English Writing Teaching Model Dependent on Computer Network Corpus Drive Model

Lei Shi
Shaanxi Institute of International Trade & Commerce, Xi'an, China
904201126@qq.com

Abstract—At present, the mainstream lexicalized English writing methods take only the corpus dependence between words into consideration, without introducing the corpus collocation and other issues. "Drive" is a relatively essential feature of words. And once the drive structure of a word is determined, it will be relatively clear what kinds of words to collocate with, hence the structure of the sentence can be derived relatively directly. In this paper, the English writing model that relies on the computer network corpus drive model is put forward. In this model, rich English corpus is introduced in the decomposition of the rules and the calculation of the probability, which includes not only the corpus dependence information, but also the drive structure and other corpus collocation information. Improved computer network corpus drive model is used to carry out the English writing teaching experiment. The experimental results show that the precision and the recall rate are 88.76% and 87.43%, respectively. The F value of the comprehensive index is improved by 6.65% compared with the Collins headword driven English modes of writing.

Keywords—Computer Network; Corpus Drive model; English Writing Teaching model; Precision

1 Introduction

English grammar writing refers to the automatic recognition of the relationship between the syntactic units contained in sentences and these syntactic units in accordance with the given grammar [1]. English grammar writing plays a very important role in the field of natural language processing. At the same time, it is also a difficult research problem publically recognized. English comprehension can generally be divided into the following steps [2]: The input of the original text, the sentence segmentation and the word attribute feature tagging, the grammar and English writing, the corpus and pragmatic and context analysis, the formation of the target forms, the sentence group, and the chapter comprehension and so on. The sentence analysis links the preceding text comprehension and the subsequent lexical analysis, serving as a connecting link between the preceding and the following [3]. Lexical analysis is the foundation, sentence analysis is the center, and text comprehension is the ultimate goal. Therefore, once the computer representation of the sentence constituents is ob-

tained, it is of practical significance no matter in the application of sentence group division, text comprehension, machine translation, machine interpretation, human-computer dialogue, information retrieval or other aspects. Studies of the English writing can be roughly divided into two ways [4]: The methods dependent on rules and the methods dependent on statistics. In order to carry out statistical English writing, first of all, certain grammar system shall be followed. In accordance with the grammar of this system, the representation form of the grammar tree is determined. Currently, the phrase structure grammar and the dependency grammar are widely used in English writing [5]. And the phrase structure English writing are generally dependent on the probabilistic context free grammar (hereinafter referred as PCFG for short) at present. In the early research work, the phrase structure English writing method that is dependent on the context free grammar reads the grammar rules directly from the manually labeled tree bank, and takes the relative frequency as the probability of the rule. Such kind of methods is quite easy to implement. However, previous work showed that the performance of these methods was not ideal. It is mainly due to the independence assumptions in the context free grammar, while these independence assumptions are often not established in the practical situation [6].

The dependency grammar was put forward in 1959. The dependency grammar is natural lexicalization that works directly in accordance with the interdependencies between the words [7]. As the interdependent nature of the words in the dependency grammar is corpus, while different languages are interlinked in the corpus level, hence the dependency grammar is a kind of syntactic theory that can objectively reveal the inherent laws of human language across the linguistic boundaries [8]. In the aspect of English, the work of dependency English writing has begun to be emphasized in recent years. The idea of partitioning is adopted, and a few rules of grammar developed are applied to first carry out partitioning processing to the sentence, so as to identify the lexical chunks with the fixed relationship. Then the dependency analysis is carried out on the entire sentence. Syntactic structure is the unity of syntactic form and corpus content. It is not only necessary to carry out form analysis on the syntactic structure, such as the syntactic hierarchy analysis, the syntactic relationship analysis as well as the sentence pattern analysis and so on, but also to conduct a variety of English writing teaching. The more comprehensive and the more profound the English writing teaching on the syntactic structure is, the more likely it is to make scientific and rational explanation to the various phenomena in the syntactic forms [9]. At present, in the lexicalized English writing, such as the headword driven English writing model, the dependency grammar only takes the corpus drive model between the words into consideration, instead of introducing more information that reflects the characteristics of the words corpus, such as the corpus type, the corpus collocation and other English corpora. However, the English corpora are crucial for the English writing and the corpus computation. For example, in the sentence "Astronomers saw stars with telescopes", the word "telescopes" is related both to its direct core word "with" in the corpus collocation and to the core word "saw" of the whole sentence. If the dependency analysis method is adopted, due to the constraints of the axiom of the dependency grammar, the dependency relationship cannot be

established between the "telescopes" and the "saw"; while this relationship is crucial for the English writing [10].

