Association Relationship Analyses of Stylistic Syntactic Structures

Haiyan Wu and Ying Liu*

Department of Chinese Language and Literature,
Tsinghua University, Beijing, China
wuhy17@mails.tsinghua.edu.cn
yingliu@mail.tsinghua.edu.cn

Abstract. Exploring linguistic features and characteristics helps better understand natural language. Recently, there have been many studies on the internal relationships of linguistic features, such as collocation of morphemes, words, or phrases. Although they have drawn many useful conclusions, some summarized linguistic rules lack physical verification of large-scale data. Due to the development of machine learning theories, we are now able to use computer technologies to process massive corpus automatically. In this paper, we reveal a new methodology to conduct linguistic research, in which machine learning algorithms help extract the syntactic structures and mine their intrinsic relationships. Not only the association of parts of speech(POS), but also the positive and negative correlations of syntactic structures, linear and nonlinear correlation are considered, which have not been well studied before. Combined with the linguistic theory, detailed analyses show that the association between parts of speech and syntactic structures mined by machine learning method has an excellent stylistic explanatory effect.

Keywords: Syntactic Structure · Part of Speech Collocation · Positive and Negative Correlation · Linear and Nonlinear Relationship · Machine Learning.

1 Introduction

In recent years, many scholars have summarized linguistic rules or features from different levels, as well as their internal relationships [12, 3, 13, 4, 26, 7]. For example, Dexi Zhu pointed out that the constituent units of the style are morphemes, words, phrases and sentences, and how morphemes form words, how words form phrases and how they form sentences [6, 21]. The process is concluded as a collocation relationship between features. As early as 1957, Firth [8] proposed the

* Corresponding Author
This work is supported by 2018 National Major Program of Philosophy and Social Science Fund "Analyses and Researches of Classic Texts of Classical Literature Based on Big Data Technology" (18ZDA238) and Project of Humanities and Social Sciences of Ministry of Education in China (17YJAZH056)

concept of collocation. He believed that the words near a word can reflect the meaning of it. Some researchers regarded the word collocation as an abstraction in syntactic level, and others thought it was vocabulary level [8]. Most of the current methods studying the linguistic features [9, 1, 19] are based on the experience of linguistic knowledge and the analyses of small-scale data. With the development of computer technology, we now can perform automatic large-scale analysis on syntax. Although its performance is not as good as linguistics experts, it is believed that in the massive corpus, some statistical conclusions can be drawn. Generally, associated relationship mining can be divided into the following two aspects.

• **Linguistic Methods**. It is regarded that linguists are the pioneers of the theory of language association. For example, as the study of the common collocation of words in Red Sorghum, WenZhu found out the distribution, semantic changes, and rhetoric effect of collocation variation in five basic syntactic structures [20]. Xianghui Cheng[26] pointed out that a register was a collection of language characteristics people used in a specific scene, including vocabulary, syntax, and rhetoric.

Linguistic-based researches usually have manually selected examples and word-for-word analyses. The advantage is that there is an excellent theoretical basis. The shortcoming is that these studies lack the verification on the large-scale corpus.

• **Statistical Methods**. With the development of computer technology, many scholars have applied computational methods to the field of linguistics research. Such as, Seretan designed the hybrid system that combined statistical methods to do multilingual parsing for detecting accurate collocational information. Experiments showed that for different languages, this system had good results [17, 14]. Besides, Seretan thought that automatic acquisition of collocations from text corpora by machine learning was essential. Then he proposed a framework for collocation identification based on syntactic criteria. It was shown that the results obtained were more reliable than those of traditional methods based on linear proximity constraints [15, 16]. Xiaoli Huo considered the stylistic features were related to the specific linguistic material information. It is necessary to express the characteristics of different registers through lexical selection, phrase construction, sentence pattern transformation, and tone adjustment, to meet the communication needs [24].

However, these methods usually are specific for some features or materials, which are not flexible enough to generalize different features of the different registers.

In this work, we focus on the collocations of POS and syntactic structures. Instead of manually annotating, we utilize machine learning methods on the corpus to help analyze the features. The results verify some interpretation of linguistic theories. Furthermore, some potential laws or internal collocations are explored. The process of our study and experiments can be concluded as a new methodology of verifying linguistic theories with the help of computer technology, which is conducive to the development of linguistic theories.

