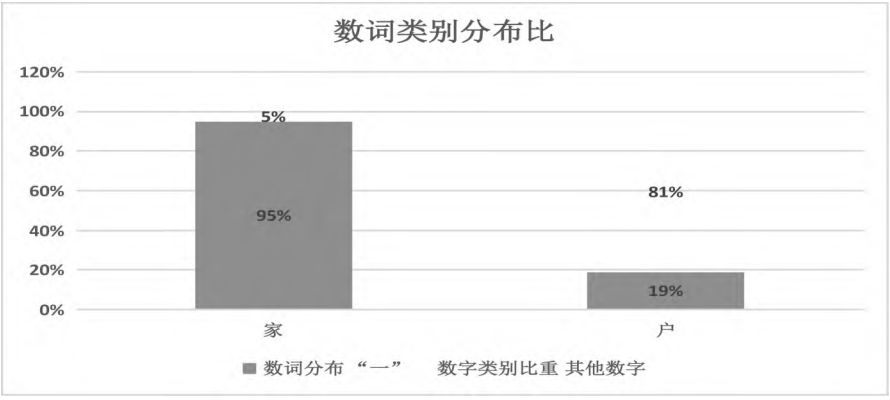


图3 “X家N”和“X户N”构式中数词搭配聚类分布

如图3所示，“X家N”构式由数词限定时，“一”占据绝对比重（95%），约有2917例，其他数字使用极少（5%），仅有153例，且多为“两”或“多”；“X户N”构式由数词限定时，“一”之外的其他数字占据主导（81%），约有2053例，且多数为大额数字，数字“一”使用较少（18%），仅有574例。这一差异值得进一步思考。据《玉篇·一部》载:“一者，数之始也”，但“一”的语义功能不仅仅是计数。根据《现代汉语词典》，“一”做定语时，与其他数词不同，其义项至少有以下六类：

①“计数一”，如“花一元钱”
②“全部”，如“奋斗一生”
③“同一”，如“万众一心”
④“另一”，如“西红柿，一名番茄”
⑤“每一”，如“一人发一件”

⑥“某一”，如“一事物对他事物的关系”

可见，计数统计只是“一”的语义功能之一。本研究中“X家N”涉及的“一家人”“一家农民”等语例也极少用于计数统计。综上，“X户N”和“X家N”构式虽然都多与数词搭配，但前者更侧重借助大额数字进行统计组织，后者的统计组织功能则较弱。

2.3 “N家”和“N户”的名词搭配原型范畴

我们再来看“N家”和“N户”构式中高关联名词的范畴分布差异。根据表2统计显示，“N户”构式的完整结构可解构并分为以下三类:

① 家庭管理集合：

猎户—(打猎的人)的(家庭)；佃户 —(租种土地的人)的(家庭)

租户—(租赁房屋的人)的(家庭)；农户 —(务农的人)的(家庭)

② 群体管理集合：

商户—(经商人)的(群体)；个体户 — (从事个体经营者)的(群体)

客户— (身份为顾客的人)的(群体)；储户—(储蓄人)的(群体)

③ 账目管理集合：

账户— (管理账目)的(户头)；社保户— (管理社保)的(户头)

税户— (管理税务)的(户头)；存户—(管理存款)的(户头)

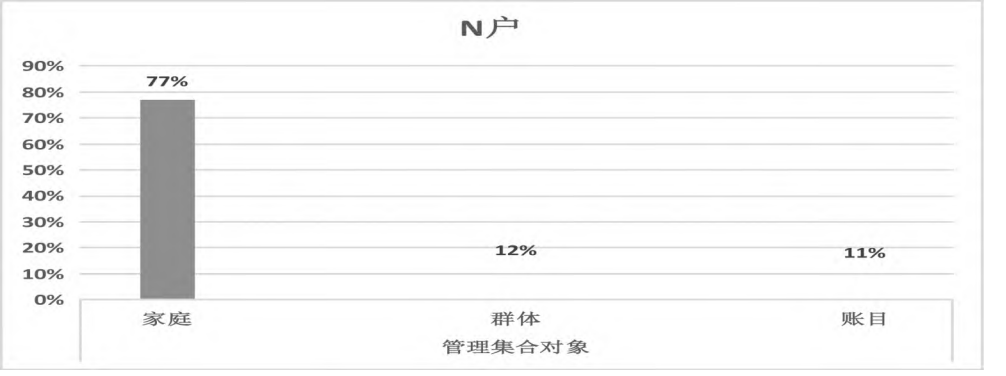


图4 “N户”构式中名词搭配范畴分布

“N家”构式主要表示行业群体集合，这一结果也再次佐证了前文有关“X家N”构式的研究结论。此外，其行业群体集合对象又可分为技能性群体集合和行为性群体集合，且前者在语料库中的出现频率要显著高于后者：

① 行业技能性群体集合：如“画家”“艺术家”“书法家”“专家”等

② 行业行为性群体集合：如“买家”“买家”“商家”“仇家”“赢家”等

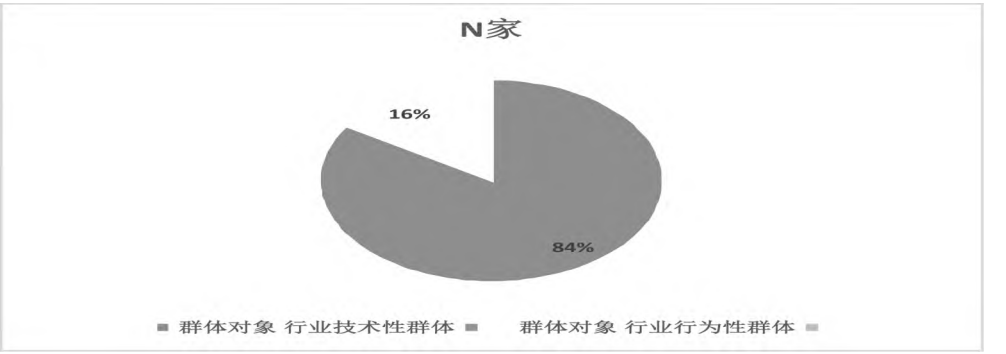


图5 “N家”构式中名词搭配聚类分布

2.4 小结

“家”与“户”均与家庭群体概念紧密相关，如“家喻户晓”“千家万户”等表达，都深刻地反映出这一特征。但二者用作量词时，仍存在明显差异。其中，“X户N”构式主要表达家庭组织集合，其高关联名词搭配主要是家庭群体，偶有行业组织，并且带有显著的管理统计特征，这与汉语中常说的“户籍”“户口”等表达在功能上保持了高度一致。而“X家N”构式中的高关联名词搭配则多是企事业单位类的非家庭组织，主要表达行业组织集合，家庭组织集合功能并不显著。基于以上分析，我们可将两类构式的原型义初步归纳如下：“X家N”构式主要表达“行业组织集合，偶表达家庭和居所组织集合”；“X户N”构式主要表达“家庭组织的管理集合，偶表达行业和居所的组织管理集合”。

3. 古代汉语构式化路径的历时考察

在考察了“X户N”和“X家N”构式在现代汉语中的名词搭配分布后，我们已在它们在“是什么”的描述维度有了整体把握，但在“怎么来”“为什么”的解释维度上，仍不清楚。量词在早期汉语中并不存在，而是基于语用需求由其他实词演化而来，发挥量化、分类及标记等功能（Aikhenvald，2000)。基于特定的认知机制，字源词的原词义或词性逐渐改变，量化特征逐渐凸显，继而形成较为固定的使用结构（构式形成），并随时间推移选择性地搭配不同名词范畴（构式演化），这一过程便是量词的构式化（蔡淑美、施春宏，2022）。

本节将基于北京大学中国语言学研究中心语料库和国家语委古代汉语语料库[①]，从两者的字源出发，考察它们在古代汉语中构式化的整体轨迹，发现其背后的认知动因及机制。构式化主要涉及两个方面的变化：功能性调整（语义和语

① 鉴于古代汉语历时研究的周期跨度较大，语料的全面性对提升研究结论的可靠性至关重要，故本研究基于两个语料库对相关古代汉语语料进行搜集和分析。

用）和结构性调整（词法和句法）（Traugott & Trousdale，2013; 文旭、杨旭，2016)。本研究主要关注量词的功能性调整。

3.1 “家”

从字源上看，据《说文》载：“家，居也”。“家”在甲骨文中写法形似“屋里关着猪”，本义为“养猪的地方”。因养猪的地方必有人，故引申为“居所”。其量词用法始见于春秋时期，表示“家庭组织”，如《庄子·徐无鬼》中“百姓悦之……至邓之虚而十有万家。”以下将从以下两个阶段考察量词“家”的构式化进程。

3.1.1 量词“家”的构式形成

“家”的量词用法主要经由以下路径演化而来[①]：

（6）上六，丰其屋，蔀其家，窥其户，阒其无人。（《周易》·周）
（7）禹平水土，主名山川；稷降播种，家殖嘉谷。（《尚书》·周）
（8）女子生，而愿为之有家。（《孟子》·春秋战国）
（9）舜有羹行，百姓悦之，故三徙成都，至邓之虚而十有万家。（《庄子》·春秋战国）
（10）济南瞗氏宗人三百余家，豪猾，二千石莫能制。（《汉书》·汉）

上古汉语早期[②]，“家”由象形义“养猪的地方”引申为“人的居所”，如“蔀其家”。据《尔雅》载:“牖户之闲谓之扆,其内谓之家”，“门”之内的地方称“家”，即“居所”（共18例）。这是因为“猪”是六畜中最早被人类驯化的动物，且在“贝”充当货币之前，“猪（肉）”是更早被使用的货币，故易与“居所”和“财产”产生关联。随后，它在“居所”基础上又可转作“家庭组织”（共22例），如“女子生，而愿为之有家。”意即希望女子成人后嫁夫成家。可见，这里的“家”是指以夫妻关系为核心的“家庭组织”，这一转换在认知上是基于“居所”和“家庭组织”间的邻近性，借助转喻思维实现的。

进入上古汉语中期，“家”的量词用法开始出现，常表示“家庭组织”的单位（共19例），前置数词通常较大，多用于统计，如“至邓之虚而十有万家”。“家”从名词到量词的演化是“家庭组织”实体中[+群体][+家人][+组织]的概念特征通过转喻被凸显脱离而实现的。在形式上，早期“家”的量词构式多为省略所修饰名词的“数+量”结构，尽管带有一定的集合名词性质，但是考虑到其显著的量化功能及“家”与“家庭组织”关联的高度规约性，这一结构应已属量词构式。

① 本文中古代汉语引证例句均取自北京大学中国语言学研究中心语料库和国家语委古代汉语语料库。
② 本研究采纳学界较为普遍接受的古代汉语分期方法：上古汉语——先秦至两汉时期、中古汉语——魏晋至隋唐五代时期及近古汉语——宋元至五四运动前（王力，1989）。

3.1.2 量词“家”的构式演化

（一）中古汉语

（11）姜维，天水之匹夫耳，获之则于魏何损，拔西县千家。（《全刘宋文》·魏晋南北朝）

（12）吾少好百家之言，身为四代之史。自开辟已来，未有爵位蝉联。（《全梁文》·魏晋南北朝）

（13）学以聚之，擅百家而诞誉；问斯辩矣，陶六艺以声。（《唐代墓志汇编》·唐）

（14）儿还教念百家诗，算应未及甘罗贵。（《敦煌变文》·五代）

量词“家”在中古汉语中除了延续上古时期的基本用法外，还可修饰“流派”，表达“流派组织”单位（共33例），但直接搭配对象通常隐现，如“百家诗”等。从认知动因上看，这一搭配变化是减损“家庭组织”中的[+家人]概念，通过隐喻机制投射[+群体]和[+组织]概念实现的，但这种用法在随后并未形成典型的“数+量+名”构式，并不属于典型的量词用法。其名词搭配的范畴分布情况如下表所示：

表3 中古汉语中量词“家”的名词搭配情况

名词范畴	高频搭配	概念特征
家庭组织（19）	千家、百家等	“+家人”“+群体”“+组织”
流派组织（33）	百家诗、百家言等	“+群体”“+组织”

（二）近古汉语

（15）叔隗轻盈饶态度，小乔妩媚足精神。风流总属一家人。（《浣溪沙·赠陈惜惜》·宋）

（16）张妈妈听得，走出来道：“早是你才来得三日的媳妇，若做了二三年媳妇，我一家大小俱不要开口了。”（《宋代话本》·宋）

（17）叫齐了一家骨肉，尽来看这家书。（《二刻拍案惊奇》·明）

（18）登州山下有一家猎户，兄弟两个，哥哥唤做解珍，兄弟唤做解宝。（《水浒传》·明）

（19）周家家大业大，村中数十家佃户，俱是他家的。（《明珠缘》·明）

（20）老残到了次日，想起一千两银子放在寓中，总不放心，即到院前大街上找了一家汇票庄，叫个日升昌字号。（《老残游记》·清）

（21）向着知州道：“本院只取此一疋，不知是那家店中的?贵州可将开店之人拘来一问。”（《五美缘》·清）

量词“家”在近古汉语中用法变化较大。在近古早期，其仍表“家庭组织”（共25例）的集合单位，但已开始多用于典型的“数+量+名”结构中，其中名词多为“人”“大小”“老小”等家庭组织成员；进入近古中期，其搭配出现了不少标记职业的家庭组织（共38例），如“猎户”“佃户”等，这种新用法是通过隐喻机制投射原“家庭组织”中[+群体][+家人][+组织]概念的同时，又进一步凸显“+职/行业”这一概念特征实现的。这一搭配用法虽然并未跳出“家庭组织”的基本范畴，但为量词“家”随后大量修饰“行业组织/机构”类对象（共52例）奠定了重要认知基础，如“一家汇票庄”“那家店”等。显然，这些对象已经完全脱离了“家庭组织”的范畴，从认知机制上看，它们是基于标记职业的“家庭组织”搭配，减损其[+群体]和[+家人]概念，仅隐喻投射[+组织]和[+职/行业]概念实现的。

此外，语料显示近古汉语中“数+量+名”结构中的数量词较先前已明显变小，量化统计功能有所弱化。这似乎与另一搭配“家庭组织”的量词“户”的功能互补存在关联。语言的经济性原则是指人们在使用语言时会以最低的消耗实现最高效的交际目的（Martinet，1975）。因此，任何不同的语言形式通常都会在语义或功能等方面存在一定差异。近义量词“家”与“户”在近古汉语后期已出现了显著分工差异，“家”一方面在修饰“家庭组织”时量化统计功能逐渐弱化，另一方面已跳出了早期与“家庭组织”搭配的功能。其名词搭配的范畴分布情况如下表所示：

表4　近古汉语中量词“家”的名词搭配情况

名词范畴	高频搭配	概念特征
无标记家庭组织（25）	人、老小等	“+家人”“+群体”“+组织”
标记职业家庭组织（38）	猎户、佃户等	“+家人”“+群体”“+组织”“+职业”
行业组织（52）	店、票庄等	“+组织”“+职业”

综上，“家”的量词构式在上古汉语已出现，早期以无名词标记的“数+量”构式形式为主，且数词较大，多用来量化统计“家庭组织”，中古汉语中还出现了修饰“流派组织”的用法，但用法未固化。近古汉语中“家”的量词用法变化最大，在形式、搭配及语义功能方面都有所演化。其中，最为显著的变化便是与“行业组织/机构”搭配的大量出现，并取代早期用法逐步成为主流用法。此外，近古汉语中的“家”即便仍与“家庭组织”搭配，量化统计功能也已不再显著，与“户”产生了一定分化。从认知机制上看，在构式形成阶段，量词“家”由“养猪的地方”义经过多次转喻运作演化而来。在构式演化阶段，其发展主要是基于隐喻机制实现的。我们可将其构式化认知路径图示如下[①]：

