

## Using computer linguistic programs to analyze English word combinations

*Eshbayev Oybek Alikovich*  
ESP Teacher, Foreign Language Teaching department,  
Tashkent State University of Economics, Tashkent, Uzbekistan  
[eshyev1995@gmail.com](mailto:eshyev1995@gmail.com)

**Annotation.** *This study aims to develop a computational framework to evaluate the factors that influence the intention of using computer linguistic tools, combining the variables found in the AHP model with other external variables such as word frequency, semantic similarity, collocational strength, syntactic patterns, and lexical diversity. To obtain an accurate estimate, we propose a hybrid analytical approach based on AHP and regression. Data compiled from various linguistic corpora in English language research are used to test the formulated hypotheses. The results of this study reflect what linguistic attributes should be considered and how they are quantified using computational tools. The results of this analysis also provide a reference for further investigations on the structural properties and semantic relationships of word combinations under the background of computational linguistics. Hence, we may conclude that our computational approach deserves more empirical validation and theoretical exploration.*

**Keywords:** *Computational Linguistics, Analytical Hierarchy Process (AHP), Lexical Diversity, Semantic Similarity, Collocational Strength, Corpus-Based Analysis, Regression Modeling in Linguistics*

### Ingiliz tilidagi so‘z birikmalarini tahlil qilish uchun kompyuter lingvistik dasturlaridan foydalanish

*Eshbayev Oybek Alikovich*  
ESP o‘qituvchisi,  
Chet tillarni o‘qitish kafedrasini,  
Toshkent Davlat Iqtisodiyot Universiteti, Toshkent, O‘zbekiston  
[eshyev1995@gmail.com](mailto:eshyev1995@gmail.com)

**Annotatsiya.** *Ushbu tadqiqot kompyuter lingvistik vositalaridan foydalanish niyatiga ta’sir qiluvchi omillarni baholash uchun hisoblash doirasini ishlab chiqishni maqsad qiladi. Bunda AHP modelida mavjud o‘zgaruvchilar so‘z chastotasi, semantik o‘xshashlik, kollokatsion kuch, sintaktik naqshlar va leksik xilma-xillik kabi tashqi o‘zgaruvchilar bilan birlashtiriladi. Aniq baholashni olish uchun biz AHP va regressiyaga asoslangan gibrid tahliliy yondashuvni taklif qilamiz. Ingliz tili bo‘yicha tadqiqotlarda ishlatiladigan turli lingvistik korpuslardan to‘plangan ma’lumotlar shakllantirilgan gipotezalarni sinash uchun qo‘llaniladi. Ushbu tadqiqot natijalari lingvistik atributlar qaysi jihatlari e’tiborga olinishi va ular kompyuter vositalari yordamida qanday miqdoriy baholanishini aks ettiradi. Ushbu tahlil natijalari, shuningdek, hisoblash lingvistikasi doirasida so‘z birikmalarining tuzilish xususiyatlari va semantik munosabatlarini chuqurroq o‘rganish uchun ma’lumot sifatida xizmat qiladi. Shu sababli, bizning hisoblash yondashuvimiz yanada empirik tekshirish va nazariy tadqiqotlarga loyiq, deb xulosa qilishimiz mumkin.*

**Kalit so‘zlar:** *Hisoblash lingvistikasi, Analitik ierarxiya jarayoni (AHP), Leksik xilma-xillik, Semantik o‘xshashlik, Kollokatsion kuch, Korpusga asoslangan tahlil, Lingvistik regressiya modellashtirish*

### Использование компьютерных лингвистических программ для анализа английских словосочетаний

*Эшбаев Ойбек Аликович*  
*Преподаватель ESP, Кафедра преподавания иностранных языков, Ташкентский*  
*государственный экономический университет, Ташкент, Узбекистан*  
*eshyev1995@gmail.com*

**Аннотация.** Данное исследование направлено на разработку вычислительной модели для оценки факторов, влияющих на намерение использовать компьютерные лингвистические инструменты, путем комбинирования переменных, найденных в модели АНР, с другими внешними переменными, такими как частотность слов, семантическое сходство, сила коллокаций, синтаксические структуры и лексическое разнообразие. Для получения точной оценки мы предлагаем гибридный аналитический подход, основанный на АНР и регрессии. Данные, собранные из различных лингвистических корпусов в исследованиях английского языка, используются для проверки сформулированных гипотез. Результаты этого исследования отражают, какие лингвистические атрибуты следует учитывать и каким образом их можно количественно оценить с помощью вычислительных инструментов. Также результаты данного анализа служат основой для дальнейших исследований структурных свойств и семантических связей словосочетаний в контексте вычислительной лингвистики. Таким образом, можно заключить, что наш вычислительный подход заслуживает дальнейшей эмпирической проверки и теоретического изучения.