2 Algorithms

2.1 Pearson Correlation Coefficient

In this paper, we employ *Pearson Correlation Coefficient* [2] to calculate the positive and negative correlation between syntactic structures. This algorithm is used to measure whether two variables have a linear correlation or not.

$$\gamma = \frac{\sum_i (x_i - \bar{x_\iota})(y_i - \bar{y_\iota})}{\sqrt{\sum_i (x_i - \bar{x_\iota})^2}\sqrt{\sum_i (y_i - \bar{y_\iota})^2}} \tag{1}$$

Where $\bar{x_\iota}$ and $\bar{y_\iota}$ represent the mean of sample X and sample Y respectively. Here, x and y can denote occurrence of different POS or syntactic structures. In following sections, we calculate the relationship between POS and the syntactic structures, and use *Heatmap* to visualize their correlation coefficients (Fig.2 and Fig.3). In *Heatmap*, the deeper the color is, the higher their correlation is.

2.2 Hierarchical Cluster

To find the linear correlation of POS and syntactic structures, we use a clustering method *Hierarchical Cluster* [11], which brings together attributes with higher correlation. Euclidean distance is used as the distance metric, shown as Formula 2.

$$\|a - b\|_2 = \sqrt{\sum_i (a_i - b_i)^2} \tag{2}$$

Where a and b represent the vector representations of two different features (POS or the syntactic structures)respectively. The clustering results are visualized together with the *Heatmap*.

2.3 LassoCV Algorithm

Most of the previous methods focus on the linear correlation of syntactic structures. But nonlinear relations are also crucial in the study of stylistic features. Therefore, we use *LassoCV* algorithm to consider the nonlinear relationship of syntactic structures. *LassoCV* algorithm judges and eliminates the collinearity between features by adding the penalty function[18], and the loss function is as follows in Formula 3.

$$\frac{\|y - Xw\|_2^2}{2n} + \gamma\|w\|_1 \tag{3}$$

where w are the parameters, n is the total number of samples, and γ is the hyper-parameter of L2-regularization. In our paper, due to differentiate three registers, we choose *MultiTaskLassoCV* which is the extension algorithm of *LassoCV*.

For all the above algorithms, we use *Python* and *Scikit-Learn Toolkit* .

https://seaborn.pydata.org/generated/seaborn.heatmap.html
https://scikit-learn.org/

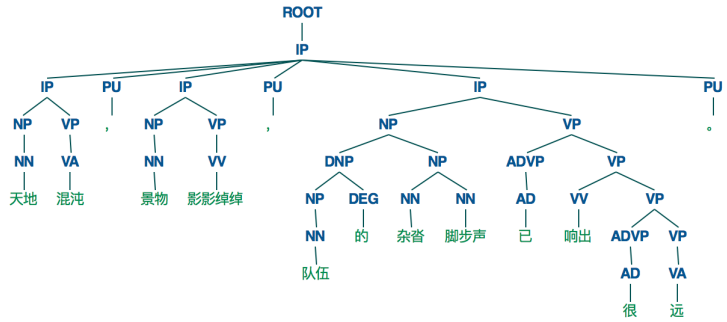

Fig. 1. Syntax Tree

2.4 Relative Definition

To better describe the syntactic structure, we give such definitions. Take VP → [VV,AS] as an example, VP on the left of the arrow is the **parent (father) node**, VV and AS on the right of the arrow are the **child nodes**, and the number of child nodes is called the **degree** of the syntactic structure VP →[VV,AS] or the **out-degree** of the phrase VP. Here, the out-degree of the phrase VP is 2.

3 Dataset and Statistic Information

3.1 Datasets

Our experiments are based on *Novel*, *News*, and *TextBook*. The detailed information is shown as follows.

• **Novel** comes from Mo Yan and Yu Hua, a total of 20 articles, of which 12 are Mo Yan and 8 are Yu Hua.

• **News** is a public dataset, which covers ten main topics including domestic and foreign news, stock news, financial news, breaking news, entertainment news and so on.

• **TextBook** consists of some Chinese textbooks in elementary, middle, and high schools. The total of 480 articles is mainly proses and novels.

