

Research Article

Analysis of Matching of Corpus Input and English Proficiency Based on the Big Data Neural Network Model

Guangli Li 

Xinyang Agriculture and Forestry University, Xinyang 464399, China

Correspondence should be addressed to Guangli Li; 2012280002@xyafu.edu.cn

Received 18 November 2021; Revised 14 December 2021; Accepted 23 December 2021; Published 10 January 2022

Academic Editor: Qiangyi Li

Copyright © 2022 Guangli Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the era of “Internet +” big data, the theory and technology of English corpus are becoming more and more mature. Corpus is an important method to reflect some language characteristics and clarify some language phenomena. In terms of cultural exchanges, Chinese students majoring in English have obvious cultural differences at home and abroad and lack the atmosphere and context for cultural exchanges. In addition, students have problems such as insufficient cultural communication skills. The big data neural network model is adopted in this paper to compare and analyze the intermediary sentences in the corpus to explore the development trend of English proficiency. Through the analysis of typical cases, it explores the weak links in the corpus teaching process and summarizes a method focusing on the combination of use of corpus and English teaching.

1. Introduction

With the continuous development of the economy, since China's entry into the WTO, many domestic enterprises and companies have begun to turn to overseas markets, and many foreign companies have also begun to flood into the Chinese market [1, 2] and gradually realize cross-border mergers and acquisitions and joint ventures, which greatly improved the comprehensive strength of the enterprise [3, 4]. However, as foreign companies enter the Chinese market, cultural conflicts exist between company operators under different cultural backgrounds in different countries, which has caused many contradictions in the business cooperation and management between the two sides, which has seriously affected the healthy development of the company. College students majoring in English are the main force of foreign employees in the future, and it is especially important to cultivate the ability of the main force to communicate with different cultures. In fact, more and more foreign language teachers begin to strengthen students' cultural ability in actual education, especially the cultivation of cultural ability.

The corpus can directly reflect a variety of discourse and pragmatic functions and can reflect different styles more

than idioms, so it can be used to evaluate students' phrase ability, language ability, and pragmatic ability. Corpus-based research methods are used to analyze the characteristics and shortcomings of the overall changes in the corpus use ability of English learners in China.

2. Big Data Neural Network Model

The English knowledge field mainly refers to the collection of knowledge units of all experience summaries and theoretical methods in the English professional field. However, in the computer field, it is possible to use a certain type of design for the knowledge contained in a specific field. This method is mainly to store the existing computer, in various aspects of system organization and management, it can conduct personalized learning and structured MACIEP personalized learning through operation and other characteristic knowledge groups, and it can provide personalized learning for MACIEP personalized learners' learning resources. The structure of the knowledge model in the English learning field requires that the English knowledge system has a good structural relationship in order to make accurate choices when recommending resource paths for MACIEP personalized learning. Generally speaking, the three dimensions of

curriculum learning, unit knowledge points, and basic knowledge level in the English knowledge field are expressed. The relationship between knowledge types is mainly the pioneer and successor relationship, parallel relationship, and related relationship. Different knowledge units and knowledge points need to include the difficulty, learning style, and learning goals of the English learning subprocess. The logical relationship between each type is shown in Figure 1.

According to the above analysis, if the structural model in the English knowledge domain is expressed as $KObject = \{Kid, Kname, Klevel, Kstyle, Kcontent, KOR\}$, then the expression Kid represents the unique mark of English knowledge points, $Klevel$ represents the difficulty coefficient of English knowledge points, $Kstyle$ represents the style of English knowledge points, $Kname$ represents the learning name of English knowledge points, $Kcontent$ represents the learning content of English knowledge points, and KOR represents a collection of corresponding relationships between English knowledge points.

Use the big data neural network model to analyze the big data information model of the material input used for the analysis and evaluation of the English level matching, and construct the raw material input control objective function for the prediction of the matching analysis ability with the English level.

$$\max_{x_{a,b,d,p}} \sum_{a \in A} \sum_{b \in B} \sum_{d \in D} \sum_{p \in P} x_{a,b,d,p} V_p, \quad (1)$$

$$\text{s.t.} \sum_{a \in A} \sum_{d \in D} \sum_{p \in P} x_{a,b,d,p} R_p^{bw} \leq K_b^{nw}(S), \quad b \in B. \quad (2)$$

The gray model is used to quantitatively and recursively evaluate the level of teaching ability supported by personalized learning. It is assumed that the historical data of teaching ability supported by personalized learning is distributed, and the initial value of the characteristic is fixed. The estimated density functional is estimated from the prediction of teaching ability supported by personalized learning.

$$u_c(t) = Kx_c(t). \quad (3)$$

In the high-dimensional feature distribution space, the corpus input and English proficiency match analysis ability prediction estimation system.

