

## КОРПУСНАЯ ЛІНГВІСТЫКА І ІНТЭРНЭТ

Уплыў камп’ютарна-апасродкаванай камунікацыі і ўласна інтэрнэту на характар даследаванняў не ва ўсіх лінгвістычных напрамках добра заўважны. Аднак ёсць галіны мовазнаўства, дзе інтэрнэт выступае спрыяльным асяроддзем існавання лінгвістыкі. Адзін з паказальных прыкладаў – корпусная лінгвістыка. Па словах Э. МакЭнэры і Э. Хардзі, “корпус неабходны лінгвісту, як астраному – тэлескоп” [12, с. 53].

Корпусная лінгвістыка – прыклад паспяховага развіцця прыкладной лінгвістыкі ў віртуальным асяроддзі і ў інтэрдысцыплінарным рэчышчы. Яна пачала хутка развівацца амаль адначасова са з’яўленнем даступных шырокаму колу навукоўцаў камп’ютарных тэхналогій і стала жыватворнай крыніцай інтэрнэт-асяроддзя.

Наяўнасць корпуса ў сучасным універсітэце ўжо даўно нікога не здзіўляе. Уласна, ва ўніверсітэцкім асяроддзі, у Браўнаўскім універсітэце, і быў створаны першы лінгвістычны корпус, знакаміты *Браўнаўскі корпус* (*Brown Standard Corpus of Present-Day American English* – Браўнаўскі стандартны корпус сучаснай амерыканскай англійскай мовы), які стаў узорам для шматлікіх перайманняў [9]. Браўнаўскі ўніверсітэт – адзін з найстарэйшых амерыканскіх універсітэтаў – знаходзіцца ў г. Правідэнс (ЗША), заснаваны ён у 1764 г. Стварыўшы ў 1964 г. *Браўнаўскі корпус*, яго аўтары, Генры Кучэра і Нэльсан Фрэнсіс, зрабілі выдатны падарунак свайму ўніверсітэту да 200-гадовага юбілею і назаўсёды ўпісалі свае імёны ў гісторыю мовазнаўства.

Канцэптуальная схема першага корпуса была немудрагелістай – 500 прыблізна двухтысячаслоўных фрагментаў разнастайных тэкстаў, 1 014 312 словаўжыванняў з разнастайных жанраў. Нягледзячы на некаторую штучнасць канструкцыі, абраны фармат даказаў сваю жыццяздольнасць. Цяпер, вядома, мільёнам словаўжыванняў нікога не здзівіш, лік ідзе на сотні мільёнаў. Але ў 1960-я гг. мільённы аб’ём корпуса здаваўся эпохальнай падзеяй.

Відавочная перавага корпуснага стандарту – яго дакладнасць у апрацоўцы колькасных паказчыкаў маўлення. Напрыклад, дзякуючы карпусам з’явіліся дакладныя звесткі пра распаўсюджанасць (частотнасць) класаў слоў, лем, асобных слоў, словазлучэнняў. Так, нягледзячы на важнасць для моўнай сістэмы самастойных часцін мовы, у рускіх тэкстах найвышэйшыя паказчыкі частотнасці маюць службовыя часціны мовы. Напрыклад, у прыназоўніка *в* частотнасць будзе ў 3,55 раза вышэйшая, чым у дзеяслова *быть*, і ў

8,1 раза вышэйшая, чым у назоўніка *год*. Менавіта службовыя часціны мовы занялі першыя пяць радкоў рэйтыngu, напрыклад, *Корпуса газетных тэкстаў* (для 11 401 479 словаўжыванняў) [4].