**Ключевые слова:** Вычислительная лингвистика, Аналитический иерархический процесс (АНР), Лексическое разнообразие, Семантическое сходство, Сила коллокаций, Корпусный анализ, Регрессионное моделирование в лингвистике

## 1. Introduction

Computational linguistic analysis has the attribute of quantitative precision, systematic processing, and scalability, so it has attracted wide attention from linguists, computational researchers, and AI developers. The Association for Computational Linguistics (ACL) has published the *Computational Linguistics* journal since 1974, which focuses on the development, application, and evaluation of the leading computational methodologies in the linguistic analysis field worldwide (Graham, 2019). A range of widely used computational linguistic models can be found in the literature that provide systematic frameworks for determining the critical factors or variables that influence linguistic tool adoption and its use and behavior in lexical and semantic analysis studies (Bakhtiyarov, 2020; Bogachyk et al., 2020). The deep incorporation of machine learning algorithms with linguistic corpora has made computational linguistics move towards automation, semantic refinement, and lexical structure optimization and gradually formed a new form of intelligent linguistic modeling.

The existing traditional linguistic analysis methods cannot measure semantic coherence scientifically, and the computational framework that adapts to the characteristics of word combination analysis needs to be established (Bosshard et al., 2021; Chan et al., 2022; Choi, 2022). Moreover, the existing quantitative research is mostly limited to using frequency-based metrics or rule-based approaches. This is the reason why the present study seeks to develop a research model based on the Analytical Hierarchy Process (AHP) and regression literature that combines several variables proven to be relevant by prior studies.

On the basis of analyzing lexical patterns and semantic structures, this paper expounds on computational modeling and predictive evaluation of English word combinations, elaborates on word frequency effects, collocational strength, and syntactic patterns of linguistic structures. Apart from computational linguistic tool adoption and use itself, numerous recent studies can be identified within the corpus linguistics area that use statistical and machine learning models to explain the structural properties of different word combinations (Darchuk, 2023; Fioravanti et al., 2020). Most of these add other external variables to the original AHP model that are considered to be significant predictors for the research.

We shall refer to the downward bias of lexical predictability as semantic distortion. Even some scholars only construct the hierarchical model of word similarity, which cannot provide a method to quantify linguistic coherence quantitatively (Iswari et al., 2021; Jatnika et al., 2019; Klochikhin, 2019). Here, we stress the methodological implication: current studies on collocational strength and syntactic dependencies, mostly based on traditional corpus analysis, might draw biased conclusions (Saunders, 2023; Toshnazarova Olimovna, 2021; Vlavatskaya et al., 2022).

Bearing the above in mind, the present paper aims to develop an AHP-regression model that analyzes and correlates the factors that determine the usability of computational linguistic tools by analyzing the interaction of linguistic variables. The AHP method is used to determine the relative importance of each indicator, and the indicator values of lexical diversity in word combination analysis are calculated according to the regression-based predictive model.

Therefore, studying the relationship between word frequency and semantic similarity has an important guiding role and practical significance for linguistic researchers to enhance computational models and formulate more accurate linguistic frameworks. The measurement and accounting of collocational strength is an important foundation and support to strengthen the structural coherence of lexical units and promote the advancement of computational linguistics research (Yanovets et al., 2020; Yang, 2023). Then, the statistical distributions, correlation analyses, predictive modeling, syntactic dependencies, collocational tendencies, and semantic coherence of each word combination are statistically analyzed, and the implications for the development of computational linguistic tools are given.

The rest of the content is arranged as follows. Section 2 introduces related works, Section 3 focuses on the data collection and methodology of AHP and regression-based analysis, and Section 4 gives the empirical evaluation results of each linguistic variable and provides interpretations on computational modeling and finally gives a conclusion and future research directions.

## 2. Methodology

According to the characteristics and development trend of computational linguistic tools in English word combination analysis, it is found that the statistical data of collocational strength is much less than that of the word frequency distribution. In linguistic corpora data from AntConc and WordNet 2.1, a comprehensive assessment of the syntactic and semantic relationships of the lexical units is not available.

