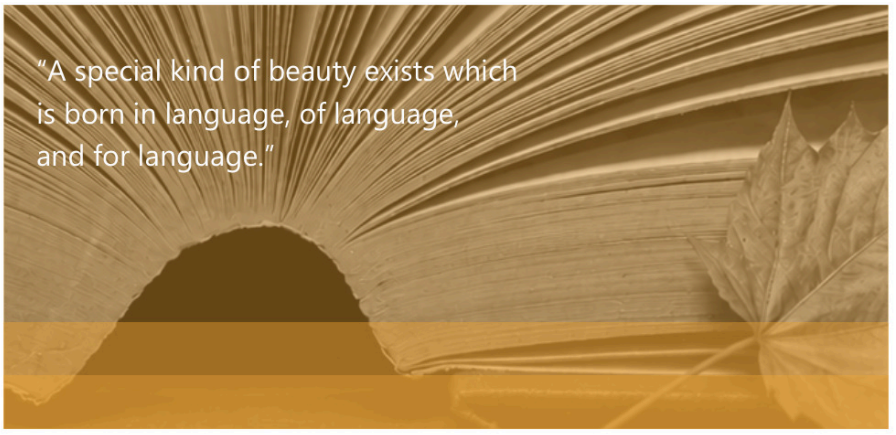
VOLUME 10, OCTOBER 2026
ISSN: 3060-4958







The Lingua Spectrum



UZBEKISTAN STATE WORLD
LANGUAGES UNIVERSITY

MORE INFO:

www.lingvospektr.uz

t.me/thelinguaspectrum

Nazarova N.B.	The Literary Representation of Women's Accomplishments and Charm in Jane Austen's Novels	319
Karimova S.S.	Connotative meaning and pragmatic interpretation in literary texts	325
Radjabova G.G.	Corpus evidence and the limits of intuition in the study of collocation	331
Adinaev O.J.	Onomastic, lexico-semantic and linguocultural foundations for the study of celestial body names	339
Muxitdinova X.S.	Sociocultural and diachronic characteristics of personal names in the Uzbek language	349
Amanov V.B.	Universal values as an object of linguocultural research in Spanish and Uzbek	358
PEDAGOGY		**378**
Kambarova D.I.	The methodology of developing oral speech competence of future foreign language teachers in a digital educational environment	379
Allayarov T.Ya.	Developing communicative competence of public health college students using the standardized patient method	389
Silanteva E.F.	Preliminary training for ESP teachers for effective integrated teaching	396
Omonova E.B., Musoyeva A.B.	Digital revision portfolios for developing English writing skills in Uzbek secondary schools	411
Paluanova Kh.D.	The relationship between pragmalinguistic analysis and linguodidactic design in preservice English teacher education	419
Adamboyeva N.Q.	Age-related development of phonological skills in children aged 7–10 and their role in reading competence	427
Alimardonov L.A.	Theoretical foundations and conceptual model of spiritual technologies in ensuring the information security of young people	435
Pazilova N.M., Abdulhakimova R.A.	Didactic potential and limitations of generative artificial intelligence in teaching Russian	443
Kadirova N.M., Tahirova Sh.Sh.	Stages of working with texts in Russian language classes at higher education institutions	450
Pulatova F.U.	Pragmadidactic features of teaching implicature and indirect speech acts in English language education	456
Sulaymonova M.A.	AI-Assisted Feedback in EFL Writing: Implications for Writing Accuracy and Learner Autonomy	464
Khodjieva Z.M.	Theoretical and Methodological Foundations and Practical Stages of Criterion-Referenced Assessment in Foreign Language Teaching at Higher Education Institutions	489
Musayeva F.M.	The role of specialized vocabulary in developing the lexical competence of oil and gas students	498

Corpus evidence and the limits of intuition in the study of collocation

Radjabova Gulnoza Giyosiddinovna
rad.gulnoza@gmail.com
PhD, Associate Professor,
Uzbekistan State World Languages University

