

БЕЛАРУСКІ КОРПУС ТЭКСТАЎ: ІНТЭРНЭТ-ДЫСКУРС

Рэзюме. Пасвяцана праблематыке становлення і развіцця беларускаязычных корпусных даследаванняў, іх прадстаўленасці ў Інтэрнэце. Абасноўваецца важнае значэнне ў навуковых даследаваннях як прыкладной, так і корпуснай лінгвістыкі. Характэрна сучаснае становленне беларускаязычнага Інтэрнэту і аналізуюцца магчымасці яго выкарыстання ў якасці асяродка стварэння і развіцця корпуснага дыскурса. Даследаванне функцыянавання мовы ў такім кантэксце звычайна асноваецца на тэкстовай форме камп'ютарна-опосродкаванай камунікацыі, часта асацыіруемай з камп'ютарна-опосродкаваным дыскурсам-аналізам. У сувязі з гэтым удзяляецца ўвага разробцы і ўвядзенню прагматычнай лінгвістыкі Інтэрнэту-камунікацыі.

Ключевые слова: камп'ютарна-опосродкаваная камунікацыя; камп'ютарна-опосродкаваны дыскурс; штучны інтэлект; прыкладная лінгвістыка; корпусная лінгвістыка; корпус тэкстаў; канкорданс; Інтэрнэт.

Abstract. The article is devoted to the evolution (formation and development) of Belarusian-language corpus studies, their representation on the Internet. The importance of applied linguistics in general as well as corpus linguistics in particular in scientific research is justified. The current state of the Belarusian-language Internet is characterized, the possibilities of its application as suitable environment for generation and elaboration of corpus discourse are analyzed. The study of language functioning in these contexts is typically based on text-based forms of computer-mediated communication, often referred to as computer-mediated discourse analysis. Within this, attention in the article is paid to the devising and implementation of pragmatic linguistics of online interactions.

Key words: computer-mediated discourse; computer-mediated communication; artificial intelligence; applied linguistics; corpus linguistics; text corpus; concordance; Internet.

Беларуская прыкладная лінгвістыка развіваецца ў рэчышчы сусветных тэндэнцый, але мае свае істотныя асаблівасці. Фундаментальнасць лінгвістычных ведаў савецкай адукацыйнай школы, якая дае сур'ёзныя перавагі студэнтам і навукоўцам у акадэмічных даследаваннях, вельмі павольна дапаўняецца неабходнымі ў ўмовах Інтэрнэту навукова-тэхнічнага прагрэсу прыкладнымі спецыялізацыямі. Але нават універсітэцкія курсы «Уводзіны ў мовазнаўства» і «Агульнае мовазнаўства» не ўтрымліваюць мінімальна-неабходнай інфармацыі аб гэтай сферы лінгвістычных даследаванняў, якая, тым не менш, паступова становіцца вызначальнай у філалагічнай практыцы. Такія прыкладныя дысцыпліны, як юрыслінгвістыка, лінгвістыка дыскурсу, нейралінгвістыка і інш., практычна нераспрацаваны ў беларускім мовазнаўстве.

Часта да прыкладной лінгвістыкі адносяць выкладанне замежных моў, што, безумоўна, мае рацыю. Хаця па меры з'яўлення новых прыкладных дысцыплін пад прыкладной лінгвістыкай усё часцей разумеецца менавіта камп'ютарна-апасродкаваная лінгвістычная практыка. З развіццём камп'ютарных тэхналогій удзельная вага камп'ютарна-апасродкаваных напрамкаў прыкладной лінгвістыкі пастаянна ўзрастае. Разам з тым кафедра прыкладной лінгвістыкі Беларускага дзяржаўнага ўніверсітэта сёння займаецца галоўным чынам навучаннем замежных студэнтаў рускай мове. Традыцыйнае значэнне *applied linguistics* як практыкі па вывучэнні замежных моў – тыповае і ў замежнай лінгвадыдактыцы, але апошнім часам усё часцей пад такім тэрмінам разумеецца спецыфічная дзейнасць па ўдасканаленні ўзаемапаразумення штучнага Інтэлекту і чалавека, карэляцыі моў чалавека і штучнага Інтэлекту.