① 方框表示实体；粗体空心箭头表示行为；圆形内为语义属性；椭圆内为名词搭配范畴；粗体线代表凸显成分；细体线代表非凸显成分。

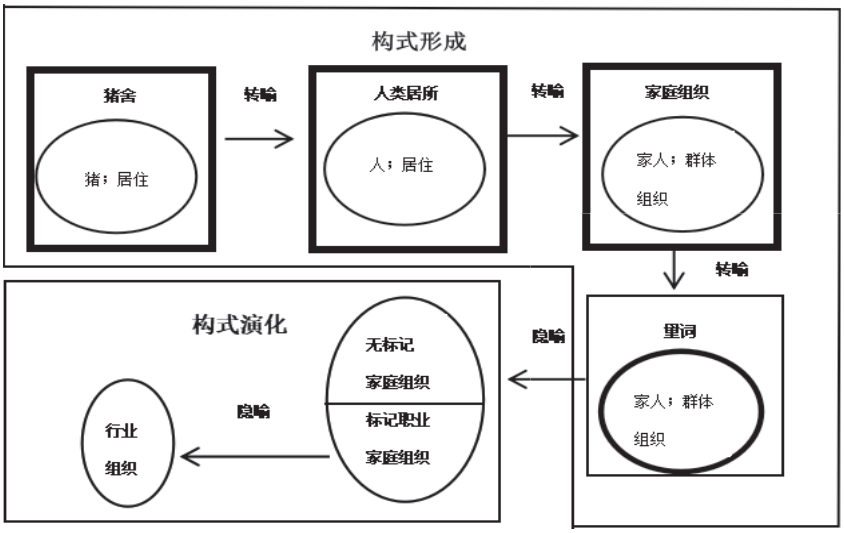


图6 量词“家”构式化的认知化认知路径

3.2 “户”

“户”在上古汉语早期就已出现，因甲骨文中象似一块带转轴的门板，故《说文》载：“户，护也。半门曰户，象形”，其本义为“单扇门”。另据《六书精蕴》载:“凡室之口曰户，堂之口曰门。内曰户，外曰门”。尽管两者就“户”的本义存在一定争议，但释义都与现代汉语中“门”的概念十分贴近。量词“户”出现较早，如《周易·讼》载：“其邑人三百户，无眚”。本部分将从以下两个阶段考察量词“户”的构式化进程。

3.2.1 量词“户”的构式形成

（22）窥其户，阒其无人。（《周易》·周）
（23）不克讼，归而逋，其邑人三百户无眚。（《周易》·周）

“户”的量词由“门”的名词义演化而来。因“门”作为一户“人家”的入口，在有关“人家”的相关概念中处于较为凸显的地位，在认知上能较自然地通过转喻代指“人家”。再借助转喻将“人家”实体中的[+家庭组织][+群体]概念特征脱离则促成了量词“户”的形成，用以表达“人家”群体单位。早期用例显示，其量词构式多呈现“数+量+（名）”结构，且数字较大，多用于统计，如“邑人三百户”等。可见，量词“户”的统计组织功能在早期就已十分显著。

3.2.2 量词“户”的构式演化

（一）上古汉语

（24）数十万户之邑，尊将军为诸侯，与天下更始。（《庄子》·春秋战国）

（25）齐宣王为大室，大益百亩，堂上三百户。（《吕氏春秋》·春秋战国）
（26）于是王大寤，出虞姬，显之于朝市，封即墨大夫以万户，烹阿大夫与周破胡。（《列女传》·西汉）
（27）徙天下豪富于咸阳十二万户。（《史记》·西汉）
（28）天子悯之，更以为大将军，益封二千户，赐钱百万。（《汉书》·东汉）

在上古汉语中，“户”做量词主要用于省略名词的“数+量+（名）”结构中，表示“人家”的单位集合（共23例）。根据布拉格学派提出的标记理论（Markedness Theory），无标记的语言，形式更简单，交流更方便，使用频率高于有标记的语言，可进一步促进其在语言系统中的固化（entrenchment）（Levinson，1983；沈家煊，1998）。据该时期文献语料显示，量词“户”除了修饰“人家”，且多省略外，并无其他搭配。这从侧面说明“户”搭配“人家”是相对规约化的用法。此外，量词“户”前的数词多为“百”“千”“万”等大额数字。可见，用于统计管理“家庭组织”仍是该时期其主要功能。

（二）中古汉语

（29）权追思之，募三州有能举知术数如吴范、赵达者，封千户侯，卒无所得。（《三国志》·魏晋南北朝）
（30）不过栎阳万户邑，夫富贵不归故乡，如衣绣夜行。（《后汉书》·唐）
（31）显祖于是率众与金共讨之于吐赖，获二万余户而还。（《北齐书》·唐）
（32）景及逆者，封万户开国公，绢布五万疋。（《旧唐书》·五代）

在中古汉语中，量词“户”延续了上一阶段的用法，仍然修饰“家庭组织”（共38例），多表统计。但不同是，在先前“数+量+(名)”结构中“户”后开始出现了如“邑”“公”“侯”等名词。但这些“数+量+(名)”结构的功能并不在于量化名词对象，而是反映其财产或官位特征，如“千户侯”，即“封地拥有千户人家的王侯”，“户”的量化对象仍是隐现的“家庭组织/人家”。这也进一步说明“户”与“家庭组织”的规约性搭配继续固化，已可用作修饰性成分。上古汉语和中古汉语中“户”的名词搭配情况如下表所示：

表5　上古和中古汉语中量词“户”的名词搭配情况

名词范畴	构式形式	高频搭配	概念特征
家庭组织（38）	“数＋量＋（名）”	千户、万户等	“+家庭组织”“+群体”“+统计”

（三）近古汉语

（33）名以为贼，州购约、允各千户侯。（《后汉书》·宋）
（34）国公三千户，实封三百户。（《金史》·元）
（35）千门万户受灾危，三市六街遭患难。（《水浒传》·明）
（36）那安老爷家的日子虽比不得在先老辈手里的宽裕，也还有祖遗的几处房庄，几户人家。（《儿女英雄传》·清）
（37）使者道："便是处子之家。那中屋另是一户人家。"（《东度记》·清）
（38）将过南北这条街，坐北向南，有一户人家，老赵就一喊叫。（《小五义》·清）

从搭配上看，这一时期"户"的名词搭配仍多为"人家"，鲜有其他搭配。从结构上看，早期近古汉语中"户"与中古汉语中并无显著变化，到近古后期出现了较为典型的"数+量+名"量词构式（共11例），但并不多见，名词隐现的"数+量+（名）"结构（约45例）仍是主流。其名词搭配情况如下表所示：

表6 近古汉语中量词"户"的名词搭配情况

名词范畴	构式形式	高频搭配	集合概念特征
家庭组织（45）	"数＋量＋（名）"	千户、万户等	"+家庭组织""+群体""+统计"
家庭组织（11）	"数＋量＋名"	人家等	"+家庭组织""+群体""+统计"

综上，"户"的量词构式在上古汉语就已形成，用来量化"家庭组织"，多呈无名词标记的"数+量"构式，且量词数额较大。一直延续到近古后期才偶现"数+量+名"构式，并以"人家"搭配居多。以上特征一方面反映出其与"家庭组织"搭配的高度规约性，具有一定集合名词特征，量词特征在古代汉语中并不显著；另一方面也反映出其对"家庭组织"显著的统计管理功能，与其在现代汉语中功能基本一致。从认知机制上看，在构式形成阶段，"户"的量词用法由其表"门"的名词通过双重转喻机制（部分<门/"+家庭组织""+群体">-整体<家庭组织>）演化而来；在构式演化阶段，它并无明显语义和搭配上的演化，多为同搭配范畴内部成员基于隐喻的类推扩展。其量词构式化的认知路径可图示如下[①]：

① 用作量词前，语料库中并未见"户"表"人家"的名词用法，为忠于语料事实，此处用虚线标记其在认知中可能潜在的名词转换过渡。

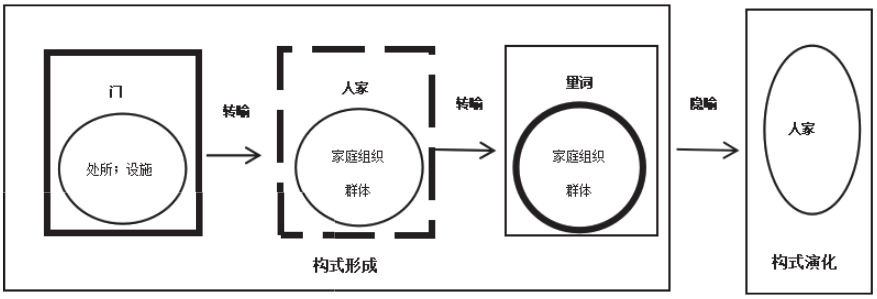


图7 量词“户”构式化的认知路径

4. 发现与启示

通过对组织量词“家”和“户”的构式搭配和构式化路径考察，本文有如下发现和启示：

（1）汉语近义组织量词的构式搭配及构式化路径异同。在现代汉语中，“X家/户N”构式的名词范畴分布同异共存：两者都可用作修饰家庭组织的单位，但不同的是，“X户N”构式多用于家庭组织，偶用行业组织和居所；“X家N”构式则多用于行业组织，偶用于家庭组织和居所。在古代汉语中，“家”和“户”的量词用法出现均较早。前者的构式形成过程比后者更复杂，但演化路径相对清晰；后者的构式形成过程相对简单，但语料层面反映出的演化路径并不清晰。两者在古代汉语前期都主要用于“家庭组织”。而到了近古汉语，“户”未产生明显变化，“家”则分化出用于“行业组织”的用法，并逐步成为主流用法延续至今。

（2）汉语近义组织量词构式化的认知机制。整体上看，转喻和隐喻是二者构式化的主要认知机制，前者在它们的构式形成阶段起主导作用，为基于“相邻”关系的不同名词演化及概念凸显提供认知助推；后者则在它们的构式演化阶段起主导作用，负责“相似”但不同类别搭配对象间的抽象类推。这一发现补充了先前Goldberg（2006）提出的构式承接关系（inheritance link）限于隐喻延伸连接（metaphorical extension link）、多义连接（polysemy link）、例示承接（instance link）及局部成分连接（subpart link）的研究结论，深化了相关理论，反映出汉语量词构建的特色。

（3）汉语量词范畴化中的精细认知特征和社会认知属性。量词是语用者对事物量化或属性标记的重要手段，是语用者思维方式和外部世界发展的反映。“家”和“户”在早期古汉语中均主要标记“家庭组织”这一范畴，但随着社会发展，尤其是工商业的发展，社会分工日趋多样，各类行业组织不断涌现。在此背景下，二者用法出现分化，“户”仍主要用于“家庭组织”，而“家”则转而主要用于“行业组织”。这既反映出语言经济性引发的不同语言分工，又反映出汉语言建构中的精细化认知特征和社会认知属性。

5. 结语

量词是以汉语为代表的汉藏语系的特色词类，具有量化和属性标记双重功能，其形成和发展既反映出汉语民族的思维方式，又带有显著的社会认知属性。本研究的意义主要有以下几个方面：首先，通过对汉语近义组织量词的演化考察，发现了汉语量词建构的特色和先前构式理论在汉语研究中的不足，深化了对构式理论和汉语特色的认识；其次，在考察汉语近义组织量词构式搭配的语言本体特征之余，进一步挖掘其构建背后的认知特征和社会学特征，为语言学、心理学及社会学的多界面融合研究提供了借鉴思路；再次，将认知语言学理论、多语料库及构式搭配分析法相结合，对汉语近义组织量词进行共时性和历时性考察，较大程度上实现了描述充分性和解释充分性的统一、静态研究与动态研究的统一，提升了研究结论的科学性；最后，现有社会语言学研究多关注语言与社会的宏观界面，本研究采用的“微观语言现实”—“宏观社会现实”考察路径拓展了该领域的研究思路，为“微观社会认知语言学”的创立带来一定启示。

参考文献

Aikhenvald, A. 2000. *Classifiers: A Typology of Noun Categorization Device*. Oxford: Oxford University Press.

Goldberg, A. 2006. *Construction at Work: The Nature of Generalization in Language*. New York: Oxford University Press.

Gries, S. 2011. Corpus data in usage-based linguistics. In M. Brdar, S. Gries & M. Fuchs.(eds.). *Cognitive Linguistics: Convergence and Expansion*. Amsterdam: John Benjamins.237-256

Levinson, S. 1983. *Space in Language and Cognition*. Beijing: World Publishing Corporation.

Martinet, S. 1975. *Studies in Functional Syntax.* Cambridge: Cambridge University Press.

Traugott, E. & G. Trousdale. 2013. *Constructionalization and Constructional Changes.* Oxford: Oxford University Press.

蔡淑美、施春宏，2022，论构式用变和构式演变。《当代修辞学》(2)：58-73。

陈伟琳，2002，多胞胎量词运用刍议。《信阳师范学院学报》(4):96-97+109。

郭锐、李知恩，2021，量词的功能扩张。《中国语文》(6):691-712+767。

李兴、李尚儒，2018，《周易》。西安：三秦出版社。

刘子平，2013，《汉语量词大词典》。上海：上海辞书出版社。

吕叔湘，1982，《中国文法要略》。北京：商务印书馆。
商务国际辞书编辑部，2017，《现代汉语词典》。北京：商务印书馆。
沈家煊，1998，实词虚化的机制——《演化而来的语法》评介。《当代语言学》(3):41-46。
王力，1989，《汉语语法史》。北京：商务印书馆。
文旭、杨旭，2016，构式化：历时构式语法研究的新路径。《现代外语》(6):731-741+872。
魏校撰，1996，《六书精蕴》。上海：上海古籍出版社。
熊加全，2020，《<玉篇>疑难字考释与研究》下。北京：中华书局。
许慎，2016，《说文解字》。杭州：浙江古籍出版社。
郑邵琳，2016，《魏晋南北朝石刻名量词研究》。上海：上海古籍出版社。
宗守云，2010，《集合量词的认知研究》。北京：世界图书出版公司。

A Study of Construction Collocation and Constructionalization of Chinese Synonymous Organization Classifiers

Abstract: Based on corpus and the R language platform, from the perspectives of construction, this paper first studies the categorization of the organizational construction "X *jia/hu* N"in modern Chinese and then explores their origin path and cognitive motivation in ancient Chinese, which expands the research perspective of Chinese classifiers, deepens the understanding of the characteristics of Chinese classifiers. The findings are as follows: (1) "X *jia* N" mainly matches nouns of "the organization of industry", occasionally matches "the organization of family or residence"; the structure of "X *hu* N" matches "the organization of family under administration", occasionally expresses "the organization industry or residence under administration". (2) Metaphor and metonymy are the main cognitive mechanisms of these construction, but metonymy plays a leading role in the formation of these constructions, while metaphor plays a leading role in the evolution of these constructions. (3) In addition to metaphorical extention link, polysemy link, instance link and subpart link, which was previously proposed by Goldberg, metonymical extention link also exists in the inheritance link of Chinese classifiers, reflecting significant refined cognitive characteristics and social cognitive attributes.