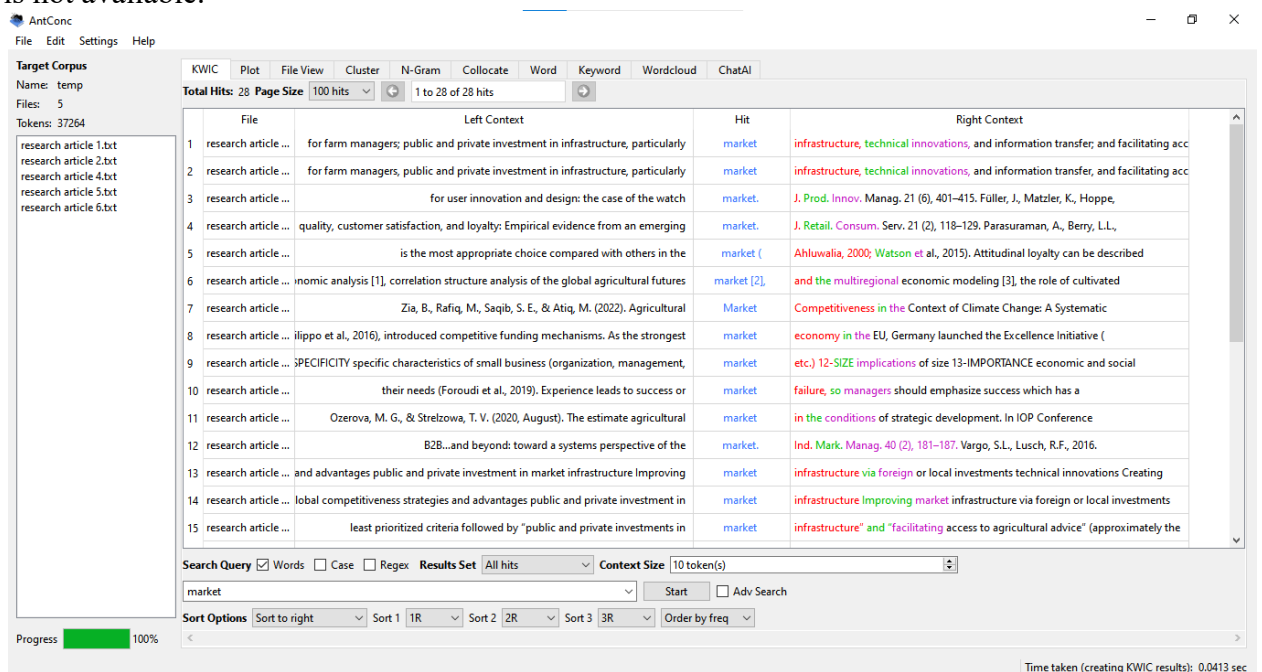
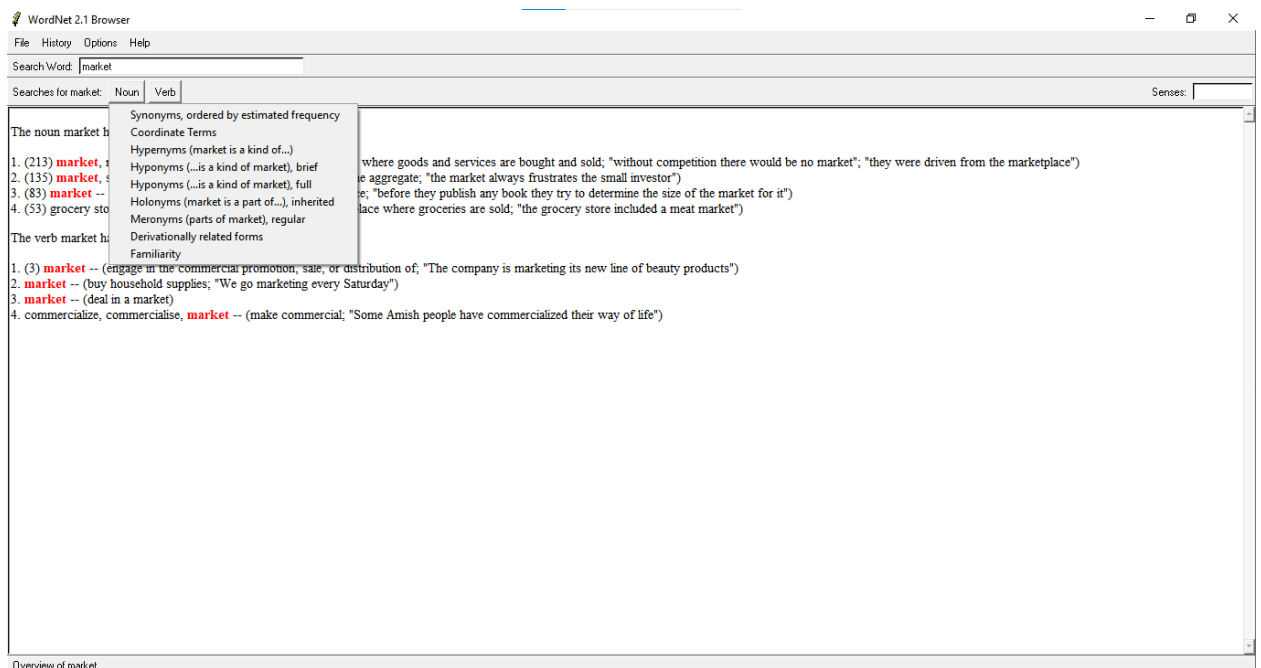


