

## ИССЛЕДОВАНИЯ СЛАВЯНСКИХ ЯЗЫКОВ

УДК 811.161.3'373

**Банько Ірына Мікалаеўна,**

аспірант аддзела лексікалогіі  
і лексікаграфіі Інстытута мовазнаўства  
імя Якуба Коласа  
ДНУ «Цэнтр даследавання беларускай  
культуры, мовы і літаратуры»  
НАН Беларусі  
г. Мінск, Беларусь

**Iryna Banko**

PhD Student of the Department of Lexicology  
and Lexicography  
Institute of Linguistics named after Yakub Kolas  
Centre of Studies of Belarusian Culture,  
Language and Literature  
of the NAS of Belarus  
Minsk, Belarus  
irina.banko9@mail.ru

ДА ПЫТАННЯ ВЫДЗЯЛЕННЯ АТРЫБУТЫЎНЫХ УСТОЙЛІВЫХ АДЗІНАК  
НЕІДЫЯМАТЫЧНАГА ХАРАКТАРУREVISITING THE IDENTIFYING ATTRIBUTIVE FIXED UNITS  
OF A NON-IDIOMATIC NATURE

У артыкуле аналізуюцца прадстаўленыя ў замежнай лінгвістыцы погляды на ўстойлівыя неідыяматычныя спалучэнні, семантыка якіх складаецца са значэнняў кампанентаў, адзін з якіх пераасэнсоўваецца, а таксама тэрміны, што выкарыстоўваюцца для абазначэння такіх адзінак. Асабліва ўвага надаецца падыходам да вычленення ўстойлівых адзінак з кантэксту, адзначаецца неабходнасць прымянення комплекснага метаду, які спалучае корпусны і лінгвістычны падыходы. Апісваюцца магчымасці вылучэння ўстойлівых неідыяматычных спалучэнняў з вялікіх масіваў тэкстаў з дапамогай Беларускага N-корпуса і праграмага сродку AntConc на беларускім моўным матэрыяле.

Ключавыя словы: *устойлівае неідыяматычнае спалучэнне; аўтаматычны пошук; корпус тэкстаў; калакацыя; фразеалагічнае спалучэнне; фразеалогія; кантэкст.*

The article analyzes the views presented in foreign linguistics on fixed non-idiomatic collocations, the semantics of which consists of the meanings of components, one of which is reinterpreted, as well as the terms used to denote such units. Particular attention is paid to approaches to isolating fixed units from the context, and the need for an integrated method combining corpus and linguistic approaches is noted. The possibilities of isolating fixed non-idiomatic collocations from large arrays of texts using the Belarusian N-corpus and the *AntConc* software tool on the Belarusian language material are described.

Key words: *fixed non-idiomatic collocation; automatic search; text corpus; collocation; phraseological combination; phraseology; context.*

Недаследаванымі ў беларускай лінгвістыцы застаюцца ўстойлівыя адзінкі, у якіх адбываецца пераасэнсаванне толькі ў адным з кампанентаў, яны не ўваходзяць у ядро фразеалогіі, але маюць з ім агульныя рысы. Можна сказаць, што такія ўстойлівыя спалучэнні знаходзяцца на перыферыі фразеалагічнай сістэмы. Атрыбутыўныя выразы *агрубелы слых, адкіды грамадства, губкі банцікам, зводны брат, дыплом з адзнакай, пачуццё абавязку* і інш., якія маюць аналітычную семантыку, спецыяльна не разглядаліся ў беларускіх навуковых крыніцах. У замежнай лінгвістыцы да такіх спалучэнняў прымяняюцца разнастайныя падыходы, але і сёння адсутнічае іх агульнапрынятае разуменне і тэрміналагічнае абазначэнне. У сваім даследаванні ўжываем тэрмін *устойлівае неідыяматычнае спалучэнне* (далей – УНС), які, на наш погляд, дастаткова поўна адлюстроўвае асаблівасці гэтых устойлівых выказаў.

Упершыню адзінкі з пераасэнсаваннем у адным кампаненце былі вылучаны В. У. Вінаградавым як *фразеалагічныя спалучэнні* – несвабодныя, замкнутыя рады слоў, дзе магчыма сінанімічная замена кампанентаў, напрыклад, *щекотливый вопрос, щекотливое положение* і інш. Фразеалагічна звязанае значэнне складана вызначыць, яно не такое выразнае, як свабоднае значэнне і драбіцца на індывідуальныя адценні, якія ўласцівы асобным фразам [1, с. 26].