Key words: organizational classifiers; collocation; "jia"; "hu"; constructionalization

（责任编辑：高彦梅）

中国戏剧．2024 (06) ：21-23 查看该刊数据库收录来源

中原人文精神的艺术呈现 评方言话剧《老家》

摘要： <正>20世纪初，话剧在中国萌芽，作为一种舶来的艺术形式，话剧不断地被吸纳、改造，并最终实现了创造性转化，如今已经成为中国戏剧体系的重要组成部分。方言话剧作为话剧民族化、地域化的产物，在戏剧舞台上独树一帜，相继涌现出《茶馆》《七十二家房客》《抓壮丁》《白鹿原》等佳作。河南方言话剧也频频发力，自2014年始，河南省重点打造“三老”系列原创方言话剧——《老汤》《老街》《老家》，分别讲述中原诚信、文化传承和根亲文化等主题，以三部曲的形式力塑中原人文精神。

基金资助： 国家社科基金(艺术学)“黄河流域道情传承研究”(项目编号：20CB163)； 2023河南兴文化工程文化研究专项项目“太康道情戏研究”的阶段性成果；

专辑： 哲学与人文科学

中原人文精神的艺术呈现

评方言话剧《老家》

文/杨志敏

20世纪初，话剧在中国萌芽，作为一种舶来的艺术形式，话剧不断地被吸纳、改造，并最终实现了创造性转化，如今已经成为中国戏剧体系的重要组成部分。方言话剧作为话剧民族化、地域化的产物，在戏剧舞台上独树一帜，相继涌现出《茶馆》《七十二家房客》《抓壮丁》《白鹿原》等佳作。河南方言话剧也频频发力，自2014年始，河南省重点打造“三老”系列原创方言话剧——《老汤》《老街》《老家》，分别讲述中原诚信、文化传承和根亲文化等主题，以三部曲的形式力塑中原人文精神。2024年新春伊始，三部曲收官之作《老家》（河南省歌舞演艺集团演出，王宏编剧，宫晓东导演，范军等主演）隆重上演，并相继入选国家艺术基金年度资助项目及第八届全国话剧优秀剧目展演，在剧坛上刮起一股强劲的“中原风”。

一

正所谓“十里不同风，百里不通音”，方言是人类社会在漫长的历史发展进程中形成的具有鲜明地域色彩的语言类型。我国地域广阔，仅汉语就有北方方言、吴方言、湘方言、赣方言、客家方言、粤方言和闽方言七大方言语系，河南方言就属于北方方言的中原官话。话剧作为一种以人物对话为主的戏剧形式，也是一门语言的艺术，因其艺术形式和表现手法更善于表现现实生活，引进之初被作为唤醒民众反帝反封建、移风易俗改良社会的工具，自然而然地使用了全国范围内较为通用的“国语”——普通话，久而久之，话剧说普通话成为思维定式和审美习惯。但方言剧绝非普通话的对立面，早在20世纪三四十年代，大众语运动引发了话剧使用国语还是方言的“应宜”论争。陈大悲、熊佛西等中国话剧的先驱都曾撰文发表自己的观点，此后的近百年中，关于方言话剧的讨论从未休止，从为方言话剧正名，到为方言话剧立命，从偏激地讨伐方言话剧，到正视方言话剧的价值，方言话剧走过多年的诘难与争议。[①]21世纪以来，随着方言话剧演出的增多以及学界对方言话剧讨论深度的拓展，人们越来越认识到方言话剧在挖掘本土文化、表现地域特色、拉近观众距离、破解话剧困境等方面的价值和意义，方言话剧的舞台演出逐渐增多，影响力也不断扩大。

河南作为文化大省，在方言话剧方面也成果丰硕。早在1963年，中国青年艺术剧院农村巡回演出队用河南方言演出话剧《李双双》，成为河南本土话剧的肇始。[②]方言话剧作品不断，近年更有《宣和画院》《老街》《老汤》《一句顶一万句》等力作问世，“三老”收官之作——话剧《老家》延续一贯的中原情怀，聚焦普通百姓日常，与《老汤》《老街》不同，话剧《老家》与现实生活更为贴近。该剧讲述了郑州一个城中村小

①贺思齐：《透视方言话剧：话语争夺、艺术困局与传播生成》，《戏剧艺术》，2022年第4期。

②陈晶晶：《方言话剧与中原文化：〈宣和画院〉的文化想象》，《殷都学刊》，2017第3期。

院，房东范二爷与各地租客之间发生的一段故事。全剧以河南方言为表现手段，展示地地道道的中原文化，开封、洛阳、漯河、杜岭街、顺河路、庆丰街等一串熟悉的地（街）名鳞次展现，胡辣汤、“猜枚”酒令、童谣等萦绕其间，营造了一个特有的中原生活场景。话剧《老家》大量使用“恁”“中”“俺”“谝”“拨楞”“龟孙”“嘟囔”“划不来”“寻（sin）媳妇”“二球”等方言词汇，还有歇后语、俗语、谚语等穿插其中，生动形象、幽默风趣，展现了中原人豪爽大气、乐观向上的群体性格，此外还灵活化用流行用语，如调侃考古“在小小的花园里面挖呀挖呀挖”、解释上楼后的单元房“孤孤单单没人缘”等，既亲切熟悉又别开生面，正如马林诺夫斯基所说：“语言是文化整体的一部分，但它并不是一个工具的体系，而是一套发音风俗及精神文化的一部分。”在话剧《老家》用方言营造的文化空间中，观众仿佛置身中原大地，在胡辣汤、烙馍豆腐串的叫卖声中，在邻里街坊家长里短的“喷空”（河南方言，意为聊天）声中，深入感受到中原人民的情感世界和生活状态，也使得话剧在表现中原人文精神时更加真实、生动，更具感染力。

二

河南有着丰富的曲艺资源，是民众文化娱乐的主要途径之一，也是中原文化的重要组成部分[①]，丰富的曲艺文化也滋养着话剧的创作。“中原人文精神三部曲”由河南省歌舞演艺集团曲艺团十年磨一剑倾力打造而成，曲艺是其鲜明的艺术烙印。三部作品的编剧均为王宏，他是话剧人，更是曲艺人，曾为济南市曲艺团创作《茶壶就是喝茶的》《泉城人家》等话剧。谈及《老家》时他这样阐述自己的创作理念：“这种曲艺与戏剧跨界交融的演出形式，让观众和演员走得更近，让角色与演员贴得更紧，让曲艺人身上的绝活有了更加开阔的舞台。”因此在剧本创作之初就为曲艺预留了充足的展示空间。在演员选择上，以范军为代表的演员大多曲艺出身，有着丰富的舞台实践经验，没有矫揉造作的“话剧腔”，曲艺表演更是信手拈来。因此，三部曲具有鲜明的曲艺色彩，如《老汤》中有“武坠子”一角，主角罗小船不止一次用坠子哼唱“想妹妹想得我吃不下去饭……”《老街》更是以民间说唱艺人的“精神圣地”——马街书会为对象，以曲艺人演曲艺史。《老家》中也有“坠子陈”一角，说书、快板等曲艺不时点缀其中。如果说《老汤》是河南曲艺与话剧融合的初次尝试，《老街》是这种艺术形式

的再度确认和成熟，那么《老家》则在曲艺的运用上更为纯熟自然。

话剧《老家》伊始，以说书人的形式开篇，“鸡也不叫了，狗也不吵了，天也不早了，人也不少了，醒木一拍大书开怼。话说大清朝光绪三十年，也就是1904年，一条铁路拉动了郑州的过去和未来……书归正章，到了今天，小院儿要拆了，故事开始了”。寥寥几句交代了小院的故事背景，比话剧旁白的形式更为新颖别致。河南坠子作为本土第一大曲种，更是作为隐含的脉络融汇全剧。小院为坠子皇后所留（与话剧《老街》梦幻联动），现在成了租客们来郑的落脚地；范二爷年轻时追奶奶，唱了三天三夜坠子，最终抱得美人归；因贪唱坠子，范二爷弄丢孙子，改变了整个家庭的命运；一曲坠子《劝世方》更是贯穿始终。

人生春暖到秋凉，
恁听俺唱一段劝世方。
都说那荣华富贵三生梦，
叫我看红尘白浪两茫茫。
是非不必争你我，
平安才有百年长。

该曲多次出现，或为背景苍凉悠长，或为情节推动发展，或为独白一唱三叹，使全剧张弛有度、富有哲思，而最后一场，小院拆迁，消失已久的百姓庙在废墟中重现，范儿爷带领众人齐唱坠子致敬。

百家姓，姓氏全，
姓啥咱都不简单。
炎黄二帝是老祖，
他有个大号叫轩辕。
我姓钱脱不了没钱花，
你姓官不一定掌大权。
谁说姓白就白干？
谁说姓王就坐金銮？
姓孙误不了当爷爷，

①张凌怡、刘景亮、李广宇：《河南曲艺史》，河南人民出版社，2007年版，第11页。

你姓郝也得有个好人缘。
姓啥不由咱，
都为敬祖先，
忘不了老根是中华，
忘不了老家在河南。

一曲《老家歌》慷慨激昂、荡气回肠，令人回味无穷。剧中曲艺与剧情的结合熨帖流畅、恰到好处，不炫技、不浮夸，使得曲艺内容及形式本身都成为话剧的亮点和特色，不仅展示了以曲艺为代表的中原优秀传统文化，而且丰富了方言话剧的艺术类型和话语表达。

三

话剧是在固定空间中表现悲欢故事、呈现世间百态，从而达到观众与演员、剧情亲密交流的艺术审美过程。方言也好、曲艺也罢，重要的是话剧本身的真实、丰满、厚重，跳脱艺术的雅俗之争，更多去呈现剧情的起伏、人物的塑造和主题的开掘。

方言话剧《老家》的剧情较为简单。全剧有两条线索，一是小院的历史。小院为坠子皇后所留，范二爷从母亲手中接过小院，因交通便利、房租低廉，小院吸引了各地租客，当下，小院面临拆迁。另一是范爷的家庭史。12 年前他贪嘴唱坠子弄丢孙子，儿子儿媳远走重洋，自此他与精神失常的老伴相依为命。两条线索合二为一、相互交织，矛盾展开不是惯常的环环相扣的写法，而是笔触撒得很开，伏笔埋得很长，话题涉及很多，如城市拆迁、自媒体整顿、网恋催婚、大学生就业等诸如此类“民生问题”，仿佛是身边随处可见的百姓日常。但全剧形散而神聚，看似漫不经心，实则独具匠心，在小院普通的日常中展现中原百姓生活的粗粝与宽广、琐碎与温情，从而带来沉浸式观剧体验。在人物塑造上，话剧《老家》聚焦小人物命运，正如编剧王宏在接受采访时所说，他曾两次撕毁《老家》的故事大纲，几乎否定了所有所谓的好人好事、大是大非，力求做到写人事、说人话、见人情。剧中有流浪少年钱来喜、患不育症的白胜夫妇、求职碰壁的研究生于子渡、开饭馆失败的梅有才、从乡下探亲的麦穗等。唯一“西装革履”的老郝是省博物院退休保安，他们都奔波在社会的底层，有生活的盘算、失意的心酸，也有彼此间的怜悯与守望。在主角范二爷的身上，更是多维度开掘了他的人生侧面，唱坠子时的骄傲、丢孙子时的忏悔、陪老伴时的耐心与帮房客时的仗义等，房东、爷爷、父亲、老伴的多重身份交织，淋漓尽致地展现了他命途多舛的一生。老伴去世时，禁口多年的他重唱坠子，“人生春暖到秋凉 / 恁听我唱段劝世方 / 老伴儿的账咱欠多少 / 回头望哪能不思量 / 当年她俏脸庞、衣红装 / 好似那一剪梅傲风霜 / 只为一句知心话呀 / 好姑娘就做了咱的新娘”，令人不禁泪目，人物形象感人至深。几位主演也生龙活虎、惟妙惟肖、真实自然，塑造了中原之子有血有肉、饱满生动的性格特征，极富戏剧张力。在主题上，全剧围绕“共享”二字呈现“老家”情怀，“共享”不仅仅是流行的共享单车、共享汽车、共享雨伞等新兴事物，更是来自不同地域租客“求子”“求爱”“求职”“求财”的生存状态，在小院这个临时大家庭里，他们可以“共享家具”“共享心事”“共享奶奶”，相互取暖、彼此慰藉。何谓老家？老家在哪？保安老郝说“老家是走多远都要回来的地方”，失智的奶奶说“家是奶奶的摇篮曲，家是树上的老鸹窝，家是被窝里烤热的砖”，流浪少年钱来喜说“奶奶在哪，哪就是我的家”，范二爷说“小院就是我的老家，我哪也不去”。他们在离开家、失去家、寻找家的过程中探究家的意义，小院（抑或是“老家”）不仅是普通人的生活空间、生活故事，也是他们的心灵空间、情感故事。这个即将面临拆迁的破落小院，安放着人们面对社会变迁热潮手足无措、不堪落寞而又焦虑不安的心。他们在城市钢筋水泥的世界里，搭建起一个短暂而又永久的乌托邦、桃花源，心安即是家。同时，小院如同一个切面展现大中原的风云巨变，结尾的《老家歌》唱出“老根是中华、老家在河南”的激昂旋律，炎黄故土精神家园、根亲文化血脉相连，小院人的热情、仗义、坚韧也升腾为中原精神的缩写和象征，展现河南作为中华民族和华夏文明重要发源地的厚重与博大。

当然，作为新搬上舞台的剧目，话剧《老家》仍有许多不足，在情节和语言上还需要打磨，如梅有才为安慰妻子麦穗谎称买房，而麦穗患癌也瞒着丈夫，这种欧·亨利《麦琪的礼物》式的剧情在呈现上还略显苍白，语言方面也存在一些过分调侃与抖包袱之处，有待进一步完善。

总之，话剧《老家》地域特色鲜明、时代气息强烈，全剧融汇中原方言、中原曲艺、中原故事弘扬中原人文精神，是一部展现中原文化、彰显黄河气派的本土话剧。“中原人文精神三部曲”《老汤》《老街》《老家》的集中亮相，不仅是方言话剧审美价值的再度确认，也是厚重中原文化的艺术突围和舞台呈现，更是话剧民族化、地域化进程中又一次艺术上的创新和求变。

［本文系国家社科基金（艺术学）“黄河流域道情传承研究”（项目编号：20CB163）、2023 河南兴文化工程文化研究专项项目“太康道情戏研究”的阶段性成果。[redacted]**士生导师］**

责任编辑：靳文泰

基于深度学习的智能文本分析技术在英语教学效果评估中的应用

报告编号：J20255001269977390

报告验真

检 索 报 告

检索主题：[redacted]

委 托 人：

数 据 库：SCI

检索时间：2025年9月20日

检索结果：

根据委托人本次委托要求，在上述数据库范围内，王金海发表论文收录情况如下表：

委托要求		检索结果
数据库	委托篇数	收录篇数
SCI	1	1

科学技术部 骑

科学技术部西南信息中心 创新中心

2025年9月20日

报告专用章

一、检索结果附表

序号	论文信息	[redacted]作者位次	收录情况	期刊指标
1	Ren Yuan, Fan Wenjuan, Wang Jinhai. Intelligent text analysis for effective evaluation of english Language teaching based on deep learning. SCIENTIFIC REPORTS, 2025, 15(1).	通讯作者	SCI	JCR影响因子(2024):3.9 JCR分区(2024):Q1区

检索员（签章）

科学技术部西南信息中心查新中心

2025年9月20日

报告专用章

西南信息中缝

二、检索结果附件

（一）SCI收录

1 record(s) printed from Clarivate Web of Science

第 1 条，共 1 条

标题: Intelligent text analysis for effective evaluation of english Language teaching based on deep learning

作者: Ren Yuan,Fan Wenjuan,Wang Jinhui

来源出版物: SCIENTIFIC REPORTS 卷: 15 期: 1 文献号: 28949 DOI: 10.1038/s41598-025-14320-5 出版年: AUG 7 2025

入藏号: WOS:001546845000006

PubMed ID: 40775439

文献类型: Article

地址: [Re
[Fan, We
[Wang, Ji
通讯作者
电子邮件

Affiliations: Zhen

ISSN: 2045-2322

心查新中心
章

scientific reports

Check for updates

OPEN

Intelligent text analysis for effective evaluation of english Language teaching based on deep learning

With the growing demand for English language teaching, the efficient and accurate evaluation of students' writing ability has become a key focus in English education. This study introduces a Hybrid Feature-based Cross-Prompt Automated Essay Scoring (HFC-AES) model that leverages deep learning for intelligent text analysis. Building on traditional deep neural networks (DNNs), the model incorporates text structure features and attention mechanisms, while adversarial training is employed to optimize feature extraction and enhance cross-prompt adaptability. In the topic-independent stage, statistical methods and DNNs extract shared features for preliminary scoring. In the topic-specific stage, topic information is integrated into a hierarchical neural network to improve semantic understanding and topic alignment. Compared with existing Transformer-based scoring models, HFC-AES demonstrates superior robustness and semantic modeling capabilities. Experimental results show that HFC-AES achieves strong cross-prompt scoring performance, with an average Quadratic Weighted Kappa (QWK) of 0.856, outperforming mainstream models. Ablation studies further highlight the critical role of text structure features and attention mechanisms, particularly in improving argumentative writing assessment. Overall, HFC-AES offers effective technical support for automated essay grading, contributing to more reliable and efficient evaluation in English language teaching.