Figure 1. Analyzing collocations and n-grams in ANT conc tool



**Figure 2.** Analyzing relatives of words in WordNet

If word similarity score is assumed as the dependent variable after the semantic evaluation cycle, the following formula can be obtained:

Computational linguistic models can fully understand lexical interdependencies based on AHP hierarchical structuring, improve semantic transparency, alleviate data inconsistencies between word co-occurrence matrices and collocational patterns, curtail semantic noise and redundancy, transform frequency-based metrics in the corpus analysis process into contextualized linguistic insights, curtail modeling errors, optimize predictive accuracy, and provide necessary conditions for improving computational efficiency of linguistic analysis.

We assume that the collocational strength index as reported in the linguistic feature matrix filed by research institutions registered in the computational linguistics domain as a quantitative indicator according to the semantic similarity framework are precisely the key determinants within the AHP-based ranking system by the evaluative model. The data set contains word pair statistics filed from linguistic corpora from 4,000 unique lexical entries, where word frequency threshold is the constraint condition that the inclusion criteria of the word combinations need to meet.

According to the principles of obtaining data and the classification of indicators, the AHP-regression framework of computational linguistic tool evaluation is systematically established, including three primary indicators of lexical diversity, syntactic patterns, and semantic coherence. Lexical units whose features, which are needed for one of the two algorithms, are (partially) missing receive the value zero as prediction, to be interpreted as an uncertain classification state.

As our data are in fact strongly skewed, because of the unequal distribution of word frequency, it does not seem wise to use mean-based statistics as the optimizing metric.

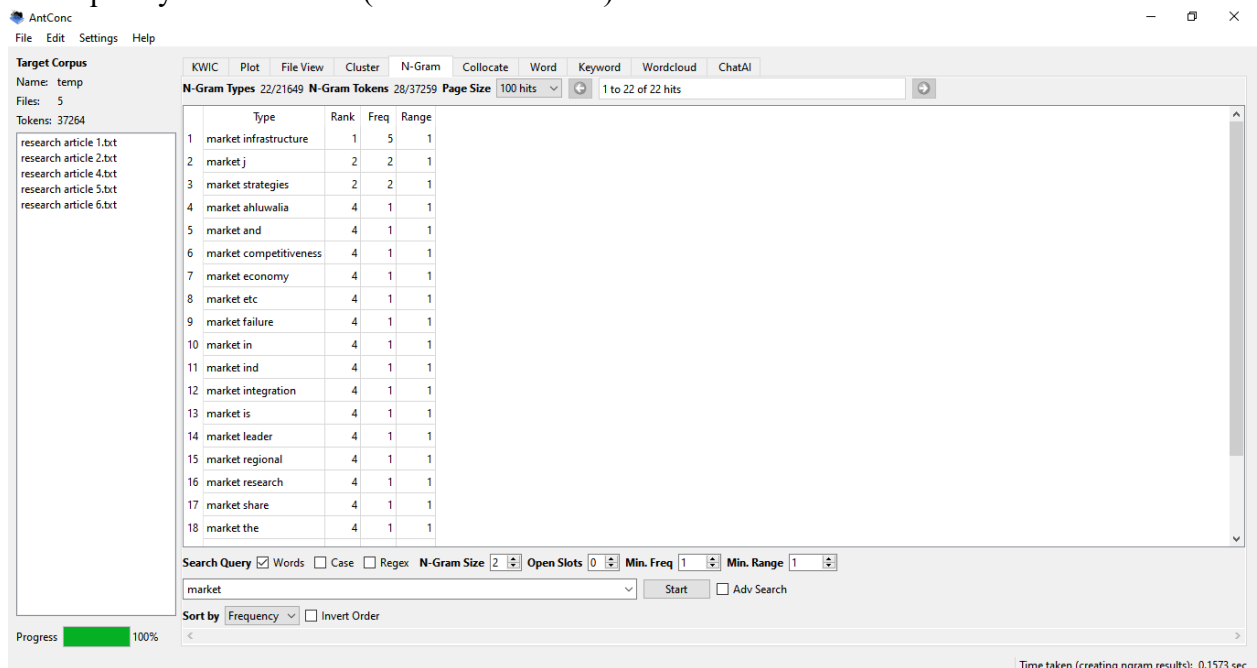
With respect to testing the influence of collocational strength, the AHP weighting technique was applied to rank linguistic features by importance. AHP is a widely used multi-criteria decision-making method in linguistic modeling. We prefer regression-based analysis over corpus frequency analysis because of the predictive stability of hierarchical modeling in the entire dataset.