У далейшым сутнасць такіх адзінак ацэньвалася па-рознаму. У межах фразеалогіі іх аналізуюць В. М. Тэлія і М. М. Шанскі. В. М. Тэлія называе іх *лексічнымі калакацыямі з аналітычным тыпам значэння* і адносіць да фразеалагізмаў, якія валодаюць ідыяматычнасцю, бо адзін з кампанентаў мае адзінкавы пераклад, магчымы для дадзенага слова толькі ў пэўным спалучэнні, а значэнне такога кампанента фразеалагічна абумоўлена [2, с. 52–53]. Свабодны кампанент такіх адзінак прадвызначае «семантычнае прачытанне» звязанага з ім па сэнсе і форме кампанента, які набывае завершанае тлумачэнне толькі ў спалучэнні з гэтым галоўным кампанентам [3, с. 65–66]. Ужыты тэрмін *калакацыя* быў адзначаны ў «Слоўніку лінгвістычных тэрмінаў» (1966) В. С. Ахманавай са значэннем «лексіка-фразеалагічная абумоўленая спалучальнасць слоў у маўленні як рэалізацыя іх полісеміі» [4, с. 199]. М. М. Шанскі адзначае, што чым шырэйшым будзе кола слоў, з якімі можа злучацца несвабодны кампанент такога фразеалагічнага спалучэння, тым бліжэй такое спалучэнне да фразеалагічных выказаў [5, с. 203].

Даследчыца Н. М. Амосава ўстрымліваецца ад ужывання тэрміна *фразеалагічныя спалучэнні* і называе УНС *узуюльна абмежаваным тыпам спалучэнняў*, а значэнне семантыкі рэалізуемага слова – *узуюльна звяза-*

ным значэннем [6, с. 69]. Па яе назіраннях, варыянтнасць ключавых слоў падобных выказаў абумоўлена не семантыкай, а узусам, традыцыяй [Там жа].

Некаторыя аўтары прапануюць характаралагічныя прыкметы такіх выказаў, сярод іх пераасэнсаванне аднаго або двух кампанентаў, ужыванне асобных кампанентаў спалучэння з абмежаваным колам слоў і ўстойлівасць ужывання [7, с. 14]. Дадатковым крытэрыем вызначана неперакладальнасць асобных кампанентаў спалучэння на іншыя мовы.

На думку У. Л. Архангельскага, які разглядае ў межах фразеалогіі ўсе ўстойлівыя спалучэнні (*фраземы*), нават са структурай сказа, утварэнне спалучэнняў з пераасэнсаваннем у адным кампаненце абумоўлена стратай або аслабленнем асноўных значэнняў слоў. Устойлівасць фразеалагічнай адзінкі вызначаецца па ўнутранай залежнасці паміж членамі фразеалагічнай адзінкі [8, с. 185] і звязана з абмежаваннем у выбары пераменных на розных моўных узроўнях [8, с. 181].

Яшчэ адным крокам у вывучэнні ўстойлівых неідыяматычных адзінак стала манаграфія А. Г. Барысавай «Калакацыі. Што гэта такое і як іх вывучаць» (1995). Тут пад *калакацыяй* разумеецца адзінка, якая з'яўляецца несвабодным неідыяматычным спалучэннем, дзе свабодны кампанент ужываецца ў сваім прамым значэнні, а несвабодны пераасэнсоўваецца.

Прыхільнікі структурнага падыходу таксама ўключаюць УНС у склад фразеалогіі. А. І. Смірніцкі параўноўвае іх з вытворнымі словамі, дзе вылучае семантычна значную і семантычна падпарадкаваную марфему [9, с. 210].