Keywords Automatic grading of english composition, Deep learning, Mixed features, Cross-topic scoring, Text analysis

With the acceleration of globalization and informatization, English education has attracted increasing worldwide attention. In English language teaching, writing serves as a critical indicator of comprehensive language ability, reflecting students' linguistic proficiency and cognitive skills[1–3]. Particularly in large-scale examinations, online education platforms, and international curricula, English writing proficiency has become a key metric for evaluating overall language competence. However, current writing instruction and assessment still rely heavily on manual grading by teachers—a process that is labor-intensive, time-consuming, and often influenced by subjective standards and inter-rater variability, making it difficult to ensure consistent and fair essay evaluation[4–6]. This issue is especially pronounced in large-scale testing scenarios, where scalable and reliable automated scoring systems are urgently needed as a replacement or complement to manual grading. Automated Essay Scoring (AES) technology has emerged as a promising solution to this challenge.

Early AES models largely depended on shallow linguistic features such as word frequency, sentence length, and spelling or grammar errors, combined with traditional machine learning methods. While these models demonstrated some effectiveness in specific tasks, their performance proved unstable when applied to essays prompted by different topics[7,8]. Changes in writing prompts, styles, or linguistic backgrounds often led to poor generalization and significant scoring bias, largely due to the models' overreliance on topic-specific features in training data[9,10]. This phenomenon, known as the "cross-prompt scoring challenge," remains one of the major obstacles in AES research. Addressing this challenge requires models that can capture general linguistic features while accurately assessing semantic alignment between essays and prompts across diverse topics, thereby improving scoring fairness and credibility.

Advancements in intelligent text analysis, particularly deep learning (DL)-based methods, enable automatic extraction of richer linguistic features from large-scale essay data. By capturing deep semantic representations

and structural patterns, DL provides a more accurate foundation for essay evaluation[11,12]. Building on these advances, this study proposes a Hybrid Feature-based Cross-Prompt Automated Essay Scoring (HFC-AES) model designed to enhance scoring accuracy and consistency in multi-topic and multi-prompt scenarios. The model integrates shallow statistical features with semantic features extracted through DL. In the topic-independent stage, shared features are derived to provide stable preliminary assessments, while in the topic-specific stage, a hierarchical neural network and cross-attention mechanism are incorporated to model semantic relationships between essays and prompts more precisely. Leveraging text structure features and attention mechanisms, the proposed approach enhances robustness and adaptability in diverse prompting conditions. Ultimately, HFC-AES offers an intelligent scoring tool that supports English language teaching and facilitates the practical application of automated scoring technology in educational evaluation.

Literature review

Over the years, the rapid development of natural language processing (NLP) has significantly advanced the field of AES. Existing research can be broadly categorized into three areas: AES models based on traditional feature engineering, AES methods utilizing DL, and AES models designed for cross-prompt and multilingual contexts.

(1) AES models based on traditional feature engineering.

Most early AES systems relied on manually designed shallow linguistic features, such as lexical density and syntactic structure, for scoring and modeling. Susanti et al. (2023) conducted a comprehensive literature review of AES systems, analyzing the use of various methods and datasets to provide methodological and dataset references for future research[13]. Li and Huang (2022) explored the influence of composition, organization, and overall quality on the evaluation of English as a foreign language writing in Chinese higher education. Through interviews with teachers and raters and a quantitative analysis of large-scale evaluation data, they identified clear differences in scoring focus. High-quality compositions were evaluated across multiple dimensions. In contrast, low-quality compositions were assessed mainly for language accuracy and content[14]. These findings highlighted the limitations of traditional AES models in constructing comprehensive scoring dimensions and underscored the need to reconsider feature selection for fairness and completeness. Although feature-engineered methods offer interpretability and computational efficiency, they struggle to capture deeper semantic relationships and contextual information. Consequently, their generalization ability is limited, and they fall short in assessing semantic alignment and overall discourse coherence, particularly for complex, variable-prompt writing tasks.

(2) AES methods utilizing deep learning.

The emergence of neural network models has led many researchers to explore DL-based approaches for automatically learning semantic and structural features in student compositions. Lim et al. (2023) developed and validated a neural network-based automated assessment system tailored for Korean second-language writing. By combining NLP techniques with pre-trained neural language models, the system improved scoring performance through analyses of linguistic features such as grammatical complexity, quantitative complexity, and fluency[15]. This work demonstrated the value of applying neural methods to non-English AES tasks, extending the applicability of DL in multilingual contexts. Beyond NLP advances, intelligent text analysis has introduced new approaches for AES. Bai and Stede (2023) reviewed recent applications of machine learning in automated evaluation of student free-text responses, including both short answers and full essays, highlighting the predominant use of feature-based and neural network architectures[16]. Compared with traditional methods, DL-based approaches excel at automatically learning complex features and modeling contextual semantic relationships and textual coherence. However, existing DL models still face notable limitations: poor transferability across prompts, vulnerability to topic bias in training data, limited interpretability due to "black box" architectures, and insufficient handling of discourse-level structures, as most focus primarily on syntactic or lexical features rather than modeling macro-level semantic organization.

(3) AES models in cross-topic and multilingual contexts.

To address the challenges posed by diverse essay prompts and the uneven distribution of language resources, researchers have explored strategies for cross-prompt and multilingual AES systems. Gao et al. (2024) reviewed the integration of AI and NLP in automated writing evaluation from an educational perspective, emphasizing the potential of large language models to improve assessment efficiency[17]. Hossain and Goyal (2024) trained pre-trained Transformer models—such as BERT, GPT, Multilingual BERT (mBERT), and Cross-Lingual Models (XLM-R)—on multilingual corpora covering over 20 languages. These models were fine-tuned for tasks including text summarization, content generation, and sentiment analysis, demonstrating strong multilingual semantic modeling and text generation capabilities, especially in coherence and fluency[18]. Li (2025) proposed a novel cross-lingual sentence similarity detection approach that combined the multilingual power of XLM-R with a stepwise weighted similarity metric integrating cosine similarity and Manhattan distance, along with language-independent embeddings from BiT-Internet and XLM-R. This method significantly improved semantic equivalence detection, setting new benchmarks in cross-lingual similarity tasks[19]. Although these studies laid a theoretical foundation for enhancing the generalization of AES systems, current models still struggle with semantic alignment and discourse-level modeling in essays—complex, highly topic-dependent text. They often fail to accurately capture semantic correspondence between essay content and prompts.

Building on these insights, this study proposes the HFC-AES model, which integrates shallow statistical features with deep neural representations. The architecture consists of two stages: a topic-independent stage that extracts stable shared features for consistent cross-prompt scoring, and a topic-specific stage that employs

a hierarchical neural network and cross-attention mechanism to precisely align essay content with prompts. By combining semantic modeling with discourse structure recognition, the model overcomes robustness limitations of existing AES systems in multi-prompt scenarios. Table 1 compares mainstream AES approaches with the proposed model in terms of feature representation, semantic alignment, discourse modeling, cross-prompt adaptability, interpretability, and model-specific enhancements:

As Table 1 illustrates, different AES approaches exhibit distinct strengths and weaknesses across feature modeling, semantic understanding, transferability, and interpretability. Traditional methods, though highly interpretable, lack the capacity to model complex semantics and discourse structures, limiting their applicability to challenging writing tasks. DL-based methods have made significant progress in semantic representation but often lack generalization beyond specific topics or datasets, resulting in unstable cross-prompt scoring. Multilingual pre-trained models demonstrate potential in handling cross-lingual semantics but remain inadequate in modeling essay-specific discourse structures. In contrast, HFC-AES systematically optimizes semantic alignment, discourse structure modeling, and robustness in cross-prompt scenarios by integrating shallow linguistic features with deep neural representations. Its cross-attention mechanism enhances the capture of key correspondences between essay content and prompts, improving both holistic discourse understanding and scoring reliability. Overall, HFC-AES balances interpretability with advanced semantic modeling, addressing limitations of existing methods and demonstrating greater generalizability and practical value.

Research model

Model overall architecture design

The HFC-AES model employs a two-stage feature extraction process. In the first stage, DL techniques are used to extract comprehensive textual features from raw essays, encompassing syntactic structure, lexical usage, and semantic information. These features are derived through pre-trained word embedding models and syntactic analysis tools. Moreover, sentence-level discourse structure features are incorporated to capture the internal logical relations and organizational framework of the composition, thereby enhancing the model's grasp of discourse-level structures. In the second stage, a cross-attention mechanism further refines feature processing by automatically learning the relative importance of various scoring criteria for the overall assessment. This mechanism effectively emphasizes critical sections of the essay and dynamically allocates feature weights in accordance with specific task requirements. The overall workflow of the HFC-AES model is illustrated in Fig. 1.

The HFC-AES model employs a dual-channel architecture. One channel is dedicated to extracting and processing global features, while the other focuses on capturing and enhancing local features. By integrating information from both channels, the model can generate scoring predictions at multiple granular levels. To enhance interpretability, a visualization technique is incorporated to display the feature weight distribution for each scoring criterion, thereby making the model's decision-making process more transparent and reproducible. The overall structure of the HFC-AES model, including the topic-independent and topic-related feature extraction stages, is illustrated in Fig. 2.

In the topic-independent stage, the model extracts shallow text features at both word and sentence levels and combines these with deep semantic features generated by DL–based text analysis methods. In the topic-related stage, a Bi-LSTM coupled with an attention mechanism constructs a hierarchical semantic network that captures semantic information relevant to both the composition and the prompt. By linking the shared layer with the task-specific layer, this stage effectively models contextual semantic relationships within the essay and integrates feature correlations across multiple tasks through a cross-attention mechanism. Finally, the scoring module combines the topic-independent and topic-related feature representations to provide precise scores for essays on the target topic, enabling comprehensive assessment of both linguistic competence and content quality in English writing. Multiple neural network architectures are employed in the design of the HFC-AES model to fully explore the multi-level semantic information of essays. Table 2 summarizes and compares the key neural network architectures used, detailing their core functions, advantages, and specific roles in this study.

From the comparison presented in Table 2, it is evident that the various neural network architectures fulfill complementary roles and collaboratively enhance the model's capability to capture multi-level semantic information within essays. The following sections provide a detailed introduction to the specific applications and implementations of these network architectures within both the topic-independent and topic-related feature extraction stages.

Method category	Feature types	Semantic alignment	Discourse modeling	Cross-prompt adaptability	Interpretability	Improvements of HFC-AES
Traditional feature engineering	Lexical, syntactic, length	Weak	Weak	Weak	Strong	Limited in modeling complex semantics
Single-prompt DL-based models	Word embeddings, CNN, RNN	Moderate	Basic	Weak	Weak	Lacks cross-prompt robustness
Multilingual pre-trained models (e.g., mBERT)	Cross-lingual semantic representations	Strong	Weak	Moderate	Weak	Lack mechanisms for essay structure modeling
HFC-AES	Hybrid shallow + deep features	Strong	Strong	Strong	Moderate	Introduces hierarchical modeling and cross-attention

Table 1. Comparison of different AES methods.

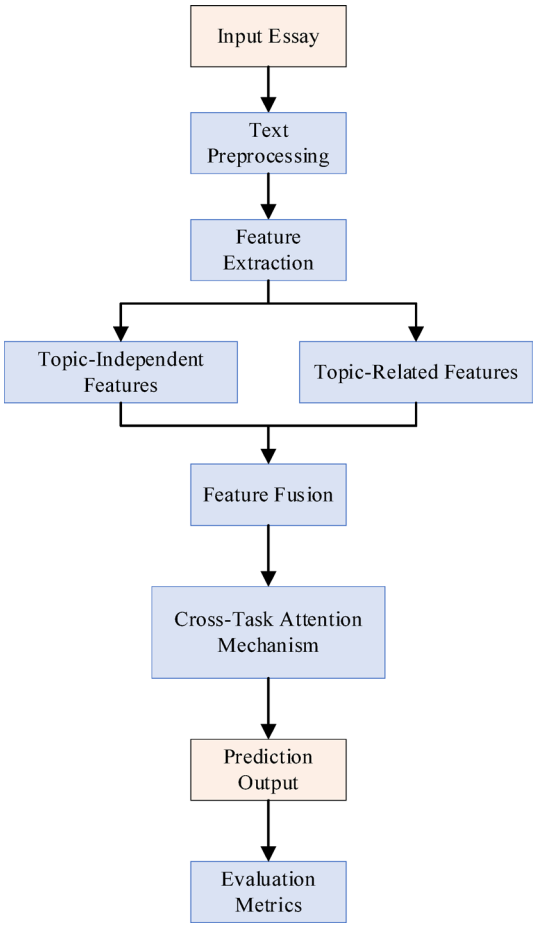


Fig. 1. The workflow of the HFC-AES model.

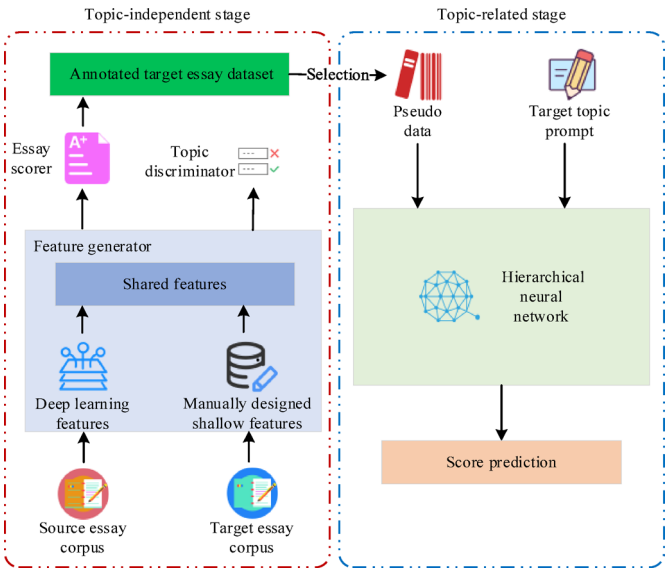


Fig. 2. Overall flow chart of the model.

Model architecture	Core function	Advantage	Application locations in this study
Convolutional Neural Network (CNN)	Extract local semantic features of text	It is good at capturing local patterns, with relatively few parameters and strong stability.	Local semantic modeling in shallow and deep text feature extraction
Long Short-Term Memory (LSTM)	Capture long-distance dependencies and global semantic relationships of text	It solves the traditional RNN gradient vanishing and is suitable for long text sequence processing	Deep feature extraction strengthens the semantic coherence between sentence sequences
Hierarchical neural network	Hierarchical modeling of the local and global structure of text	It preserves text hierarchies and enhances topic-related semantic understanding	In the topic-related feature extraction stage, the relationship between the topic and the topic of the composition is processed.
Attention mechanism	Dynamically adjust the weights of different features to focus on key semantic information	It enhances the model's ability to identify important information and improves task adaptability	The theme-related stage supports the multi-task feature fusion of the cross-attention mechanism.

Table 2. Main neural network architectures in the HFC-AES model.