Annotation *Drawing on data from the Corpus of Contemporary American English (COCA) and the Russian National Corpus, the article examines how reliably native-speaker intuition describes collocation and semantic prosody. It is shown that speakers can usually judge whether a word combination is acceptable, but they are much less successful in assessing its frequency, its typical lexical environment, and the positive, negative or neutral evaluative colouring that a word acquires from its habitual contexts. Such features become visible only when large amounts of text are analysed. A contrastive analysis of English and Russian material demonstrates that these patterns are language-specific and cannot be transferred from one language to another through translation or introspection. On this basis, conclusions are drawn for Uzbek: the need to build and use corpora, to verify phraseological and lexicographic descriptions against corpus evidence, and to take collocational and prosodic information into account in teaching Uzbek and in translation.*

Keywords *Corpus linguistics, collocation, semantic prosody, linguistic intuition, phraseology, COCA, Russian National Corpus, Uzbek language*

Kollokatsiyalarni o‘rganishda korpus dalillari va intuitsiyaning cheklari

Radjabova Gulnoza Giyosiddinovna
rad.gulnoza@gmail.com
PhD, dotsent,
O‘zbekiston davlat jahon tillari universiteti

Annotatsiya *Maqolada Amerika milliy korpusi (COCA) va Rus tili milliy korpusi ma’lumotlari asosida ona tilida so‘zlashuvchining til sezgisi (lisoniy intuitsiya) kollokatsiya va semantik prosodiyani tavsiflashda qanchalik ishonchli ekanligi tahlil qilinadi. Ko‘rsatilishicha, so‘zlovchi so‘z birikmasining qabul qilinishi mumkinligini odatda to‘g‘ri baholaydi, biroq uning qo‘llanish chastotasini, tipik leksik muhitini hamda so‘zning odatiy kontekstlari tufayli orttirgan ijobiy, salbiy yoki neytral baho tusini aniqlashda xato qiladi. Bunday xususiyatlar faqat katta hajmdagi haqiqiy matnlarni tahlil qilish orqali ochiladi. Ingliz va rus tillari materiallarini qiyosiy tahlil qilish bu qonuniyatlar har bir tilga xos ekanini, ularni tarjima yoki introspeksiya yo‘li bilan bir tildan boshqasiga ko‘chirib bo‘lmasligini isbotlaydi. Shu asosda o‘zbek tili uchun xulosalar beriladi: reprezentativ korpuslarni yaratish va ulardan foydalanish, frazeologik va leksikografik tavsiflarni korpus dalillari bilan tekshirish, shuningdek, o‘zbek tilini o‘qitish va tarjimada kollokatsion hamda prosodik ma’lumotlarni hisobga olish zarur. Shuningdek, grammatik jihatdan maqbul birikma bilan nutqda odatiy qo‘llanadigan birikma o‘rtasidagi farq ham alohida batafsil ko‘rib chiqiladi.*

Kalit so‘zlar *Korpus lingvistikasi, kollokatsiya, semantik prosodiya, lisoniy intuitsiya, frazeologiya, COCA, Rus tili milliy korpusi, o‘zbek tili*

The Lingua Spectrum
Cogito, ergo sum

Jurnal Oliy Attestatsiya Komissiyasi tomonidan 2025-yil 4-yanvar kuni 366/5-sonli qaror bilan tavsiya etilgan milliy ilmiy nashrlar ro'yxatiga kiritilgan

Корпусные данные и границы интуиции в изучении коллокаций

Раджабова Гулноза Гиёсиддиновна
rad.gulnoza@gmail.com
PhD, доцент,
Узбекский государственный университет мировых языков

Аннотация *На материале Корпуса современного американского английского (COCA) и Национального корпуса русского языка в статье анализируется, насколько надёжна языковая интуиция носителя при описании коллокаций и семантической просодии. Показано, что говорящий, как правило, верно оценивает допустимость сочетания слов, но гораздо чаще ошибается в определении его частотности, типичного лексического окружения, а также положительной, отрицательной или нейтральной оценочной окраски, которую слово приобретает благодаря привычным контекстам употребления. Эти свойства выявляются только при анализе больших массивов аутентичных текстов. Сопоставительный анализ английского и русского материала демонстрирует, что подобные закономерности специфичны для каждого языка и не могут быть перенесены из одного языка в другой ни путём перевода, ни путём интроспекции. На этой основе формулируются выводы для узбекского языка: необходимость создания и использования репрезентативных корпусов, проверки фразеологических и лексикографических описаний корпусными данными, а также учёта коллокационной и просодической информации при обучении узбекскому языку и в переводе. Особое внимание уделяется различию между формальной допустимостью словосочетания и характером его реального употребления.*