Падобная сітуацыя, па звестках *BNC* (Брытанскага нацыянальнага корпуса), назіраецца і ў англійскай мове [8]. Сярод дваццаці самых частотных слоў англійскай мовы – толькі тры дзеясловы: *be* ‘быць’, *have* ‘мець’ і *do* ‘рабіць’. “Індэкс рэпрэзентатыўнасці”, або  $I_r$  (колькасць словаўжыванняў корпуса, падзеленая на колькасць словаўжыванняў лемы), у першай дваццатцы найбольш частотных слоў *BNC* (для 110 691 482 словаўжыванняў) будзе вагацца ад  $\approx 17,89$  для *the* да  $\approx 214,03$  для *by*.

Пра дакладную корпусную статыстыку для сістэмы беларускай мовы пакуль можна толькі марыць. Але новыя магчымасці, якія апошнім часам дае даследчыкам з’яўленне ў інтэрнэце беларускамоўных корпусных рэсурсаў *Corpus Albaruthenicum* і *Паралельнага беларуска-рускага корпуса ў складзе НКРМ* (*Нацыянальнага корпуса рускай мовы*), – безумоўна, станоўчая адзнака. Калі першы з узгаданых карпусоў мае дачыненне да мовы навуковага стылю (350 027 слоў), то ў складзе *Паралельнага беларуска-рускага корпуса НКРМ* прадстаўлены 2 839 268 словаўжыванняў у выглядзе шэрагу твораў беларускай мастацкай літаратуры і некалькіх перакладаў на беларускую мову. У першай палове 2013 г. у інтэрнэце з’явіўся яшчэ адзін рэсурс: *Беларускі N-корпус*. Інфармацыя на сайце пра яго вельмі сціслая: першы варыянт корпуса меў 1945 тэкстаў (~15 млн словаўжыванняў) мастацкай літаратуры (проза і вершы) з граматычнай разметкай і пашпартызаванай (аўтар, назва твора і г. д.) [3]. Па апошніх звестках, “корпус мае больш за 50 тысяч тэкстаў (~30 млн словаўжыванняў)” [3].

Першы працаздольны рускамоўны корпус быў створаны ў 1990-я гг. па аналагічнай “браўнаўскай” мадэлі ў найстарэйшым, вядомым з 1477 г., скандынаўскім універсітэце – Упсальскім. Прафесар Л. Лёнгрэн быў непасрэдным кіраўніком выкананай у Інстытуце славістыкі Упсальскага ўніверсітэта працы па стварэнні корпуса, адным з вынікаў якой у 1993 г. стаў *Частотны слоўнік сучаснай рускай мовы* [7]. Упсальскі корпус складаецца з прыблізна 600 рускіх тэкстаў, якія змяшчаюць адзін мільён словаўжыванняў (токенаў), у роўнай меры размеркаваных паміж публіцыстыкай і прозай. Публіцыстычныя тэксты адносяцца да перыяду паміж 1985 і 1989 гг., у той час як

проза, мова якой старэе не так хутка, ахоплівае больш працяглы перыяд: 1960 – 1988 г. У корпус не ўключаны паэзія і драматургія [15].

Неабходна адзначыць, што выданне частотных слоўнікаў мовы стала надзвычай актуальным у сярэдзіне ХХ ст. У гэты перыяд з’явілася шмат частотных слоўнікаў, напрыклад рускай мовы: Г. Ёсельсана (1953), Э. А. Штэйфельт (1963), Н. Вакара (1966), І. П. Харакоза (1971), Г. П. Паляковай (1971) і інш. Не ўсе праекты суправаджаліся арганізацыяй класічных корпусных баз, але так ці інакш засноўваліся на тэкставых масівах. У некаторай ступені такую дзейнасць можна разглядаць як корпусную па сутнасці.

Наколькі важны факт прысутнасці ці адсутнасці якой-небудзь адзінкі ў мове – пытанне досыць папулярнае. З’яўленне корпусных тэхналогій не дало адназначнага адказу на яго, але значна спрасціла працэдуру яго пошуку: “...адсутнасць моўнай рысы ў прыватнай сукупнасці моўных фактаў не можа быць даказана яго адсутнасцю ў корпусе. Корпус – толькі ўзор, і дадзеная рыса можа праявіцца ў большай сукупнасці тэкстаў, пакуль не прадстаўленых у корпусе... У той жа час корпусныя лінгвісты ўпэўнены, што частотныя характарыстыкі – цікавыя і важныя...” [17].