З пазіцый лексікаграфіі спалучэнні з пераасэнсаваннем у адным кампаненце былі разгледжаны К. А. Іваннікавай. Усе іменныя спалучэнні даследчыца падзяляе паводле функцыі ў маўленні (свабодныя спалучэнні, тэрміналагічныя спалучэнні і сродкі выразнасці) і паводле спосабу абазначэння паняцця (спалучэнні, значэнне якіх можна вывесці са значэння яго кампанентаў, і фразеалагічныя спалучэнні) [10, с. 40–41]. УНС згодна з прапанаванай класіфікацыяй па функцыі ў маўленні даследчыца адносіць да сродкаў выразнасці, а па спосабе абазначэння паняцця – да фразеалагічных (або лексічна непадзельных) спалучэнняў [Там жа].

У той жа час Б. А. Ларын, кіруючыся гістарычным прынцыпам, выводзіць фразеалагічныя спалучэнні за межы фразеалогіі. На яго думку, такія адзінкі валодаюць мінімальнай ступенню ідыяматычнасці, знаходзяцца на перыферыі сапраўдных фразеалагічных злучэнняў і не маюць характэрных прыкмет [11, с. 223]. Этап іх фарміравання навуковец лічыць пераходным і кароткачасовым.

Як відаць, погляды мовазнаўцаў на сутнасць і прыкметы УНС адрозніваюцца, аднак прапанаваны лексіка-семантычны падыход да выдзялення такіх адзінак да гэтага часу пашыраны ў лінгвістыцы.

З развіццём ідэй машыннага перакладу паўстала пытанне аб спосабах вычлянення ўстойлівых адзінак з тэксту. УНС у гэтым аспекце асобна не аналізаваліся і разглядаліся ў межах агульнай праблемы. Адзін з першых падыходаў прапанаваны ў рамках тэорыі «Сэнс ↔ Тэкст» (І. А. Мельчук), дзе несвабоднае спалучэнне разумеецца як спалучэнне семантычнай дамінанты, якая вызначае выбар іншага кампанента для перадачы ўсяго спалучэння [12, с. 231]. Распрацаваная даследчыкам тэорыя лексічнай функцыі атрымала пашырэнне, яна прымяняецца і ў сучасных прыкладных даследаваннях.

Так, А. М. Баранаў і Д. А. Дабравольскі распрацавалі класіфікацыю УНС, якія называюць *калакацыямі*, на аснове лексічных функцый, апісаных І. А. Мельчуком і А. К. Жалкоўскім. Калакацыі – гэта слабаідыяматычныя фразеалагізмы пераважна са структурай словазлучэння, у якіх вылучаюць семантычна галоўны кампанент (базу), што ўжываецца ў прамым значэнні, і дапаможны кампанент (калакатар) [13, с. 73]. Выбар базы залежыць ад інтэнцыі моўцы, а выбар калакатара – ад базы. Тэрмін *калакацыя* становіцца найбольш пашыраным для абазначэння ўстойлівых спалучэнняў такога тыпу, аднак у залежнасці ад падыходу яго змест мяняецца.

У межах лексіка-граматычнага падыходу (А. І. Бальшакова, А. Г. Барысава, В. А. Мітрафанова) калакацыі разглядаюцца як семантыка-сінтаксічныя адзінкі, ці лексічна вызначаныя элементы граматычных структур. Для іх вычлянення прымяняюцца лексіка-граматычныя шаблоны – структурныя мадэлі моўных канструкцый з указаннем істотных граматычных прыкмет лексем і сінтаксічных умоў іх ужывання. Сёння лексіка-граматычны падыход выкарыстоўваецца разам з іншымі пры распрацоўцы метадаў вычлянення ўстойлівых спалучэнняў з кантэксту.

Да лексіка-граматычнага падыходу блізкі сінтаксічны падыход, у межах якога калакацыя разумеецца як паслядоўнасць слоў, што ўтварае сінтаксічную канструкцыю (назоўнік + прыметнік, дзеяслоў + назоўнік (аб'ект), назоўнік + прыназоўнік + назоўнік) [14, с. 17]. Для вылучэння адзінак з кантэксту прапанавана захаванне пэўнага парадку размяшчэння кампанентаў і адсутнасць семантычных трансфармацый. З пункту гледжання сінтаксічнага падыходу да калакацый могуць быць прыменены крытэрыі сінтаксічнай звязанасці, кантактнага размяшчэння ў тэксце і наяўнасці падпарадкавальнай сувязі паміж кампанентамі.