The existing mainstream English writing theories fail to effectively depict the essential characteristics of the English, which results in the relatively huge difference in the effect of the current English writing and the corpus calculation results compared with the English. In the English writing, the drive structure can describe the relationship between the English sentence syntactic structure and the corpus constituents relatively well. Therefore, it is necessary to investigate and study the formal grammar theory more systematically and extensively, especially the computer network corpus, and to establish the English writing model on this basis.

Most of the studies on the computer network corpus at present focus on the research of the drive characteristics of English without taking the drive structure of the whole sentence into consideration. It is expected that a kind of sentence drive structure can be defined, and this drive structure should be able to reflect the drive relationship between all the words in the sentence.

2   Computer Network Corpus

The computer network corpus, like the dependency grammar, is also considered to be derived from the linguistic thinking of the French linguist Tesnieree. In accordance with the description of Mr. Lu Jianming in the preface of the "Research on Modern English Drive Structure Corpus (Shenyang, Editor in chief Zheng Dingou), the term "Valency/valenz" (also known as "Drive"/"Direction") is borrowed from chemistry. In chemistry, the concdpt of "Valency" is used to illustrate the proportional relationship between the numbers of the atoms of each element in the molecular structure; while Tesniere's introduction of the concept of "Valency" into the syntactics is intended to show how many noun phrases that a verb can control. For example, "Eat" is a bivalent verb, which needs to control two noun phrases, to illustrate "Who eat" and "what to eat"; while "Give" is a trivalent verb, which needs to control three noun phrases, to illustrate "Who give", "Give to whom" and "What to give". It is easy to see that the computer network corpus and the sentence level corpus calculation (especially the corpus role labeling) are closely linked. At present, the studies on drive are not only limited to verbs, the drive of adjectives and nouns is also studied by a lot of people. For example, the adjective "Young" and the noun "Sister" are both univalent, which need to control a noun phrase respectively to illustrate "Who is young" and "Whose sister".

The corpus of computer network emphasizes that the verb is the center of a sentence; hence it is the verb centism. The drive of the verb is closely related to the generation of a sentence. And the key to generate or understand a sentence lies in the verb. The verb core stricture formed by verbs (or referred to as the predicate core structure) is the basis for the generation of a sentence. Any sentence is generated by the combination of the verb core structure with a certain syntactic structure into the sentence pattern by adopting the grammatical approaches and endowing it with pragmatic value. In accordance with the number of the germ-verbs that a verb is associat-

ed with (the compulsory corpus associated with the verbs), verbs can be classified into three categories, that is, the "Valency" classification of the verbs, including the monovalent verbs, the bivalent verbs and the trivalent verbs. The idea of drive hierarchy refers to dividing the monovalent concept into the drive hierarchy that is composed of four planes, including association, terms, bits and elements. "Association" refers to the number of different corpora roles that a verb can associate with in various sentences. "Term" refers to the number of nominal components (including nominal components guided by the prepositions) that a verb can be associated with in a sentence. "Bit" refers to the number of nominal elements that a verb can be associated with in a sentence without making use of the preposition. And "Element" refers to the number of nominal elements that a verb can be associated with in a simple basic sentence. The combination and control ability of the verbs at different levels and different syntactic frameworks can be fully reflected through this kind of hierarchy relationship.

The drive of the adjectives is classified. In accordance with the number of complement words of the predicate adjectives, it can be divided into monovalent, bivalent and trivalent; in accordance with the compulsion degree of the predicate adjectives to the complement words, it can be divided into required valency and optional valency; in accordance with whether there are marked prepositions in the complement words, it can be divided into marked valency and unmarked valency. In accordance with the degree of stability of complement words of the predicate adjectives, it can be divided into static valency and dynamic valency. On the basis of the direct effect of the studies on the drive of the English verbs, the studies on the drive of the English nouns are carried out. From the drive point of view, modern English nouns can be divided into two major categories, that is, the non-valent nouns (or zero valent nouns) and the valent nouns, which is classified according to whether the nouns have the drive requirement. And the valent nouns can be divided into two categories: One is derived from the predicate, and the other is not derived from the predicate. It often contains a downgraded predication structure, which can then be further divided into two minor categories, monovalent noun and bivalent noun, in accordance with the control ability.