Не ўсе корпусныя праекты былі завершаны, напрыклад, толькі частка запланаваных прац была выканана ў рамках стварэння *Машинынага фонду рускай мовы*, пачатага ў 1985 г. [1]. Хоць, безумоўна, назапашаны вопыт паспрыяў паспяховаму развіццю рускай корпуснай лінгвістыкі ў ХХІ ст.

Ідэя *корпуса* не ўзнікла нечакана, тэксты чалавечтва збіралі з даўніх часоў. Аднак менавіта са з’яўлення камп’ютарных тэхналогій карпусы пачалі паспяхова ствараць у выглядзе велізарных тэкставых масіваў. У англійскай мове *корпуснае* значэнне ‘калекцыі фактаў або рэчаў’ вядомае з 1727 г. [13]. Традыцыя раскрываць адценні значэння канкрэтнага слова праз кантэкст пашырылася даволі даўно. Класічны прыклад – “Оксфардскі слоўнік англійскай мовы”, па аналагічнай мадэлі ён даступны і ў рэжыме анлайн [14]. У шматлікіх слоўніках, асабліва тлумачальных, такая практыка атрымала шырокае распаўсюджанне. Пасля з’яўлення камп’ютарных тэхналогій прырода тэкстаў не змянілася, змяніліся магчымасці іх апрацоўкі.

Існаванне ў Сярэднявеччы розных перакладаў *Бібліі*, у тым ліку шматмоўных *Біблій*, абумовіла не толькі сучасны фармат паралельных карпусоў, але і практыку іх “выраўноўвання” па сказах. Больш за тое, менавіта ў розных варыянтах *Бібліі* часта пранумаравана кожнае слова. Выданне шматмоўных тэкстаў друкарскім спосабам таксама можна разглядаць як правобраз сучасных карпусоў.

Між іншым, першым досведам выкарыстання камп’ютара для корпуснай апрацоўкі тэксту лічыцца пачатак у 1946 г. езуітам Р. Буза працы над складаннем канкардансаў твораў Фамы Аквінскага. Сваёй ідэяй Р. Буза зацікавіў кіраўніцтва кампаніі *IBM*, дзякуючы якой праца выконвалася з дапамогай камп’ютараў. Працэс заняў больш за 30 гадоў, а праз без малага 60 гадоў, у 2005 г., функцыя *Index Thomisticus* з’явілася ў інтэрнэце на сайце *Corpus Thomisticum* [11].

Фармат корпуса даўно перарос першапачатковы стандарт не толькі колькасна, але і якасна. З аднаго боку, лацінскае паходжанне тэрміна і класічная семантыка дазваляюць выкарыстоўваць яго для абазначэння практычна любой моўнай сукупнасці – ад “корпуса прыназоўнікаў” да “корпуса твораў”. З другога боку, паспяхова збіраюцца калекцыі ўзораў вуснага маўлення, відэа- і мультымедыяныя карпусы – “корпус жэстаў”, напрыклад.

У шырокім сэнсе словам *корпус* упэўнена можна назваць любую сукупнасць тэкстаў у электронным выглядзе. Электронныя бібліятэкі не валодаюць, вядома, уласна лінгвістычнай разметкай, але, тым не менш, у тым ці інакшым выглядзе структуравання і суправаджаюцца метамоўнай інфармацыяй.