3.2 Syntactic Structures

Avram Noam Chomsky pointed out that each sentence should conform to its syntactic rules[5]. Since we study the association of syntactic structures, we use *Syntax Tree* to convert each sentence in the corpus into a corresponding syntactic structure representation, as shown in Fig. 1.

For each sentence, we use *Stanford CoreNLP* to construct its corresponding syntax tree. Take a sentence from *Novel* as an example, e.g. "天地混沌，景物影影绰绰，队伍的杂沓脚步声已响出很远。(It is too dark to see the scenery, and

https://www.sogou.com/labs/resource/cs.php
https://github.com/stanfordnlp/CoreNLP

the footsteps of the team have been far away.)", whose corresponding syntactic tree is shown in Fig. 1. In Fig. 1, the syntax structures are extracted IP → [IP, PU, IP, PU,IP,PU], IP →[NP,VP], IP → [NP, VP], IP → [NP, VP], NP → NN, VP → VA, VP → VV, NP →[DNP, NP] and so on. The abbreviations of these syntactic structures are marked by the Pennsylvania tree library [25]. The meaning of the abbreviation of the parent nodes are shown in Table 1.

Table 1. Penn Treebank Tags for Phrase Structures

Tag	**Description**	**Tag**	**Description**
ADJP	Adjective phrase	LCP	Phrase formed by "XP+LC"
ADVP	Adverbial phrase headed by AD	LST	List marker
CLP	Classifier phrase	NP	Noun phrase
CP	Clause headed by C (complementizer)	PP	Preposition phrase
DNP	Phrase formed by "XP+DEG"	PRN	Parenthetical
DP	Determiner phrase	QP	Quantifier phrase
DVP	Phrase formed by "XP+DEV"	UCP	Unidentical coordination phrase
FRAG	Fragment	VP	Verb phrase
IP	Simple clause headed by INFL	...	...

By analogy, the same operation is performed for each sentence of each register, and the statistical information is shown in Table 2.

Table 2. Dataset Statistic Information

Dataset	**Novel**	**News**	**TextBook**
Total # of Sentences	117,353	29,754	24,119
Average Length of Sentences	19.36	27.85	17.41
Total # of Syntactic Structures	2,371,399	800,338	431,792
# of Different Syntactic Structures	5,404	2,502	3,188
# of Different NP-Phrase	1,570	707	1,261
# of Different VP-Phrase	1,612	769	821
# of Average Species	15.5081	20.0250	14.3592
Average Height	9.6474	11.3023	9.3987
Average Width	19.1002	27.2606	17.2163
Syntactic Richness	0.5934	0.5855	0.6031

From Table 2, we find that the number of sentences in *News* is close to *TextBook*, but the number of syntactic structure in *News* is larger than that in *TextBook*. It is because the sentence length of *News* is the longest, which leads to more syntactic structures in the sentence. Besides, since NP and VP phrases appear in large numbers in our corpus, we count their number for further analyses.

4 Association Relationship Mining

In a syntax structure, children POS co-occurrence can reflect the meaning of their father nodes. The positive and negative collinearity of POS and syntax structures reflect the characteristics of the language.