Similarly, in the English writing, most of the valent nouns are monovalent nouns. For example, the word "food" and "volume" in the noun phrase "pet food volume" are both monovalent nouns. The word "food" is a modifier for the headword "volume", and the word "pet" is a modifier for the word "food" instead of the headword "volume". And the probability of the grammar rule NPB → NN NN NN is as the following:

$$P_l\left(L_i\left(l_i\right)\right)|H,P,h,L_1\left(l_1\right)\mathrm{L}\ L_{i-1}\left(l_{i-1}\right)=P_l\left(L_i\left(l_i\right)|P,L_{i-1}\left(l_{i-1}\right)\right) \tag{1}$$

$$P_l\left(R_i\left(r_i\right)\right)|H,P,h,R_1\left(r_1\right)\mathrm{L}\ R_{i-1}\left(r_{i-1}\right)=P_l\left(R_i\left(r_i\right)|P,R_{i-1}\left(r_{i-1}\right)\right) \tag{2}$$

Part of the valent nouns are bivalent nouns, such as the word "cream" in the noun phrase "vanilla icecream" is a bivalent noun, and both the word "vanilla" and "ice"

are the modifiers for the headword "cream". The probability of the grammar rule NPB → NN NN NN is as the following:

$$P_l\left(L_i\left(l_i\right)\right)|H,P,h,L_1\left(l_1\right)\mathrm{L}\ \ L_{i-1}\left(l_{i-1}\right)=P_l\left(L_i\left(l_i\right)|H,P,h\right) \tag{3}$$

$$P_l\left(R_i\left(r_i\right)\right)|H,P,h,R_1\left(r_1\right)\mathrm{L}\ \ R_{i-1}\left(r_{i-1}\right)=P_l\left(R_i\left(r_i\right)|H,P,h\right) \tag{4}$$

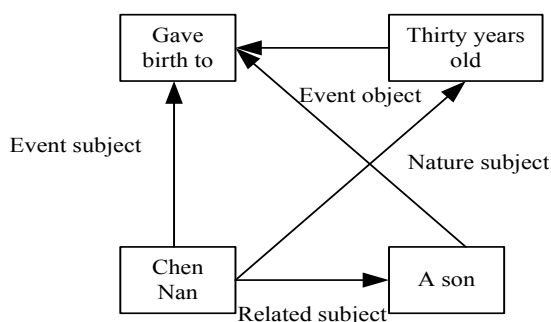## 3   English Writing Teaching Model

In the computer network corpus, possessor-subject possessee-object sentence refers to the sentence like "Wang Mian's father died when he was seven". Compared with the general sentence pattern, it can be seen that such syntactical structure has the following characteristics: (1) There is no direct corpus relation between the subject of the sentence and the verb, and it is not a necessary component of the corpus of the predicate verb, which is represented by the atypical nature of the subject type; (2) The object in the sentence is mostly the agent of predicate verb, which is represented by the atypical nature of the object type; (3) The relationship between the subject and the object is not dependent on the verb but on the "Possession-affiliation" relationship between the two components in the lexical corpus, and predicate verb (or the adjective) that accompanies this characteristic is monovalent (or mono directed).

The following figure provides the dependency tree and the proposed drive structure with the sentence "Chen Nan gave birth to a son when she was thirty years old". Among them, Figure 2 shows a possible form of the drive structure. Changes and improvement may also occur in this kind of structure during the research process. It can be seen that the dependency trees for the two sentences have the same form. However, they have different structures. Therefore, it can be seen that compared with the traditional phrase structure tree and the dependency tree, the drive structure has reflected more corpus characteristics. At the same time, the drive information of the words is relatively stable. Finally, from the form point of view, the drive structure is not a tree, but a directed graph. Therefore, the drive structure has even stronger expression power than the phrase grammar and the dependency grammar. In addition, it also has the potential to obtain the higher syntax corpus calculation performance. Since in the sentence "Chen Nan gave birth to a son when she was thirty years old", "Chen Nan" is zero valent, "son" and "thirty years old" are both monovalent. In addition, the "thirty years old" is the modifier for the verb "gave born to" as a time adverb. And the verb "gave birth to" is bivalent. With the drive information of all these words, it is possible to obtain the aforementioned drive structure with relative accuracy.

