Research Article

Construction of a Multimodal Corpus of College Students’ Spoken English Based on Semantic Concepts

Xianhui Tian

School of Foreign Studies, Henan Polytechnic University, Jiaozuo, Henan 454000, China

Correspondence should be addressed to Xianhui Tian; xianhuitian@hpu.edu.cn

Received 24 March 2022; Revised 8 April 2022; Accepted 11 April 2022; Published 11 May 2022

Academic Editor: Liping Zhang

Copyright © 2022 Xianhui Tian. 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.

With the rapid development of multimedia technology, multimodal corpus came into being, which also opened up a new idea for oral English grammar teaching. In view of the problems existing in the current traditional English grammar textbooks, this paper applies the concepts of “corpus driven” and “multimodality” to grammar textbooks and tries to describe language use in a more authentic, reliable, and up-to-date form. In view of the problems existing in the current traditional English grammar teaching materials, this paper applies the semantic concept to the grammar teaching materials and tries to describe the language use in a more authentic, reliable, and up-to-date form. Based on the semantic concept, this paper studies the construction of college students’ oral English multimodal corpus. After the experiment, the results of the language fluency, accuracy, and complexity of the subjects in the experimental group and the control group are tested. The differences between the two groups were analyzed by one-way ANOVA as follows: in terms of language fluency, \( F(1, 62) = 0.835, P > 0.04 \), which showed that there was no significant difference between the two groups after the experiment. In terms of accuracy, \( F(1, 62) = 10.693, P < 0.04 \). There was significant difference between the two groups after the experiment. In this study, by analyzing and counting the distribution of conjunctions in students’ oral discourse, the types of conjunctions, their frequency of use, and the types of errors in use, we hope to understand the level and limitations of students’ use of conjunctions in oral discourse, so that students can better apply semantic conceptual cohesion and speak a coherent discourse. First of all, we should help students establish the awareness of textual cohesion and coherence.

1. Introduction

English grammar textbooks have always been an indispensable tool for English learners in China to learn grammar. Traditional English grammar textbooks mainly focus on language structure, but with the improvement of learners’ practical requirements for language knowledge and the reform of English grammar teaching, spoken English grammar has gradually become an important part of English grammar and spoken English teaching [1]. With the wide application of computer technology, the rapid development of Internet and the vigorous development of multimedia, corpus research has moved from Corpus 3.0 to Corpus 4.0. Corpus is an advanced language teaching and research tool in college students’ spoken English. However, most existing corpora are text corpora, which are limited in context, authenticity and richness of language, and do not meet the needs of language teaching in the “multimedia” era [2, 3]. The multimodal corpus of spoken English for college students is a new corpus developed from text corpus and spoken corpus. It has the advantages of contextualization, abundant storage of audio and video content, high reliability of corpus, and easy confirmation of corpus source, but it also has the disadvantages of troublesome transcription and annotation of corpus content and difficult processing. Multimodal corpus is a multimedia corpus based on speech theory, which takes speech activities as the research object, extracts information and knowledge from raw data as the means, and is driven by context model, and covers the language, sound, images, and actions of the whole speech activities [4]. With the rapid development of artificial intelligence technology [5, 6], multimodal corpus came into being, which also opened up new ideas for oral English grammar teaching. In view of the problems existing in the current traditional
English grammar textbooks, this paper applies the concepts of "corpus-driven" and "multimodal" [7] to grammar textbooks and tries to describe language use in a more authentic and reliable way.