У межах корпуснага падыходу адбылася змена паняцця *калакацыя*, якое сёння разумеецца як спалучэнне двух і больш слоў, што сустракаецца часцей, чым гэта можна было чакаць выпадкова; вылучаюць фразеалагізаваныя і свабодныя калакацыі [15, с. 10]. Даследаванне ўстойлівых адзінак праводзіцца з апорай на кантэкст – збор тэкстаў, у якім калакацыя разумеецца як суправаджэнне слова, або слоўны матэрыял, у якім яно часцей за ўсё сустракаецца [16, с. 97].

Аднак некаторыя даследчыкі такі корпусна-арыентаваны падыход лічаць выключна статыстычным. У межах статыстычных даследаванняў адбываецца замена лінгвістычных крытэрыяў статыстычнымі або матэматычнымі. Пры гэтым корпусна-арыентаваны падыход мае шэраг пераваг, сярод якіх можна вылучыць зніжэнне рэсурсаёмістасці працэсу экстракцыі, большае пакрыццё тэксту, перавагі ў вывучэнні варыятыўнасці словазмяняльных формаў і сінтаксічных канструкцый і выяўленне калакацый, якіх няма ў закрытых спісах і слоўніках [16, с. 98]. Асноўная праблема статыстычнага падыходу – гэта павышэнне працэнта памылак у створаных аўтаматычна рэестрах калакацый. Акрамя таго, наяўныя падыходы распрацаваны толькі для двухслоўных спалучэнняў і не даюць магчымасці вызначыць «ядро» – галоўны кампанент спалучэння [Там жа].

Для паспяховага вырашэння праблемы вычлянення ўстойлівых адзінак з кантэксту даследчыкамі прымяняецца комплексны метады, які камбінуе ў пэўнай паслядоўнасці некалькі вышэйадзначаных. Недасканаласць прапанаваных метадаў патрабуе абавязковага рэдагавання лінгвістамі даных, атрыманых аўтаматычна машынным спосабам [16, с. 99; 17], таму лексіка-семантычны падыход да ўстойлівых адзінак не губляе сваёй актуальнасці.

Сёння корпусны падыход займае адно з вядучых месцаў у мовазнаўчых даследаваннях, ён шырока прымяняецца ў лексікаграфічнай практыцы, у лінгвістычнай педагогіцы і інш. У беларускім мовазнаўстве падобныя даследаванні таксама прадстаўлены, многія з іх праводзяцца на базе Беларускага N-корпуса [18], распрацаванага супрацоўнікамі Інстытута мовазнаўства імя Якуба Коласа НАН Беларусі.

Беларускі N-корпус забяспечаны праграмнымі сродкамі, з дапамогай якіх вычляненне ўстойлівых неідыяматычных спалучэнняў з тэкстаў розных жанраў уяўляецца магчымым. У корпусе прадугледжаны выбар спалучэнняў за кошт «двухпраходнага» пошуку: спачатку ў базе выконваецца пошук абзацаў па словах і граматычных прыкметах, затым ідзе праверка знойдзеных абзацаў на адпаведнасць больш складаным крытэрыям [19, с. 76].

Беларускі N-корпус можа выкарыстоўвацца для пошуку УНС з ужо вядомымі кампанентамі з кантактным або дыстанцыйным размяшчэннем ва ўсіх магчымых граматычных формах. Напрыклад, пры пошуку спалучэння *зялёны чай* корпус прадстаўляе спіс кантэкстаў з гэтым выразам, але не дае колькасных паказчыкаў, падлік ужыванняў магчымы ўручную. Аўтаматычны падлік ужыванняў лексічных адзінак і выказаў магчымы толькі, калі іх колькасць большая за 300.

Пэўны патэнцыял мае пошук у корпусе па адным слове з выкарыстаннем граматычных паказчыкаў. Так, выбарка спалучэнняў “прыметнік + чай” у корпусе прадстаўлена 614 ужываннямі: *травяны чай, завараны чай, гарачы чай, салодкі чай, рамонкавы чай, духмяны чай, зялёны чай, чорны чай* і інш. у розных склонавых формах. Аднак колькасныя паказчыкі для кожнага спалучэння магчыма атрымаць таксама толькі ўручную.