Syntactic structure is the unity of the syntactic form and the corpus content. For the syntax structure, it is not only necessary to carry out the formal analysis, such as the syntactic hierarchy analysis, the syntactic relation analysis, the sentence pattern analysis and so on, but also necessary to conduct a variety of English writing teaching.
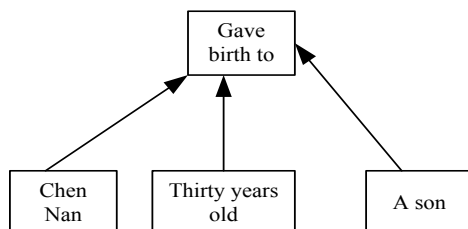
Fig. 1. Dependency tree of the sentence "Chen Nan gave birth to a son when she was thirty years old"

Fig. 2. A possible drive structure of the sentence "Chen Nan gave birth to a son when she was thirty years old"

The more comprehensive and more profound the English writing teaching for the syntactic structure, the more likely it is to make the scientific and rational explanation of the various phenomena in the syntactic form: And the basic ideas are as the following: On the basis of the structure of the sentence phrases or the structure of dependencies, the sentence drive structure is obtained through the analysis of the corpus dictionary. In turn, it makes use of the dependence relationship in the sentence drive structure to make the necessary correction. And the relatively more detailed approaches are as the following:

1. In accordance with the drive information of the words and the structure of the sentence phrases, some derivation rules of the sentence drive structure can be obtained as the following (For the discussion of the relationship between the sentence form and the word drive, please refer to book "Research on the English Drive Structure Corpus" by Yuanyu Lin):

   (a) Sentence phrase structure is an implementation of the word drive. And the number of the drive of the words must be met in the sentence structure.

   (b) For the words at the same level of a sentence, the headword can control the other words, while the words other than the headword shall be subjected to the control of a certain word at the same;

   (c) For the nouns at the same level of a sentence, the latter nouns can control the preceding ones.

2. In Collins's headword driven English writing model, when the probability of the grammatical rules is calculated, it is assumed that the modified components are independent of each other in the model. And in the model, in accordance with the English corpus of the drive relationship of the words that are extracted the drive structure; there shall be no independence from each other between the modified components that have the drive relationship.

In the English writing model, rich English corpus is introduced, which includes not only the corpus dependence information that is determined by the syntax tree or the dependency tree, but also includes the corpus collocation information that is determined by the drive structure tree diagram that is corresponding to the sentence drive tree.

In the following figure, the "Astronomers saw stars with telescopes" is used as an example to show the structure of the phrase structure tree and the drive structure that is envisaged. Among them, Figure (5) and Figure (6) are the diagrams of two different syntax trees. Figure (7) and Figure (8) show the possible drive structure form that is obtained from the analysis on the basis of the phrase structure tree. And it is possible that this kind of form may also be subjected to changes and improvement in the process of the studies.