Semantic concept, as opposed to semantic arbitrariness, is a special word and core idea of cognitive linguistics. As Braniz said, "semantic motivation refers to the relationship between language form and language meaning, which is a reflection of human thinking and world structure." The semantic concept is due to the orderly cycle of starting turn, maintaining turn and giving up turn. Interactive turn is usually the completion of semantic and grammatical sequence, but there are still some unfinished turn [8]. On the one hand, the unfinished turn is caused by interruption. In the process of interaction, the listener wants to compete for the right to speak and forcibly interrupt the speaker's discourse process. The current speaker's turn becomes the unfinished turn. On the other hand, when the semantic and grammatical sequences are not completed, there is a signal to give up the turn, and the listener receives the signal and then takes over the turn [9]. As a semantic concept, cohesion embodies that the interpretation of one component in discourse depends on the semantic relationship of the interpretation of another component. Cognitive research on semantic concepts can not only reproduce the cognitive background of semantic formation but also help human beings understand their own way of thinking. Therefore, it is necessary to study English semantic concepts from the perspective of experience view.

In the construction of multimodal corpus of college students' spoken English, the clues to give up turn are often not used alone. The end of the same turn often contains many clues, such as the decrease of pitch and intensity attached to the completion of grammar and semantic sequence, which is often accompanied by pause and silence [10, 11]. In turn-taking, verbal forms and nonverbal forms have a synergistic effect to jointly promote the conversation process. Connectivity is a semantic concept, formally embodied by conjunctions and connectives [12, 13]. In this study, by analyzing and counting the distribution, types, frequency, and errors of conjunctions in students' oral discourse, we hope to understand the level and limitations of students' use of conjunctions in oral discourse. Complete the use of rising or falling tones at the end of sentences in the multimodal corpus of college students' spoken English; completion of a grammatical sequence; completion of a semantic sequence; turn constitutes pause and silence when the unit is completed; extension of last syllable and stressed syllable in sentence; modal words with additional loudness, accompanied by pause after vowel elongation of modal words; slow down; the decrease of pitch and intensity; semantically repeated statements; summative statement; and stop, change of body posture, and tense relaxation of various parts of the body [14, 15].

This paper studies and innovates the above problems from the following aspects:

1. A multimodal corpus model based on semantic concepts is proposed for college students' oral English. The multimodal corpus's rich corpus provides a real and vivid context for English learning, allowing students to broaden their knowledge and become familiar with common vocabulary, sentence patterns, and stylistic features under relevant topics, thereby improving their reading comprehension ability. Multimodal corpus can play a strong supporting and auxiliary role in both teacher-led classroom teaching and students' autonomous learning to improve students' comprehensive English application ability in college English vocabulary teaching, listening and speaking teaching, and reading teaching.

2. The construction of a multimodal spoken English Corpus annotation system is based on semantic concepts. The oral English multimodal corpus of college students can be used in the teaching of semantic concepts in foreign language listening, speaking, reading, writing, and translation, and research in a particular field should be expanded. We should first help students establish awareness of textual cohesion and coherence so that they can better apply semantic conceptual cohesion and speak a coherent text. Grammar has always been primarily taught at the sentence level in traditional English classes.