Ключевые слова *Корпусная лингвистика, коллокация, семантическая просодия, языковая интуиция, фразеология, COCA, Национальный корпус русского языка, узбекский язык*

Introduction

Anyone who has marked essays written in a foreign language knows a particular kind of error that is hard to explain to the student. Take a sentence such as *The rain was very strong yesterday*. Nothing in it breaks a rule of English grammar, and the meaning is perfectly clear, yet an English reader pauses over it, because in English rain is *heavy* rather than *strong*. Asked why, the teacher can usually say only that this is how it is said. That is true, but it satisfies neither the student nor the linguist.

For most of the twentieth century the matter rested more or less there. Word combinations of this kind were recognised, described and classified, but the evidence for them came mainly from the intuition of native speakers and the judgement of lexicographers. Intuition is a remarkable instrument. It tells us at once that a sentence is ill-formed, that a phrase sounds old-fashioned, that a pun works. What it does badly, as this article will try to show, is tell us how often something occurs and what attitudes a word habitually carries with it. These happen to be exactly the properties on which collocation depends.

Corpus linguistics changed the terms of the discussion by making usage observable

(Biber et al., 1998; McEnery & Hardie, 2012). A reference corpus such as the Corpus of Contemporary American English (COCA), with about one billion words, or the Russian National Corpus (RNC) lets the researcher see not what a speaker believes is said but what has actually been written and spoken, in what quantities and in what company. The aim of this article is to consider what that shift means for the study of collocation, and in particular to identify the points at which intuition and corpus evidence part ways. The discussion draws on English and Russian corpus data and brings in Uzbek as a third point of comparison.

The question is a timely one for Uzbek philology. Work on a national corpus of the Uzbek language has been going on for several years and has already been the subject of a dedicated international conference (O'zbek tilining milliy korpusi, 2022). Once such a resource is in wide use, philologists will have to decide what kinds of questions it is worth asking of it, and the experience of English and Russian corpus research offers some guidance.

Theoretical background

The idea that the meaning of a word lies partly in its habitual neighbours is older than the corpus. Firth's much-quoted remark that "you shall know a word by the company it keeps" (Firth, 1957, p. 11) placed collocation among the levels at which meaning is made, alongside phonology and grammar. Halliday (1966) sharpened the point with a pair of examples that has since become classic. English speakers say *strong tea* but not *powerful tea*, and *powerful car* rather than *strong car*, although the two adjectives are close in meaning. The restriction cannot be derived either from the meaning of the words or from any grammatical rule. It simply has to be observed.

The Russian phraseological tradition reached a related conclusion by a different route. Vinogradov's classification of phraseological units singled out phraseological combinations (*фразеологические сочетания*), in which one component carries a meaning that is available only in combination with a limited set of other words (Vinogradov, 1977). Russian *крепкий чай*, where *крепкий* means neither 'solid' nor 'durable', is a textbook case. This tradition, however, worked mostly with material the researcher had selected, and it tended to treat such combinations as a special layer of the lexicon rather than as a property that runs through language as a whole.

It was Sinclair who drew the more radical conclusion. On the basis of the first large computer corpora he proposed that text is produced according to two principles. The open-choice principle treats each position in a sentence as a slot to be filled by anything the grammar allows. The idiom principle recognises that speakers have at their disposal a very large number of semi-preconstructed phrases, each of which amounts to a single choice (Sinclair, 1991). In Sinclair's view it is the idiom principle that predominates in ordinary text. Hoey (2005) later gave this a psychological interpretation. Every time we meet a word we are primed to expect it in certain contexts and combinations, and the accumulated effect of these primings is what we experience as a sense of naturalness.

This last point explains why intuition is at once indispensable and unreliable. Our sense that *heavy rain* is right and *strong rain* is odd is itself a product of exposure, but the process that produces the judgement is not open to introspection. We can report the verdict, not the record of encounters behind it. Hunston (2002) notes that intuition is a particularly poor guide to frequency, and several of the best-known findings of corpus linguistics are patterns that native speakers had never noticed and, when shown them, often found surprising.