The continuous function of the calculation model is $u: I \times IR^d \rightarrow IR$, after $k-1$ iterations, $k \geq 1$, the corpus input and the English level matching analysis and evaluation of the gray-scale sequence satisfy $N(k) < L$, and the big data neural network model is used to obtain personalization. The output index of learning-supported teaching analysis and evaluation is the K adjacent sample values of the distributed big data information flow.

$$P_{IJ} = \sum_{d_i \in KNN} \text{Sim}(x, d_i) y(d_i, C_j). \quad (4)$$

The fusion method of big data information is used to construct personalized learning-supported teaching, and the

objective function is used to construct, analyze, and evaluate the interdomain classification of the information flow of large-distributed data. That is, the objective function of the big data cluster is

$$J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c \mu_{ik}^m (d_{ik})^2. \quad (5)$$

Analyze the studied English courses supported by personalized learning, quantitatively analyze and evaluate the exponential correlation distribution sequence, find an excellent method of K value, and obtain the quantitative recursive feature extraction result of education analysis and evaluation:

$$x_n = a_0 + \sum_{i=1}^{M_{AR}} a_i x_{n-i} + \sum_{j=0}^{M_{MA}} b_j \eta_{n-j}. \quad (6)$$

Here, a_0 is the sampling range of the initial personalized learning support teaching ability education evaluation; x_{n-i} is a scalar time series; b_j is the vibration attenuation value of personalized learning support arrangement and analysis evaluation.

3. Big Data Neural Network Model Analysis

The ultimate goal of education big data development is to return to the essence of education and achieve it. Teach what you can. The biggest drawback of the unified education model with one cut and two ends is that it ignores individual differences among students. Learner's personality characteristics mainly include knowledge level, mistakes/misunderstandings, emotional characteristics, cognition, and metacognitive abilities (Figure 2).

3.1. Knowledge Level. English knowledge level is not only the level that English learners have mastered of the current basic knowledge in this field, but also the most basic and personalized learning characteristics of English learners—the intuitive manifestation of MACIEP's personalized learning effect. This learning mode is mainly based on the target classification of the learner's learning content, mainly by analyzing the cognitive level of individual learning knowledge points of each English learner, and the difficulty of the English learning content has been adjusted in real-time coefficients and learning order to construct a personalized learning method. At present, with the continuous deepening of the intelligence of MACIEP personalized learning, the capacity of the database in the system has gradually increased, and the difficulty of data processing has gradually increased. The English learner model adopted by the main technology of the system can accurately measure learners. Mastered English knowledge levels, for example, the MAEVIF and SoNITS systems, respectively, use semantic network technology to construct the knowledge level model of English learners. Some people propose the main design rules of the user model, based on the knowledge level and learning style, which is the MACIEP personalized learning system.