More knowledge of the corpus can be obtained from the drive structure diagrams as shown in Figure (7) and Figure (8) in combination with the syntax tree diagrams as shown in Figure (5) and Figure (6): There is the corpus drive model between the word "telescopes" and the word "with" in the sentence drive tree as shown in Figure (5), while there is also a corpus collocation relationship with the word "saw" at the same time. However, in the sentence drive tree as shown in Figure (6), there is the corpus drive model between the word "telescopes" and the word "with", while there is a corpus collocation relationship with the word "stars" at the same time.
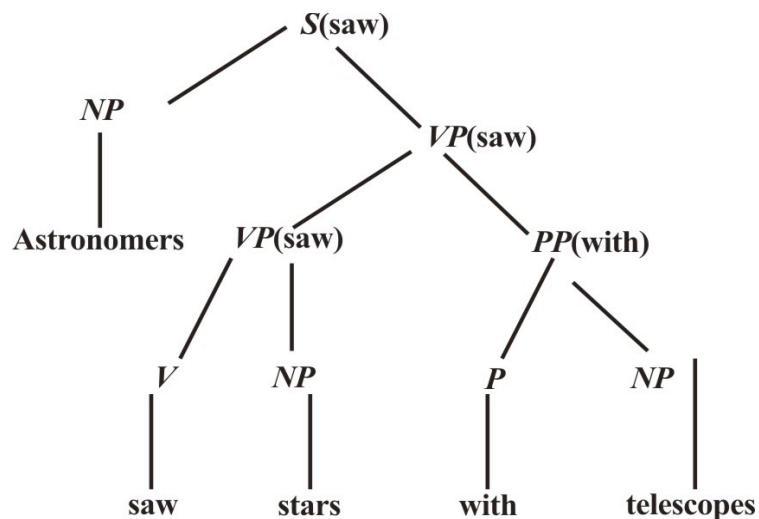
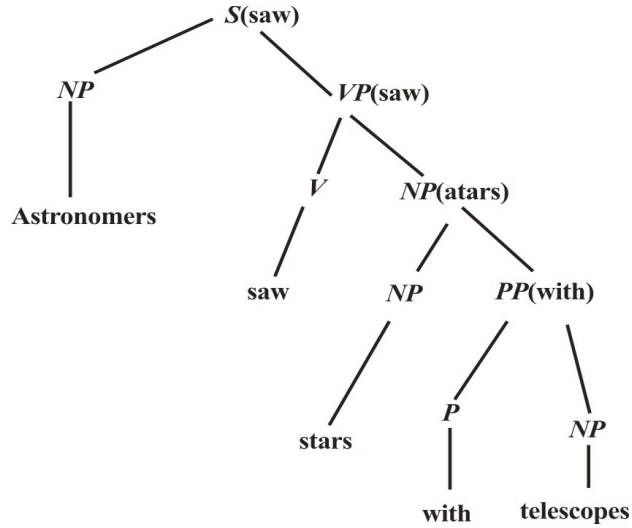**Fig. 3.** Sentence "Astronomers saw stars with telescopes" driver tree 1

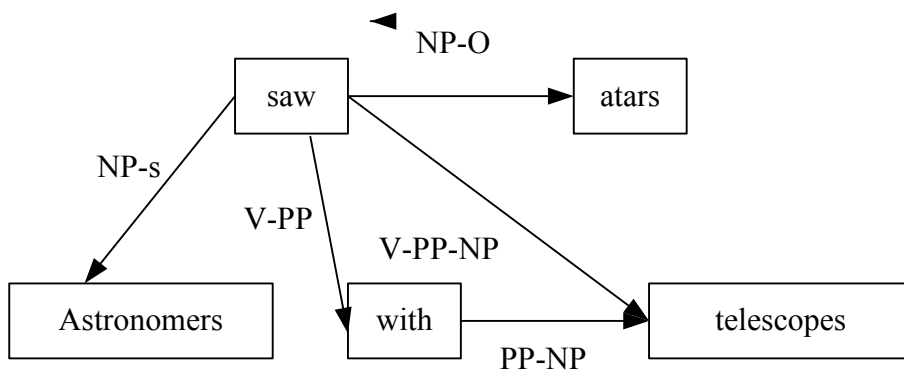**Fig. 4.** Sentence "Astronomers saw stars with telescopes" drive tree 2

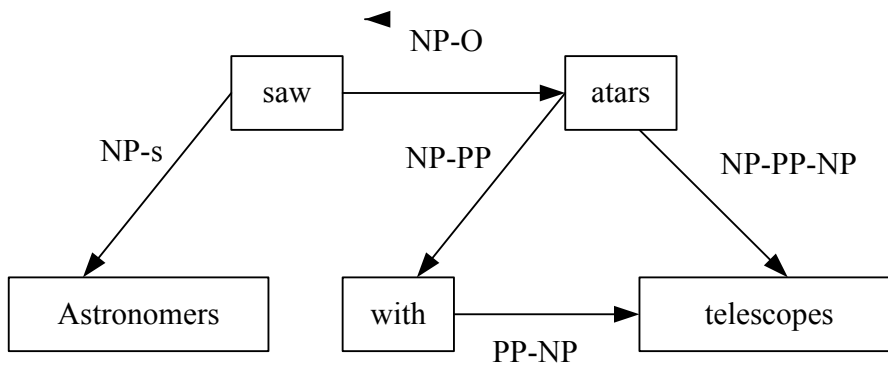**Fig. 5.** Drive structure corresponding to the sentence drive tree 1