Інтэрдyscyплінарнасць і сінкрэтычнасць сучасных моўных даследаванняў – адметныя рысы, якія з'яўляюцца своеасаблівымі маркёрамі іх прыкладнага характару. Але выніковасць намаганняў тут у значнай ступені абмяжоўваецца інерцыяй канонаў апісальнай традыцыі. Тэорыі ў даследаваннях такога кшталту па прычыне недастатковай распрацаванасці – мінімум, таму разважанні, напрыклад, пра беларускую корпусную лінгвістыку звычайна абмяжоўваюцца азначэннямі. Практыка беларускай прыкладной лінгвістыкі – гэта *terra incognita*. Невыпадкова Сусветнае павучэнне, якое дазваляе распаўсюджаць веды мабільна і свечасова, паступова набывае рысы ўніверсальнага і эфектыўнага асяроддзя існавання і функцыянавання самых розных дадзеных. Свядомае выкарыстанне Інтэрнэту ў такой камп'ютарна-апасродкаванай дыскурсіўнай якасці, безумоўна, мае вялікі патэнцыял, асабліва ў дачыненні да прыкладных лінгвістычных даследаванняў.

Выкарыстанне Інтэрнэту ў якасці дыскурсіўнай практыкі прадстаўляецца надзвычай перспектыўным. Існаванне тэкстаў у Інтэрнэце сведчыць аб ужо рэалізаванай магчымасці іх выкарыстання, як мінімум, для прыватных і даследчых мэт.

Прыметы фарміравання новай віртуальнай рэчаіснасці – не толькі экспансія энцыклапедычных рэсурсаў па ўзоры *Вікіпедыі*, але і стварэнне ўніверсальных калекцый тэкстаў – *Tatoeba*, перакладчыкаў і слоўнікаў – *Мультитран* (гл. Вікіпедыя; *Tatoeba project*; Словарь Мультитран).

З іншага боку, пошукавыя менеджары Інтэрнэту – гэта не толькі *Google* і *Yandex*: большасць сучасных Інтэрнэт-парталаў дазваляе ажыццяўляць пошук па зададзенай тэматыцы. І нават такі нескладаны, ён дае магчымасць паспяхова вырашаць цэлы шэраг задач. Большасць беларусаў у Інтэрнэце звяртаюцца не толькі да вузканакіраваных сайтаў, а актыўна выкарыстоўваюць пошукавыя сервісы (*Google*, *Yandex*), разам з тым усё большым попытам карыстаюцца шырокаразгалінаванымі парталамі (*TUT.BY*, *Mail.ru*).

Дасягненні ў галіне штучнага Інтэлекту знаходзяць прымяненне ў сучасным навуковым дыскурсе амаль адначасова з іх з'яўленнем. Інтэрнэт становіцца гіперкрыніцай ведаў, своеасаблівай бэта-версіяй

матрыцы спазнання сусвету. У артыкуле *Private traits and attributes are predictable from digital records of human behavior* (Прыватныя рысы і асаблівасці з'яўляюцца прадказальнымі па лічбавых запісах чалавечых паводзін) часопіса *PNAS (Proceedings of the National Academy of Sciences of the United States of America)* прыводзяцца звесткі аб паспяховай ідэнтыфікацыі ўдзельнікаў сацыяльнай сеткі *Facebook* з дапамогай функцыі *Like* паводле іх этнічнай, палітычнай, рэлігійнай прыналежнасці, характару, інтэлекту, узросту, полу і г. д. (гл. Kosinski, Stillwell, Graepel).