4.1 Parts of Speech(POS) Collocation Mining

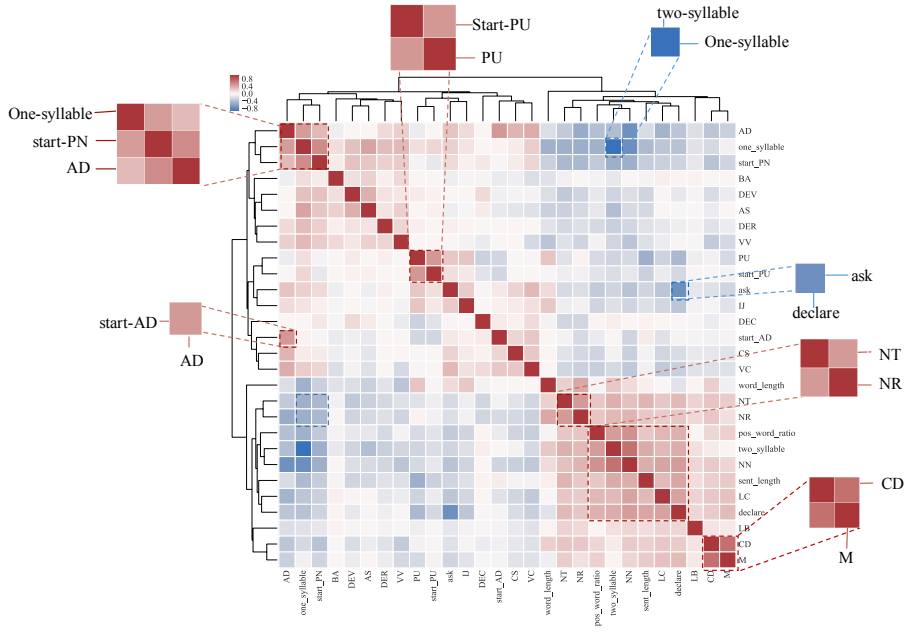

Fig. 2. POS Collocation Relationship

To study the syntactic structure, we utilize the hierarchical clustering in Section 2.2 to cluster the correlation values of POSs. The result is shown in Fig. 2, in which the meaning of these acronyms corresponds to the mark of the Pennsylvania Tree Library[22]. In Fig. 2, we conclude that some POSs have higher associated values, e.g., *CD+M*, *AD+VV*, *VA+DEC*, *VC+SP*, *P+NN*, *MSP+VV+DEC/DEG*, *NP/VV+AS*, *VV/NN+SP*, *NT+NR*, *VV + P + NN/NR + LC*, etc.

We can find that the POSs having positive collocations usually appear in the same syntactic structure. Using the collocation of *VV + AS* as an example, in Chinese, we often encounter the following sentence represents a certain state [23], e.g. "VP + 了" ("下雨了"(it rained)). In this case, only if *VV* and *AS* appear at the same time can they indicate a certain state. They form a syntactic structure VP → VV AS. The linear relationships between the POSs recover some specific syntax structures. It can help us understand the internal association of the syntactic structures.

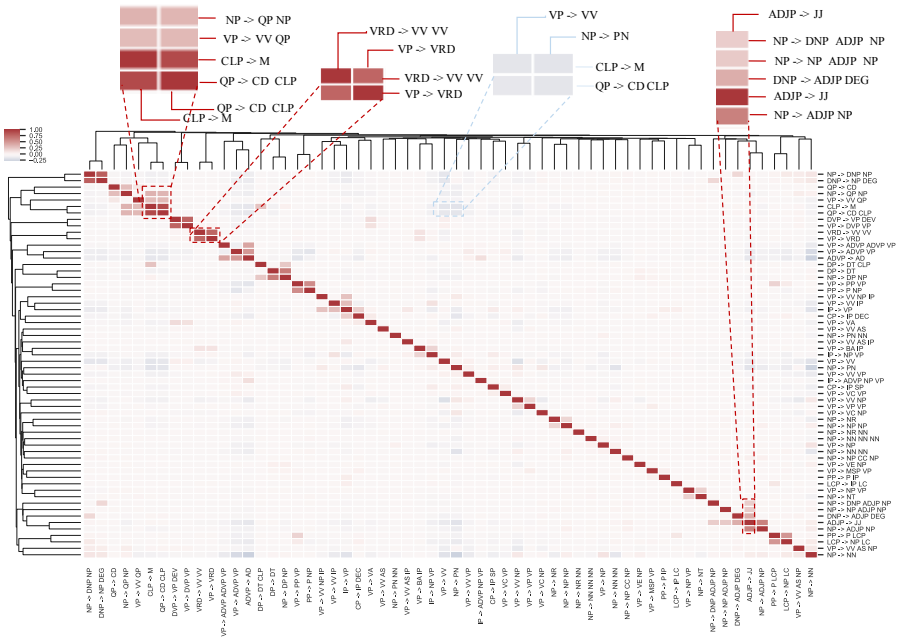

Fig. 3. Association Relationship of Syntactic Structures

4.2 Positive and Negative Correlation Mining

The internal relations of the syntactic structure are from the POS, which have been studied by many linguists. What interests us are the relationship between the syntactic structures, which has not been well studied. We consider the two groups of correlations between syntactic structures: positive and negative, linear and non-linear. Similarly, we use *Hierarchical Clustering*. The result is shown in Fig. 3.