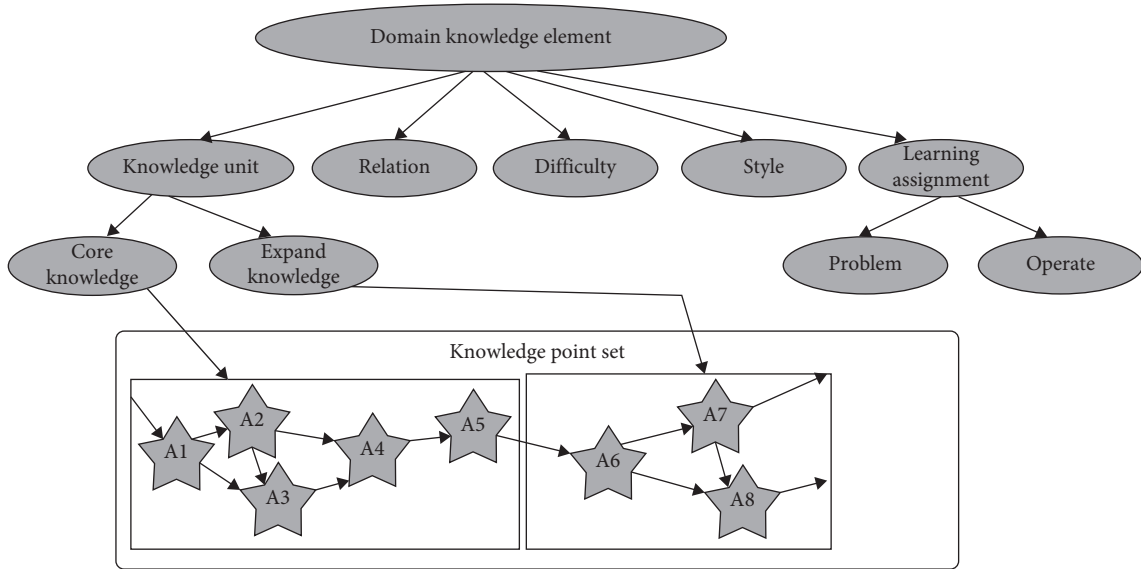


FIGURE 1: Diagram of the English domain knowledge model.

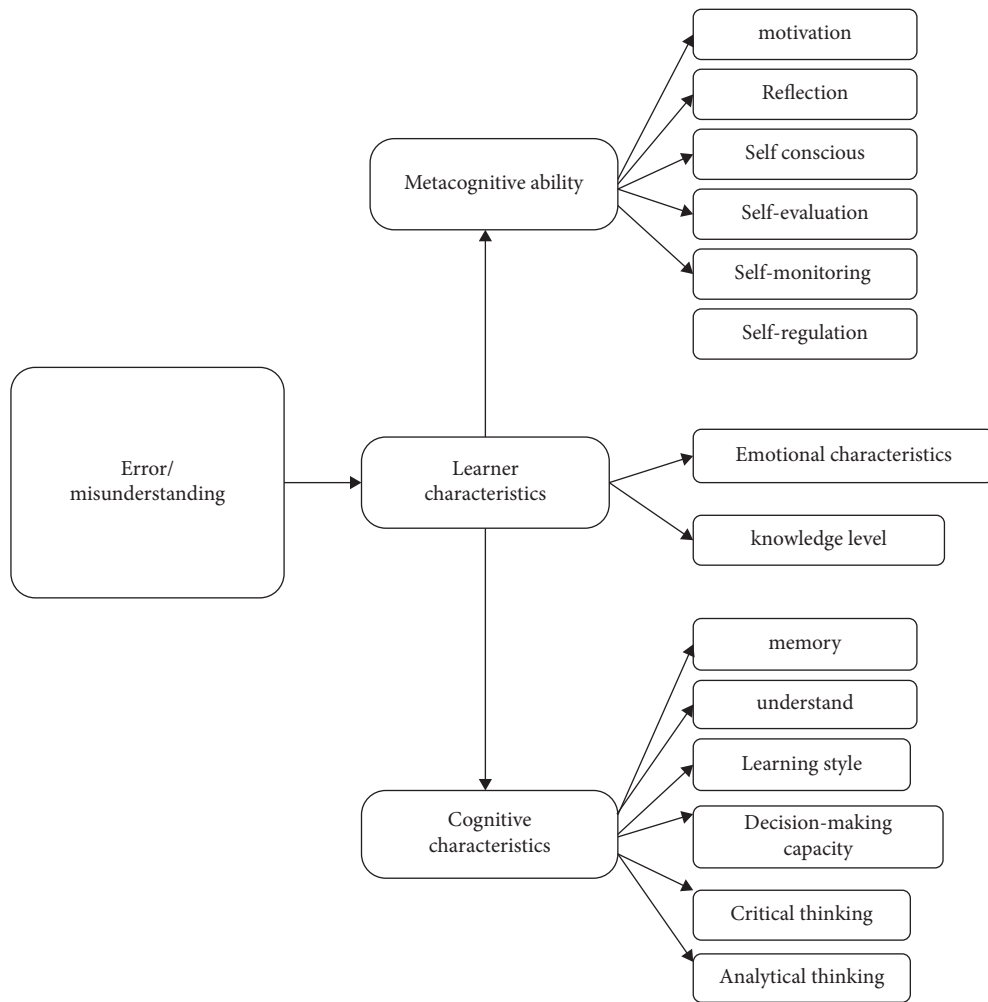


FIGURE 2: Personality characteristics of adaptive English learning.

MACIEP personalized learning can not only integrate different modeling technologies, but also make models based on the learner's knowledge level. The Web Easy Math and GIAS systems construct an English learning model by effectively combining the lead model and mechanical learning technology to effectively evaluate the knowledge level of learners. The MACIEP system can effectively combine the coverage model and the lead model to accurately determine the proficiency of each unit of English learning. At the same time, it can provide English learners with different grammatical rules used in different environments.