Так ці інакш Інтэрнэт – гэта асноўны на сённяшні дзень спажывец тэхналагічных, у тым ліку камунікацыйных, дасягненняў. Прама ці апасродкавана – іншае пытанне. Але, называючы рэчы сваімі імёнамі, камп'ютарна-апасродкаваная камунікацыя сёння рэвалюцыйнымі тэмпамі стварае новае навуковае і культурнае асяроддзе. Для таго каб зразумець, што 90 % моладзі зарэгістраваны ў сацыяльных сетках, многія – у некалькіх, неабавязкова праводзіць сацыялагічнае даследаванне. Змены ў кансерватыўнай культурнай сферы жорстка дыктуюцца прыярытэтамі моўнага функцыянавання.

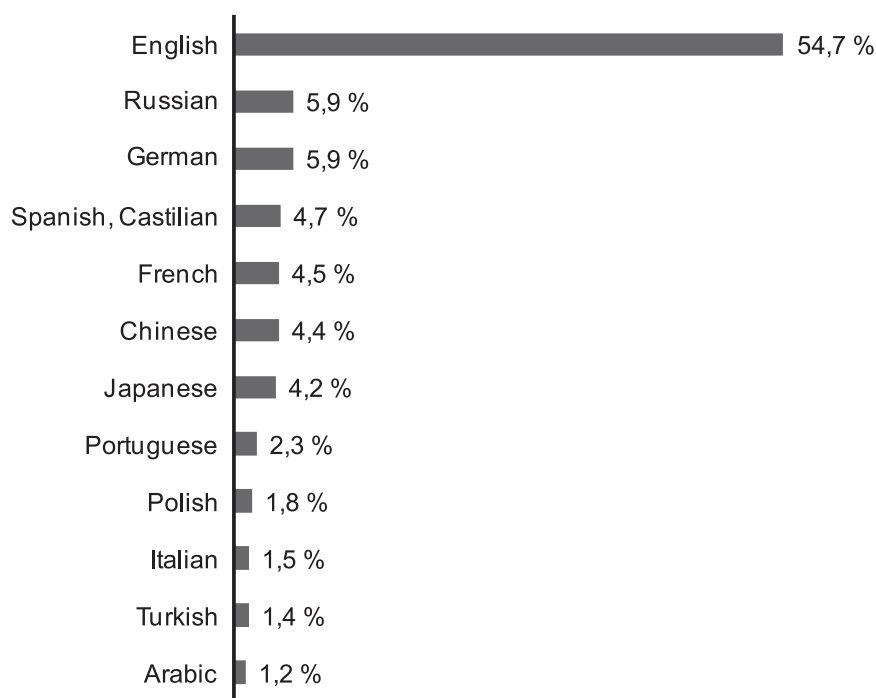
Па дадзеных сацыяльнага апытання, 83 % (!) беларусаў звяртаюцца да Інтэрнэту практычна штодзённа. У гэтым сапраўдная ілюстрацыя сучаснага моўнага функцыянавання. А вось і структура кантэнту (цяжка ўявіць, што Інтэрнэт не будзе рэагаваць на аб'ектыўныя запатрабаванні): беларусаў цікавяць звесткі адносна вучобы і працы (64 %), па даведкі звяртаецца 48 %, бавяць час – 41,5 %, даведваюцца аб навінах – 34 % (гл. На пресс-конференции ИАЦ).

Аб'ектыўнае разуменне сучаснай навуковай парадыгмы працаўнікамі Сусветнага павуціння, як правіла інжынерамі, матэматыкамі, дазваляе ім упэўнена ідэнтыфікаваць сферу сваёй дзейнасці як прыкладную лінгвістыку (гл. Прыкладная лінгвістыка).

У бліжэйшым славянскім асяроддзі, у Расіі, Польшчы і на Украіне, развіццю шырокага кола напрамкаў прыкладной лінгвістыкі ўдзяляецца больш адрасная і належная ўвага – і ў выніку шмат чаго зроблена. Паспяховая распрацоўка і ўдасканаленне нацыянальнага корпуса рускай мовы пазіцыяніравала яго на ўзровень прызнаных узораў (гл. Национальный корпус русского языка, *далей – НКРЯ*).