З дапамогай інструмента «Кластар» можа выконвацца пошук словазлучэнняў з паказам колькасці ўжыванняў у тэкстах корпуса. «Кластар» дае магчымасць па адным слове ў пошукавым запыце з тэкстаў корпуса выявіць словазлучэнні. Гэта выконваецца праз пазначэнне колькасці слоў злева і справа ад пошукавага слова. Атрыманыя вынікі абавязкова павінны быць дадаткова прааналізаваны з-за недасканаласці сучасных праграмных сродкаў, якія праводзяць пошук толькі па частотных крытэрыях. Так, пры пошуку праз «Кластар» спалучэнняў са словам *чай* з умовай 1 слова злева прапануюцца такія спалучэнні, як *чорны чай, зялёны чай, гарачы чай, моцны чай, кубак чаю, шклянка чаю* і інш. у розных граматычных формах, спалучэнні лексем і службовых слоў *на чай, з чаем, і чай* і г.д. Акрамя таго, у выніках пошуку прадстаўлены спалучэнні частак слоў з-за памылак друку, недастатковай расчытанасці тэксту і пераносу (*звы чай, твор чай* і інш.). Пры пошуку на слова *зялёны* з умовай 1 слова справа прадстаўлены іншы спіс спалучэнняў, дзе першае месца займае 497 словаўжыванняў выразу *зялёнай эканомікі*, а *зялёны чай* стаіць на 27 пазіцыі і прадстаўлены 67 ужываннямі.

Вынікі пошукавага запыту па «Кластары» адлюстроўваюць кантактнае размяшчэнне кампанентаў УНС. Пры пошуку выказаў з дыстанцыйным размяшчэннем кампанентаў павінна быць зададзена большае значэнне для лексем злева і справа. Такім чынам, Беларускі N-корпус дае магчымасць выбару спалучэнняў, аднак прапанаваныя падыходы працаёмкія і патрабуюць значнай апрацоўкі лінгвістамі атрыманых вынікаў.

Для выяўлення калакацый з тэкстаў мэтазгодна прымяненне корпусных менеджараў, у якіх прадугледжаны меры асацыяцыі, з дапамогай чаго можна праводзіць падлік частот слоў або словаформ і частот сумеснай

сустракаемасці. Шырокія магчымасці для лінгвістычных даследаванняў мае праграма AntConc, распрацаваная Лоўрэнсам Энтані (Laurence Anthony; Waseda University, Japan) [20]. Праграма мае англамоўны інтэрфейс, прадугледжвае стварэнне ўласнага корпуса і можа выкарыстоўвацца для аналізу тэкстаў, у тым ліку і на беларускай мове.

Для вычлянення УНС з тэкстаў у AntConc магчыма выкарыстанне наступных інструментаў: Clusters, які па зададзеным слове дазваляе выявіць пары слоў, што «прыцягваюцца» адно да аднаго; N-Gram таксама выяўляе пары слоў, але без пошукавага слова; Collocates – інструмент для пошуку калакацый у тэкстах з дапамогай мер асацыяцыі.

Для аналізу намі быў падрыхтаваны корпус, што ўключае тэксты розных жанраў 143 беларускіх аўтараў XX – пачатку XXI ст. Агульная колькасць словаўжыванняў склала 26 036 965.

Праілюструем работу з праграмай AntConc на прыкладзе калакацый з закрытага спіса са структурай *прыметнік + назоўнік і назоўнік + назоўнік*. Выкарыстаем папярэдне загрузаныя ў праграму мастацкія тэксты і звернемся да інструмента Cluster. Для выбаркі пазначым калакацыі з мінімальнай частатой сустракаемасці (Min. Freq) 1, памер калакацыі (N-Gram Size) будзе складаць 2 словы. У выпадку пошуку калакацый *прыметнік + назоўнік* з дыстанцыйным размяшчэннем патрэбна будзе пазначыць памер калакацыі роўны або большы за 3. У графе пошуку ўвядзём прыметнік *малады*. Для выяўлення спалучэнняў з прыметнікам ва ўсіх магчымых склонавых формах неабходна праводзіць пошук па аснове слова, пазначыўшы патэнцыяльны канчатак прыметніка зорччай: *малад\**. Таксама інструмент Cluster дазваляе праводзіць пошук слоў злева і справа, што выконваецца з дапамогай функцыі Search Term Position: OnLeft; OnRight; OnLeft / Right.