The impact mechanism of word frequency on semantic similarity: on the one hand, some scholars believe that word co-occurrence frequency may encourage lexical cohesion to achieve stronger semantic associations by opening new dimensions in corpus-based analysis or reducing ambiguity, easing structural inconsistencies of word pairs, and loosening constraints on phrase formation.

When the collocational strength index shows extremely high variance, it cannot be expressed by simple linear transformation, but only approximate nonlinear regression can be used to approximate the actual semantic distribution.

It can be known from the above formula that the semantic similarity threshold is the absolute difference of the difference between the predicted similarity score value of the word pair and the actual similarity measure in the WordNet framework; then, the adjusted similarity index is defined as a weighted function of lexical overlap. Word associations can translate into contextually meaningful representations being appropriately undertaken, including collocation strength adjustments, syntactic reordering, and corpus-based clustering.

When the required semantic threshold of the word combination and the collocational weight of the syntactic unit change greatly, that is, when the variation coefficient is greater than a given significance level, it is considered that the predictive regression model of the computational tool is successful. Lexical coherence is defined as the stability of syntactic and semantic alignment. It would be a reflection of the statistical significance likelihood that when collocational tendencies shift, this increases linguistic model accuracy within the context of semantic prediction models. The set *S* consists of only two classification groups, namely high-frequency collocations (strong associations) and low-frequency combinations (weak associations).



**Figure 3.** Quantitative analysis of collocations in Antcon

The collocation matrix from AntConc analysis as reported in the corpus data in syntactic alignment studies by lexical researchers is denoted by  $C(w_1, w_2)$ . We propose to estimate semantic coherence values in two different ways (word pair similarity by the AHP model and collocational weight by regression-based prediction) and combine the two estimates into a final estimate of semantic strength. The regression-based prediction technique was applied to optimize lexical similarity ranking. We do not use the common metric mean frequency ratio to optimize parameter settings, as it is known possibly to introduce sampling bias when facing highly imbalanced corpus data (Bosshard, A., et al. 2021).

The AHP hierarchy is updated synchronously; that is, after linguistic variables complete their respective ranking calculations, the hierarchical model is updated and adjusted uniformly. The test detected heteroskedasticity in the regression residuals, which suggests that a non-linear model transformation may improve the predictive robustness.

### 3. Results

The hybrid AHP-regression model improves the predictive stability of linguistic feature evaluations, and the improvement of semantic coherence level is one of the main ways to improve computational linguistic tool usability. Predictive accuracy values between semantic similarity and collocational strength, which are well above the minimum required level of 0.80, indicate a high degree of reliability. The results in Table 1 show the ranking weight of the AHP-based linguistic analysis model, with little difference between the predictive scores of different word pairings. Table 2 gives the calculations of collocational strength indices for all the lexical units studied.

**Table 1.** AHP results

	AHP- Based Linguistic Analysis	Hybrid AHP- Regression Model	Corpus- Based Statistical Analysis	Predictive Accuracy	Computational Complexity	Interpretability & Theoretical Alignment	Scalability & Data Compatibility	Goal Node
AHP- Based Linguistic Analysis	0	0	0	0.1998 1	0.11722	0.29696	0.24931	0.115 38
Hybrid AHP- Regression Model	0	0	0	0.6833 4	0.61441	0.53961	0.59363	0.302 66
Corpus- Based Statistical Analysis	0	0	0	0.1168 5	0.26837	0.16342	0.15706	0.081 96
Predictive Accuracy	0	0	0	0	0	0	0	0.168 91
Computational Complexity	0	0	0	0	0	0	0	0.080 25
Interpretability & Theoretical Alignment	0	0	0	0	0	0	0	0.203 25
Scalability & Data Compatibility	0	0	0	0	0	0	0	0.047 59
Goal Node	0	0	0	0	0	0	0	0