3.2. Cognitive Characteristics. Cognitive features are difficult for learners to recognize in learning and, at the same time, are a more complex feature that changes according to the changes of learners. Cognitive characteristics include memory, collaboration, problem-solving, judgment, analysis, and thinking. Learning characteristics will determine the learner's efficiency of learning books, including some cognitive characteristics such as perception, collection, absorption, and feedback. For example, visual learning is interested in graphics learning, and auditory learning is interested in audio materials. Those who rely on group learning belong to the field-dependent type of learners, and those who feel that the independent learning effect is good belong to the field-independent learners. The learning habits and characteristics of each learner are directly related to the sorting and display style of learning content in MACIEP.

English grammar knowledge learning system Web PTV combines movable type models and machine learning technology to measure learners' attention in grammar exercises and cultivate learners' good learning habits. The F-CBR-DHTS system uses fuzzy logic technology and lead model to analyze learners' cognitive characteristics, speculate on the ability to understand historical documents and materials, and then use Pexa and Sossa to apply subject technology to construct a learner model to obtain learner's knowledge and provide reasonable learning materials based on characteristics such as level, personal quality, and learning tendency. Mahnane et al. used an adaptive hypermedia system AHS-TS, which classifies learners according to a cardboard model and adjusts English teaching content to adapt to learners' individual thinking styles.

3.3. Emotional Characteristics. Emotion is the subjective experience of external stimuli accompanied by changes in physiological indicators such as facial expressions, movements, heartbeats, blood pressure, and brain waves. The learner can judge the emotional state of a moment through the subtle changes of the external and internal stimuli. At present, the focus of online education platform and big data support is still to improve the knowledge level and ability of learners, but it often ignores the learner's emotional state.

In fact, the learning efficiency of learners is closely related to their emotional state. The emotional state is mainly related to deep motivational factors. In actual teaching, experienced teachers and experts observe the emotional state of learners, provide corresponding feedback, and encourage

learners to learn. The learner's emotional state is happiness, excitement, concentration, enthusiasm, sadness, anger, anxiety, fear, fatigue, distraction, confusion, fatigue, indifference, etc. A positive state is happy. You can concentrate on promoting learning. In addition, the negative state has a negative impact on the learning process, such as burnout and distracted attention. Therefore, MACIEP is to monitor the emotional state of learners, increase learners' enthusiasm for participating in learning activities, and improve the effectiveness of learning. There is a need to use different technical methods.

The big data neural network model is a new tool for optimizing English learning problems. It is especially suitable for the research of prediction and comprehensive evaluation [5,6]. Its essence is based on optimization theory, assuming the use of linear functions to spatially build learning algorithms in high-order feature spaces.

Assuming linear separability, the sample set is $x \in R^d$, $y \in \{-1, +1\}$. In the d -dimensional space, the general form of the linear classification function is $g(x) = w \cdot x + b$, and the specific classification surface equation is

$$w \cdot x + b = 0. \quad (7)$$

To ensure that the hyperplane can be classified correctly for all samples, it must meet

$$y_i [(w \cdot x) + b - 1] \geq 0, \quad i = 1, 2, \dots, n. \quad (8)$$

In summary, satisfying the above conditions, the smallest hyperplane $\|w\|^2$ is the optimal hyperplane. You can convert the best classification solution for string similarity. That is, the function is calculated under the constraint of satisfying expression (2).

$$\varphi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} (w \cdot w), \quad (9)$$

$$L(w, b, a) = \frac{1}{2} (w \cdot w) - \sum_{i=1}^n a_i \{y_i [(w \cdot x_i) + b] - 1\}.$$

The optimal classification function obtained by solving is

$$f(x) = \text{sgn}\{(w \cdot x) + b\} = \text{sgn}\left\{\sum_{i=1}^n a_i^* y_i \cdot (x_i \cdot x) + b\right\}. \quad (10)$$

The exact linear separability of the sample can be processed in accordance with (1)~(5); however, if the sample cannot be linearly separated, that is, the relaxation variable $\xi_i \geq 0$ is added to (2) to satisfy the condition. This style (2) is available:

$$y_i [(w \cdot x) + b - 1] + \xi_i \geq 0, \quad i = 1, 2, \dots, n. \quad (11)$$