Наколькі вялікім павінен быць корпус – не вельмі істотна. У камп’ютарна-апасродкаваным або віртуальным асяроддзі лічбавае прадстаўленне як старонкі тэксту, так і шматтомнага збору твораў запатрабуюць зусім мала месца. Так, аб’ём стомільённага корпуса *BNC* з яго 5,2 гібайтамі па сілах любому сучаснаму персанальнаму камп’ютару. Зразумела, гіперкорпус *інтэрнэт* – гэта таксама корпус, гэтак жа, як тэксты інтэрнэту – гэта таксама тэксты, а пошукавыя сістэмы інтэрнэту, па сутнасці, мала чым адрозніваюцца ад класічных корпус-менеджараў.

Лінгвістычныя карпусы па класічнай схеме ўключаюць у сябе звесткі ў выглядзе, як правіла, тэкстаў і праграмнага забеспячэння, з дапамогай якога можна даследаваць і выкарыстоўваць гэтыя звесткі. Пытанне “Якая частка корпуса больш важная?” будзе рытарычным. Больш за тое, толькі ў сукупнасці лінгвістычныя звесткі і інфармацыйны інструментарый ствараюць *корпус* [12]. Сёння асобныя корпусныя рэсурсы могуць дасягаць некалькіх мільярдаў адзінак і многія з іх ёсць у інтэрнэце.

Даступнасць корпусных рэсурсаў праз сетку *інтэрнэт* служыць своеасаблівай гарантыяй іх функцыянальнасці, развіцця і запатрабаванасці. Наколькі важны факт прысутнасці сучасных лінгвістычных рэсурсаў у інтэрнэце, можна зразумець па гісторыі першага беларускага корпуса – базы *Машинынага фонду беларускай мовы*. Вышэйзга-

даны рэсурс, створаны ў лабараторыі інтэлектуальных інфармацыйных сістэм на факультэце прыкладной матэматыкі і інфарматыкі Беларускага дзяржаўнага ўніверсітэта па замове Міністэрства інфармацыі Рэспублікі Беларусь у 2006 г., вядомы толькі вузкому колу спецыялістаў.

Вядома, лінгвістычныя параметры корпуса часткова апісаны [2]. Акрамя таго, элементы дадзенага корпуса былі паспяхова інкарпараваны *Яндэксам* і, адпаведна, *Нацыянальным корпусам рускай мовы*. Аднак публічнай падзейі ў беларускай лінгвістыцы ён так і не стаў. Доступ да яго абмежаваны, прычым ужо не толькі яго ведамаснай прыналежнасцю, але і састарэлым праграмным суправаджэннем. Меркаванне, што знаходжанне названага рэсурсу ў сетцы *інтэрнэт* змагло б забяспечыць непасрэдным доступ да яго шырокага кола карыстальнікаў, не патрабуе доказаў. Альтэрнатыва – маральная састарэласць праграмага забеспячэння і сціплае месца на паліцы ў музеі.

Паўвекавы тэрмін корпуснай дзейнасці абумовіў цікавую асаблівасць адпаведнай дыскурсіўнай практыкі: рэарганізацыю ўжо зробленай раней працы. Карпусы, якія працуюць на базе маральна састарэлага праграмага забеспячэння, цалкам рэальна забяспечыць больш дасканалымі сродкамі праграмага суправаджэння. Новыя тэхналагічныя магчымасці дапамагаюць набыць “другое дыханне” і могуць істотна падоўжыць паўнаватасную функцыянальнасць корпуса. Прыклад – *Цюбінгенскія карпусы рускіх тэкстаў*, якія забяспечылі анлайн-доступ да тэкстаў 1980-х гг., сабраных у знакамітым *Упсальскім корпусе рускіх тэкстаў*. Дапоўнены фрагментамі інтэрв’ю, новы сінтэтычны праект сведчыць пра дастаткова высокую філалагічную каштоўнасць некалі сабраных тэкстаў як самадастатковую і ўнікальную сукупнасць [6]. Доступ да *Цюбінгенскіх карпусоў рускіх тэкстаў* арганізаваны Цюбінгенскім універсітэтам. Там жа, акрамя рускамоўных, у спісе карпусоў прысутнічаюць яшчэ 30 рэсурсаў.