Можна канстатаваць, што прыкладная лінгвістыка ў Беларусі развіваецца на перыферыі тэарэтычнай і апісальнай лінгвістычных традыцый. Адбываюцца круглыя сталы, семінары, пачалі праводзіць спецыялізаваныя канферэнцыі ў Гомелі (гл. Баркович, Ван Цин 2012). Ёсць цікавыя сведчанні ўдзелу ў міжнародных праектах (гл. Nooj 2013), але прыярытэт у прыкладных даследаваннях на Беларусі пакуль належыць ужо добра распрацаваным замежным спецыялістам і самадастатковым напрамкам, такім, напрыклад, як сінтэз маўлення.

З іншага боку, вельмі патрэбныя беларускай мове айчыныя корпусныя распрацоўкі, як правіла, датычацца рускай, англійскай, радзей французскай, нават кітайскай моў (гл. Баркович, Ван Цин 2012, 8–13). Перавага прыкладных даследаванняў на англійскай ці рускай мове над беларускамоўнымі прама прапарцыянальна распаўсюджанасці моў у свеце. Дадзеныя віртуальнай рэчаіснасці аб'ектыўна сведчаць аб моўных прыярытэтах (малюнак).



Дадзеныя выкарыстання моў у Інтэрнэце

Згодна з прыведзенымі дадзенымі, руская мова мае абсалютную перавагу не толькі ў Інтэрнэце Расіі: 79,0 % на Украіне, 86,9 % у Беларусі, 84,0 % у Казахстане, 79,6 % ва Узбекістане, 75,9 % у Кыргызстане і 81,8 % у Таджыкістане (гл. Russian is now).

Стандартызацыя і кааперацыя на міжнародным узроўні, між іншым, адбываюцца незалежна ад свядомых намаганняў беларускіх філолагаў. Сведчанне таму – уключэнне ў НКРЯ руска-беларускага модуля разам з англійскім, нямецкім, італьянскім, польскім і інш. Прыемна, што беларуская мова ўяўляецца актуальнай для замежных корпусных лінгвістаў. Але, на жаль, сітуацыя з прыцягненнем беларускай мовы без актыўнага ўдзелу беларускіх лінгвістаў нагадвае фарміраванне музейнай калекцыі.

Дарэчы, расійскія складальнікі *Паралельнага беларуска-рускага корпуса* ў складзе НКРЯ сутыкнуліся з пакуль нявырашанымі пытаннямі па карэктнаму распазнаванню граматыкі і лексікі беларускай мовы. Канкардансы (англ. *concordance*), напрыклад, утрымліваюць разам і рускія, і беларускія прыклады, памыляюцца з часцінамі мовы, лексемамі і г. д. Зробленая расійскімі даследчыкамі самастойная праца заслугоўвае вялікай павагі. Пра магчымыя памылкі складальнікі НКРЯ адкрыта заяўляюць у рубрыцы *Ошибки в корпусе*. Трэба адзначыць, што складанасць корпусных праектаў абумоўлівае пэўны час на іх адладку і ўдасканаленне, але нават у фактычна тэставым рэжыме *Паралельны беларуска-рускі корпус*, безумоўна, карысны.

Акрамя жыццёва неабходных праграмных сродкаў развіццё корпуснай лінгвістыкі немагчыма без «крытычнай масы» рэлевантнага матэрыялу: слоўнікаў і тэкстаў у электронным фармаце. Акадэмічныя слоўнікі сучаснай беларускай мовы ёсць, але для корпусных распрацовак патрабуюцца іх прадстаўленасць на машынных носьбітах і максімальна спрошчаны да іх доступ (праз Інтэрнэт, напрыклад) неабмежаванага кола навукоўцаў і проста патэнцыйных карыстальнікаў.