4. Research Design

4.1. Research Method. The big data neural network model is a widely used method for retrieving new and interesting knowledge. The big data neural network model represents of generalizing from labeled examples: the big data neural

network model is a generalization of the big data neural network model in a fuzzy environment [7–10]. The knowledge represented by the big data neural network model is more natural for human thinking. Classical neural network models are widely used in pattern recognition, English matching analysis and learning, and data mining. The big data neural network model is introduced to generalize the classification model, and the samples can be classified. This method allows us to use numerical and symbolic values to represent fuzzy modes in the learning phase (tree construction) or generalization phase. The problems existing in the matching analysis of corpus input and English proficiency are analyzed using big data analysis technology. Because there are many influencing factors involved in the study of English proficiency matching analysis mode, special improvement and data analysis need to be carried out on the level of English matching analysis. By constructing related variables and influencing factors that restrict the level of English proficiency matching analysis level, and using data integration and clustering data processing methods to match the English proficiency analysis model, the research ability of the innovative model of English curriculum teaching can be improved [1, 11, 12].

English proficiency matching analysis ability estimation model based on the big data neural network model can be used to complete the index parameter clustering and integration of English proficiency matching analysis ability and realize the research quantitative planning of corpus input and English proficiency matching analysis and accurate assessment of English proficiency matching analysis capabilities.

4.2. Corpus Research. The Chinese Students' Spoken English Corpus and the International English Corpus are used in this paper. These two corpora are mainly for large-scale English learners developed by research scholars well known in the Chinese education industry (Table 1). It contains English composition records of the students majoring in English from more than 10 colleges and universities in China. In order to ensure that the scale of the research corpus is representative and practical, English corpus completed within a time frame is used in this paper. A total of about 7000 samples totaling about 170 words are selected. The corpus of these words extracted is relatively similar, and so is the subject matter, with very strong representativeness and contrast. In this research, the students majoring in English in grades 1-2 are classified as lower-grade students and the students majoring in English in grades 2-4 are classified as higher-grade students. Two corpus databases for the composition of lower grades and higher grades are constructed. After contrastive analysis, the difference from the two corpora is that the students majoring in English in China can measure the usage and quality development of the English corpus during the four-year study period.

4.3. Search and Statistics Tools. From the search to the final definition of the object, the filter parameters involved in this study include frequency of occurrence, multixtext

reproduction rate, and mutual information value. There is no corpus search software yet, which can meet the search requirements of these three parameters simultaneously. Therefore, the research of WordSmith 5.0 chose to find the corpus of all 3-5 words and their frequency and the reproduction rate of multiple texts and then run Collocate 2.0 to obtain the word count and mutual information value of all 3-5 words. The result can be retrieved in this way and used for various verifications. By comparing the samples, the fitting of the search results of the word material and frequency of the two software reached more than 95%.

The Mann-Whitney U test function of the statistical software SPSS is used to compare whether the correctness of the corpus used by the two groups of English learners is statistically significant in this study. This takes into account the number of random samples in the index row. On the other hand, it reduces the measurement error in order to improve the reliability of the inspection results.

4.4. Research Process. In order to achieve an accurate assessment of the matching analysis ability of English proficiency, an information sampling model of the constrained parameters of the matching analysis ability of English proficiency needs to be built first. Combining nonlinear information fusion method and time series analysis method, statistical analysis of English level matching analysis ability is carried out. Constraint index parameters of English proficiency matching analysis ability are a set of nonlinear time series.

5. Research Results and Discussion

5.1. Corpus Search Results and Definition of Target Corpus. The 32835 and 18895 3-5 corpora are obtained, respectively, through the preliminary search of the two corpora of the lower-grade group and the higher-grade group. Since the frequency of use of the search results is very low and contains many corpora that only appear in a few texts and other noisy corpora, the target corpus needs to be defined based on stricter definitions and benchmarks to improve the operability and comparison of the research.

The screening of presearch results is based on 3 parameters. First of all, the words with frequencies below 40 times per million words are filtered to ensure that the target language meets the basic characteristics of high-frequency use. Secondly, the corpus whose recurrence rate is less than 5% of the total text is filtered. In order to avoid the object, the language used by a small number of English learners is repeated. Thirdly, words with a mutual information value less than 3.0 are filtered to ensure the ideal combination strength of the target word material. After filtering, there are 40 and 71 objects that meet the requirements in the two corpora, respectively.

5.2. Analysis Based on Frequency. Due to the different capacities of the two corpora, the researchers converted the original frequencies of all target corpora and the total number of the two groups of corpora to obtain the

TABLE 1: The capacity and representativeness of the two corpora.