Дзейнасць па арганізацыі корпуса тэкстаў можа быць досыць разнастайнай. Крытэрыі адбору тэкстаў для кожнага асобна ўзятага корпуса вызначаюцца суб’ектам у адпаведнасці са спісам навуковых або практычных задач. *Парог рэпрэзентатыўнасці* будзе вызначацца неабходнасцю дастаткова поўна падаць у корпусе запланаваны моўны аб’ект. Здольнасць корпуса тэкстаў адэкватна адлюстроўваць асаблівасці *праблемнай галіны*, рэлевантныя для пэўнага тыпу лінгвістычнага даследавання, вызначаецца яго рэпрэзентатыўнасцю.

Аб’яднанне адабраных па адпаведных крытэрыях тэкстаў адбываецца, як правіла, з улікам папярэдне вызначанай мэты. Умовы стварэння і функцыянавання абранага матэрыялу заўсёды ва-

лодаюць комплексам прыкмет у кантэксце металінгвістыкі. Разам з тым, пачынаючы з Браўнаўскага корпуса (амаль адразу выкарыстанага для сацыялагічных даследаванняў), практычная значнасць прафесійна створанага корпуснага рэсурсу, як правіла, перавышае запланаваныя магчымасці.

Існуе досыць шырокая класіфікацыя карпусоў – па многіх лінгвістычных і інфармацыйных параметрах. Паводле крытэрыяў адбору тэкстаў і рэпрэзентатыўнасці карпусы прынята дзяліць на два класы. У першы клас уваходзяць *універсальныя карпусы* тэкстаў, мэта якіх – аб’ектыўна адлюстроўваць шырокую плынь маўленчай дзейнасці. Толькі суцэльны, максімальна свабодны ад суб’ектыўнага ўплыву корпус, напрыклад *нацыянальны корпус*, дазваляе атрымаць поўную, сістэмную карціну мовы. Традыцыя называць вялікія, максімальна прадстаўнічыя рэсурсы “нацыянальнымі”, магчыма, ідзе ад *British National Corpus*. Ён быў першым у шэрагу карпусоў, якія прэзентуюць мову адной краіны, Вялікабрытаніі, пры шырокай распаўсюджанасці англійскай мовы ў глабальным маштабе.

На практыцы корпусныя даследаванніносяць лакальны па пэўных паказчыках характар. Метадалогія пабудовы карпусоў першага тыпу такці інакш грунтуецца на прынцыпе дэдукцыі: ад агульнага (аб’ектыўнай маўленчай практыкі) да канкрэтных тэкстаў, якія адлюстроўваюць гэтае агульнае. Ва ўніверсальных карпусах высокая дакладнасць вынікаў забяспечваецца высокімі колькаснымі паказчыкамі. Неабходнасць улічваць пэўную прапарцыянальнасць і структураванасць тэкстаў абумоўлена частотнасцю той ці іншай з’явы ў праблемнай галіне. Гэтае патрабаванне абумоўлівае папярэдняю арыентацыю рэсурсу па ўзроўневым прынцыпе: фанетычным, марфалагічным, сінтаксічным, лексічным, семантычным, дыскурсіўным і інш. Што зусім не выключае ўніверсальнай прэзентацыі моўнага матэрыялу.

У другі клас уваходзяць *ілюстрацыйныя карпусы*, пабудаваныя для даследавання пэўнай моўнай заканамернасці. У гэтым выпадку пры стварэнні корпуса даводзіцца абмяжоўвацца тымі ці іншымі выбаркамі з агульнай сукупнасці моўных звестак і ў рэчышчы індукцыі рухацца да абагульненняў. Гэтыя выбаркі павінны адпавядаць крытэрыям якаснай рэпрэзентатыўнасці матэрыялу для лінгвістычнага даследавання. Разгляд праблемнай галіны з улікам дыскурсіўнага аспекту дазваляе паспяхова вырашаць многія лінгвістычныя задачы на аснове кампактнага матэрыялу. Працэдура пабудовы карпусоў другога класа павінна забяспечваць высокую ступень карэктнасці адлюстравання асобнага лінгвістычнага феномена.