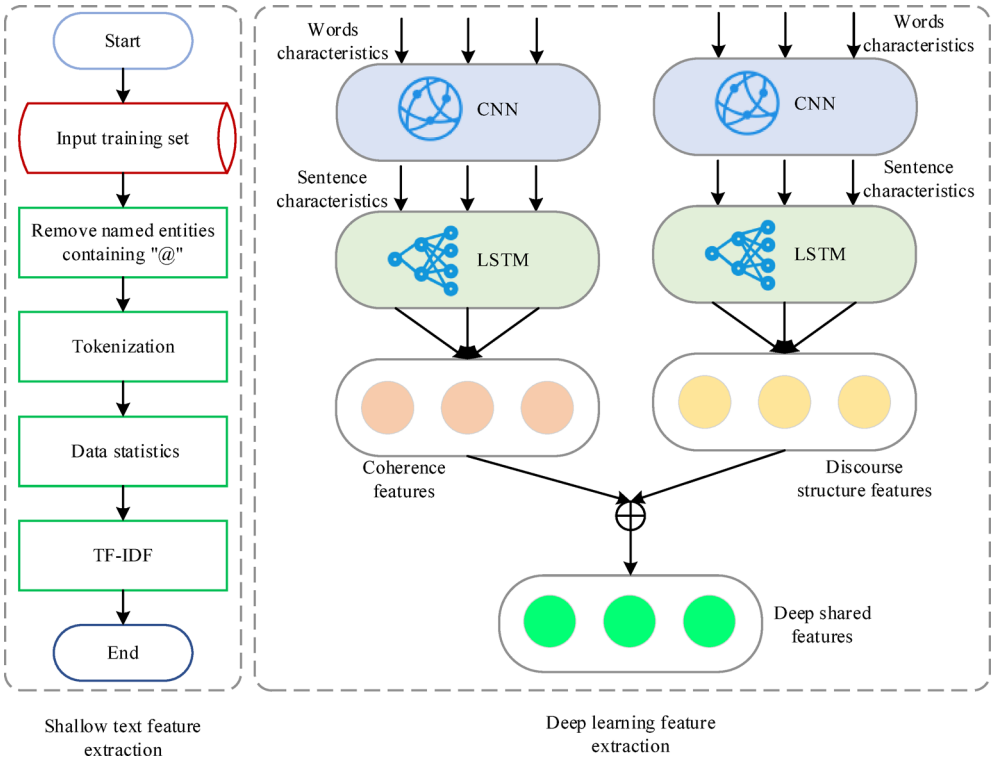


Fig. 3. Feature Extraction in topic-independent stage.

Topic-independent feature extraction stage

In the topic-independent stage, the primary objective is to extract shared features between the source and target essay datasets. These features are then employed to build a preliminary scoring model for initial evaluation of the target essays. The shared features consist of shallow text features and DL features, as illustrated in Fig. 3. Shallow text features are manually designed using statistical methods and capture fundamental textual information such as vocabulary usage frequency and sentence structure. In contrast, DL features are automatically learned by deep neural networks (DNNs), which identify complex patterns and semantic relationships within the text. The hybrid approach of combining shallow text features with DL features aims to leverage the strengths of both, enhancing the accuracy and robustness of the scoring model. Shallow features provide intuitive, easily computable information—such as lexical richness and sentence length—that directly reflect students' language proficiency and offer high interpretability. Meanwhile, DL features model text at a deeper level, capturing intricate grammatical and semantic relationships, including logical sentence connections, discourse structure, and underlying semantic intentions. These high-level features are critical for assessing essay coherence, organization, and semantic precision. By employing intelligent text analysis, this extraction approach integrates traditional lexical and syntactic information with deeper semantic and contextual insights[20]. The combination of shallow and DL features mitigates the potential information loss that may occur if either feature set is used alone, thereby improving the model's capacity for comprehensive evaluation of students' writing skills.

(1) Shallow text feature extraction.

Shallow text features are extracted at the word and sentence levels using statistical methods to capture students' vocabulary proficiency and sentence structure skills. To improve feature relevance, the Term Frequency-Inverse Document Frequency (TF-IDF) method measures each word's importance within an essay by balancing its frequency against its rarity in the entire corpus. This effectively filters out common but less meaningful words and highlights key terms, enhancing lexical feature discrimination and scoring accuracy.

At the word level, features include composition length, average word length and its variance, number of spelling errors, and ratios of prepositions and conjunctions[21,22]. These are extracted with tools like SpellCheck and NLTK and analyzed alongside vocabulary profiles. Additionally, intelligent text analysis identifies spelling error types and word usage frequency in context, providing richer semantic insights[23].

At the sentence level, features describe structural and coherence aspects, such as sentence count, average sentence length, grammatical errors, and overall coherence. Sentence count and the sentence-to-word ratio reflect essay complexity, while average sentence length indicates structural sophistication. Grammatical error counts serve as an accuracy metric[24,25]. Sentence coherence is calculated using the following weighted formula:

$$p = 0.45\times\ \bar{m} + \bar{l}\ -21.05 \tag{1}$$

p is the sentence coherence score; $\bar{m}$ denotes the average number of characters in a word; $\bar{l}$ refers to the average length of a sentence. To calculate these features, *Language*, a localized tool, is used to detect grammatical errors in sentences and calculate the coherence of sentences.

(2) Deep text feature extraction.

Deep text features are extracted using a DNN that vectorizes and models essay text to capture high-level semantic attributes such as coherence and discourse structure. To better extract these deeper semantic features, the model combines Convolutional Neural Network (CNN) and Long Short-Term Memory Network (LSTM), leveraging their complementary strengths. First, essay text is converted into word vectors using the Word2Vec method. These word vectors are dynamically weighted through intelligent text analysis to automatically identify key topic words and important expressions. The CNN extracts local features from the word vectors, producing sentence-level representations. Subsequently, the LSTM captures temporal dependencies and global semantic information across sentences, generating features related to coherence and text structure. Finally, these DL features are combined with shallow text features to form the model's input, which is then fed into the essay scorer and topic discriminator modules.

To further enhance performance, the feature extraction process employs a joint optimization mechanism involving the feature generator, essay scorer, and topic discriminator. The feature generator aims to produce features that benefit the scorer while confusing the topic discriminator. The essay scorer predicts the essay score accurately, and the topic discriminator attempts to identify the topic source of the features. The loss function for the feature generator is defined as follows:

$$Loss_{\theta_f} = Loss_{\theta_y} - \alpha\, Loss_{\theta_d} \tag{2}$$

$Loss_{\theta_y}$ and $Loss_{\theta_d}$ represent the loss function of the composition evaluator and the topic discriminator. α refers to a hyperparameter to weigh the two objectives. The loss function of the topic discriminator is:

$$Loss_{\theta_d} = L_d(\theta_f, \theta_d) \tag{3}$$

$L_d(\theta_f, \theta_d)$ uses cross entropy loss to measure the interference degree of features generated by feature generator to topic discriminator. The loss function of the composition scorer is:

$$Loss_{\theta_y} = L_y(\theta_f, \theta_y) \tag{4}$$

$L_y(\theta_f, \theta_y)$ utilizes mean square error (MSE) to measure the deviation between the predicted and real scores. To realize joint optimization, the parameter updating rules are as follows:

$$\theta_f \leftarrow\ \theta_f - \mu \left(\frac{\partial L_y}{\partial \theta_f} - \alpha \frac{\partial L_d}{\partial \theta_f} \right) \tag{5}$$

$$\theta_y \leftarrow\ \theta_y - \mu \frac{\partial L_y}{\partial \theta_y} \tag{6}$$

$$\theta_d \leftarrow\ \theta_d - \mu \left(\alpha \frac{\partial L_d}{\partial \theta_d} \right) \tag{7}$$

μ is the learning rate. θ_f, θ_y and θ_d represent the parameters of feature generator, composition scorer, and topic discriminator, respectively.

Topic-related feature extraction stage

Although grammatical, lexical, and coherence features are extracted during the topic-independent stage, the composition's topic information has yet to be fully incorporated. Topic relevance plays a critical role in accurate scoring, especially in cross-topic tasks where aligning the essay content with the prompt is essential. The topic-

related feature extraction stage focuses on capturing features closely tied to the prompt from the essay text. By integrating topic information into the scoring framework via a neural network, the model improves its semantic understanding and ability to judge topic alignment. Hierarchical neural networks effectively capture multi-level semantic information, enhancing the model's overall grasp of essay semantics and better handling the complex relationship between prompts and compositions. Compared to traditional flat neural networks, hierarchical architectures preserve the text's structural hierarchy, reducing risks of information loss or misinterpretation. This strengthens the model's accuracy and reliability in topic-specific scoring. Text typically contains multiple semantic levels—from local words to sentences and overall discourse. By modeling these levels, hierarchical networks simultaneously attend to local and global information, deepening the model's understanding of topic-related content. Therefore, this stage employs a hierarchical neural network to extract topic-related information layer by layer, enabling the model to capture topic elements and their semantic connections within the essay. As illustrated in Fig. 4, this stage divides the model into shared and task-specific layers. The shared layer extracts general semantic features, while the task layer focuses on features specific to the particular scoring task.

(1) Shared layer.

The sharing layer extracts general semantic features through word embedding, word-level convolution, and attention pooling operations, ensuring broad applicability of the feature representations.

The word embedding layer encodes each word in the essay into a high-dimensional vector that captures its semantic and grammatical properties. This study uses a pre-trained BERT model for word vectorization. Specifically, let the essay $E = \{sent_1, sent_2, \cdots, sent_n\}$, where n denotes the number of sentences, and each sentence $sent_i = \{d_1, d_2, \cdots, d_m\}$, with mmm representing the number of words. After encoding with BERT, each word vector is represented as follows:

$$w_i = BERT^{\text{represent}(d_i)} \tag{8}$$

where represent denotes the encoding method applied to the word d_i. Leveraging BERT's pre-training capabilities, the word embedding layer effectively captures contextual semantic dependencies and lexical-level information.

In the word-level convolution layer, a one-dimensional CNN processes the word embeddings to extract sentence-level semantic features. Subsequently, attention pooling aggregates these features into a comprehensive sentence representation. Specifically, for each word w_i in a sentence, the convolution operation extracts its part-of-speech feature representation g_i:

$$g_i = f(W_g \cdot [w_i : w_{i+h_w-1}] + b_g) \tag{9}$$

The function f denotes a nonlinear activation function; W_g represents the convolution kernel weight matrix; b_g is the bias term; h_w refers to the convolution kernel size; and $[w_i : w_{i+h_w-1}]$ indicates the set of words within the current sliding window.

Next, the attention vector a_i and attention score v_i for each word are computed using the attention mechanism as follows:

$$a_i = \tanh(W_a \cdot g_i + b_a) \tag{10}$$

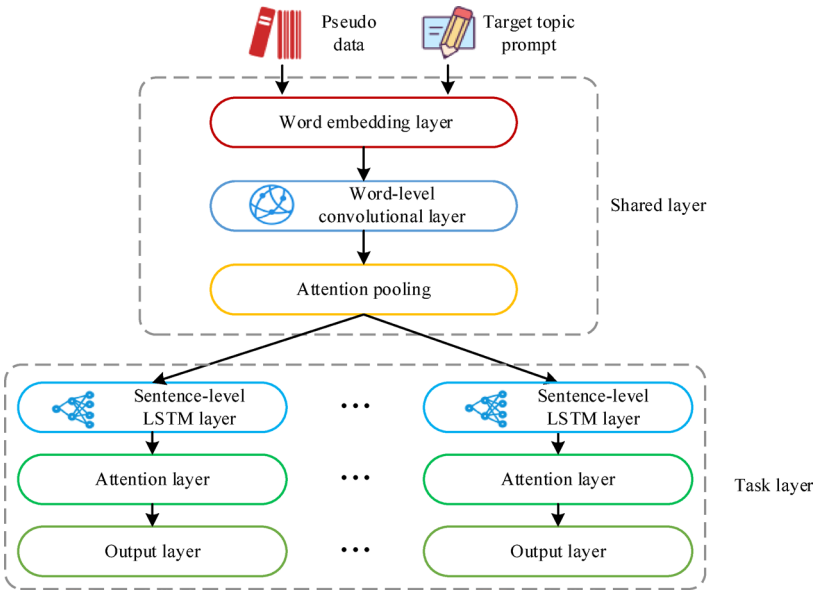


Fig. 4. Neural network structure for topic-related feature extraction.

$$v_i = \frac{e^{W_v \cdot a_i} \exp(W_v \cdot a_i)}{\sum \exp(W_v \cdot a_j)} \tag{11}$$

where W_a and W_v are trainable weight matrices and b_a is the bias vector. The sentence representation s is obtained via the weighted sum of all word features:

$$s = \sum v_i \cdot g_i \tag{12}$$

where s represents the final sentence-level semantic feature vector.

(2) Task layer.

The task layer performs feature modeling using a Bi-LSTM, sentence-level attention mechanism, and an output layer to complete feature extraction and topic-related scoring.

Sequence information is vital for semantic modeling in compositions. Compared to traditional LSTM, Bi-LSTM captures contextual information from both past and future states, allowing more comprehensive modeling of sentence semantics[26]. For each task j, the Bi-LSTM processes the input sentence representations s_t^j and outputs h_t^j, following these equations:
Input gate calculation:

$$i_t^j = \sigma\left(a_{i,t}^j + r_{i,t-1}^j + b_i^j\right) \tag{13}$$

$$a_{i,t}^j = W_i^j s_t^j \tag{14}$$

$$r_{i,t-1}^j = U_i^j h_{t-1}^j \tag{15}$$

Forget gate calculation:

$$f_t^j = \sigma\left(a_{f,t}^j + r_{f,t-1}^j + b_f^j\right) \tag{16}$$

$$a_{f,t}^j = W_f^j s_t^j \tag{17}$$

$$r_{f,t-1}^j = U_f^j h_{t-1}^j \tag{18}$$

Candidate unit state:

$$\tilde{c}_t^j = \tanh(a_{c,t}^j + r_{c,t-1}^j + b_c^j) \tag{19}$$

$$a_{c,t}^j = W_c^j s_t^j \tag{20}$$

$$r_{c,t-1}^j = U_c^j h_{t-1}^j \tag{21}$$

Unit state update:

$$c_t^j = i_t^j \odot \tilde{c}_t^j + f_t^j \odot c_{t-1}^j \tag{22}$$

Output gate:

$$o_t^j = \sigma\left(a_{o,t}^j + r_{o,t-1}^j + b_o^j\right) \tag{23}$$

$$a_{o,t}^j = W_o^j s_t^j \tag{24}$$

$$r_{o,t-1}^j = U_o^j h_{t-1}^j \tag{25}$$

Hidden state update:

$$h_t^j = o_t^j \odot \tanh\left(c_t^j\right) \tag{26}$$

where s_t^j represents the input sentence representation of the j-th task at time t. h_t^j refers to the corresponding output vector. Weight matrices W_i^j, W_f^j, W_c^j, W_o^j, U_i^j, U_f^j, U_c^j, U_o^j and offset vectors b_i^j, b_f^j, b_c^j and b_o^j are all model parameters, and σ is the activation function. $\odot$ represents element-level multiplication.

To strengthen task relevance, a sentence-level attention mechanism assigns weights to sentence features for each task j. The attention vector q_t^j and weight a_t^j are computed as:

$$q_t^j = \tanh(W_q^j h_t^j + b_q^j) \tag{27}$$

The attention weight is calculated as:

64

$$a_t^j = \frac{\exp(W_a^j \cdot q_i^j)}{\sum_k \exp(W_a^j \cdot q_i^j)} \tag{28}$$

The weighted sentence summary o^j is:

$$o^j = \sum_t a_t^j h_t^j \tag{29}$$

where W_q^j and W_a^j are training matrix parameters; b_q^j denotes the bias vector. q_t^j and a_t^j represent attention vector and attention weight respectively; o^j is the final sentence feature representation vector of the current task *j*.