Corpus	Text number	Number of pictographs	Representativeness
Lower-grade students	3965	1009966	Lower-grade corpus applied writing
Higher-grade students	2055	677765	Higher-grade corpus applied writing
Total	5989	1687645	Applied writing of corpus majoring in English

standardized frequency of each corpus and the frequency of uses per million in the two corpora. Visually compare the data usage of the two groups of students. According to the conversion result, the object input by the lower-grade students is 40 per million words, and the target language input by the higher-grade students is 105 per million words. This shows that the number of words used by the higher-grade group is more than twice that of the lower-grade group.

The ratio of the symbolic forms of the object class can reflect the diversity of language use from the side. The total of all objects in the two corpora is the number of pictograph, and the number of words that are not repeated is the number of class symbol. The larger the class pictograph, the lower the repetition rate of the same vocabulary, and the abundant usage of vocabulary facilitates the change. According to Figure 3, the sentence input of the higher and lower students was 1/183 and 1/376, respectively. The higher students used more language materials in their essays, in which fewer words and sentences are used repeatedly. Generally speaking, compared with the lower grades, the higher-grade student grades use not only more vocabulary, but also a wider variety of vocabulary.

According to surveys conducted by many researchers in China, the amount of raw materials used by students majoring in English is proportional to their language level. However, the results of research on the relationship between language usage and language proficiency of English learners of two languages in China and other countries are different [13, 14]. Among researchers of other countries, low-level English learners are more likely to rely on language to express English than high-level English learners. English learners in countries outside of English spoken countries speak more languages than native English speakers. The inconsistent results are probably due to differences in the research objects or the definition datum of terms. This is also the raw material input for us to investigate English learners. We should focus on qualitative analysis instead of quantity. The former is more trustworthy.

5.3. Classification and Distribution Characteristics of Target Corpus. In order to classify various objects, Altensberg's structural classification and Biberet al and Hyland's functional classification are adopted to count the total number and distribution of each corpus. The word materials are divided into three types: complete subsections, clause components, and incomplete sentence by structural classification method according to language functions according to the completeness of the structure. The language materials are divided into three types according to language functions by functional classification method. Factual vocabulary that expresses time and space, quantity, behavior, experience, etc.

and the language material that expresses the function of result, conversion, and limitation are language material that expresses the interpersonal relationship between the views and positions of both parties.

According to the structure classification results of the two groups of corpus, proportion of clause components is the highest, 85% and 83%, respectively, in the corpus used by lower grades and higher grades. The proportion of incomplete phrases is 18% and 17%. The complete sentences used are least, 4% are used by the lower-grade group, and they are not used by higher-grade students. From the results of the functional classification of raw materials, the use ratios of fact-based materials, discourse-based materials, and interpersonal relationship materials in the lower-grade group were close to 36%, 36%, and 31%, respectively. Among the language materials used by the higher-grade group, factual language materials, vocabulary language materials, and interpersonal relationship language materials accounted for 57%, 28%, and 18%, respectively.

Generally speaking, the categorical distribution of materials in the lower and higher grades of the English department in China is not balanced for writing essays, and the characteristics of excessive use of several language materials exist. It is noteworthy that even if there are few students, there is no need for language materials that express sentence types, some important vocabulary functions, and interpersonal relationship functions. English learners will also find deficiencies and defects in the input diversity, which is consistent with the conclusions of related studies.

5.4. Accuracy and Misuse of Corpus

5.4.1. Interpretation of Index Rows and Accuracy of Corpus. According to the retrieval result of the object index row, there are thousands of index rows for each lexical material, which is difficult to interpret one by one. If the index row is selected artificially, it will be too subjective and cannot guarantee the random representativeness of the sample. Therefore, in this study, we first used the hierarchical extraction method to select 30 sample words from the two objects to retrieve the index row and then used Concendance Randomizer to randomly extract 20 from the index rows of each raw material to obtain 450s sample of index row.

The definition and criterion of the correctness of raw materials are one of the difficult points of this research. Manually judging whether the materials are input correctly one by one, distinguishing between linguistic and nonlinguistic errors, has brought great inspiration to this research. However, whether the use of language materials is correct is defined not only based on the language materials themselves, but based on the larger language unit composed of language

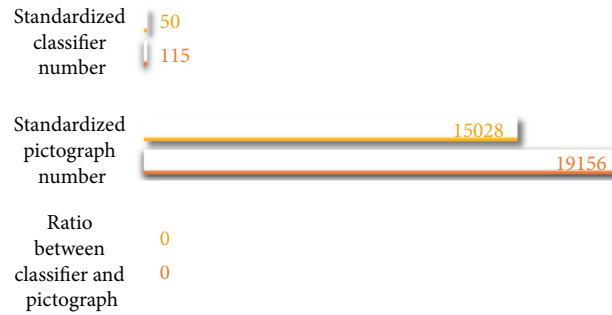


FIGURE 3: Statistics of the frequency of corpus used by students in high and low grades.