Fig. 6. Drive structure corresponding to the sentence drive tree 2

4 English Writing Teaching Model Dependent on the Computer Network Corpus Drive model

4.1 Computer Network Corpus Driven

The headword driven English writing model is the most representative lexicalization model. In order to give full play to the role of the lexical information, the headword drive model introduces the core word / part of speech information for each and every non terminal in the grammatical rule. And due to the introduction of the information of the words and phrases, it will be inevitable that the serious sparsity problem will take place. In order to alleviate this problem, the headword driven mode breaks down the right hand side of each grammar rule into three major parts, which include: a central component; a number of modifier components to the left of the central component; a number of modifier components to the right of the central component, respectively. In a formal manner, the aforementioned major parts can be written in the following form:

$$P_l\left(ht, hw\right) - L_m\left(lt_m, lw_m\right) \mathrm{L}\ L_1\left(lt_1, lw_1\right) H\left(ht, hw\right) R_1\left(rt_1, rw_1\right) \mathrm{L}\ R_n\left(rt_n, rw_n\right) \qquad (5)$$

In which, P stands for the non terminal symbol, H stands for the central component, R_1 stands for the modifier components to the right. Hw, lw and rw stand for the central words of the components, respectively. And ht, lt and rt stand for their parts of the speech, respectively. Further assumptions are made. First of all, the core component H is generated by P, and then all the modifier components on the left and right sides are generated independently with H as the center, respectively. In this way, the probability of the grammar rule with the form as shown in the equation (1) is as the following:

$$P_h\left(H\middle|\left(ht, hw\right)\right) \mathrm{o} \prod_{i=1}^{m+1} P_i\left(L_i\left(lt_i, lw_i\right)\middle|H, P, h, \mathrm{V}_l(i-1)\right) \mathrm{o} \prod_{i=1}^{n+1} P_i\left(R_i\left(rt_i, rw_i\right)\middle|H, P, h, \mathrm{V}_l(i-1)\right) \quad (6)$$

In which, L_{m+1} and R_{n+1} stand for the stop symbols on the left and right sides, respectively. $\mathrm{V}_l(i-1)$ stand for the distance function, and the compensation structural information is absent. The distance information takes three kinds of situations into consideration: 1 Whether there is a component that is preceding the component; 2 Whether there is a verb that appears preceding the component; 3 Whether there is punctuation mark that appears preceding the component.

4.2 English Writing Model Driven by Computer Network Corpus

Let $P(h)$ stand for the core words at the upper hierarchy that the current core word h in the syntax tree is dependent on, and the other symbols stand for the same as the above section. In the English writing model, each grammar rule can be written in the form as the following:

$$P_h\left(ht,hw\middle|P(h)\right)-L_m\left(lt_m,lw_m\right)\cdots L_1\left(lt_1,lw_1\right)\circ H\left(ht,hw\middle|P(h)\right)R_1\left(rt_1,rw_1\right)\cdots R_n\left(rt_n,rw_n\right) \quad (7)$$

The probability of the grammar rule in the form as shown in the equation (7) is as the following:

$$P_h\left(H\middle|(ht,hw),P(h)\right)\circ\prod_{i=1}^{m+1}P_i\left(L_i\left(lt_i,lw_i\right)\middle|L_{i-1}\left(lt_{i-1},lw_{i-1}\right),\mathrm{L}\ ,L_1\left(lt_1,lw_1\right),(ht,hw)P(h)\right)$$
$$\circ\prod_{i=1}^{n+1}P_i\left(R_i\left(rt_i,rw_i\right)\middle|R_{i-1}\left(rt_{i-1},rw_{i-1}\right),\mathrm{L}\ ,R_1\left(rt_1,rw_1\right)(ht,hw)P(h)\right) \quad (8)$$

In which, L_{m+1} and R_{n+1} stand for the stop symbols on the left and right sides, respectively. And the probability in the equation (8) is as the following

$$P_i\left(R_i\left(rt_i,rw_i\right)\middle|R_{i-1}\left(rt_{i-1},rw_{i-1}\right),\mathrm{L}\ ,R_1\left(rt_1,rw_1\right)(ht,hw)P(h)\right) \quad (9)$$

It can be broken down into two probabilities as the following

$$P_i\left(rt_i\middle|rt_{i-1},rt_{i-2},\mathrm{L}\ ,rt_1,ht,rw_i\right) \quad (10)$$

$$P_i\left(rw_i\middle|rw_{i-1},rw_{i-2},\mathrm{L}\ ,rw_1,hw,P(h)\right) \quad (11)$$