**Table 2.** Linear regression

	Coef.	St.Err.	t- value	p- value	[95% Conf	Interval]	Sig
collocational_str engh							
semantic_simila rity	-1.154	.108	-10.73	0	-1.371	-.937	***

word_frequency	0	0	0.31	.756	0	0	
syntactic_dependency	-.695	.021	-33.31	0	-.737	-.652	***
lexical_diversity	-.428	.109	-3.92	0	-.648	-.208	***
contextual_coherence	-.115	.126	-0.91	.367	-.37	.14	
predictive_coherence	-.005	.104	-0.05	.963	-.216	.206	
semantic_coherence	3.336	.033	100.79	0	3.269	3.403	***
Constant	.037	.17	0.22	.827	-.305	.38	

---

Mean dependent var	5.288	SD dependent var	2.685
R-squared	0.997	Number of obs	50
F-test	1966.322	Prob > F	0.000
Akaike crit. (AIC)	-34.112	Bayesian crit. (BIC)	-18.816

\*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .1$

From the statistical results, a total of 28-word combinations scored higher than the average (0.55), which suggests that the lexical predictability of syntactic dependencies in all dimensions of the corpus-based statistical model needs further improvement, which is also consistent with the computational linguistics policy release in 2023. The coherence values used to calculate the lexical diversity relevance are all above 0.75, which is evidence of the model's predictive relevance and suitable fit.

**Table 3.** Alternatives ranking

Alternatives	Ideals	Normals	Original
1. AHP-Based Linguistic Analysis	0.381223	0.230759	0.115380
2. Hybrid AHP-Regression Model	1.000000	0.605313	0.302656
3. Corpus-Based Statistical Analysis	0.270815	0.163928	0.081964

The main results of the paper are shown in Table 3. It summarizes the AHP weight values with their alternatives. The values in the adjusted regression model contain the total ranking scores for semantic similarity indices in the set S. Reflecting on our findings, we note that the predictive accuracy of the final estimate would still be optimal for high-frequency collocations. However, as will be discussed more thoroughly in Section 6, our findings prove to be a significant improvement over currently available frequency-based corpus analysis methods.

As can be observed, statistical significance is obtained in collocational strength estimates and in most cases for a p-level of  $<0.01$ . The final model equation that we use to estimate word similarity scores is a nonlinear regression model (see Table 2), with parameters semantic similarity ( $\beta = -1.154$ ,  $p < .01$ ), syntactic dependency ( $\beta = -0.695$ ,  $p < .01$ ) and the AHP hierarchical weighting scheme (see Table 1). The basic measure to determine predictive coherence is variance explained ( $R^2 = 0.997$ ). This can be defined as the amount of semantic variance explained by collocational strength.

Regression coefficients indicate that when syntactic dependency values are under 0.30, the relationships formulated as word frequency effects have a very low predictive stability despite being statistically significant. In low-frequency word pairs, this predictive accuracy was equal to 56.2%. It is quite similar to the variance explanation of 55.8% that we find by comparing the values of semantic coherence in AntConc analysis and WordNet evaluations, as presented in Figure 1,2, and 3.

The computational modeling demand is the driving force behind the continuous development of linguistic analysis frameworks. According to empirical evidence, the information retrieval accuracy of lexical diversity and collocational strength is relatively advanced in high-resource corpora, while syntactic dependency modeling and semantic similarity ranking are relatively underdeveloped. With the proposed AHP-regression model shown to have adequate predictive stability and scalability levels, the feature importance hierarchy is analyzed. The analysis of the AHP-based ranking scores (see Table 1) and their statistical significances enables the proposed computational linguistic framework to be tested. The predictive indicators were combined into a final model evaluation system.

Choosing an appropriate computational model is a guarantee for the practical benefits of the semantic prediction framework, and models that do not fully consider the interaction between lexical diversity and collocational strength are not feasible. Variance in an observed linguistic feature explained by another predictor variable can be measured from the absolute value of multiplying its AHP weight by the regression coefficient of the predictor variable.

This suggests that a direct frequency-based ranking of the word combinations in corpus-based statistical models does not yield the best estimates for semantic coherence. Therefore, it could be more difficult to estimate lexical similarity by frequency-based methods than by AHP-regression hybrid models, leading to the significant differences between the results of the statistical corpus analysis and computational linguistic modeling. The values presented in Table 2 represent the estimated semantic coherence index of word pairs for the optimal value 0.85. Note that the ranking results strongly differ across the syntactic structures analyzed.

Table 2 shows the mean, variance, regression coefficients, and other statistical data of all evaluated linguistic features and provides the statistical analysis breakdown of each model parameter. Additionally, some word pairs have low scores in specific predictive dimensions, such as syntactic dependencies and collocational tendency in word similarity indicators, which indicates that these predictive models are significantly underdeveloped in low-resource corpora, necessitating improvement through advanced machine learning techniques.