Сучасны трэнд лінгвістыкі, эмпірычныя даследаванні абумоўліваюць цікавасць да збору і сістэ-

матызацыі велізарных корпусных масіваў, у якіх парадак лічбаў словаўжыванняў – сотні мільёнаў. Напрыклад, у НКРМ – прыкладна паўмільяра да токенаў [5]. Найлепшым чынам адпавядаюць такой тэндэнцыі карпусы першага тыпу – універсальныя, нацыянальныя. Але і спіс карпусоў, створаных лінгвістамі для пэўных лакальных даследаванняў, проста бясконцы. Пры гэтым пераважная частка з іх, што характэрна, пазней выкарыстоўваецца для выканання больш шырокага кола задач, чым планавалася першапачаткова.

Немалаважны аспект камп’ютарна-апасродкаванай камунікацыі – мінімальнасць дыстанцыі паміж стварэннем тэкстаў ці іх калекцый, напрыклад у выглядзе корпуса, – і доступам да іх адрасатаў ці даследчыкаў. Корпус-менеджары чацвёртага пакалення – у карпусах *BNC*, *PELCRA*, *Hellenic National Corpus*, *НКРМ* і іншых – забяспечваюць доступ да карпусоў у рэжыме рэальнага часу. Такая арганізацыя корпуснай дзейнасці вызначаецца вельмі істотнымі перавагамі: “Гэтыя [інтэрнэт-] інтэрфейсы абумоўлены, вядома ж, не толькі меркаваннямі прававога парадку. Яны таксама дазваляюць стваральнікам корпуса рабіць сваю працу даступнай аўтаматычна, з дапамогай ужо добра знаёмага ўсім карыстальнікам камп’ютараў праграмнага забеспячэння (вэб-браўзер). Дасягальныя ў сусветным павуцінні, яны без абмежаванняў даступныя карыстальнікам любой аперацыйнай сістэмы – у супрацьлегласць такім інструментам трэцяга пакалення як *WordSmith*, даступнаму толькі праз *Microsoft Windows*, або *Xkwc*, даступнаму толькі ў сістэмах тыпу *Unix*...” [12, с. 44]. Даступнасць корпусных дадзеных у інтэрнэце, безумоўна, будзе садзейнічаць іх верыфікацыі.

Метамоўныя і дыскурсіўныя асаблівасці корпуснай дзейнасці абумоўліваюць развіццё і функцыянаванне ключавых асаблівасцей саміх карпусоў. Рэпрэзентатыўнасць любога корпуса, у прыцыпе, не абмяжоўваецца ўласна лінгвістычнымі параметрамі. Так, у кожным канкрэтным выпадку можа ўзнікнуць неабходнасць улічыць стылістычны, часавы, аўтарскі і іншыя аспекты тэкставага масіва праблемнай галіны. Наяўнасць *разметкі* тэкстаў – своеасаблівай абалонкі корпусных дадзеных – дазваляе дапоўніць працу даследчыкаў вопытам і кампетэнцыяй складальнікаў корпуса.

Успрыманне корпуснай сукупнасці дадзеных у кантэксце, з улікам дыскурсіўных якасцей, стварае аснову для сістэмнага даследавання маўленчай дзейнасці. І такая парадыгматычная абумоўленасць корпусных рэсурсаў ператварае сукупнасць тэкстаў у шматузроўневы культурны феномен. Раней ці пазней дыскурсіўны ўзровень корпусных дадзеных павінен быў стаць аб’ектам самай пільнай цікавасці лінгвістаў: “Дыскурсіўныя ўласцівасці тэксту даў-

но разглядаліся як крытычныя для лінгвістычнай тэхналогіі, і больш за сорок гадоў наша разуменне і імкненне выкарыстоўваць дыскурсіўныя асаблівасці тэксту ўзрасталі па многіх прычынах” [16].