In Fig. 3, generally, the correlation between syntactic structures is not as strong as POSs. But there are still some associations between syntactic structures. Taking ADJP $\rightarrow$ JJ as an example, it has a positive relationship with NP, DNP (upper right in Fig. 3). ADJP $\rightarrow$ JJ is a sub-tree of the associated syntactic structures. For example, ADJP $\rightarrow$ JJ is an extention of the ADJP in the NP $\rightarrow$ [DNP, ADJP]. Other correlations between syntactic structures are similar, and their co-occurrences usually reflect specific sentence patterns.

Since VP and NP phrases are the most, it is essential to analyze them better. We calculate the correlation of different syntactic structures with NP $\rightarrow$ NN, shown in Fig.4. The red part indicates negative correlations, and the blue part indicates positive correlations. We only list a part based on the value of the correlation value. The positively related structures are determiner or adjectival phrases and verb-object constructions, which usually co-occur with NP phrase.

Similarly, we list the positive (regular font) and negative (bold italics) related phrases of the other syntactic structures with the largest number of categories, shown in Table 3.

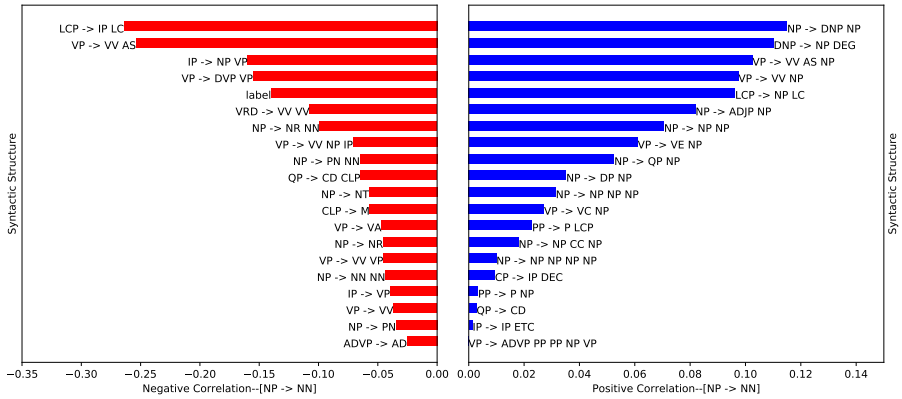

Fig. 4. NP → NN Related Syntactic Structures

Table 3. Positive and Negative Correlation of Syntactic Structures

VP→VV	**IP→NP VP**	**QP→CD**
VP→DVP VP	NP→PN	NP→QP NP
VP→VV VP	VP→VV AS IP	NP→NP PRN
VP→VP VP	CP→IP DEC	NP→NR PRN
VP→MSP VP	VRD→VV VV	NP→NP NP NP
VP→VV NP IP	LCP→IP LC	NP→NP CC NP
VP→VA	***IP→VP***	***NP→PN***
NP→QP NP	***CLP→M***	***ADVP→AD***
VP→VV AS NP	***PP→P NP***	***VP→VV***
QP→CD CLP	***ADVP→AD***	***IP→VP***
DNP→NP DEG	***DNP→NP DEG***	***NP→DNP NP***
NP→DNP NP	***QP→CD CLP***	***CP→IP DEC***
NP→NN	***LCP→NP LC***	***IP→NP VP***
P→VV NP	***NP→DNP NP***	***VP→VA***

Through statistical analysis, we draw the following two conclusions:

1. Positively related syntactic structures usually can form a larger syntactic tree, and usually are father and children nodes;
2. Negatively related syntactic structures are conflicting and usually can not be in the same syntax tree.

4.3 Linear and Nonlinear Correlation Mining

Since the positive and negative correlation of the syntactic structures is based on *Pearson Correlation Coefficient*, it is a linear correlation. However, for complex registers, it is necessary to study the nonlinear relationship between syntactic structures. Next, we use *MultiTaskLassoCV* in Section 2.3 to get the nonlinear correlation of the syntactic structure as shown in Table 4.