The model further incorporates a cross-task attention mechanism to exploit semantic correlations across tasks in multi-task learning. The attention score u_i^j for the *i*-th feature in task *j* with respect to other task features $A_{-j,l}$ is calculated by:

$$u_i^j = \frac{\exp\left(score(o^j, A_{-j,i})\right)}{\sum_l \exp\left(score(o^j, A_{-j,l})\right)} \tag{30}$$

Cross-task information integrates via weighted sum:

$$p^j = \sum_i u_i^j A_{-j,i} \tag{31}$$

The final representation vector is formed by concatenating o^j and p^j:

$$z^j = [o^j; p^j] \tag{32}$$

where u_i^j represents the attention weight of the *i*-th feature in the *j*-th task. $A_{-j,i}$ denotes the feature set of other tasks. $score$ refers to the attention score function. p^j means the integrated cross-attention feature. Finally, o^j and p^j are spliced to form the final task feature vector z^j.

Finally, the task layer predicts essay scores through the output layer, which applies a sigmoid activation function to map the feature vector z^j to the range [0,1]:

$$\widehat{y}^j = sigmoid(W_y^j z^j + b_y^j) \tag{33}$$

where W_y^j and b_y^j are weights and biases, and $\widehat{y}^j$ is the predicted score for task *j*.

In summary, the proposed HFC-AES model achieves multi-level collaborative modeling of essay content and topic semantics through two stages: topic-independent and topic-related feature extraction and modeling. It fully integrates shallow textual features with deep semantic features and combines shared and task-specific layers. Moreover, the cross-task attention mechanism enhances the model's adaptability to semantic variations across topics, improving scoring accuracy. The next section evaluates the model's performance on cross-topic AES tasks using multiple public datasets. Results on scoring effectiveness, ablation studies, and practical applications demonstrate the model's effectiveness and usability.

Experimental design and performance evaluation

Dataset collection

This experiment uses the Automated Student Assessment Prize (ASAP) dataset, which contains a large number of English compositions primarily designed to evaluate students' writing proficiency. The ASAP dataset includes eight distinct topics, each corresponding to a subset of compositions labeled with overall scores. To protect privacy and reduce scoring bias, sensitive information such as specific names and locations in the compositions is anonymized by replacing them uniformly with the placeholder "@entity." Text preprocessing also involves removing non-standard characters and special symbols, and converting all letters to lowercase to minimize noise during model training. For text segmentation, the NLTK toolkit is employed to perform sentence- and word-level tokenization, supporting subsequent hierarchical semantic modeling. Given the differing scoring intervals across topics, all original scores are normalized to the range [0,1] to ensure fairness and comparability in cross-topic scoring. To prevent data leakage in cross-topic experiments, prompt words and keywords explicitly related to the composition topics are removed. This step avoids the model "cheating" by learning the prompt content directly and helps ensure the scoring model's generalization truly reflects writing quality. The dataset is split into training and test sets in a 3:2 ratio. For each experiment, one topic serves as the test set while the remaining seven topics form the training set. The model performance is evaluated using 50% cross-validation.

Experimental environment and parameters setting

Table 3 lists the software, hardware, and development environment used in this experiment, along with key parameter settings. Model parameters are primarily determined using empirical rules and optimized based on validation set performance.

The Quadratic Weighted Kappa (QWK) is used as the evaluation index, which mainly measures the consistency between the model and the real rater, and considers the square penalty of the scoring deviation. The calculation equation of QWK is as follows:

Software and hardware environment/parameters	Configuration/value
Operating system	Window10
Processor	AMD Ryzen7
Display card	NVIDIA GeForce RTX 2060
Programming language	Python
DL framework	Tensorflow2.0
Optimization function	RMS-prop
Initial learning rate	0.005
Batch size	64
Weight initialization method	Xavier initialization
Dropout	0.5
Epochs	30
Number of CNN output channels	128
Number of hidden units in LSTM	256
Convolutional kernel size	3, 4, 5 (parallel convolution)
Activation function	ReLU

Table 3. Experimental environment and parameter settings.

$$QWK = 1 - \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}O_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}E_{ij}} \tag{34}$$

N refers to the total number of rating levels. O_{ij} is the number of the actual score i and the predicted score j. E_{ij} represents the expected frequency calculated according to the rater's score distribution and the predicted score distribution. w_{ij} is the weight based on the difference of scores, and the square difference weight is usually adopted:

$$w_{ij} = \frac{(i-j)^2}{(N-1)^2} \tag{35}$$

The value range of QWK is [−1,1], where 1 means complete consistency, 0 means random consistency, and a negative value means poor consistency.

Performance evaluation

(1) Comparison of model scoring results.

To evaluate the effectiveness of the HFC-AES model, this study compared it with five established AES models: the Hierarchical Attention Model (Hi-att)[27], Co-attention[28], Temporary Deep Neural Network (TDNN)[29], Siamese Enhanced Deep Neural Network (SEDNN)[30], and Cross-Task Scoring Model (CTS)[31]. Hi-att and Co-attention target single-topic scoring, while TDNN, SEDNN, and CTS address cross-topic scoring. To further strengthen the results, two additional mainstream Transformer-based models were included: the BERT-based AES model (BERT-AES) and the GPT-based generative AES model (GPT-AES). BERT-AES uses multi-task fine-tuning to emphasize sentence-level semantic consistency, while GPT-AES incorporates prompt information and generates scoring predictions by producing rating sequences. Figure 5 presents the QWK results of all models in cross-topic evaluation.

Figure 5 shows that in cross-topic AES, single-topic models such as Hi-att and Co-attention perform worse than cross-topic AES models. Among all models, HFC-AES achieves the highest performance, with an average QWK of 0.856, surpassing other cross-topic approaches and confirming its effectiveness. GPT-AES and BERT-AES achieve mean QWK scores of 0.810 and 0.791, respectively, outperforming traditional RNN and CNN models but still falling short of HFC-AES. These results indicate that while Transformer architectures excel at feature extraction, HFC-AES gains further advantages through structural optimization and cross-task modeling. Its multi-level semantic modeling and accurate topic-related feature extraction enhance the alignment between compositions and prompts. By integrating the task and shared layers with a cross-task attention mechanism, the model effectively handles semantic differences between topics, improving cross-topic scoring accuracy. Furthermore, the joint optimization mechanism enhances robustness and scoring consistency, enabling HFC-AES to achieve superior performance. To investigate the reasons behind HFC-AES's performance advantage, a comparison was conducted with two-stage cross-topic AES models, TDNN and SEDNN, focusing on QWK results for pre-scoring compositions in the topic-independent stage, as shown in Fig. 6.

Figure 6 shows that the HFC-AES model achieved a higher QWK than TDNN and SEDNN in pre-scoring compositions during the topic-independent stage. Its average QWK across eight prompts was 0.769, outperforming TDNN (0.546) and SEDNN (0.681). These results confirm that the first stage of HFC-AES is critical for improving pre-scoring quality. Unlike the comparison models, HFC-AES better incorporates prompt

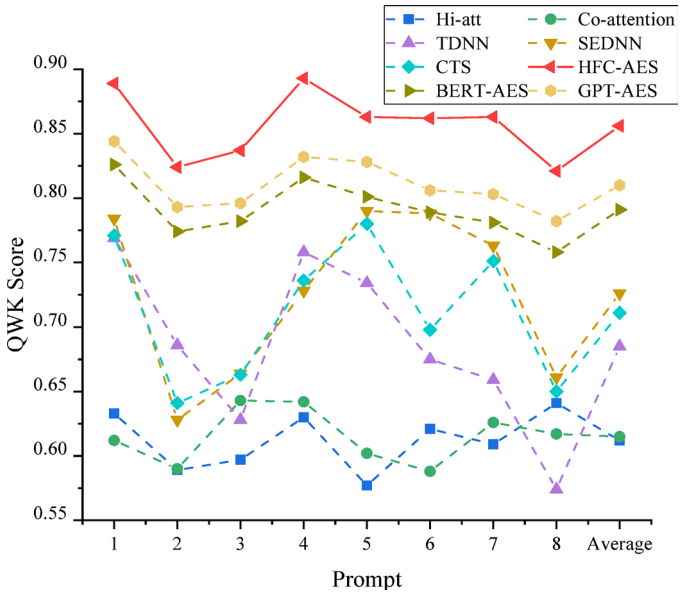


Fig. 5. Comparison of QWK values of various models in cross-topic scenes.

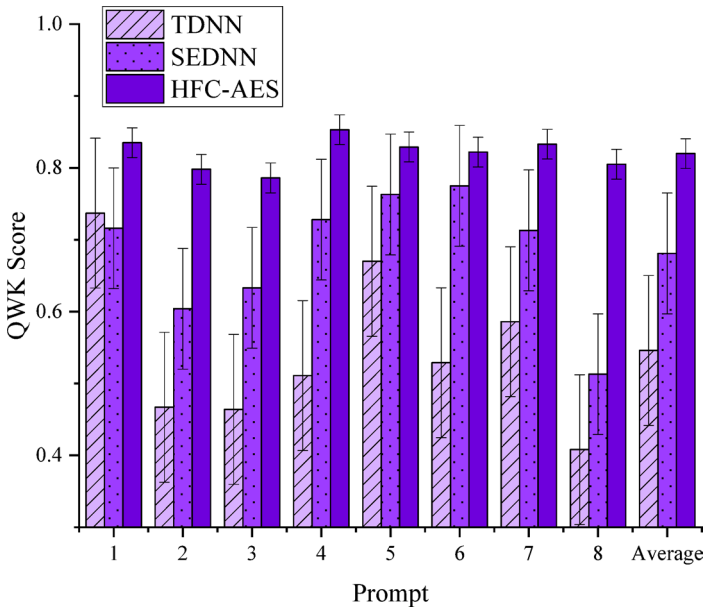


Fig. 6. QWK Comparison of Three Cross-Topic Models in the Topic-Independent Stage.

information, leading to stronger cross-topic performance. In the topic-related stage, HFC-AES again performed best. Its hierarchical neural network structure effectively captured the complex semantic relationships between compositions and prompts. By extracting general semantic features in the shared layer and emphasizing topic-relevant information in the task layer, the model improved topic alignment and scoring accuracy. The Bi-LSTM and attention mechanisms further enhanced contextual modeling and feature extraction, enabling superior results in cross-topic scoring.

To assess the model's generalization across different writing types and datasets, supplementary experiments were conducted on the publicly available TOEFL11 and International Corpus of Learner English (ICLE) datasets. TOEFL11 contains compositions from 11 groups of non-native English speakers, and ICLE consists of academic texts from multiple non-English-speaking countries. To ensure robust and unbiased evaluation, tenfold cross-validation with repeated verification was applied to each dataset to minimize overfitting. Figure 7 reports the QWK scores of all models.

Figure 7 presents the QWK evaluation results of all models on the TOEFL11 and ICLE datasets. HFC-AES consistently outperformed the comparison models, achieving a QWK of 0.852 on TOEFL11 and demonstrating

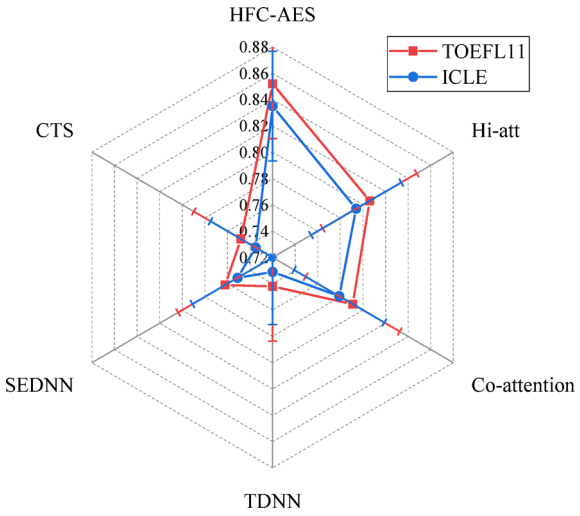


Fig. 7. The QWK evaluation results of each model on TOEFL11 and ICLE datasets.

Sample ID	S1	S2	S3
Prompt	Should governments invest more in public transportation?	Is it better to live in a city or a rural area?	Should college education be free?
Excerpt	"While some may argue that cars offer greater freedom, I firmly believe that investing in public transportation leads to a greener, more efficient society. Isn't it better to reduce traffic jams and pollution?"	"Living in a city has many benefits. You can go to museums, restaurants, or hospitals easily. Everything is close."	"College education should be free so that everyone can access knowledge. However, the government needs a sustainable plan to fund it."
Human Score	4.5	3.0	4.8
Model Score	3.7	4.1	4.5
Score Gap	−0.8	+1.1	−0.3
Analysis of Deviation	The model misinterpreted the rhetorical question and contrastive reasoning, underestimating the strength of the author's stance and giving a lower score.	Despite the fluent language and clear structure, the essay lacked critical analysis. The model over-weighted surface fluency and failed to penalize the lack of argument depth, resulting in an inflated score.	Minor spelling and grammar issues were over-penalized by the model, leading to a slight underestimation of the overall quality.

Table 4. Sample analysis of the differences between model and human scoring.

strong adaptability to non-native English writing. In contrast, traditional models such as Hi-att and Co-attention delivered lower accuracy and weaker consistency, highlighting the superiority of HFC-AES in handling compositions from diverse linguistic backgrounds. These results confirm the model's robust generalization capability, particularly in scoring tasks involving non-native writers.

To further examine performance differences in real scoring scenarios, a qualitative error analysis was conducted on representative samples. Table 4 lists three compositions with their prompts, human-assigned scores, model predictions, and explanations for scoring discrepancies.

The discrepancies primarily stem from the model's limited capacity to interpret rhetorical devices, nuanced tone, and deeper reasoning. Essays with complex structures or implicit meaning were more prone to misjudgment. This highlights an area for improvement: integrating advanced discourse reasoning modules or fine-tuning pre-trained language models at the discourse level to enhance recognition of implicit semantics and rhetorical strategies.

(2) Ablation experiments.

Systematic ablation experiments were designed to evaluate the contributions of different features and mechanisms in the HFC-AES model to scoring performance. Two categories were examined: feature-level ablation (discourse structure, topic-independent features, and topic-related features) and mechanism-level ablation (e.g., attention mechanisms). Each feature or mechanism was removed individually and in combination to assess its impact on performance.

For feature-level ablation, the following configurations were tested: Structural features: discourse structure removed; Topic-independent features: all topic-independent features removed; Topic-related features: all topic-related features removed; Structural + topic-independent features: both discourse structure and topic-independent features removed. The results are presented in Fig. 8.

Figure 8 shows that each feature type contributes differently to model performance. Removing discourse structure features reduces the average QWK to 0.827, confirming their value in capturing overall organization and logical coherence. The impact is greater when topic-independent features are excluded, with the QWK dropping to 0.765, highlighting the importance of basic linguistic indicators such as vocabulary and syntax

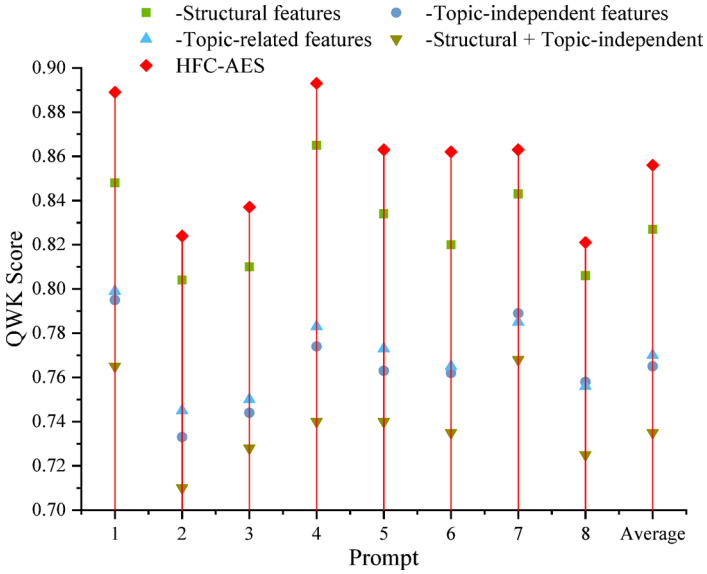


Fig. 8. Results of the feature-level ablation experiment.

in modeling text complexity and writing style. Eliminating topic-related features results in a similar decline, with the QWK decreasing to 0.770, underscoring their role in assessing how well a composition aligns with its prompt. The largest drop occurs when both discourse structure and topic-independent features are removed, with the QWK falling to 0.735. This demonstrates that each feature type supports the others: removing one weakens performance, and removing both amplifies the effect. For example, tasks using Prompts 1–4 show the steepest degradation under this combination. Compared to the full HFC-AES model, which achieves an average QWK of 0.856, this ablation produces a 0.121 loss, emphasizing the need for diverse feature inputs. Overall, these results confirm that discourse structure, basic linguistic features, and topic-semantic matching work together to enable accurate scoring.