2. Related Work

2.1. Research Status at Home and Abroad. Albu pointed out that in traditional foreign language teaching, teachers often find it difficult to create learning environment due to the lack of information technology support. Using multimodal corpus to support language learning makes it possible to reproduce or copy the real context in teaching, and enriching the real context and corpus input to drive teachers' classroom teaching and students' autonomous learning is helpful to improve the autonomy and constructiveness of language learning [16]. Green and Birdsong put forward that many European countries have carried out large-scale multimodal corpus construction and related research, among which the concepts, standards, and tools have become important topics for scholars [17]. Sidd et al. put forward that with the support of multimodal retrieval technology, multimodal corpus can reflect the pronunciation in real context, which is conducive to driving self-discovery language knowledge construction and improving students' multireading ability, autonomous learning ability, and comprehensive language application ability [18]. Guichon put forward the application of multimodal corpus in foreign language teaching in 2000. Anthony Baldry, an Italian expert in systemic functional linguistics and multimodal discourse analysis, and other scholars have jointly developed a multimedia information retrieval tool "multimodal corpus tagging system" [19]. Le and Miller put forward the concept of oral interactive ability based on interactive view and thought that we can build the ability of verbal interaction with others from language knowledge and interactive skills for the purpose of communication, including the degree of understanding others'
speech, conversation cooperation ability, and topic organization and management ability. Two dimensions define oral communicative competence, which brings new ideas to the study of oral communicative competence and attracts more and more scholars' attention [20]. Demetriou et al. put forward that language learning needs more situations, that is, contexts, and only a large number of real contexts can promote language cognition and learning, which is the fundamental idea of multimodal corpus-driven language learning [21]. Wagner put forward that the communication theory model is developing continuously and tends to be systematic and comprehensive, and the division of communicative competence is gradually refined. These theories and models have a far-reaching impact on language testing and teaching. However, each model has some shortcomings, and its components overlap and are difficult to distinguish. For example, in Canal and Swain's communicative competence model, "social language competence" is difficult to distinguish from "grammatical competence," because properly understanding and expressing discourse in different social environments must be inseparable from the role of grammatical competence [22]. Doliana et al. put forward the definition of multimodal corpus: "a multimedia corpus based on "speech theory," which takes speech activities as the research object, extracts information and knowledge from original data as the means, and takes context model as the driving force, including language, sound, image, and action of the whole speech activities [23]. Kolkmann and Falkum proposed that the establishment of multimodal corpus provides new learning resources and tools for college students' oral English. It can effectively support the independent discovery and construction process under the guidance of teachers and plays a positive role in promoting the pan centralization of teaching resources, the diversification of teaching methods, and the equality of teachers and students [24]. Qin and Kong proposed that multimodal corpus refers to "a corpus that integrates text corpus, audio corpus and static and dynamic image corpus, and users can carry out retrieval, statistics and other operations through multimodal mode" [25].

2.2. Research Status of College Students' Oral English Based on Semantic Concept. This paper investigates the use of semantic concepts to construct a multimodal corpus of spoken English from college students. The multimodal corpus captures the behaviors and situations of teachers and students in language class in multiple dimensions, including words, images, audio, and video, and reconstructs the entire picture of interpersonal interaction. This monomodal college students' spoken English corpus has lost some modes that have information value, such as facial expressions and body movements, real language context, and pronunciation and intonation, lowering the corpus' use value. However, an English grammar textbook for college students that is based on a multimodal corpus and uses a semantic concept effectively solves these issues. The existing spoken corpus' primary function is to analyze speech errors using transcribed texts, but little attention is paid to labeling, symbols, and analysis of nonverbal communication, which is crucial in everyday communication. Multimodal corpora are more difficult to create than traditional text corpora, and theoretical issues related to them must be resolved as soon as possible. Another aspect that requires attention is how to combine the characteristics of China's information and media technology development, as well as how to construct
a local multimodal corpus foreign language teaching theory. Oral grammar is a grammatical phenomenon that occurs in spoken discourse and is primarily based on the study of informal conversational English. Oral English grammar instruction focuses on sentence and text-level grammar, requiring students to apply it in context rather than memorize it. The multimodal corpus of spoken English in college is based on the semantic concept, which refines the labeling indicators from many dimensions, such as pronunciation, vocabulary, syntax, discourse, and nonverbal communication, in order to make an all-round and multilevel analysis of this type of students’ oral English ability and provide useful feedback for oral English classroom teaching and students’ independent learning after class.