From Table 4, we use black bold italics to represent nonlinear relationships, and the rest are linear correlations. We find that most of the syntactic structures

Table 4. Linear and Non-Linear Relationship of Syntactic Structures

NP→NN NN	**NP→NP NP**	**NP→NR**	**VP→VV AS**
NP→NT	NP→NP CC NP	NP →NR NR NR	NP→QP NP
NP→NR PRN	VP→VV VP	VP→NP VP	NP→NP PRN
VP→VV AS NP	VP→VP VP	VP→VV	NP→NR PRN
NP→NN	VP→MSP VP	VRD→VV VV	NP→NP NP NP
LCP→NN LC	NP→ADJP NP	LCP→NP LC	NP→NP CC NP
ADVP→AD	VP→VA	NP→PN	NP→NN PN
NP→PN	NP→QP NP	CLP→M	ADVP→AD
VP→VA	VP→VV AS NP	PP→P NP	VP→VV
NP→DT DEG	DNP→NP DEG	DNP→NP DEG	NP→DNP NP
VP→VV VP	NP→DNP NP	QP→CD CLP	NP→NR
LCP→IP LC ***CP→IP DEC***	***IP→NP VP*** ***PP→P IP***	***IP→VP***	***IP→IP ETC***

are linearly related, and the nonlinear correlations are the phrases of IP and CP, as shown in Fig.5. We take IP phrases as an example to analyze.

In Fig.5, both sentences belong to the IP phrase, and in which IP phrase can be expressed in these forms: IP→[IP, PU], IP→[NP, VP], IP→VP, IP→[NP, VP, PU,PU] ,and IP→[PU,VP],etc. In Fig.5, since some IP phrases are located in the same layer and some are nested in the inner layer. Recursive loops can occur in the tree to form complex IP phrases. Therefore, the relationship between such IP phrases is non-linear.

By statistical analysis, *Novel* contains *IP*, *CP*, *NP*, etc. and their corresponding percents are 77.59%, 7.7%, 6%. *News* includes 84% *IP*, 5% *CP* and 7.6% *NP*, etc. Similarly, *Textbook* mainly includes *IP*, *CP* and *NP*, and the corresponding percents are 74.55%, 8.93%, 6.43%. For the phrases of *IP*, *CP* and *NP* in these three registers, *IP* in *Novel* is mainly IP → VV, of which most are momentary verbs e.g. "笑"(Smile), "说"(Say); *IP* of *News* is mainly IP → [NP,VP] and IP →[NP,NP], e.g. "政府指出 (Government Points)……", "市场经济 (Market economy)……"; and *Textbook* is IP → VP, e.g. "踢这个蓝色的球"(Kick this blue ball). For the three registers, whether from the height of *IP* in the syntax tree or the length of the sentence involved, *IP* phrase of *News* is the highest, with an average height of 2.8 (*TextBook*:1.84;*Novel*: 1.81). Usually, the higher the syntax tree is, the more intermediate nodes restrict the central words. Besides, The average length of the sentence involved in *IP* phrase of *Novel* is 9.19, *News* is 15.92, *TextBook* is 12.83. Combined with the lengths, we can conclude that *News* contains the most restrictive words followed by *TextBook*, and *Novel*.

Except for the phrases *IP* and *CP*, the other structures usually have linear correlations, and the results are consistent with the clustering results.

5 Linguistic Relevance Analyses

In the previous sections, we analyze the relationship of syntactic structures from the perspective of machine learning. In this part, we will explain the conclusions

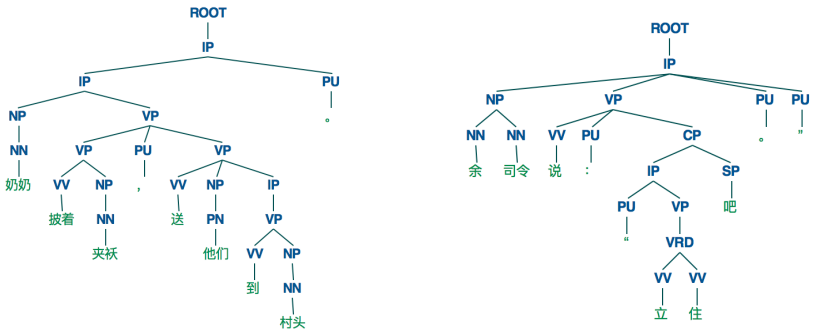

Fig. 5. Syntax Trees of IP and CP

we have obtained from the registers itself. Combined with Fig. 3, Table 3, and Table 4, we choose the following sets of syntactic structures for analysis and explanation. These groups of structures come from *Novel*, *News*, and *TextBook*.