The next step evaluates the role of the attention mechanism. Three configurations are tested: Attention: removal of the attention mechanism; Attention + topic-related features: removal of both attention and topic-related features; Structural + topic-independent + attention: removal of discourse structure, topic-independent features, and attention. The outcomes are shown in Fig. 9.

Figure 9 shows that removing the attention mechanism alone lowers the model's average QWK to 0.818, only a slight decrease from the complete model. This indicates that the attention mechanism, though secondary, still contributes meaningfully, especially in handling compositions with complex structures or inter-sentence relationships. Its impact becomes more pronounced when combined with other features. For instance, removing both the attention mechanism and topic-related features reduces the average QWK to 0.792, a much larger drop than removing either alone, highlighting their interdependence. The attention mechanism enhances the modeling of semantic alignment between compositions and prompts, ensuring accurate topic matching. When discourse structure, topic-independent features, and the attention mechanism are all removed, the QWK further falls to 0.778, resulting in a loss of 0.078 compared with the full model (0.856). Performance declines are especially evident in Prompts 4 and 7, which require high-level semantic abstraction and contextual reasoning. Prompt 4 involves balancing ethical concerns and scientific progress, often using metaphors, concessions, and dual-argument structures that demand strong semantic and structural comprehension. Prompt 7 calls for critical analysis of social phenomena and technological impacts, with frequent logical reasoning and subjective expression. Without topic-related features and the attention mechanism, the model struggles to determine whether a composition stays focused on the prompt, reducing scoring consistency.

Overall, the attention mechanism is not the sole determinant of performance, but its synergy with semantic features significantly improves topic understanding and contextual semantic capture, making it a vital component in cross-topic scoring. To further evaluate the HFC-AES model under different feature configurations, additional ablation experiments were conducted on shallow learning (SL) and DL features. By removing each type separately, the SL-only and DL-only models were obtained, and their effects on cross-topic composition scoring are presented in Fig. 10.

Figure 10 shows that the overall scoring performance of the HFC-AES model drops when either shallow learning (SL) or DL features are removed. The average QWK decreases to 0.821 without SL features and to 0.812 without DL features, both lower than the complete model's 0.856. This indicates that both feature types are essential for accurate scoring. Shallow features, such as word frequency, sentence length, and syntactic diversity, provide intuitive and stable indicators of linguistic complexity and writing style, helping the model assess basic language quality. DL features, by contrast, capture richer semantic representations and contextual relationships through neural networks, improving the model's ability to evaluate semantic coherence and logical flow. Together, they form a complementary multi-level semantic representation of each composition, making their joint use a key factor in achieving high-precision scoring. Figures 8 and 9 further reveal that among all features, topic-

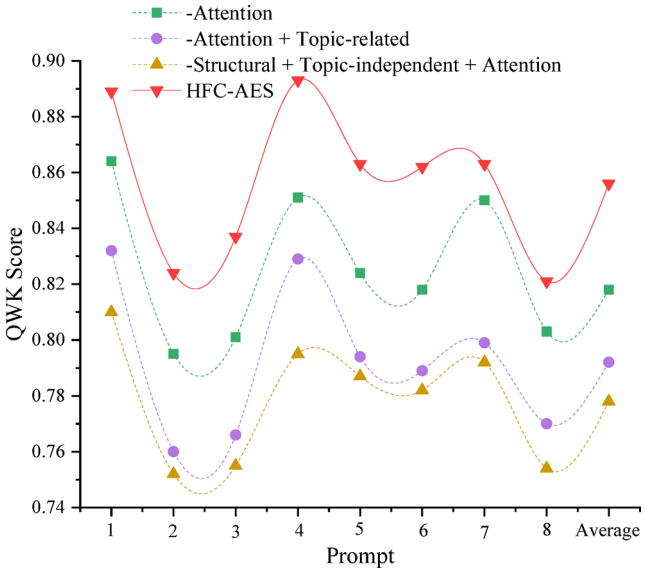


Fig. 9. Results of mechanism-level ablation experiments.

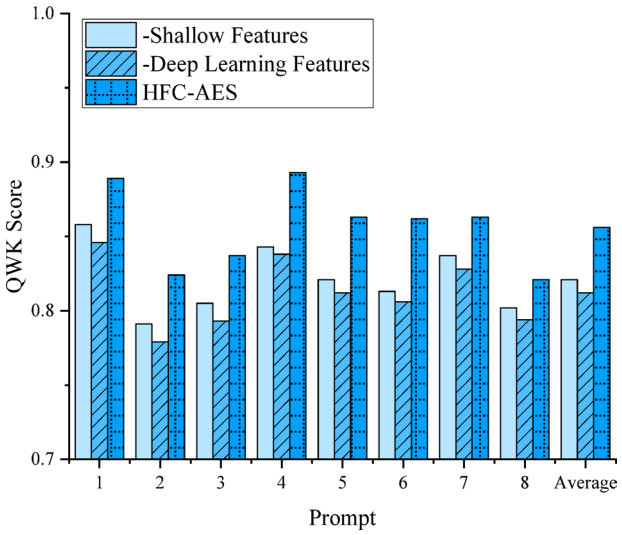


Fig. 10. Experiments on ablation with different feature types.

related features have the greatest impact. Removing them lowers the average QWK from 0.856 to 0.827, with marked performance drops on Prompts 4 and 7. These features directly model the semantic alignment between compositions and their prompts—an especially challenging aspect of cross-topic scoring—allowing the model to more accurately judge topical relevance. In HFC-AES, this is accomplished through bidirectional LSTM and attention mechanisms in the task layer, substantially improving scoring consistency and accuracy across topics.

To enhance interpretability, attention weight distributions and feature importance were further analyzed to provide deeper insights into the model's decision-making. Table 5 presents the attention weights assigned to specific features across different scoring dimensions.

To further reveal how the model assigned attention weights within specific texts, the attention distribution for the sentence "College education should be free so that everyone can access knowledge. However, the government needs a sustainable plan to fund it." was visualized. The visualization is shown in Fig. 11.

The intra-sentence attention distribution reveals that the model assigns higher weights to phrases like "sustainable plan" and "government needs," indicating its focus on the practical feasibility issues raised in the essay. This focus is crucial for evaluating the logical completeness of argumentative writing. However, the attention on the phrase "should be free so that everyone can access knowledge" is more dispersed, reflecting the model's lower sensitivity to idealistic or emotional expressions compared to factual statements. This difference further highlights the model's limitation in handling subjective stances and shifts in tone. This word-level visualization based on attention aids in explaining specific scoring discrepancies and represents a promising

Prompt	Grammar	Semantics	Vocabulary usage	Organization structure	Total score
1	0.22	0.19	0.18	0.41	0.89
2	0.20	0.21	0.22	0.37	0.82
3	0.25	0.20	0.17	0.38	0.83
4	0.24	0.21	0.19	0.36	0.89
5	0.23	0.22	0.18	0.37	0.86
6	0.21	0.23	0.19	0.37	0.86
7	0.26	0.19	0.18	0.37	0.87
8	0.22	0.20	0.20	0.38	0.82

Table 5. Interpretability analysis of attention mechanisms.

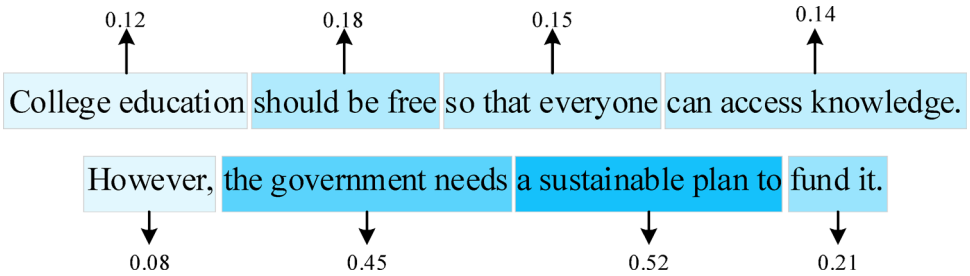


Fig. 11. Visualization of attention weight distribution for the example sentence.

Feature	Importance scoring
Grammar	0.25
Semantics	0.22
Vocabulary usage	0.18
Organization structure	0.20
Text structure	0.15

Table 6. Feature importance assessment.

direction for improving model interpretability. The feature importance assessment quantifies the contribution of each feature to the scoring decisions. The results are presented in Table 6.

Tables 5 and 6 reveal that, in the interpretability analysis of the attention mechanism, the model places greater emphasis on organizational structure during scoring. This suggests that the HFC-AES model prioritizes logical coherence and structural quality when evaluating compositions. Regarding feature importance, grammatical and semantic features hold significant weight, underscoring their critical role in determining final scores. In contrast, discourse structure shows relatively lower importance, possibly due to its reduced influence in certain composition types. These findings indicate that the model's scoring decisions largely depend on grammar and semantic quality, while its attention to organizational structure supports effective assessment of coherence and logical consistency.

(3) Influence of the cross-attention mechanism on the scoring model.

The HFC-AES model incorporates a cross-attention mechanism to evaluate both overall composition quality and specific scoring dimensions, including semantics, grammar, vocabulary usage, and organizational structure. The impact of this mechanism on overall scoring performance is assessed, with results presented in Fig. 12.

Figure 12 illustrates how the HFC-AES model dynamically adjusts the weight assigned to various features for different scoring tasks after incorporating the cross-attention mechanism. This adjustment notably enhances scoring accuracy and consistency. For the overall composition score, the cross-attention mechanism allocates weights thoughtfully across scoring dimensions. Semantic and grammatical features receive weights of 0.159 and 0.168, respectively, highlighting the model's emphasis on semantic coherence and grammatical accuracy—aligning well with human scoring criteria. Vocabulary usage is weighted at 0.133, reflecting its importance in scoring, particularly in terms of diversity and precision. When predicting the organizational structure score, the mechanism concentrates the majority of the weight (0.173) on organizational features, significantly down-weighting other aspects. This selective focus enables the model to prioritize key features relevant to specific scoring tasks, thereby improving the accuracy of individual dimension scores. In summary, the cross-attention

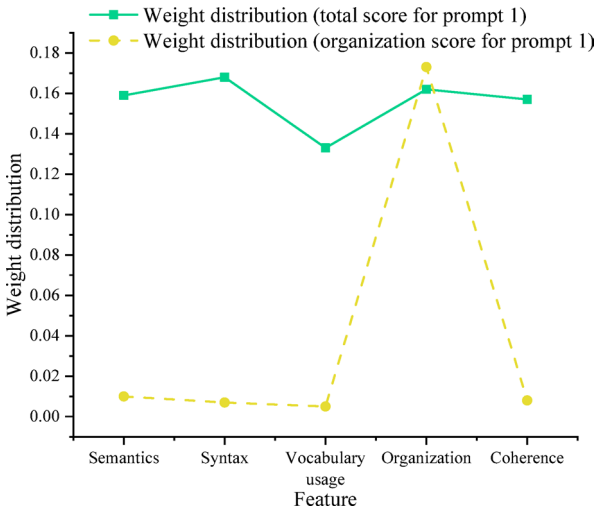


Fig. 12. Feature Weight Distribution for Topic 1 in Predicting Overall and Individual Scores.

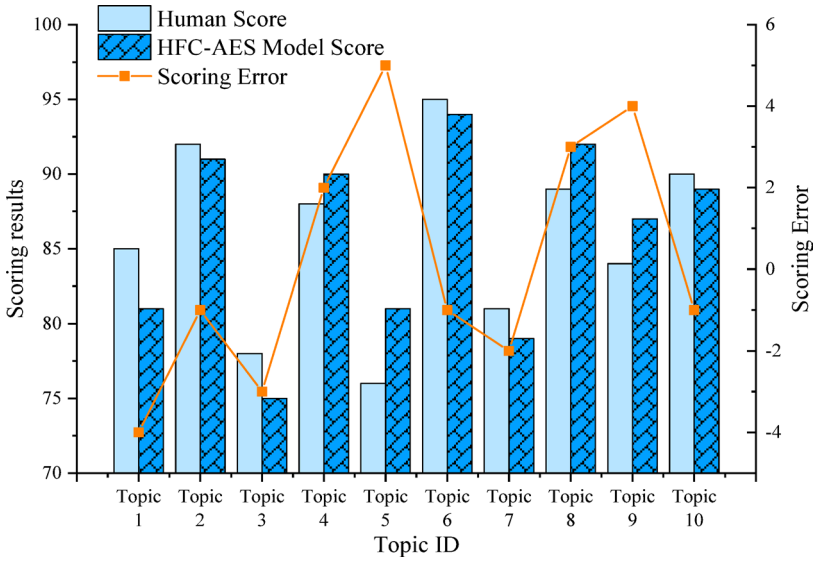


Fig. 13. Comparison of Human Score and HFC-AES model score.

mechanism allows the model to flexibly reweight features depending on the scoring task, enhancing the precision and rationale of composition evaluation.

(3) Practical application of the HFC-AES model.

To evaluate the practical utility of the HFC-AES model, its automatic scoring results are compared with human evaluator scores. This comparison helps verify the model's accuracy and feasibility in real-world settings. Figure 13 presents this comparison, with scores normalized to a maximum of 100 points.

Figure 13 shows that the differences between the HFC-AES model's scores and human ratings are minimal, with most errors falling within a 3-point range. This indicates that the HFC-AES model closely approximates human scoring standards, making it well-suited for practical automatic composition scoring tasks. While minor discrepancies may occur in individual cases, the model generally performs reliably, effectively supporting automatic scoring needs in real-world applications and demonstrating strong feasibility and potential.

To further assess the model's practical applicability, its processing time was evaluated by measuring the average scoring time per composition. All experiments were conducted on a consistent hardware and software platform. Comparative models included HFC-AES, TDNN, SEDNN, CTS, BERT-AES, and GPT-AES. The results are summarized in Table 7.

Table 7 shows that the processing time of the HFC-AES model is slightly longer than that of traditional DL models. This is primarily due to its integration of shallow features, deep semantic representations, discourse structure information, and a multi-module collaborative training mechanism. However, its processing time

Models	TDNN	SEDNN	CTS	BERT-AES	GPT-AES	HFC-AES
Average time consumption (s)	0.19	0.25	0.32	0.98	1.22	0.87

Table 7. Comparison of processing time in model rating stage (Unit: seconds/article).

remains significantly shorter than that of GPT-AES and BERT-AES, which rely on large-scale pre-trained models and suffer from considerable time bottlenecks in practical applications due to their vast parameter sizes and complex inference procedures. Overall, HFC-AES achieves a strong balance between high scoring accuracy and acceptable processing efficiency, making it well-suited for scenarios that demand precise grading. In practical educational settings, this means the HFC-AES model can score approximately 69 essays per minute. For instance, in a medium-sized high school where 3,000 essays need to be graded in a single exam, the model can complete the task within 45 min. This level of performance offers a feasible and effective solution for classroom assessments, online writing platforms, and large-scale standardized testing.