Вынікі аўтаматычнага пошуку двухкампанентных спалучэнняў з прыметнікам *малады* ў праграме AntConc прадстаўлены на мал. 1.

У аўтаматычна створаны рэестр калакацый увайшло 1331 словазлучэнне з прыметнікам *малады* і спалучэнні прыметніка *малады* са службовымі словамі. Атрыманы спіс павінен быць адрэдагаваны з мэтай адбору граматычна правільных словазлучэнняў з падпарадкавальнай сувяззю і далейшага вычлянення з іх УНС на падставе лінгвістычных крытэрыяў. Так, сярод выяўленых спалучэнняў з прыметнікам *малады* адзначаюцца граматычна правільныя канструкцыі *малады чалавек, малады хлопец, малады пісьменнік, малады твар* і г.д., спалучэнні, што з'яўляюцца часткай больш шырокіх па структуры спалучэнняў (*малады прыгожы, малады здаровы* і пад.), а таксама частотныя спалучэнні лексем, што не з'яўляюцца

словазлучэннямі (*малады і, малады з, малады як і інш.*). Да патэнцыяльных УНС з групы граматычна правільных спалучэнняў можна аднесці *малады хлеб, малады век, малады месяц*.

AntConc

File Edit Settings Help

Target Corpus

Name: my\_corpus

Files: 5469

Tokens: 26036965

Cluster Types 1331 Cluster Tokens 4546 Page Size 100 hits 1 to 100 of 1331 hits

Cluster	Rank	Freq	Range
1 малады чалавек	1	565	236
2 малады і	2	220	164
3 малады хлопец	3	147	100
4 малады яшчэ	4	125	102
5 малады а	5	90	79
6 малады мужчына	6	70	56
7 малады з	7	59	50
8 малады але	8	53	49
9 малады быў	9	49	43
10 малады настаўнік	10	39	21
11 малады не	11	38	37
12 малады хлапец	11	38	33
13 малады у	13	34	32
14 малады пісьменнік	14	33	19
15 малады паэт	15	32	23
16 малады прыгожы	16	31	29
17 малады ён	17	28	26
18 малады сад	18	26	25
19 малады твар	18	26	26
20 малады здаровы	20	24	21
21 малады як	21	23	22
22 малады голас	22	21	20

Search Query  Words  Case  Regex Cluster Size 2 Min. Freq 1 Min. Range 1

малады Start  Adv Search

Sort by Frequency  Invert Order Search Term Position  On Left  On Right  On Left/Right

Progress 100%

Using Cluster cache Time taken (creating cluster results): 0.0843 sec

Мал. 1. Вынікі аўтаматычнага пошуку двухкампанентных спалучэнняў з прыметнікам *малады*

Такім жа чынам пошук можа выконвацца і па граматычна галоўным кампаненце УНС. На мал. 2 прадстаўлены вынікі пошуку УНС з кампанентам-назоўнікам *голас*. З 2879 атрыманых калакацый як патэнцыяльныя УНС можна вылучыць спалучэнні *разважлівы голас, жывы голас*.

Аналагічна праводзіцца пошук атрыбутыўных спалучэнняў са структурай *назоўнік + назоўнік*. Аўтаматычна створаны спіс калакацый з субстантывам *даніна* ўключае 25 спалучэнняў (мал. 3), сярод якіх падпарадкавальныя словазлучэнні, спалучэнні назоўніка з прыназоўнікам (*даніна за*), назоўніка з прыметнікам, граматычна залежным ад наступнага субстантыва (*даніна незабыўнай*). Да патэнцыяльных УНС могуць быць аднесены *даніна модзе, даніна часу, даніна павагі, даніна абавязку, даніна памяці, даніна ўмоўнасці*.