Адсутнасць (у электронным выглядзе і свабодным доступе) шырокага кола нарматыўных слоўнікаў – у першую чаргу тлумачальнага слоўніка, на жаль, перашкаджае стварэнню ў тым ліку інтэрнэт-рэсурсаў, адпаведных крытэрыям нацыянальнага корпуса. Ужо створаным рэсурсам не хапае паслядоўнасці (гл. Slounik.org). Наяўнасць у Інтэрнэце неўнармаванай беларускай мовы і адсутнасць электроннага эталона аказваюць негатыўны ўплыў на стан беларускай гутарковай мовы, размываюць норму і ў далейшым, між іншым, звязяць магчымасці выкарыстання сучаснай беларускай інтэрнэт-мовы як крыніцы любога корпуса ўзуальнай беларускай мовы.

Збіранне тэкстаў на беларускай мове ў прынцыпе не ўяўляе сабой невырашальнай задачы з пункту гледжання аўтарскага права. Па-першае, тэрмінавыя абмежаванні аўтарскага права ўжо амаль не закранаюць класіку беларускай літаратуры. Па-другое, ёсць электронныя архівы перыядычных выданняў, выдавецтваў. Ужо сёння праграма пошуку па беларускіх тэкстах, размешчаных, напрыклад, на сайтах газеты *Звязда* альбо *Беларускай палічкі*, дае тыя ж магчымасці, што і любы неразмечаны (неанатаваны) корпус тэкстаў, а значыць, можа выкарыстоўвацца для разнастайных лінгвістычных мэт (гл. *Звязда*; *Беларуская палічка*). Уключэнне надрукаваных у дзяржаўных сродках масавай інфармацыі і выдавецтвах артыкулаў і мастацкіх твораў, відавочна, пры выкарыстанні такіх тэкстаў у агульнанацыянальным корпусным праекце не будзе выклікаць асаблівых цяжкасцей з пункту гледжання аўтарскага права. Па-трэцяе, надзвычай перспектывным з'яўляецца пашырэнне ў Інтэрнэце так званых «ліцэнзій»: як свабоднага праграмнага забеспячэння (англ. *free software licence*), так і ліцэнзій, якія дазваляюць дастаткова свабоднае выкарыстанне кантэнту (гл. *Creative Commons Attribution-ShareAlike*).

Выкарыстанне для складання корпусаў суцэльных тэкстаў не з'яўляецца невырашальным пытаннем ужо па прычыне корпуснай практыкі выкарыстання тэкставых рэсурсаў па алгарытме пошуку кантэкстаў з дапамогай так званых «канкардансераў». Канкардансы могуць складацца як з фрагментаў на ўзроўні сказаў, так і ўтрымліваць часткі сказаў, фіксіраваную колькасць лексем.

Паказальнай з'яўляецца папулярнасць тэхналогіі «вэб-краўлінга» (англ. *web crawling*), якая выкарыстоўвае для стварэння корпусаў матэрыялы, выкладзеныя ў Інтэрнэце. Прыкладам мільярднага лінгвістычнага корпуса, створанага па інтэрнэт-тэхналогіях, стаў праект *WaCky project (Web as Corpus kool ynitiative)*, у выглядзе *ukWaC*, *deWaC* і *itWaC* – адпаведна англійскай, нямецкай і італьянскай мовам тэкстаў (гл. Baroni, Bernardini, Ferraresi, Zanchetta 2013).

Стварэнне корпусных рэсурсаў на базе юрыдычных ці навукова-тэхнічных тэкстаў стала рэальнасцю дастаткова даўно. Для беларускай мовы праблема тут толькі адна – невялікая колькасць такой дакументацыі, асабліва ў Інтэрнэце. Першым беларускім корпусам тэкстаў, даступным у Інтэрнэце, якраз і стаў навукова-тэхнічны (гл. *Corpus Albaruthenicum*).