3. Principle and Model of Semantic Concept

As a new cognitive perspective, experiential view closely connects the meaning of language with human physical experience. In the past 20 years, the experiential study of language has attracted the attention of many scholars at home and abroad. “Experiential thought,” “experiential behavior,” and “experiential cognition” provide new ideas for people to study language. Studying semantic motivation from the perspective of experience view can help people further understand the mechanism of semantic formation. The compilation of traditional dictionaries is mostly language processing through the editor’s introspection. In this way, the processed learning corpus will be processed for teaching or learning, highlight the grammatical rules and semantic concepts of the language, and deviate from the authenticity of the language. Multimodal corpus can also be used in reading teaching. The rich corpus in the multimodal corpus, such as audio and video materials such as dramas, speeches, and news, has a variety of styles and a wide range of topics, which provides a real and vivid context for English learning, helps to expand students’ knowledge and make them familiar with the common vocabulary, sentence patterns, and stylistic characteristics of relevant topics, so as to improve their reading comprehension ability. In college English vocabulary teaching, listening and speaking teaching, and reading teaching, multimodal corpus can play a strong supporting and auxiliary role in both teacher-led classroom teaching and students’ autonomous learning, so as to improve students’ comprehensive English application ability. The multimodal model of spoken English under the semantic concept is shown in Figure 1.

This paper adopts the method of college students’ oral English multimodal corpus based on semantic concepts. Firstly, the features of a group of words are selected, and then, each word is compared with the features of this group of words to obtain a relevant feature vector. The similarity is calculated by calculating the angle cosine of the vector.

The vector of text in semantic topic space is defined as

$$\vec{V}_{di} = \sum_{k=1}^{n} f_{ij} \cdot k \cdot t_k.$$  \hspace{1cm}(1)$$

Among them, $f_{ij}$ is the number $t_1, t_2, \cdots, t_n$ of feature items appearing in the text, which is the corresponding vector of each feature item in the semantic topic space $P$. 

![Figure 2: Multimodal spoken English corpus annotation system.](image-url)
Standardize the feature vectors to get vectorized representation of the text $d_i$ in the semantic topic space.

$$
\overline{d_i} = \frac{1}{V_i} V_i \cdot d_i.
$$

After the feature vectors of the text are obtained, the cosine of the included angle between the feature vectors $d_i$ and $d_j$ of the text is used to calculate the similarity. The similarity measurement method is as follows:

$$
S(d_i, d_j) = d_i^T \cdot d_j^T = |d_i| \cdot |d_j| \cdot \cos (w_{ij}) = \cos (w_{ij}).
$$

Given the training word sequence $w_1, w_2, ..., w_T$, the objective function constructed according to the Skip-Gram principle is

$$
1 \frac{1}{T} \sum_{i=1}^{T} \sum_{\forall j < i \land j \neq 0} \sigma(p(w_{i+j} | w_i)).
$$

Usually, the hierarchical softmax function is used to represent the language probability set $P(w_{i+j} | w_i)$, and the Huffman tree coding is used to represent the sentences with the length of $L$ according to the word frequency. This data structure can quickly find high-frequency words and greatly reduce the computational complexity

$$
p(w | w_j) = \frac{\exp (v_{w}^T v_{w_j})}{\sum_{w} \exp (v_{w}^T v_{w})} = \prod_{j=1}^{L} \sigma \left( \left( m(w, j + 1) = \exp (\exp (-x)) \right) \left( n(w_{i+j}) \right) \left( v_{w_{i+j}}^T v_{w_j} \right) \right).
$$

where $v_{w}$ and $v_{w}^T$ are the input vector and output vector of the word $w$; $w$ is the total number of all words; $\sigma(x) = 1/(1 + \exp(-x))$; the path from $j$ node to root node is $n(w, j)$, especially $n(w, 1) = root$ and $n(w, L(w)) = w$.

Predict the output word vector $v_{w}^T$ according to the input word vector $v_{j}$, i.e.,

$$
v_{w}^T = b + W v_{j} + U \tanh (d + Hv_{w}).
$$

tanh is the neuron activation function and $b, d, W, U, H$ is the parameter to be solved.