The product of the above two probabilities are denoted as $S(rw_i)$, which stand for the word $rw_{i-1},rw_{i-2},\mathrm{L}\ ,rw_1,P(h)$ has the related words with the corpus collocation relationship with the current word rw_i (which is determined by the drive structure diagram that is corresponding to the sentence drive tree) as the following:

$$P_i\left(rw_i\middle|rw_{i-1},rw_{i-2},\mathrm{L}\ ,rw_1,hw,P(h)\right)=P_i\left(rw_i\middle|hw,\mathrm{V}_r(i-1)S(rw_i)\right) \quad (12)$$

It is further assumed that hw, $S(rw_i)$ is dependent on the condition of rw_i as the following:

$$P_i\left(rw_i\middle|hw,\mathrm{V}_r(i-1)S(rw_i)\right)=\frac{P_i\left(rw_i\middle|hw,\mathrm{V}_r(i-1)\right)\circ P_i\left(rw_i\middle|S(rw_i)\right)}{P_i\left(rw_i\right)} \quad (13)$$

In the equation (13), the probability $\dfrac{P_i\left(rw_i\middle|\Phi(rw_i)\right)}{P_i\left(rw_i\right)}=\dfrac{P_i\left(rw_i,S(rw_i)\right)}{P_i\left(rw_i\right)\circ P_i\left(S(rw_i)\right)}$ is established, that is, rw_i, $S(rw_i)$ mutual information, therefore, the probability of the overall equation (13) is very clear, and it is in line with the linguistic phenomenon.

It can be said that the independent hypothesis that is assumed in the lexicalized context free grammar is not applicable to the language phenomenon, which is neither suitable for the language, not suitable for the English, more importantly. In the Eng-

lish writing model, the conditional independence assumption is adopted to take the place of the conditional independence hypothesis in the headword drive model of the English writing. From the statistical point of view, it is the assumption of the independence in the relative condition. And the assumption of the independence is excessively strong in the hypotheses, and it is not consistent with the language phenomena either. Therefore, the English writing model is more in line with the actual physical process of the language. Through the conduction of decomposing and modification of the rules of the Collins model, the lexicalized English writing model that is dependent on the drive structure and in combination of the headword drive mode (including the corpus dependent information that is determined by the syntax tree, and also including the corpus collocation information that is determined by the drive structure diagram corresponding to the sentence drive tree) and other language knowledge, the precision of the English writing can be improved.

5 Instance Analysis

The test data is extracted from the Penn Treebank Tagset (CHTB) version 5.0, and the majority of the materials are collected from the news of the Xinhua News Agency, the Sinorama news magazine and the Hong Kong News. CTB is a corpus that is publicly published by the Language Data Alliance (hereinafter referred to as LDC for short), which has provided a public training and testing platform for the English writing. The tree bank includes a total of 507,222 words, 824,983 Chinese characters, 18,782 sentences and 890 data files. In order to achieve the balance among the various corpora sources in the training set, the development set and the test set, the corpus is segmented into the following parts: The files 301 ~ 320 and 611-630 are taken as the debug set, the files 271 ~ 300 and 631-660 are taken as the test set, and the rest of the files are taken as the training set. In all the experiments in this paper, all the parameters of the model are obtained from the estimation in the training set by adopting the maximum likelihood method.

Four evaluation indicators that are commonly used are applied in the test results, that is, the precision rate, the recall rate, the comprehensive index value *F* and the crossed brackets CB. And the definitions of them are as the following

Precision is an index that measures the proportion of the correct components in all the components that are analyzed in the English writing system. And recall is an index that measures the proportion of all the correct components in the actual components that are analyzed in the English writing system.

Composite index: $F = (P \times R \times 2)/(P + R)$.

Crossed brackets CB: This index provides the average number of the components in one tree that are intersecting with the boundary of the components of the other trees.

The Baseline system used in the experiment is DanielM.

DBParser is implemented by the Bikel relying on the Collins mode. And Table 1 sets out the experimental results of the baseline system and the improved model of the English writing.

It can be seen from Table 1 that, both the corpus based dependency information that is determined by the syntax tree or the dependency tree and the corpus collocation information that is determined by the driver structure diagram that is corresponding to the sentence drive tree are used in the decomposition of the rule and the calculation of the probability. Therefore, the precision rate P , the recall rate R , the comprehensive index value F and the crossed brackets of the improved model have been significantly improved when it is compared with the headword driven English writing in the Collins. The experimental results show that the application of the knowledge of the linguistic features has shown a great effect on the writing of the statistical English. This has also pointed out one direction of the statistical English writing in the English from the side: To look for more grammatical, corpus, pragmatic and other characteristic knowledge from the perspective of linguistics.