KWIC	Plot	File View	Cluster	N-Gram	Collocate	Word	Keyword
Cluster Types 2879		Cluster Tokens 10647		Page Size 5000 hits		1 to 2879 of 2	
	Cluster	Rank	Freq	Range			
84	падаць голас	80	16	15			
85	пачуць голас	80	16	15			
86	потым голас	80	16	14			
87	радасны голас	80	16	16			
88	тут голас	80	16	16			
89	ты голас	80	16	16			
90	чуўся голас	80	16	16			
91	ўжо голас	80	16	15			
92	павышаючы голас	92	15	14			
93	панізіў голас	92	15	13			
94	пачула голас	92	15	13			
95	першы голас	92	15	12			
96	разважлівы голас	92	15	15			
97	самы голас	92	15	14			
98	яму голас	92	15	13			
99	грубы голас	99	14	14			
100	наш голас	99	14	13			
101	нейчы голас	99	14	14			
102	пазнала голас	99	14	12			
103	прагучаў голас	99	14	13			
104	прыемны голас	99	14	12			
105	слабы голас	99	14	14			
106	сцішыў голас	99	14	12			

Search Query  Words  Case  Regex Cluster Size 2 Min. Freq 1

Sort by Frequency  Invert Order Search Term Position  On Left  On Right

Мал. 2. Вынікі пошуку УНС з кампанентам-назоўнікам *голас*

KWIC	Plot	File View	Cluster	N-Gram	Collocate	Word	Keyword
Cluster Types 25		Cluster Tokens 32		Page Size 100 hits		1 to 25 of 25 hits	
	Cluster	Rank	Freq	Range			
1	даніна модзе	1	4	4			
2	даніна часу	2	3	2			
3	даніна за	3	2	2			
4	даніна павагі	3	2	2			
5	даніна абавязку	5	1	1			
6	даніна адна	5	1	1			
7	даніна бутэлька	5	1	1			
8	даніна бяроз	5	1	1			
9	даніна добрым	5	1	1			
10	даніна згадкам	5	1	1			
11	даніна зжытаму	5	1	1			
12	даніна любасцям	5	1	1			
13	даніна незабыўнай	5	1	1			
14	даніна непараўнальны	5	1	1			
15	даніна памяці	5	1	1			
16	даніна самаздаволення	5	1	1			
17	даніна светаадчуванню	5	1	1			
18	даніна светлым	5	1	1			
19	даніна трывожнаму	5	1	1			
20	даніна шырока	5	1	1			
21	даніна яго	5	1	1			
22	даніна і	5	1	1			
23	даніна ў	5	1	1			

Search Query  Words  Case  Regex Cluster Size 2 Min. Freq 1

Sort by Frequency  Invert Order Search Term Position  On Left  On Right

Мал. 3. Вынікі пошуку УНС з кампанентам-назоўнікам *даніна*

Для ўстанаўлення прыналежнасці такіх патэнцыяльных выразаў да УНС неабходна правядзенне спецыяльнага лінгвістычнага аналізу на прадмет адпаведнасці крытэрыям, уласцівым устойлівым неідыяматычным спалучэнням.

Сучасныя тэхналогіі могуць паспяхова прымяняцца пры выдзяленні УНС з тэкстаў. З дапамогай Беларускага N-корпуса могуць аналізавацца тэксты, якія прадстаўлены ў корпусе, а такія праграмныя сродкі, як AntConc, дазваляюць праводзіць аналіз тэкстаў на выбар карыстальніка. Праблема ўключэння ў спісы службовых слоў можа быць вырашана праз дабаўленне стоп-слоў у праграме AntConc. У цэлым спосабы вылучэння УНС з дапамогай электронных сродкаў дзейсныя і значна спрашчаюць працу па пошуку прадстаўленых адзінак у любых тэкставых крыніцах. Асноўным недахопам сёння з'яўляецца немагчымасць уліку праграмамі семантыкі аналізуемых адзінак, таму павінна праводзіцца далейшая праверка і рэдагаванне аўтаматычна створаных спісаў.