For the comment $C_i$ of the $i$ composition, the TF-IDF method is used to calculate the weight value of each comment short sentence $TF-IDF$, and a group of comment short sentence sequence $K_{TF-IDF}$ is obtained by sorting from large to small

$$
TF-IDF = tf_{i,j} \times idf_j = \frac{n_{i,j}}{\sum_{k} n_{k,j}} \times \log \left( \frac{|D|}{|\{j : t_i \in d_j\}|} \right).
$$

Text rank method is used to calculate the TR weight value of each comment short sentence, and a group of comment short sentence sequence $K_{i}^{textrank}$ is obtained by sorting from large to small

$$
TR = (1 - d) + d \times \sum_{V \in \text{Out}(v_j)} \sum_{V \in \text{In}(v_j)} \frac{w_{i,j}}{W_{i,j}} \times \text{TR}.
$$

Calculate the similarity between the comprehensive vector $v_{i}^{all}$ of the composition $v_{i}^{all}$ to be evaluated and the comprehensive vector $cc$ of each composition in the training

<table>
<thead>
<tr>
<th>Test items</th>
<th>Group</th>
<th>Average</th>
<th>Standard deviation</th>
<th>Number of people</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1 fluency</td>
<td>2</td>
<td>86.2451</td>
<td>5.05924</td>
<td>35</td>
</tr>
<tr>
<td>Test 2 accuracy</td>
<td>2</td>
<td>86.4251</td>
<td>4.57846</td>
<td>35</td>
</tr>
<tr>
<td>Test 3 complexity</td>
<td>1</td>
<td>.7905</td>
<td>0.03735</td>
<td>35</td>
</tr>
<tr>
<td>Test 2 accuracy</td>
<td>2</td>
<td>.7624</td>
<td>0.03284</td>
<td>35</td>
</tr>
<tr>
<td>Test 3 complexity</td>
<td>1</td>
<td>1.2564</td>
<td>0.06508</td>
<td>35</td>
</tr>
<tr>
<td>Test 3 complexity</td>
<td>1</td>
<td>1.1742</td>
<td>0.05078</td>
<td>35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test item</th>
<th>Freedom</th>
<th>Mean square</th>
<th>Variance ratio</th>
<th>Significant level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1 complexity</td>
<td>Interblock</td>
<td>1</td>
<td>1.001</td>
<td>.042</td>
</tr>
<tr>
<td>Within the group</td>
<td>61</td>
<td>23.283</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Test 2 complexity</td>
<td>Interblock</td>
<td>1</td>
<td>.012</td>
<td>10.693</td>
</tr>
<tr>
<td>Within the group</td>
<td>61</td>
<td>.002</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Test 3 complexity</td>
<td>Interblock</td>
<td>1</td>
<td>.092</td>
<td>27.291</td>
</tr>
<tr>
<td>Within the group</td>
<td>61</td>
<td>.004</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
library, and sort it from large to small

\[ \text{Similarity}_{ij} = \cos \left( v^\text{all}_i, v^\text{all}_j \right) = \frac{v^\text{all}_i \cdot v^\text{all}_j}{\sqrt{\sum_{m=1}^{M} (v^\text{all}_{i,m})^2} \sqrt{\sum_{m=1}^{M} (v^\text{all}_{j,m})^2}}. \]
\[ (9) \]

Solving the objective function of each parameter,

\[ \text{Loss} = \sum_{i} \left( y^\wedge_{i} - y_{i} \right)^2 + \sum_{k} \Omega(f_k) = \sum_{i} \left( y^\wedge_{i} - y_{i} \right)^2 + \sum_{k} \left( y^T + \frac{1}{2} \lambda \| w \|^2 \right). \]
\[ (10) \]

\( \Omega(f_k) \) includes two parts: parameter \( \gamma \) reflects the influence of the number of node \( T \) on the error; parameter \( \lambda \) reflects the influence of node weight \( w \) on the error.