Сфарміраваліся з нагоды корпусных інтэрнэт-практык нават напрамкі даследаванняў *Гуглалогія (Googleology)* і *Яндэксалогія*. Называць такую пошукавую ідэалогію навукай, відавочна, яшчэ зарана, але аб'ектам лінгвістычнага аналізу яна ўжо стала (гл. Kilgariff). А калі стыхійны на першы позірк інтэрнэт-пошук з дапамогай *Google* штотыднёва аналізуецца на *Euronews* – гэта ўжо тэндэнцыя. Дарэчы сам тэлевізійны канал *Euronews* выходзіць паралельна на васьмі мовах, што з'яўляецца выдатным увасабленнем мультымедычнага корпуса (гл. *Euronews*).

Беларуская корпусная лінгвістыка не мае пакуль эпахальных дасягненняў. Але ёсць прыклады паспяховай працы беларускіх даследчыкаў над корпусным праектам, які ёсць прыклады недаацэненасці айчынных дасягненняў у галіне прыкладной лінгвістыкі. Гэта – вядомы толькі спецыялістам корпус, створаны ў навукова-даследчай лабараторыі інтэлектуальных інфармацыйных сістэм на факультэце

прыкладной матэматыкі і інфарматыкі Беларускага дзяржаўнага ўніверсітэта па заказе Міністэрства інфармацыі. Хочацца спадзявацца, што названы корпус калі-небудзь з'явіцца ў Інтэрнэце.

Ствараюцца ў асноўным невялікія лінгвістычныя корпусы ў Нацыянальнай акадэміі навук Беларусі, Мінскім дзяржаўным лінгвістычным універсітэце, анансавана праца па стварэнні корпуса ў Гродзенскім дзяржаўным універсітэце імя Янкі Купалы. Корпус тэкстаў пад эгідай НАН Беларусі плануецца выкарыстоўваць як крыніцу для стварэння новага акадэмічнага тлумачальнага слоўніка беларускай мовы.

Новыя магчымасці, якія апошнім часам прадстаўляе для даследчыкаў з'яўленне ў Інтэрнэце беларускамоўных корпусных рэсурсаў *Corpus Albaruthenicum* і *Паралельнага беларуска-рускага корпуса* ў складзе НКРЯ, безумоўна, станоўчая адзнака. Калі першы з узгаданых корпусаў мае дачыненне да мовы навуковага стылю (350 027 слоў), то ў складзе *Паралельнага беларуска-рускага корпуса* 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*. Звесткі аб ім вельмі сціслыя: першы варыянт корпуса мае 1945 тэкстаў (~15 млн словаўжыванняў) мастацкай літаратуры (проза і вершы) з граматычнай разметкай і пашпартызацыяй (аўтар, назва твора і г. д.) (гл. Беларускі N-корпус 2013).

Выкарыстанне большасцю еўрапейскіх моў лацінцы вельмі спрыяе інтэнсіўным даследаванням з прымяненнем універсальных інструментаў, такіх як *MonoConc*, *WordSmyth Tools* і інш. (гл. Моноконс; WordSmyth Tools). Дарэчы, некаторыя з іх працуюць і з кірыліцай (*MonoConc*). Сімптаматычна, што працаздольныя алгарытмы часта выкарыстоўваюцца ў пазнейшых корпусных праектах, у тым ліку іншамовных. Выкарыстанне беларускай мовай кірыліцы накладвае пэўныя абмежаванні на выкарыстанне корпусных дасягненняў рамана-германскай філалогіі, але на прыкладзе НКРЯ можна бачыць, як праблемы праграмага забеспячэння паспяхова вырашаюцца для роднасных рускай і беларускай моў. Такім чынам, ужо створаныя і адпрацаваныя беларускамоўныя рэсурсы, безумоўна, паспрыяюць стварэнню нацыянальнага корпуса беларускай мовы (гл. Совпель 2002).