## ЛІТАРАТУРА

1. *Виноградов, В. В.* Русский язык: грамматическое учение о слове : учеб. пособие / В. В. Виноградов. – Л. : М. : Учпедгиз, тип. Печат. двор, 1947. – 784 с.
2. *Телия, В. Н.* Что такое фразеология / В. Н. Телия. – М. : Наука, 1966. – 86 с.
3. *Телия, В. Н.* Русская фразеология: семантический, прагматический и лингвокультурологический аспекты / В. Н. Телия. – М. : Шк. «Яз. рус. культуры», 1996. – 288 с.
4. *Ахманова, О. С.* Словарь лингвистических терминов : ок. 7 000 терминов / О. С. Ахманова. – Изд. 2-е стер. – М. : Совет. энцикл., 1969. – 607 с.
5. *Шанский, Н. М.* Лексикология современного русского языка : пособие / Н. М. Шанский. – М. : Просвещение, 1964. – 316 с.
6. *Амосова, Н. Н.* Основы английской фразеологии / Н. Н. Амосова. – Л. : Изд-во Ленингр. ун-та, 1963. – 208 с.
7. *Регинина, К. В.* Устойчивые словосочетания русского языка : учеб. пособие / К. В. Регинина, Г. П. Тюрина, Л. И. Широкова ; под ред. Л. И. Широковой. – М. : Рус. яз., 1976. – 296 с.
8. *Архангельский, В. Л.* Устойчивые фразы в современном русском языке: Основы теории устойчивых фраз и проблемы общей фразеологии / В. Л. Архангельский ; Рост. гос. пед. ин-т. – Ростов н/Д : Изд-во Рост. ун-та, 1964. – 315 с.
9. *Смирницкий, А. И.* Лексикология английского языка / А. И. Смирницкий ; подг. к печати и отред. канд. филол. наук В. В. Пассек. – М. : Изд-во лит. на иностр. яз., 1956. – 260 с.
10. *Иванникова, Е. А.* Устойчивые именные словосочетания как предмет фразеологии и лексикографии / Е. А. Иванникова // Современная русская лексикология : сб. ст. / Акад. наук СССР, Отд-ние лит. и яз., Науч. совет по лексикологии и лексикографии ; редкол.: С. Г. Бархударов (гл. ред.) [и др.]. – М., 1966. – С. 23–45.
11. *Ларин, Б. А.* Очерки по лексикологии, фразеологии и стилистике : [сб. ст.] / отв. ред. Б. А. Ларин. – Л. : Изд-во Ленингр. ун-та, 1956. – 232 с. – (Учен. зап. Ленингр. гос. ун-та им. А. А. Жданова ; № 243. Сер. филологических наук ; вып. 42).
12. *Иорданская, Л. Н.* Смысл и сочетаемость в словаре / Л. Н. Иорданская, И. А. Мельчук ; общ. ред. Е. Н. Саввиной. – М. : Яз. славян. культур, 2007. – 665 с.

13. Баранов, А. Н. Основы фразеологии: краткий курс : учеб. пособие / А. Н. Баранов, Д. О. Добровольский. – 3-е изд., стер. – М. : Флинта, Наука, 2016. – 307 с.
14. Гладка, В. А. Структурносинтаксичний підхід у вивченні колокацій: на матеріалі французької мови / В. А. Гладка // Наук. зап. Нац. ун-ту “Острозька акад.”. Сер. Філологічна. – 2013. – Вип. 39. – С. 16–20.
15. Хохлова, М. В. Исследование лексико-синтаксической сочетаемости в русском языке с помощью статистических методов : на базе корпусов текстов : автореф. дис. ... канд. филол. наук : 10.02.21 / М. В. Хохлова ; С. Петерб. гос. ун-т. – СПб., 2010. – 26 с.
16. Бобкова, Т. Извлечение колокаций из корпуса украинских текстов / Т. Бобкова // Studies about Lang. – 2015. – Vol. 27. – P. 93–105.
17. Wielki słownik języka polskiego PAN : zasady opracowania / zest. P. Źmigrodzki. – Kraków : Wydanie szóste, 2023. – URL : <https://wsjp.pl/page/zasady> (date of access: 12.06.2024).
18. Беларускі N-корпус. – URL : <https://bnkorporus.info/korporus.html> (дата звароту : 12.06.2024).
19. Кошчанка, У. Нацыянальны корпус беларускай мовы ў кантэксте корпуснай лінгвістыкі славянскіх краін / У. Кошчанка // Мовазнаўства. Літаратуразнаўства. Фалькларыстыка : XVI Міжнар. з'езд славістаў, Бялград, 19–27 жн. 2018 г. : дакл. беларус. дэлегацыі / НАН Беларусі, Беларус. кам. славістаў ; адк. рэд. А. А. Лукашанец. – Мінск, 2018. – С. 60–79.
20. AntConc // Laurence Anthony's Website. – URL : <https://laurenceanthony.net/software/antconc> (date of access: 09.07.2024).

*Поступила в редакцию 05.09.2024*