When no one speaks after the completion sentence in grammatical and semantic terms, it is usually within 2 seconds; silence refers to when no one speaks for a long time after the turn is completed. All turn taking positions can have pauses associated with them. The powerful topic retrieval function of multimodal corpus can provide a large number of corpus based on the real context for listening and speaking teaching or be used to introduce the topic of listening and speaking classroom, improve students’ interest in participation, deepen their understanding of the topic, and expand their ideas, or as a language input before listening and speaking task, activate and enrich students’ relevant language reserves, or as a language input before listening and speaking task, activate and enrich students’ relevant language reserves. Oral English learners should master standard pronunciation and intonation as much as possible in addition to learning the grammar points of everyday communication terms. Traditional grammar teaching materials clearly cannot provide information other than in text mode, whereas oral grammar teaching materials based on multimodal corpus can directly convey the complete pronunciation and picture to learners, thereby compensating for this shortcoming. At the moment, the speaker will release the signals of abandoning the turn when the discourse process is completed or if the speaker does not want to continue speaking. According to these signals, the listener will judge the turn conversion correlation position in order to
smoothly connect the turn and continue the conversation. The following are the most common: grammatical or semantic sequence completion, pause and silence after the completion of a turn, semantically repeated sentences, summary sentences, pitch and intensity reduction, gaze, designated receiver, and so on.

(1) The completion of most turn-taking is marked by the completion of grammar or semantic sequence, and the speaker’s discourse says after that, stop talking and give up the turn.

(2) Summative statement

(3) Semantically repeated statements

(4) Pause or silence after turn completion

The current speaker actively selects the next speaker and transfers the turn to the other party, in addition to expressing that they have completed the discourse process and given up the turn through the above methods. There are two main ways to signal the other party to take over the turn: directly specifying the next speaker through verbal form and nonverbal gaze and gestures. The processing of human cognitive activities, which are inseparable from the cognitive experience abstracted by the body in space activities, is intimately linked to the relationship between semantic concepts and the world. As a result, cognitive linguists argue that “meaning is rooted in the speaker’s knowledge and belief and comes from people’s physical experience, categorization, and conceptual system.” It is extremely important in the teaching of languages. In contrast to traditional text-based corpora, multimodal corpora can not only provide text retrieval results for vocabulary, but also audio and video playback, making the teaching context more real and vivid. A set of Chinese science and engineering college students’ oral English corpus annotation system is extracted based on the multimodal discourse media system and combined with the actual situation of Chinese science and engineering college students’ oral English output, as shown in Figure 2.

Semantic concepts can help Chinese college students learn to associate from the perspectives of physical and spatial experience, as well as improve their oral English expression ability. At the same time, this paper summarizes common expressions from the perspective of physical and spatial experience, provides a new idea for Chinese students to learn English, and assists them in discovering the fun of learning English. The system of labeling is divided into two types: verbal and nonverbal. Pronunciation and writing are both part of speech. “Pronunciation” looks at syllable and stress pronunciation; “character” looks at the ability to choose words and construct sentences and grades it on three levels: word use, sentence, and text. Companion language and body language are examples of nonverbal communication. The tone of a sentence can be divided into five categories: rising tone, falling tone, flat tone, rising and falling tone, and rising and falling tone. The main focus of body language annotation is on the activities of various parts of the body, such as eye communication, gestures, expressions, head movement, and expression naturalness. The successful application of corpus linguistics to the compilation of English grammar textbooks serves as a crucial foundation for the use of multimodal corpus in oral English grammar textbooks. Many corpus-based English grammar textbooks have been published in the West since the early days of corpus linguistics. Multimodal corpus, in comparison to traditional corpus, retains more information in the language and has a higher application value in oral grammar textbooks.

4. Implementation of College Students’ Oral English Multimodal Corpus

4.1. A Multimodal Corpus of Spoken English Based on Semantic Concepts. Synonyms and antonyms are rarely used in words. A vocabulary chain is formed when several words appear at the same time in a discourse around a single topic. These words are from the same lexicon. When students come across or think of one or more words in the vocabulary set, they are likely to think of other words in the set as well. However, statistical data shows that vocabulary storage in students’ brain language databases is disorganized, and it cannot be systematically extracted to form a vocabulary.