1. LCP →[NP,LCP] + PP →[PP,LCP], from the linguistic point of view, LCP →[NP, LCP] literally means "the inside of *NP*", and PP →[PP,LCP] means "*PP*(in)...", so the phrase LCP →[NP, LCP] combines with PP →[PP,LCP] together to form PP →[PP,NP,LCP]. From this perspective, the formation process of PP →[PP,NP,LCP] is a linear association. In the register, they often form a larger prepositional phrase to form a sentence adverbial by co-occurrence. Combined with Fig. 3, we find that the phrases LCP →[NP,LCP], NP →[NP,NP] and PP →[PP,LCP]are clustered together, and are red, which indicates that they have a positive correlation. In other words, the probability of co-occurrence is relatively high. This is consistent with the interpretation of linguistics, the sentences with such syntactic structures often appear in the register, e.g.
   a. "他的两头羊在羊堆里拱出来。"(His two sheep were arched out in the sheep.) from Hua Yu's *To Live*, the phrase *CLP*: "在 (*PP*) 羊堆 (*NP*) 里 (*LCP*)" .
   b. "她在父亲身边跪下，轻轻地把父亲的裤子褪下来。" (She kneels beside her father and gently fades her father's pants down.) from Mo Yan's *Red Sorghum*, the phrase *CLP*: "在 (*PP*) 父亲 (*NP*) 身边 (*LCP*)".
2. VP →[ADVP,VP]+ VP →[VV,VV], from the perspective of mathematical formulas, the phrases VP →[ADVP,VP] and VP →[VV,VV] can be combined into VP →[ADVP,VV,VV]. In linguistics, such a phrase structure is reasonable. Besides, KeshengLi held that the two VV in this phrase have the same syntactic status[10], e.g.
   a. 打算回家 (Going home) [Predicate and Object Structure]
   b. 研究结束 (End of study) [Subject-Predicate Structure]
   c. 挖掘出来 (Dig out) [Predicate-Complement Structure]

   From Fig.3, we find that these two phrases are clustered together and are red, which is consistent with their semantic interpretation.
   In our chosen corpus, there are indeed a large number of sentences with these structures, e.g.

a. “中国证监会今日正式发布实施……”(The China Securities Regulatory Commission officially released the implementation today...), the phrase VP is 正式 (ADVP) 发布 (VV) 实施 (VV).
b. “大力地促进发展农产业…” (Vigorously promote the development of the agricultural industry). the phrase VP is 大力地 (ADVP) VV(促进)VV(发展).

It can be seen from these two examples, and we think that in language expressions, such a combination is reasonable. From Fig. 3, VP → [ADVP,VP] and VP →[VV,VV] are clustered together, which shows these two phrases have a positive correlation,ie, co-occurrence.

3. DP → [DT,QP] + DP → [DT,CLP] + NP →[DP,CP], the phrases DP →[DT,QP] and DP → [DT,CLP] act as modifiers, while *NP* is the central language. They can combine two groups of phrases NP → [DT,QP,CP] and NP → [DT,CLP,CP], where NP → [DT,QP,CP] emphasizes quantity and NP → [DT,CLP,CP] emphasizes location,e.g.

   a. “把那十件红色的裙子拿个过来。”(Take the ten red skirts over.), where NP → [DT,QP,CP] is “那 (DT) 十件 (QP) 红色的裙子 (CP)”
   b. “那个放在椅子上的书叫什么名字？ ”(What is the name of the book on the chair?), where NP → [DT,CLP,CP] is “那个 (DT) 放在椅子上的书 (CLP) 叫什么名字 (CP)?”