Research methods

English data were taken from COCA (Davies, 2008–present) and Russian data from the main corpus of the RNC. For each verb or adjective under discussion, collocates were retrieved with the corpus's own collocation function, using a span of four words to the right of the node. Collocates were treated as

significant if they occurred at least five times and had a mutual information score of 3 or higher, a threshold widely used in corpus studies (Hunston, 2002; McEnery & Hardie, 2012). For the analysis of semantic prosody, the 50 strongest noun collocates of each node were classified by hand as negative, neutral or positive. The classification was based on the concordance lines rather than on dictionary definitions, since a noun that looks neutral in isolation may be used almost exclusively in unfavourable contexts. All data were retrieved in January 2026.

The Uzbek material is treated differently. To the author's knowledge, no balanced reference corpus of Uzbek with publicly accessible collocation statistics comparable to those of COCA or the RNC is yet available to the individual researcher, so the Uzbek examples below come from standard usage and bilingual dictionaries. They serve as illustrations and as hypotheses for future corpus testing, not as quantitative evidence. This asymmetry between the three languages is not merely a limitation of the study. As will become clear, it is part of its argument.

Results and Discussion

Collocational Restriction Across Three Languages

Table 1 brings together five combinations that are common in all three languages but are built on different lexical choices. The English and Russian expressions are examined with reference to COCA and the RNC. The last column gives the English phrase that results from rendering the Uzbek combination, and in most cases the Russian one as well, word for word.

English (COCA)	Russian (RNC)	Uzbek	Word-for-word English
make a decision	*принять решение*	*qaror qabul qilmoq*	*accept a decision*
pay attention	*обратить внимание*	*e'tibor bermoq*	*give attention*
ask a question	*задать вопрос*	*savol bermoq*	*give a question*
heavy rain	*сильный дождь*	*kuchli yomg'ir*	*strong rain*
strong tea	*крепкий чай*	*quyuq choy*	*thick tea*

Table 1. Selected collocations in English, Russian and Uzbek

The rows of the table are not equally instructive. *Give a question* and *thick tea* are simply unidiomatic, and any competent speaker of English rejects them at once. Here intuition does its job well. The more interesting case is *accept a decision*. The phrase is grammatical, and COCA contains 6 instances of it, so a learner who checks it against a search engine, or even against the corpus, will find that it exists. The concordance lines tell a different story. In 3 of the 6 lines examined, the subject accepts a decision taken by someone else, usually an authority, as in *accept the court's decision* or *accept the referee's decision*. The combination is attested, but it does not mean what the Uzbek or Russian speaker intends by it. Intuition, which asks only whether a phrase is possible, lets it through. The corpus answers a different question, namely what the phrase usually does, and that answer exposes the error. Learner corpus research has shown that verb-noun combinations cause trouble even for advanced learners and that many of the errors can be traced to the first language (Nesselhauf, 2005). The present example suggests that the most persistent of them are likely to be those that happen to coincide with a real, but differently used, English phrase.

Russian and Uzbek point in the same direction here. *Принять решение* and *qaror qabul qilmoq* are both built on a verb of accepting. Whether the Uzbek phrase arose under Russian influence or independently is a question for historical study. For the learner of

English it hardly matters, since transfer from either language leads to the same mistake.

The case of *strong rain* is different again. It is not ungrammatical, and it is not absent from COCA either: the corpus contains 9 examples, against 1195 instances for *heavy rain*. A dictionary or the answer key of a textbook would simply mark it as wrong. The corpus does something more honest. It shows that one option is roughly 133 times as frequent as the other, and it lets the reader examine the handful of contexts in which the rarer option does occur. What the researcher receives is not a verdict but a distribution, and a distribution is a more accurate description of how a language actually behaves.

Semantic Prosody and What Intuition Misses

Collocation concerns which words occur together. Semantic prosody concerns the evaluative colouring that a word acquires from its habitual collocates. The term was brought into wide use by Louw (1993), who built on Sinclair's observations, and it is here that the gap between intuition and evidence is widest.