![Statistical change chart of freshmen-seniors feedback items.](image)
chain when expressing, so it has to be pieced together temporarily, leading to a vocabulary poverty, misuse, and simplification trend. Multireading or multimodal teaching advocates mobilizing learners’ senses and getting them to cooperate to participate in language learning through various channels such as the Internet, pictures, role-playing, and various teaching methods. As a form of diversified reading teaching, language learning driven by a multimodal corpus of college students’ spoken English based on semantic concepts supports students in searching multimodal information according to learning tasks or personal interests and mobilizes students’ multiple senses to learn, which is conducive to improving language productivity. Simultaneously, multimodal teaching allows students to interact with real cultural scenes while learning semantic concepts, experience cultural differences, cultivate cross-cultural awareness, and improve cross-cultural communication skills. Multimodal financial English corpus can be used as a tool reference book in college English teaching, whether for teachers or students. Unlike traditional dictionaries, corpus possesses a dialectical unity of regularity and variability. The multimodal corpus of spoken English from college students can be used in the teaching of semantic concepts in foreign language listening, speaking, reading, writing, and translation, and research in a specific field should be expanded. Students’ characteristics in verbal and nonverbal dimensions, for example, can be studied further in the context of spoken English. We should first help students establish awareness of textual cohesion and coherence in order for them to make better use of semantic conceptual cohesion and speak coherent texts. Grammar instruction in traditional English classes has always focused on the sentence level, such as the relationship between a single sentence and a single sentence within a segment.

The use of oral English multimodality among college students in semantic concept teaching can help to mitigate the problem of teaching corpus distortion to some extent. The corpus’ index retrieval collocation function provides learners with not only real and rich financial English resources, but also realistic context and learning opportunities to observe financial English language phenomena. One after the other, the English Chinese literature corpus, Chinese English parallel corpus, Chinese English corpus, military English corpus, and New Horizon College English textbook corpus were established. Simultaneously, relevant corpus-assisted teaching software has been developed one after the other. The corpora listed above are text-based corpora, and the majority of them are for semantic concept language research. Foreign language teachers should introduce and analyze texts using relevant discourse linguistics theories. To begin, they should introduce relevant concepts such as reference, substitution, ellipsis, cohesion, repetition, and cooccurrence; at the same time, they should teach the application principles and rules of these cohesive mechanisms in discourse to help students establish the necessary knowledge framework. Finally, gradually teach them how to use it in language communication and to speak a coherent and cohesive discourse. Currently, the multimodal corpus of college students’ oral English under the semantic concept is used in foreign language teaching with college students, but it can also be used with primary school, junior high school, and senior high school students. The content can be expanded to include multimodal textbook design, teacher training, network courseware development, outline creation, teaching mode and teaching evaluation method, textbook evaluation, and exercise design reform, among other things.

4.2 Experimental Results and Analysis. This experimental test is mainly quantified from three aspects: language fluency, accuracy, and complexity. The experimental data were statistically analyzed by the SPSS13.0 software, and the significant level of differences in oral English ability between the experimental group and the control group was detected. The variance analysis of the experimental data is shown in Tables 1 and 2.

Table 1 showed the results of language fluency, accuracy, and complexity of subjects in the experimental group and the control group after the experiment. The average scores of the subjects in the experimental group were higher than those in the control group in terms of language accuracy and complexity, while test 1 showed that there was no significant difference in language fluency between the two groups.
The test scores of the two groups were tested by the SPSS13.0 software, and the differences between the two groups were analyzed by one-way ANOVA as follows: in terms of language fluency, \( F(1, 62) = 0.835, P > 0.04 \), which showed that there was no significant difference between the two groups after the experiment. In terms of accuracy, \( F(1, 62) = 10.693, P < 0.04 \). After the experiment, there was a significant difference between the two groups. In terms of language complexity, \( F(1, 62) = 27.291, P < 0.04 \), which shows that there is a significant difference between the test scores of the two groups after the experiment.