The problem of data sparseness is another important factor that can seriously affect the performance of the English writing system. The improved model adopts the smoothing techniques which includes the dependent corpus type and the variable length model. It has successfully solved the problem of the data sparseness and greatly improved the system performance. The English writing experimental results after the smoothing techniques are applied are shown in Table 2. And a more detailed discussion in this area will be published separately.

Table 1. Experimental results of the English writing

Model	Precision%	Recall%	F%〗	Crossed brackets
Baseline	82.76	80.17	81.44	2.05
Improve model	86.13	85.21	85.66	1.83

Table 2. Experimental results of the English writing

Model	Precision%	Recall%	F%〗	Crossed brackets
Baseline	82.76	80.17	81.44	2.05
Improved model after the smoothing technique is adopted	86.13	85.21	85.66	1.83

6   Conclusions

In English writing teaching, the computer network corpus drive model can describe the syntactic structure and corpus composition of the English sentences relatively well. Therefore, the related theories of computer network corpus are systematically investigated and studied in this paper. And on this basis, a kind of computer network corpus drive model is established. In the English writing model that is dependent on computer network corpus drive model, for the decomposition of the rule and the calculation of the probability, not only the corpus dependent information that is determined by the syntax tree or the dependency tree is utilized, the corpus collocation information that is determined by the driver structure diagram corresponding to the sentence drive tree is also used, so that the performance has been significantly im-

proved. Most of the studies on the computer network corpus at present focus on the drive characteristics of the English without taking the structure of the whole sentence into consideration. It is expected that a kind of sentence drive structure can be defined. And such kind of drive structure should be able to reflect the drive relationship between all the words in sentence. And the drive relationship that is expected to be marked is not only involving the direct relationship between the verbs and the noun phrases, but is also involving the relationship between the nouns and the noun phrases, the adjectives and the noun phrases, and even the adverbs and the verb adjective phrases, that is, the drive structure should be a complete, syntactic structure.

7　References

[1] Xia, K. (2015). Exploration on the scaffolding instruction mode in higher vocational oral english teaching guided by the constructivism theory. Bioresource Technology, 191(1), 146-156.

[2] Salerno, S. M., Jackson, J. L., & O'Malley, P. G. (2010). Interactive faculty development seminars improve the quality of written feedback in ambulatory teaching. Journal of General Internal Medicine, 18(10), 831-834. https://doi.org/10.1046/j.1525-1497.2003.20739.x

[3] Chen, H., & Chen, H. (2015). The influence of world englishes on chinese english and english teaching in college. Journal of Organometallic Chemistry, 192(1), 1-15.

[4] Brunner, B. J. (2013). Is the trend of using apostrophe's to form the plural's of noun's in english of concern to professor's of writing?. Oceanology, 53(3), 352-364.

[5] Rahayu, A. D. (2015). Grammatical errors made by the freshmen of the english department of state university of malang in paragraph writing. Febs Letters, 351(3), 433–436.

[6] Sugiono, M. (2014). An error analysis in using english tenses on english sentence writing of the second year students of sltp negeri 7 jember in the academic year of 2001/2002 /oleh muhammad sugiono. Macromolecules, 15(3), 831-834.

[7] Ip, N., & Lee, J. (2015). Helping students find the 'write' way: an evaluation of a new english-in-the-discipline course in dissertation writing. Journal of Cardiovascular Electrophysiology, 14(4), 443-445.

[8] Turnbull, S. J., Jones, A. E., & Allen, M. (2010). Identification of the class characteristics in the handwriting of polish people writing in english. Journal of Forensic Sciences, 55(5), 1296-1303. https://doi.org/10.1111/j.1556-4029.2010.01449.x

[9] Anna, C., & Jane, V. (2010). The evaluation of undergraduate students' written english language skills. Medical Education, 34(8), 642-647.

[10] [Iñiguez, G., Barrio, R. A., Kertész, J., & Kaski, K. K. (2011). Modeling opinion formation driven communities in social networks. Computer Physics Communications, 182(9), 1866-1869. https://doi.org/10.1016/j.cpc.2010.11.020

8　Author

Lei Shi is with the Shaanxi Institute of International Trade & Commerce, Xi'an, China.