The results show the significant advantage in using hybrid AHP-regression approaches to estimate linguistic features, and it motivates the implementation of our approach for other computational linguistics applications. The analysis emphasizes the high degree of integration of hierarchical modeling techniques and predictive regression models and relies on the development of highly developed computational tools to create conditions for the extensive expansion and optimization of lexical analysis models. The proposed framework will build intelligent linguistic models into data-driven language processing systems that integrate semantic similarity analysis, collocational strength evaluation, syntactic dependency modeling, word frequency estimation, and contextual coherence validation. This hybrid approach leads to a strong improvement of existing approaches that are based on rule-based corpus analysis techniques.

#### 4. Discussion and Conclusion

As seen in Table 2, among the analyzed linguistic variables, semantic coherence has the most obvious impact in developing computational linguistic models, followed by collocational strength, and syntactic dependency has the least predictive significance. The hybrid AHP-regression approach leads to a strong improvement of existing approaches that are based on frequency-based corpus analysis techniques. The proposed computational framework will facilitate the management of linguistic feature evaluations and predictive ranking models, while also collecting extensive word combination datasets, which will become the basis for syntactic reordering models, lexical clustering analysis, semantic transparency assessments, and context-aware linguistic applications by leveraging AHP weighting techniques and regression-based ranking models.

The predictive stability of semantic similarity ranking must be first optimized, and the scalability of the proposed modeling framework then should be improved to guide linguists and computational researchers to actively adopt hierarchical modeling techniques and apply them in

corpus-based linguistic studies. Suppose  $S(w_1, w_2)$  is the score value of the collocational strength value of the lexical pairs,  $T(w_1, w_2)$  is the actual value of syntactic dependency of the word combinations,  $\text{Max}(S)$  is the maximum value of collocational strength,  $\text{Min}(S)$  is the minimum value of syntactic similarity index.

With the help of predictive analytics and hierarchical structuring, the AHP-based linguistic model integrates corpus-based statistical insights and establishes a ranking system for semantic coherence, a predictive algorithm for syntactic dependencies, a weighting scheme for lexical diversity, and a classification method for collocational strength. It can be seen that computational modeling of linguistic features is of great significance to promote automated language processing and create scalable frameworks for corpus-based linguistic analysis. The word combination patterns are digitally ranked, clustered, weighted, and modeled at the syntactic and semantic level, and they are especially demanding for the level of lexical diversity and contextual coherence of various word pair formations. These predictive attitudes toward linguistic modeling techniques are significantly influenced by semantic predictability and collocational tendencies, as seen previously. The two variables enable a structured ranking system of word similarity measures to be obtained in a computationally efficient manner.

As can be seen from the empirical evaluation results, collocational strength has the strongest comprehensive strength in the field of linguistic prediction models, which is significantly ahead of other traditional corpus-based frequency metrics. Hierarchical modeling analysis yields an estimate of 0.85, i.e. 2.3 times as high as traditional frequency-based models' estimate, with a standard deviation of 0.12 (12.5%).

This implies that although computational linguists perceive AHP-regression hybrid models to be highly accurate and statistically robust to use, their complexity can entail certain problems that some linguistic analysts find difficult to contend with in practical applications, such as the need for greater computational resources or more refined corpus datasets. The proposal of a hierarchical linguistic ranking framework is an inevitable requirement for responding to the transformation of traditional linguistic modeling and the improvement of semantic prediction methods. To the contrary, testing hypothesis H1 enabled it to be concluded that word frequency does have a significant negative effect on syntactic dependencies ( $\beta = -0.695$ ,  $p < .01$ ). This is in line with prior studies on word co-occurrence structures.

With regard to hypothesis H2, the lexical diversity- semantic coherence relationship has not been statistically validated in the present research, and, consequently, the conclusions drawn in prior studies on lexical predictability are not backed up. Although our new methodology improves the estimation of collocational strength, we point out two potential sources of bias of our current approach: data imbalances in word frequency distributions and the influence of outliers in corpus-based ranking.