Sinclair (1991) found that the phrasal verb *set in* is used almost exclusively with subjects denoting unpleasant states of affairs, of the kind represented by *rot* or *decay*. Nothing in the dictionary meaning of *set in*, roughly 'begin and seem likely to continue', predicts this, and native speakers asked to explain the verb rarely mention it. Stubbs (1995) described a similar tendency in the verb *cause*. Dictionaries define it neutrally as 'make something happen', and in the abstract most speakers would probably agree that one can cause anything at all. The corpus paints a different picture. Among the 50 strongest noun collocates of CAUSE in COCA, 45 denote something unpleasant, such as *problems*, *cancer* or *death*, while 2 are neutral and only 3 are positive. A writer who says that a new policy *caused great improvements* has produced a grammatical sentence that nevertheless sounds faintly ironic or awkward to an English reader, and neither the writer nor, very often, the teacher will be able to say why.

Russian divides the same semantic territory differently. *Причинять* is restricted to harm, as in *причинять вред*, *ущерб*, *боль*, *неудобства*, and the RNC data show that all strongest noun collocates are negative. *Вызывать*, by contrast, combines freely with both pleasant and unpleasant states, so that *вызывать интерес* and *вызывать восхищение* are as normal as *вызывать тревогу*. Russian thus lexicalises a distinction that English leaves to prosody. The consequence for translation is immediate. A translator who renders *вызывать интерес* as *cause interest* produces a phrase that COCA records only 12 times, against 12 for *arouse interest* and 26 for *attract interest*. The Russian verb is neutral, but its most obvious English equivalent is not.

Uzbek offers a third arrangement. *Qiziqish uyg'otmoq*, literally 'to awaken interest', is close in structure to English *arouse interest*, and the same verb combines with *shubha* 'suspicion' and *xavotir* 'concern'. *Zarar yetkazmoq* is plainly negative in its object. The verb *keltirib chiqarmoq* 'to bring about' is harder to judge. An Uzbek speaker's intuition may suggest that it leans towards problems and illnesses rather than achievements, but this is precisely the kind of claim the present article cautions against. Without a corpus it remains an impression, and the history of research on semantic prosody suggests that impressions of this sort need to be checked before they are written into dictionaries or textbooks.

What Uzbek Asks of a Corpus

The Uzbek examples raise a practical point that goes beyond the present shortage of data. Even when a large Uzbek corpus is in everyday use, methods developed for English cannot be carried over to it without adjustment, for at least three reasons.

The first reason is morphology. The English noun *decision* appears in two forms, *decision* and *decisions*. The Uzbek noun *qaror*, by contrast, surfaces as *qarorni*, *qarorlar*,

qarorlarni, *qarorga*, *qarorimiz*, *qaroringizni* and many other forms, because number, possessive and case suffixes attach to the stem one after another. A search for the bare form will miss most of the evidence. Collocation statistics for Uzbek will therefore be only as good as the lemmatisation behind them, and a reliable morphological analyser is not an optional extra but the foundation of the whole enterprise.

The second reason is word order. English verb–object collocations are normally found to the right of the verb, which is why a right-hand span works well for *make a decision*. In Uzbek the object precedes the verb and can be separated from it by other material, as in *qarorni oxir-oqibat qabul qildi* 'finally made the decision'. A span designed for English would look in the wrong direction. Word order is a known difficulty even for closer language pairs. In a study of English–Russian parallel corpora, Radjabova (2025a) found word order variation to be the largest single source of automatic misalignment, accounting for 24% of errors, ahead of ellipsis and idiomatic translation. Since Uzbek departs from English word order more consistently than Russian does, the problem can be expected to be sharper still in English–Uzbek material. Many Uzbek verbal collocations are, moreover, built on light verbs such as *qilmoq*, *bermoq* and *olmoq*, so that *qabul qilmoq* is already a two-word unit before any object is added. A corpus that does not treat such compound verbs as single lexical items will break up exactly the combinations the phraseologist most wants to study.

The third reason is orthography. Uzbek texts circulate in both Latin and Cyrillic script, and even within the Latin script the letters *o'* and *g'* and the apostrophe known as *tutuq belgisi* are typed with several different characters. A glance at almost any official announcement will show the same letter entered in two or three ways. For a computer these are different words, and normalisation has to be solved before any frequency count can be trusted. None of these problems is new to the teams building the national corpus (O'zbek tilining milliy korpusi, 2022). They are mentioned here not as criticism but as a reminder that the quality of the linguistic questions we can put to an Uzbek corpus will depend on how they are solved.