Discussion

In summary, the proposed HFC-AES model integrates shallow textual features with DL representations in a two-stage framework that includes both topic-independent and topic-related feature extraction and modeling. This design significantly improves scoring consistency and robustness compared to existing approaches. For example, Li et al. (2023) developed an AES method that combined multi-scale features with Sentence-BERT embeddings and shallow linguistic and topic-related features, achieving a QWK of 0.793[32]. Wang (2023) extracted semantic features via CNN and LSTM and topic features through TF-IDF, which resulted in a neural network-based AES model with a QWK of 0.816[33]. Dhini et al. (2023) proposed an AES model based on semantic and keyword similarity using Sentence Transformers; by incorporating multilingual Paraphrase-Multilingual-MiniLM-L12-V2 and DistilBERT-Base-Multilingual-Cased-V1 models, their approach improved evaluation scores by 0.2 points[34]. In contrast, this model enhances the understanding of composition content and semantics and strengthens the robustness and adaptability of topic information through a cross-task attention mechanism. Consequently, it offers a more comprehensive and effective technical solution for intelligent evaluation in English language teaching.

In practical applications, computational efficiency is crucial for automated scoring systems deployed at scale. This study evaluates the HFC-AES model's performance in processing thousands of essays in near real-time. On a single GPU machine, the model achieves an inference throughput of approximately 200 compositions per minute, satisfying the demands of most online education platforms. To increase throughput further, distributed computing and data parallelism can be employed to distribute scoring tasks across multiple servers for near-linear acceleration. Additionally, asynchronous batch processing can substantially improve overall system capacity while maintaining scoring latency within seconds. These features meet the low-latency, high-concurrency requirements of large-scale educational environments. To address scenarios with limited computing resources, lightweight model alternatives are explored. Recent advances in DL have produced compressed pretrained models like DistilBERT and TinyBERT, which maintain strong semantic understanding while greatly reducing parameter counts and computational overhead. These distilled models can be efficiently deployed on edge devices or resource-constrained classroom settings. By integrating the HFC-AES multi-stage feature fusion strategy with these lightweight models as substitutes for deep semantic extractors, the system retains high scoring accuracy while lowering latency and computational costs. This makes the scoring system more practical for large-scale real-world education. Future work will focus on systematically evaluating and optimizing these lightweight versions to further enhance the model's applicability in educational contexts.

Although the HFC-AES model demonstrates strong efficiency and scoring consistency, deploying automated scoring systems raises important ethical concerns. The model may place excessive emphasis on surface-level features like language fluency and syntactic accuracy, potentially undervaluing creativity and critical thinking. This could lead to a bias favoring style over substance. Moreover, compositions reflecting significant differences in gender, cultural background, or language variants risk being unfairly scored due to imbalances in the training data, which can introduce algorithmic bias. To address these issues, future work should focus on enhancing training mechanisms to promote diversity, inclusiveness, and fairness—for example, by integrating fairness correction modules and improving the recognition and understanding of non-standard linguistic expressions. Additionally, quality control should be enforced through manual audits and human-in-the-loop processes to ensure that automated systems complement rather than fully replace human evaluators, thus mitigating risks of misuse or overreliance on technology.

Conclusion

Research contribution

This study proposes a cross-topic automatic English composition scoring model, HFC-AES, which integrates DL features with shallow text features through both topic-independent and topic-related feature extraction. The model aims to enhance the accuracy and reliability of English composition scoring, thereby improving the evaluation of English language teaching effectiveness. Experimental results validate the model's effectiveness, leading to the following conclusions: (1) In cross-topic composition scoring, HFC-AES achieves the best performance with an average QWK of 0.856, surpassing other cross-topic models. Its pre-scoring QWK also outperforms TDNN and SEDNN, indicating a key role in improving pseudo-data quality. (2) Ablation experiments reveal that discourse structure features significantly impact argumentative composition scoring, with a 13.32% drop

in score when these features are removed. In contrast, removing the attention mechanism has a relatively minor effect. Overall, the HFC-AES model excels when both text structure features and the attention mechanism are included, underscoring their importance in cross-topic scoring. (3) The introduction of the cross-attention mechanism substantially enhances scoring performance, aligning model predictions more closely with human judgments. By combining DL and shallow text features, HFC-AES demonstrates clear advantages in cross-topic English composition scoring, providing robust technical support and a practical foundation for evaluating English teaching effectiveness. The introduction of HFC-AES not only represents a performance breakthrough in automated scoring but also opens new possibilities for fairness, consistency, and efficiency in English teaching assessment. Traditional manual grading suffers from subjectivity and scalability limitations, while HFC-AES shows strong potential to reshape educational assessment through data-driven approaches. Its adaptability and scalability in scoring cross-topic, multi-task, and linguistically complex essays highlight its broad applicability. Moreover, the model's interpretability modules offer transparent, visualized evidence for teachers, facilitating applications in instructional feedback, writing assistance, and educational diagnostics. This paves the way for future "human–machine collaborative" educational assessment.

Future works and research limitations

Although the HFC-AES model has demonstrated strong performance in cross-topic English composition scoring, several limitations remain. First, its generalization to different genres—such as narrative, reflective, or creative writing—needs improvement. These genres often feature nonlinear structures, subjective experiences, and emotional expression, which may not align well with the model's current focus on structure and logical coherence. Future work could explore genre-adaptive modules or multi-genre scoring branches to enhance flexibility and applicability. Second, while preliminary consideration has been given to multilingual extension, the model's potential cultural biases, linguistic preferences, and adaptation to geographically diverse corpora have not been systematically examined. Future research should emphasize fairness by integrating sociolinguistic and educational assessment theories to evaluate scoring consistency across students from varied socioeconomic and educational backgrounds. Strategies such as balanced training data, fairness-aware regularization, and bias mitigation techniques will be critical for reducing potential disparities. Third, despite its superior performance, the model's computational demands remain relatively high, posing challenges for deployment in resource-constrained school settings or large-scale online examination platforms. Subsequent efforts will focus on model compression, knowledge distillation, and the development of lightweight, edge-compatible versions to improve efficiency and practical usability. Finally, to address trust and transparency issues inherent in automated scoring, future work may explore "human–machine hybrid scoring systems" where model outputs serve as decision-support tools or initial screening aids for human graders. This approach could safeguard scoring quality while enhancing efficiency and feedback speed, facilitating deeper integration of AI technologies into educational practice and advancing the intelligent transformation of composition assessment.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author Jinhai Wang on reasonable request via e-mail coolwang001@zua.edu.cn.

Received: 24 December 2024; Accepted: 30 July 2025
Published online: 07 August 2025

References

1. Kostka, I. & Toncelli, R. Exploring applications of ChatGPT to english Language teaching: Opportunities, challenges, and recommendations. *Tesl-Ej* **27** (3), n3 (2023).
2. Zhao, X. Leveraging artificial intelligence (AI) technology for english writing: Introducing wordtune as a digital writing assistant for EFL writers. *RELC J.* **54** (3), 890–894 (2023).
3. Sharadgah, T. A. & Sa'di, R. A. A systematic review of research on the use of artificial intelligence in english Language teaching and learning (2015–2021): What are the current effects? *J. Inform. Technol. Educ. Res.* **95**(3), 215–228 (2022).
4. Chan, K. K. Y., Bond, T. & Yan, Z. Application of an automated essay scoring engine to english writing assessment using many-facet Rasch measurement. *Lang. Test.* **40** (1), 61–85 (2023).
5. Ramesh, D. & Sanampudi, S. K. An automated essay scoring systems: A systematic literature review. *Artif. Intell. Rev.* **55** (3), 2495–2527 (2022).
6. Lim, C. T. et al. A comprehensive review of automated essay scoring (AES) research and development. *Pertanika J. Sci. Technol.* **29** (3), 1875–1899 (2021).
7. Mizumoto, A. & Eguchi, M. Exploring the potential of using an AI Language model for automated essay scoring. *Res. Methods Appl. Linguistics* **2** (2), 100050 (2023).
8. Darwish, S. M., Ali, R. A. & Elzoghabi, A. A. An automated english essay scoring engine based on neutrosophic ontology for electronic education systems. *Appl. Sci.* **13** (15), 8601 (2023).
9. Uto, M. A review of deep-neural automated essay scoring models. *Behaviormetrika* **48** (2), 459–484 (2021).
10. Zhang, F. Design and application of an automatic scoring system for english composition based on artificial intelligence Technology. *Int. J. Adv. Comput. Sci. Appl.* **14**(8), 195–205 (2023).
11. Wang, Z. Computer-assisted EFL writing and evaluations based on artificial intelligence: A case from a college reading and writing course. *Libr. Hi Tech.* **40** (1), 80–97 (2022).
12. Phoophuangpairoj, R. & Pipattarasakul, P. Preliminary indicators of EFL essay writing for teachers' feedback using automatic text Analysis. *Int. J. Educ. Methodol.* **8** (1), 55–68 (2022).
13. Susanti, M. N. I., Ramadhan, A., Warnars, H. & L H Automatic essay exam scoring system: A systematic literature review. *Procedia Comput. Sci.* **216**, 531 (2023).
14. Li, J. & Huang, J. The impact of essay organization and overall quality on the holistic scoring of EFL writing: Perspectives from classroom english teachers and National writing raters. *Assess. Writ.* **51**, 100604 (2022).
15. Lim, K. T., Song, J. & Park, J. Neural automated writing evaluation for Korean L2 writing. *Nat. Lang. Eng.* **29** (5), 1341–1363 (2023).

16. Bai, X. & Stede, M. A survey of current machine learning approaches to student free-text evaluation for intelligent tutoring. *Int. J. Artif. Intell. Educ.* **33** (4), 992–1030 (2023).
17. Gao, R. et al. Automatic assessment of text-based responses in post-secondary education: A systematic review. *Comput. Educ. Artif. Intell.* **6**(000), 100206 (2024).
18. Hossain, M. Z. & Goyal, S. Advancements in natural Language processing: Leveraging transformer models for multilingual text Generation. *Pac. J. Adv. Eng. Innovations.* **1** (1), 4–12 (2024).
19. Li, Y. Research on bilingual translation sentence similarity detection method based on Cross-Linguistic Mapping. *Int. J. High Speed Electron. Syst.* **25**, 2540537 (2025).
20. Wang, D., Su, J. & Yu, H. Feature extraction and analysis of natural Language processing for deep learning english Language. *IEEE Access.* **8**, 46335–46345 (2020).
21. Maamuujav, U. Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays. *Assess. Writ.* **49**, 100540 (2021).
22. Crossley, S. A. Linguistic features in writing quality and development: An overview. *J. Writ. Res.* **11** (3), 415–443 (2020).
23. Lu, S. et al. Multiscale feature extraction and fusion of image and text in VQA. *Int. J. Comput. Intell. Syst.* **16** (1), 54 (2023).
24. Geluso, J. Grammatical and functional characteristics of preposition-based phrase frames in english argumentative essays by L1 english and Spanish speakers. *J. Engl. Acad. Purp.* **55**, 101072 (2022).
25. Goh, T. T., Sun, H. & Yang, B. Microfeatures influencing writing quality: The case of Chinese students' SAT essays. *Comput. Assist. Lang. Learn.* **33** (4), 455–481 (2020).
26. Yanfi, Y. et al. Multi-head attention based bidirectional LSTM for Spelling error detection in the indonesian language. *IEEE Access* **12**, 188560–188571 (2024).
27. Dowell, N. & Kovanovic, V. Modeling educational discourse with natural Language processing. *Education* **64**, 82 (2022).
28. Zhou, H. et al. CANet: Co-attention network for RGB-D semantic segmentation. *Pattern Recogn.* **124**, 108468 (2022).
29. Ridley, R. et al. Prompt agnostic essay scorer: A domain generalization approach to cross-prompt automated essay scoring. arXiv preprint arXiv:2008.01441 (2020).
30. Li, X., Chen, M. & Nie, J. Y. Shared and enhanced deep neural network model for cross-prompt automated essay scoring. *Knowl. Based Syst.* **210**, 106491 (2020).
31. Do, H., Kim, Y. & Lee, G. G. Prompt-and trait relation-aware cross-prompt essay trait scoring. arXiv preprint arXiv:2305.16826 (2023).
32. Li, F. et al. Automatic essay scoring method based on multi-scale features. *Appl. Sci.* **13** (11), 6775 (2023).
33. Wang, J. A study of scoring english tests using an automatic scoring model incorporating Semantics. *Autom. Control Comput. Sci.* **57** (5), 514–522 (2023).
34. Dhini, B. F. et al. Automatic essay scoring for discussion forum in online learning based on semantic and keyword similarities. *Asian Assoc. Open. Universities J.* **18** (3), 262–278 (2023).

Additional information

Correspondence and requests for materials should be address

HANYU JINYI LIANGCI GOUSHI JINGXI FANCHOUHUA DE RENZHI YANJIU

汉语近义量词构式精细范畴化的认知研究

吉林大学出版社

图书在版编目（CIP）数据

汉语近义量词构式精细范畴化的认知研究 /
. -- 长春 : 吉林大学出版社, 2023.10
ISBN 978-7-5768-2220-5

Ⅰ. ①汉… Ⅱ. ① Ⅲ. ①汉语—同义词
—数量词—研究 Ⅳ. ①H146.2

中国国家版本馆CIP数据核字(2023)第193854号

书　　名：汉语近义量词构式精细范畴化的认知研究
HANYU JINYI LIANGCI GOUSHI JINGXI FANCHOUHUA DE RENZHI YANJIU

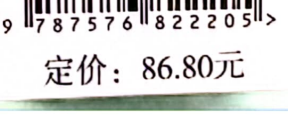
作　　者：
策划编辑：付晶淼
责任编辑：付晶淼
责任校对：李婷婷
装帧设计：雅硕图文
出版发行：吉林大学出版社
社　　址：长春市人民大街4059号
邮政编码：130021
发行电话：0431-89580036/58
网　　址：http://www.jlup.com.cn
电子邮箱：jldxcbs@sina.com
印　　刷：长春市中海彩印厂
开　　本：787mm × 1092mm　1/16
印　　张：16.75
字　　数：320千字
版　　次：2023年10月　第1版
印　　次：2023年10月　第1次
书　　号：ISBN 978-7-5768-2220-5
定　　价：86.80元

目 录

HANYU JINYI LIANGCI GOUSHI JINGXI FANCHOUHUA DE RENZHI YANJIU

汉语近义量词构式精细范畴化的认知研究

ISBN 978-7-5768-2220-5



定价：86.80元


文库
河南省高等学校哲学社会科学优秀著作资助项目
[redacted]
著
公民道德责任研究
河南大学出版社
HENAN UNIVERSITY PRESS

图书在版编目(CIP)数据

公民道德责任研究 / [redacted]. -- 郑州 : 河南大学出版社, 2023.4

ISBN 978-7-5649-5451-2

Ⅰ. ①公… Ⅱ [redacted] Ⅲ. ①公民教育-社会公德教育-研究-中国 Ⅳ. ①D648.3

中国国家版本馆 CIP 数据核字(2023)第 076707 号

公民道德责任研究

GONGMIN DAODE ZEREN YANJIU

策划统筹 杨国安 谌洪波
责任编辑 张雪彩
责任校对 林方丽
封面设计 陈盛杰

出 版 河南大学出版社
地址:郑州市郑东新区商务外环中华大厦 2401 号 邮编:450046
电话:0371-86059715(高等教育与职业教育分公司) 网址:hupress.henu.edu.cn
0371-86059701(营销部)
排 版 河南大学出版社设计排版部
印 刷 广东虎彩云印刷有限公司
版 次 2023 年 4 月第 1 版 **印 次** 2023 年 4 月第 1 次印刷
开 本 710 mm×1010 mm 1/16 **印 张** 25.25
字 数 455 千字 **定 价** 68.00 元

(本书如有印装质量问题,请与河南大学出版社营销部联系调换。)