   From Fig.3, we discover that the phrases DP → [DT,QP], DP →[DT,CLP] and NP →[DP,CP] are clustered together, which proves that there is a correlation between them. This fully explains the rationality of linguistic interpretation.

Through mining the associations of syntactic structures, we can conclude that studying the relevance of syntactic structures helps us to analyze registers thoroughly and their differences and connections at the syntactic structure level.

6 Conclusion

In this paper, we study the syntax structures of the different registers. Instead of manually efforts, we use machine learning methods to explore the associations inside and between syntactic structures, including positive and negative correlations, linear, and nonlinear correlations. Combined with the theory of linguistics, we carry out a detailed analysis of the syntactic structures and part of speech(POS). Through analyses, we find that the associations between POS and syntactic structure explored by machine learning methods have a good interpretation effect, which provide a insight to study the theory of stylistic features. However, current discoveries by the machine learning methods are more preliminary and plain than those linguistic theories proposed by linguistics. Our further work includes a combination of more powerful machine learning methods with more profound linguistic theories.

References

1. Allen, J.F.: Towards a general theory of action and time. Artificial intelligence **23**(2), 123–154 (1984)
2. Benesty, J., Chen, J., Huang, Y., Cohen, I.: Pearson correlation coefficient. In: Noise reduction in speech processing, pp. 1–4. Springer (2009)
3. Bernstein, J.B.: Topics in the syntax of nominal structure across romance. (1994)
4. Chang, H.W.: The acquisition of chinese syntax. In: Advances in psychology, vol. 90, pp. 277–311. Elsevier (1992)
5. Chomsky, N.: Aspects of the Theory of Syntax, vol. 11. MIT press (2014)
6. DexiZhu: Grammar Printed lecture. Commercial Press (1982)
7. Ferguson, C.A.: Dialect, register, and genre: Working assumptions about conventionalization. Sociolinguistic perspectives on register pp. 15–30 (1994)
8. Firth, J.R.: Papers in Linguistics 1934-1951: Repr. Oxford University Press (1961)
9. Foley, W.A., et al.: Functional syntax and universal grammar. Cambridge University Press (2009)
10. KeshengLi, HaixiaMan, et al.: Boundedness of VP and linked event structure. Ph.D. thesis (2013)
11. Langfelder, P., Zhang, B., Horvath, S.: Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for r. Bioinformatics **24**(5), 719–720 (2007)
12. Pollock, J.Y.: Verb movement, universal grammar, and the structure of ip. Linguistic inquiry **20**(3), 365–424 (1989)
13. Rorat, T.: Plant dehydrins—tissue location, structure and function. Cellular & molecular biology letters **11**(4), 536 (2006)
14. Seretan, V.: Induction of syntactic collocation patterns from generic syntactic relations (2005)
15. Seretan, V.: Collocation extraction based on syntactic parsing. Ph.D. thesis, University of Geneva (2008)
16. Seretan, V.: Syntax-based collocation extraction, vol. 44. Springer Science & Business Media (2011)
17. Seretan, V., Wehrli, E.: Accurate collocation extraction using a multilingual parser. In: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. pp. 953–960. Association for Computational Linguistics (2006)
18. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) **58**(1), 267–288 (1996)
19. Watson, D., Tellegen, A.: Toward a consensual structure of mood. Psychological bulletin **98**(2), 219 (1985)
20. WenZhu: A study on the syntactic structure of the study of word collocation variation in "red sorghum". Chinese Teaching (5), 153–155 (2014)
21. Wright, W.: Lectures on the comparative grammar of the Semitic languages, vol. 43. University Press (1890)
22. Xia, F.: The part-of-speech tagging guidelines for the penn chinese treebank (3.0). IRCS Technical Reports Series p. 38 (2000)
23. XiaoFan: Sentence meaning. Chinese Linguistics (3), 2–12 (2010)
24. XiaoLiHuo: Research on Stylistic Features and Its Influence Variables. Master's thesis, Nanjing University (2014)
25. Xue, N., Xia, F., Chiou, F.D., Palmer, M.: The penn chinese treebank: Phrase structure annotation of a large corpus. Natural language engineering **11**(2), 207–238 (2005)
26. Yip, M.J.: The tonal phonology of Chinese. Ph.D. thesis, Massachusetts Institute of Technology (1980)