Інавацыйны патэнцыял корпусных даследаванняў цесна звязаны з інфармацыйнымі магчымасцямі камп’ютарна-абумоўленай навуковай дзейнасці. Эмпірычныя даследаванні прапануюць тую базу дадзеных, якая абнаўляе лінгвістычныя ўніверсаліі, дазваляе інтэграваць філалагічна-апасродкаваную гуманітарную сферу і сінтэзаваць новую якасць лінгвістычнай тэорыі.

### Спіс літаратуры

1. Андрющенко, В. М. Концепция и архитектура машинного фонда русского языка / В. М. Андрющенко. – М.: Наука, 1989. – 200 с.
2. Барковіч, А. А. Лексічны патэнцыял беларускай мовы ў святле корпуснай лінгвістыкі / А. А. Барковіч. – Мінск: РІВШ, 2012. – 158 с.
3. Беларускі N-корпус [Электронны рэсурс]. – Рэжым доступу: <http://bnkorporus.info>. – Дата доступу: 30.01.2014.
4. МГУ-ЛОКЛІ: Компьютерный корпус текстов русских газет конца XX-ого века [Электронный ресурс]. – Режим доступа: [http://www.philol.msu.ru/~lex/corpus/corp\\_descr.html](http://www.philol.msu.ru/~lex/corpus/corp_descr.html). – Дата доступа: 29.10.2013.
5. Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru>. – Дата доступа: 29.10.2013.
6. Тюбингенские корпуса русских текстов [Электронный ресурс]. – Режим доступа: <http://www.sfb441.uni-tuebingen.de/b1/rus/korpora.html>. – Дата доступа: 29.10.2012.
7. Частотный словарь современного русского языка [Ред.: Л. Лёнгрен] // Acta Universitatis Upsaliensis, Studia Slavica Upsaliensis 32. Uppsala, 1993. – 188 p.
8. British National Corpus [Electronic resource]. – Mode of access: <http://www.natcorp.ox.ac.uk>. – Date of access: 29.10.2013.
9. Brown corpus manual [Electronic resource]. – Mode of access: <http://khnt.aksis.uib.no/icame/manuals/brown>. – Date of access: 29.10.2013.
10. Corpus Albaruthenicum [Electronic resource]. – Mode of access: <http://grid.bntu.by/corpus>. – Date of access: 29.01.2014.
11. Corpus Thomisticum [Electronic resource]. – Mode of access: <http://www.corpusthomisticum.org>. – Date of access: 29.10.2013.
12. McEnery, T. Corpus Linguistics: Method, Theory and Practice / T. McEnery, A. Hardie. – Cambridge: Cambridge University Press, 2012. – 294 p.
13. Online Etymology Dictionary [Electronic resource]. – Mode of access: <http://www.etymonline.com/index.php?search=pager>. – Date of access: 29.10.2013.
14. Oxford English Dictionary [Electronic resource]. – Mode of access: <http://www.oxforddictionaries.com>. – Date of access: 29.10.2013.
15. The Uppsala Russian Corpus – Uppsala universitet [Electronic resource]. – Mode of access: <http://www.moderna.uu.se/slaviska/ryska/corpus/description>. – Date of access: 29.10.2013.
16. Webber, B. Discourse structure and computation: past, present and future / B. Webber, A. Joshi // Proceeding ACL ‘12 Proceedings of the ACL–2012 Special Workshop on Rediscovering 50 Years of Discoveries. – P. 42 – 54.
17. Wynne, M. Searching and concordancing / M. Wynne // Corpus Linguistics: An International Handbook. – Berlin – New York: Walter de Gruyter. – P. 706 – 737.

Аляксандр БАРКОВІЧ,  
кандыдат філалагічных навук.

Аўтар ахвярае ганарар на развіццё часоніка.