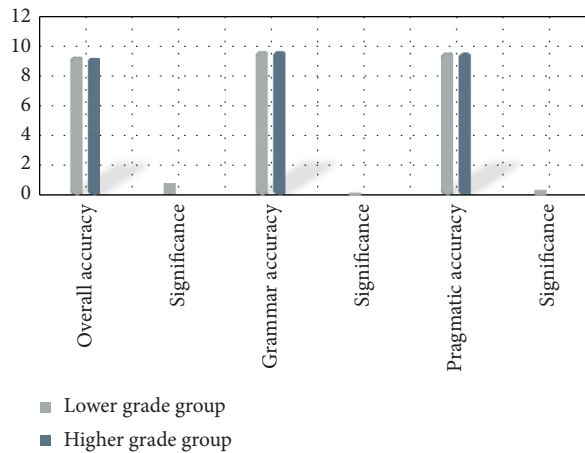


FIGURE 4: Comparison of the accuracy of corpus use by higher- and lower-grade students.

materials. Focus on the situation where English learners already know and can input the corpus, but cannot use it correctly and appropriately in the context. Based on the correctness, the point of view of accurately using objects for linguistic communication and the essential feature of using lexical materials as a form-function complex are the two dimensions included by the correctness of verbal materials. Grammaticality refers to the correct structure of the language material itself and the more complete language unit and coincidences with the grammatical norms. Secondly, appropriateness refers to the proper realization of pragmatic functions in the discourse by the language material itself and the more complete language unit. Therefore, the misuse of word blocks includes two types that are not grammatical and inappropriate. In particular, more detailed subcategory can be used according to the characteristics of the misuse.

Combining context, a dictionary, and a large-scale reference corpus, the researchers determined the grammatical and semantic accuracy of each corpus in the index row and calculated the use accuracy of each language material and the average accuracy of the two groups of phraseology of students. Figure 4 shows the descriptive and inferential statistical results of the correctness of the materials of the two languages used. There are certain differences in the grammatical correctness, pragmatic correctness, and overall correctness used in the corpus of the lower-grade group and the higher-grade group, but the difference has not yet reached statistical significance. In other words, the accuracy

of the raw materials of the higher-grade students is not very high, and it is even better than that of the lower-grade students. The correctness of raw materials for students majoring in English has not improved significantly during the four years of study.

5.4.2. *Misuse of Corpus.* Among the 400 index rows, according to the judgment results of grammatical and pragmatic errors and the classification statistics (refer to Table 2), there are 12 cases of grammatical errors used by higher-grade students, of which consistency and infinitive errors are the most, with 6 cases and 3 cases, respectively. The lower grades use more language errors, a total of 19 cases, all of which belong to inappropriate subclasses.

Generally speaking, the misuse of English learners may be affected by factors such as language ability and negative transfer of mother tongue. The form of the raw material itself, that is, the two characteristics of functional complex and structural incompleteness, will also affect the correct use of language materials. From the perspective of the grammatical language ability related to the corpus input in Figure 5, as a form-function complex, the corpus is restricted by grammar and pragmatic rules, and the grammar and language ability of English learners will definitely affect the use of the language materials. Secondly, because most of the corpus is structurally incomplete, only in the context of forming a larger unit with other languages and linguistic

TABLE 2: Types of corpus misuse by high- and low-grade students.

	Grammar error	Subcategory	Pragmatic error	Subcategory
Lower-grade group	8	AGRE3 INFI2 INEX1	10	COLL10
Higher-grade group	6	ARTI1 AGRE3 INFI1 ARTI1	11	COLL11
Total	13	AGRE6 INFI3 ARTI2 INEX1	20	COLL20

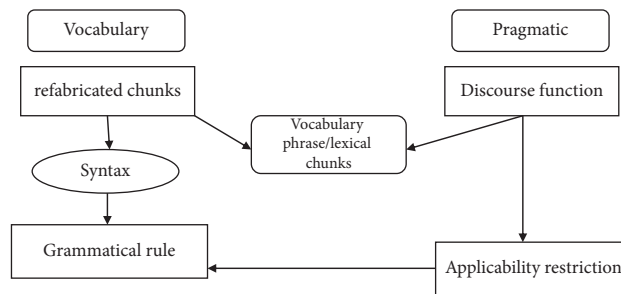


FIGURE 5: Grammatical and pragmatic competence involved in corpus input.

materials can a complete semantic function be realized. In the process of using the corpus, incorrect or inappropriate grammar may occur.