Conclusion

The comparison of English, Russian and Uzbek material leads to three conclusions. First, intuition is a reliable judge of possibility and an unreliable judge of typicality. It rejects *give a question* but accepts *accept a decision*, because the second phrase is possible even though it does not mean what the speaker intends. Second, the evaluative colouring of words, their semantic prosody, is largely hidden from introspection and differs from language to language, as the comparison of English *cause* with Russian *причинять* and *вызывать* shows. Third, the value of a corpus lies less in the answers it provides than in the evidence it presents. A dictionary, a textbook key and, increasingly, an AI assistant each give the user a single answer. A concordance gives a distribution of attested usage and leaves the interpretation to the reader. Both kinds of tool have their place, but only the second lets the researcher, or the student, see why an answer is right. This is the principle behind Johns's (1991) data-driven learning, in which learners work directly with corpus evidence, and experimental studies have on the whole found the approach effective (Boulton & Cobb, 2017). In Uzbekistan this line of work has been developed by Radjabova, who has described the methodological characteristics of corpus technologies in foreign language teaching (Radjabova, 2022) and their use in teaching academic writing (Radjabova, 2023), and who has tested COCA-based exercises, including work on semantic prosody, with B2-level students at the Uzbekistan State World Languages University (Radjabova, 2025b).

None of this makes the philologist's intuition redundant. Concordance lines have to be read, and reading them is an act of interpretation that no frequency count can

perform. Sinclair's (2004) well-known call to trust the text is not a call to distrust the reader. What corpus linguistics asks for is more modest and at the same time more demanding, namely that intuition be treated as a source of hypotheses rather than as evidence. For Uzbek linguistics, which is now building its own corpus resources, this suggests that the most useful questions will concern not the size of the national corpus but whether it allows researchers to see collocation at all, through sound lemmatisation, a sensible treatment of word order and compound verbs, and consistent spelling. When it does, many of the claims now made about Uzbek phraseology on the strength of intuition can be put to the test, and some of them, no doubt, will turn out to be true.

References:

1. Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
2. Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393.
3. Davies, M. (2008–present). *The Corpus of Contemporary American English (COCA)*. https://www.english-corpora.org/coca/
4. Firth, J. R. (1957). A synopsis of linguistic theory, 1930–1955. In *Studies in linguistic analysis*. 1–32. Blackwell.
5. Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In memory of J. R. Firth*. 148–162. Longman.
6. Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.
7. Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.
8. Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal, 4*, 1–16.
9. Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair*. 157–176. John Benjamins.
10. McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
11. Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins.
12. O'zbek tilining milliy korpusi: muammo va vazifalar. (2022). *Xalqaro ilmiy-amaliy konferensiya materiallari (2022-yil 31-may)*. ToshDO'TAU.
13. Radjabova, G. G. (2022). Methodological characteristics of corpus technologies in teaching foreign language. *International Journal on Integrated Education, 5*(1), 157–163. https://doi.org/10.31149/ijie.v5i1.2645
14. Radjabova, G. G. (2023). Corpus technologies in teaching academic writing. *Foreign Languages in Uzbekistan, 1*(48), 92–103. https://doi.org/10.36078/1679549918
15. Radjabova, G. G. (2025a). Linguistic aspects of automatic text alignment in parallel corpora. *Western European Journal of Linguistics and Education, 3*(5), 121–125.
16. Radjabova, G. G. (2025b). Using Corpus of Contemporary American English [COCA] to enhance writing competence of B2 level students. *Foreign Languages in Uzbekistan, 11*(6), 110–124. https://doi.org/10.36078/1767689405
17. Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

18. Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge.
19. Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language, 2*(1), 23–55.
20. Виноградов, В. В. (1977). Об основных типах фразеологических единиц в русском языке. В *Избранные труды. Лексикология и лексикография*. 140–161. Наука.
21. Национальный корпус русского языка. (б. г.). *Национальный корпус русского языка*. https://ruscorpora.ru