In this experiment, the research on multimodal corpus foreign language teaching is divided into two categories: theoretical discussion and empirical research. The theoretical discussion is divided into three categories: literature review, teaching model, and discourse analysis model. The empirical research is divided into four categories: retrieval tools, corpus construction, oral teaching, and multimodal metaphor ability. Two experiments were carried out for comparison. The experimental results are shown in Figures 3 and 4.

As can be seen from Figures 3 and 4, there are 36 literatures on multimodal corpus-based foreign language teaching, including 15 literatures on theoretical discussion, and the remaining 21 are empirical studies, which shows that domestic multimodal corpus-based foreign language teaching research has just started, and theoretical research is insufficient, and theories are supplemented in the developing empirical research. After two experiments, the proportion of theoretical research is lower than that of empirical research. Therefore, the construction of corpus, including the construction of multimodal corpus of college students’ spoken English and the empirical research of corpus-driven assisted English teaching, is an area that Chinese researchers continue to explore and explore.

According to the standard and classification of feedback items, we have made statistics. In second language learning, whether the feedback is appropriate or not can be used as the standard of language ability evaluation. Among native language learners, feedback items can express attitude, reception, and understanding to promote the smooth progress of conversation, and feedback items can also be used as the index of oral interaction grading. Three experiments were conducted to compare the number and frequency of feedback items in oral interaction among freshmen, sophomores, juniors, and seniors. The experimental results are shown in Figures 5–7.

From Figure 5 to Figure 7, it can be seen that the average frequency of feedback items of senior students is higher than that of freshmen. Between 30 and 40 hours, the average frequency of feedback items of sophomores is higher than that of juniors, and that of seniors is higher than that of seniors. Therefore, it can be concluded that seniors are the best in general. During the establishment of multimodal corpus of college students’ spoken English, attention should be paid to strengthening the design and planning of corpus construction, the collection and processing of corpus, and the standardization of corpus management system construction, promoting the coconstruction and sharing of corpus resources, improving the use efficiency and avoiding the waste of resources caused by repeated construction. According to the basic principles of corpus building, that is, pertinence, representativeness, and scale, consider that the users of multimodal corpus of college public English are mainly English teachers and non-English majors at the basic stage of college, according to the teaching content, requirements of college public English, and students’ learning foundation and interest.

5. Conclusions

In college English teaching, multimodal corpus can provide students with real language materials and multiple contexts, help to realize the new teaching mode of “students first, teachers second,” and is of great significance to promote the reform of college English teaching in China. Preliminary progress has been made in the establishment of multimodal corpus for oral English grammar teaching in semantic concepts. However, due to the complexity of multimodal corpus collection, annotation, and segmentation, the results still have great defects. Based on the theoretical basis and practical significance of college students’ oral English multimodal corpus based on semantic concept, this paper summarizes...
the construction of corpus at home and abroad, especially the research status of multimodal corpus and language teaching application. On this basis, this paper constructs college students' oral English multimodal corpus. Finally, it makes a multidimensional exploration on college English teaching based on multimodal corpus. After the experiment, the language fluency, accuracy, and complexity of the subjects in the experimental group and the control group were tested. The differences between the two groups were analyzed by one-way ANOVA as follows: in terms of language fluency, $F(1, 62) = 8.35, P > 0.04$, which showed that there was no significant difference between the two groups after the experiment. In terms of accuracy, $F(1, 62) = 10.693, P < .04$. There was significant difference between the two groups after the experiment. The application of college students' oral English multimodal corpus under semantic concept in foreign language teaching in China is still a field to be further studied. At present, the multimodal corpus of college students' oral English under the semantic concept is applied to college students in foreign language teaching and can also be used for primary school, junior, and senior high school students. The content can be extended to the fields of multimodal textbook design, teacher training, network courseware development, outline setting, teaching mode and teaching evaluation method, textbook evaluation, exercise design reform, and so on.

**Data Availability**

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

**Conflicts of Interest**

The author does not have any possible conflicts of interest.

**References**