6. Conclusions

Intermediary corpus is used in this paper to carry out comparative analysis and trend comparison, to randomly test the sample corpus retrieval and index to evaluate the overall change characteristics of the corpus input characteristics and corpus application ability mastered by English learners in China at different stages of learning. Through the phrase ability test of the English learner, according to the corpus learning situation, it can enhance the learning level of the English learner for the phrase learning, find the problems in the use in time, summarize the diversity and accuracy of the corpus during the learning process, provide a strong basis for the continuous development of English education in China, and cultivate the language and phrase skills of English learners.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This study was sponsored by Xinyang Agriculture and Forestry University.

References

- [1] S. Maniraho, "Attitudes and motivation of teacher training college teachers and students toward English learning and use as medium of instruction in Rwanda," *European Journal of Mineralogy*, vol. 27, no. 1, pp. 31–50, 2014.
- [2] J. Xu and J. Zhuang, "Modeling costly learning and counter-learning in a defender-attacker game with private defender information," *Annals of Operations Research*, vol. 236, no. 1, pp. 271–289, 2016.
- [3] B. Winegarden, D. Glaser, A. Schwartz, and C. Kelly, "MCAT verbal reasoning score: less predictive of medical school performance for English language learners," *Medical Education*, vol. 46, no. 9, pp. 878–886, 2012.
- [4] H. Park, "Production and perception skill developments of Korean coronal obstruents by inexperienced English-speaking learners of Korean," *Journal of the Acoustical Society of America*, vol. 138, no. 3, pp. 1946–1947, 2015.
- [5] C. A. Liebert, L. Mazer, S. Bereknyci Merrell, D. T. Lin, and J. N. Lau, "Student perceptions of a simulation-based flipped classroom for the surgery clerkship: a mixed-methods study," *Surgery*, vol. 160, no. 3, pp. 591–598, 2016.
- [6] Y. Wang, F.-Y. Zhao, J. Kuckelkorn, H. Spliethoff, and E. Rank, "School building energy performance and classroom air environment implemented with the heat recovery heat pump and displacement ventilation system," *Applied Energy*, vol. 114, pp. 58–68, 2014.

- [7] M. H. Zeller, J. Reiter-Purtill, and C. Ramey, "Negative peer perceptions of obese children in the classroom environment," *Obesity*, vol. 16, no. 4, pp. 755–762, 2012.
- [8] B. Eren, L. Marot, M. Wisse et al., "In situ evaluation of the reflectivity of molybdenum and rhodium coatings in an iter-like mixed environment," *Journal of Nuclear Materials*, vol. 438, pp. 852–855, 2013.
- [9] J. Bouffard, L. Renault, S. Ruiz, A. Pascual, C. Dufau, and J. Tintoré, "Sub-surface small-scale eddy dynamics from multi-sensor observations and modeling," *Progress in Oceanography*, vol. 106, no. 6, pp. 62–79, 2012.
- [10] D. Schultz, S. Duffield, S. C. Rasmussen, and J. Wageman, "Effects of the flipped classroom model on student performance for advanced placement high school chemistry students," *Journal of Chemical Education*, vol. 91, no. 9, pp. 1334–1339, 2014.
- [11] Dongyoung and Kim, "An investigation of the benefits of a corpus-based ESP course for military English," *The Journal of Linguistics Science*, vol. 82, pp. 47–69, 2017.
- [12] L. L. Aull, D. Bandarage, and M. R. Miller, "Generality in student and expert epistemic stance: a corpus analysis of first-year, upper-level, and published academic writing," *Journal of English for Academic Purposes*, vol. 26, pp. 29–41, 2017.
- [13] M. Yeldham, "Second language English listeners' relative processing of coherence-based and frequency-based formulas: a corpus-based study," *Applied Linguistics Review*, vol. 138, no. 3, pp. 1946–1947, 2019.
- [14] I. Muneer, M. Sharjeel, M. Iqbal, M. Rao, and P. Rayson, "CLEU-a cross-language English-Urdu corpus and benchmark for text reuse experiments," *Journal of the Association for Information Science and Technology*, vol. 160, no. 3, pp. 591–598, 2018.