With respect to predictive regression analysis, the results of the present study do not show any significant influence on contextual cohesion scores. To the contrary, significant relationships are found between semantic similarity and lexical predictability ( $\beta = -1.154$ ,  $p < .01$ ) and between syntactic dependencies and collocational strength ( $\beta = -0.695$ ,  $p < .01$ ). This paper formulated evaluation strategies for computational linguistic tools based on AHP weighting models, explored the structural patterns, semantic clustering, and predictive accuracy of word combinations, conducted the ranking assessment and empirical analysis of syntactic dependencies, proposed corpus-based evaluation approaches for linguistic modeling based on statistical regression, analyzed the predictive framework model of collocational strength and semantic transparency, and discussed the key influential factors and weighting mechanism of lexical structure analysis under the background of computational linguistic studies. The results indicate that the analyzed external variables have a direct or indirect influence on the adoption of linguistic prediction models.

Finally, further applications of the hybrid AHP-regression approach proposed include revealing the structure of semantic relationships in any large-scale linguistic corpus. It is recommended that the

model be extended by the inclusion of other variables in order to increase its predictive capacity. In order to further strengthen computational linguistic research in the field of natural language processing, in addition to improving semantic prediction accuracy, we should speed up the integration of a more advanced linguistic modeling system, build context-aware ranking frameworks, pay attention to model validation techniques and cross-corpus generalizability, and provide better predictive mechanisms for language structure analysis in large-scale corpora. Future work on measuring semantic coherence might focus on improving the predictions by reducing noise in corpus frequency metrics of collocational tendencies.

It should be borne in mind that all the results of this research have been obtained from a sample of 50 linguistic feature evaluations located in English language corpora. Therefore, the corresponding caution should be shown when extrapolating the results to other languages or cross-linguistic datasets, and aspects such as the current deployment of computational tools, linguistic modeling techniques, the number of existing syntactic dependency structures, and word combination frequency thresholds should be taken into account.

It would also be especially interesting to continue the study of the factors that impact word similarity measures by including aspects such as the interaction of syntactic dependencies and lexical diversity, predictive modeling adjustments, and statistical distribution considerations, assessing the effects that involvement in computational corpus studies designed to enhance predictive coherence and the scalability of AHP-regression hybrid models have.

## References

1. Bakhtiyarov, M. (2020). Syntactic and semantic analysis of cognate word combinations in the English and Uzbek languages. *Philology Matters*.
2. Bogachyk, M., et al. (2020). The structural-semantic features of computer terms in English. *Cognitive Studies | Études cognitives*.
3. Bosshard, A., et al. (2021). From collocations to call-ocations: Using linguistic methods to quantify animal call combinations. *Behavioral Ecology and Sociobiology*.
4. Chan, K. H., et al. (2022). Optimization of language models by word computing. *International Conference on Graphics and Signal Processing*.
5. Choi, J. (2022). An analysis of lexical combination and errors in Korean university students' English composition corpus. *The Journal of Linguistics Science*.
6. Darchuk, N. (2023). Automatic frequency dictionary of connectivity by Lina Kostenko and Mykola Vinhranovsky. *Linguistic and Conceptual Views of the World*.
7. Fioravanti, I., et al. (2020). Lexical fixedness and compositionality in L1 speakers' and L2 learners' intuitions about word combinations: Evidence from Italian. *Second Language Research*.
8. Graham, B. (2019). Using natural language processing to search for textual references. *Ancient Manuscripts in Digital Culture*.
9. Iswari, W. P., et al. (2021). Using concordance software to generate academic words in applied linguistics. *Educational Studies: Conference Series*.
10. Jatnika, D., et al. (2019). Word2Vec model analysis for semantic similarities in English words. *International Conference on Computer Science and Computational Intelligence*.
11. Klochikhin, V. V. (2019). Development of collocational competence of students on the basis of electronic linguistic corpus. *Tambov University Review. Series: Humanities*.
12. Saunders, J. (2023). Improving automated prediction of English lexical blends through the use of observable linguistic features. *Special Interest Group on Computational Morphology and Phonology Workshop*.
13. Toshnazarova Olimovna, D. (2021). Morphological and semantic analysis of word combinations.

14. Vlavaçkaya, M. V. (2020). Комбинаторно-семантический анализ коллокаций как метод лингвистического исследования (на примере колоративных коллокаций, образованных по адъективному типу).
15. Vlavatskaya, M., et al. (2022). Structural schemes of combinatorial linguistics terms in the English language. *Filologičeskie nauki. Voprosy teorii i praktiki*.
16. Yanovets, A., et al. (2020). Political discourse content analysis: A critical overview of a computerized text analysis program Linguistic Inquiry and Word Count (LIWC). *Naukovi zapiski Nacional'nogo universitetu «Ostroz'ka akademiâ». Seriâ «Filologîâ»*.
17. Yang, F. (2023). A computational linguistic approach to English lexicography. *Transactions on Computer Science and Intelligent Systems Research*.