У доўгім чаканні нацыянальнага корпуса беларускай мовы ёсць выдатныя магчымасці для корпусных даследаванняў інтэрнэт-дыскурсу. Менавіта Інтэрнэт пры слабай насычанасці спецыялізаванымі лінгвістычнымі корпусамі паспяхова выконвае ролю эксперыментальнай пляцоўкі для корпусных даследаванняў. Больш за тое, беларускамоўны Інтэрнэт ужо з'яўляецца крыніцай адлюстравання і фарміравання беларускага маўлення. Выпрацаваная лінгвістычная кампетэнцыя, стабільнае функцыянаванне ў адкрытым доступе нарматыўных слоўнікаў і тэкстаў, анатаваных у сучасным лічбавым фармаце, напрыклад у выглядзе рэпрэзентатыўнага корпуса, ствараць надзейны базіс паспяховага развіцця прыкладной лінгвістыкі ў Беларусі.

БІБЛІАГРАФІЧНЫ СПІС

- Баркович А. А., Ван Цин. Изучение лексического значения многозначных слов на базе лингвистических корпусов // Компьютерная лингвистика: научное направление и учебная дисциплина: сб. науч. ст. Гомель, 2012. Вып. 2. С. 8–13.
- Беларуская палічка [Электронны рэсурс]. URL: <http://knihi.com>. Дата доступу: 29.03.2013.
- Беларускі N-корпус [Электронны рэсурс]. URL: <http://bnkorus.info>. Дата доступу: 30.05.2013.
- Вікіпедыя [Электронны рэсурс]. URL: <http://be.wikipedia.org/wiki>. Дата доступу: 29.03.2013.
- Звязда [Электронны рэсурс]. URL: <http://www.zvyazda.minsk.by/ru/search/ext.php>. Дата доступу: 29.03.2013.
- На пресс-конференции ИАЦ рассказали о белорусках в Интернете [Электронный ресурс]. URL: <http://it.tut.by/341338>. Дата доступа: 29.03.2013.
- Национальный корпус русского языка [Электронный ресурс]. URL: <http://www.ruscorpora.ru>. Дата доступа: 30.05.2013.
- Прикладная лингвистика: на пути к искусственному интеллекту [Электронный ресурс]. URL: <http://events.dev.by/prikladnaya-lingvistika-na-puti-k-iskusstvennomu-intellektu>. Дата доступа: 29.03.2013.
- Словарь Мультитран [Электронный ресурс]. URL: <http://www.multitrans.ru/c/m.exe?a=1&SHL=2>. Дата доступа: 29.03.2013.
- Совпель И. Лингвистические ресурсы белорусского языка и их приложения // Беларуская мова. Інтэрнэт і камп'ютэр: матэрыялы круглага стала Міжнароднага сімпозіума «Разнастайнасць моў і культур у кантэксце глабалізацыі» (Мінск, 20 мая 2002 г.). Мінск, 2002. С. 7–15.
- Baroni M., Bernardini S., Ferraresi A., Zanchett E.: The WaCky wide web: a collection of very large linguistically processed web-crawled corpora [Electronic resource]. URL: http://wacky.sslmit.unibo.it/lib/exe/fetch.php?media=papers:wacky_2008.pdf. Date of access: 29.03.2013.
- Corpus Albaruthenicum [Electronic resource]. URL: <http://grid.bntu.by/corpus>. Date of access: 29.03.2013.
- Creative Commons Attribution-ShareAlike 3.0 Unported License [Electronic resource]. URL: <http://creativecommons.org/licenses/by-sa/3.0>. Date of access: 29.03.2013.
- Euronews [Electronic resource]. URL: <http://ru.euronews.com/news>. Date of access: 29.03.2013.
- Kilgariff A. Googleology is bad science [Electronic resource]. URL: <http://www.kilgariff.co.uk/Publications/2007-K-CL-oogology.pdf>. Date of access: 29.03.2013.
- Kosinski M., Stillwell D., Graepel T. Private traits and attributes are predictable from digital records of human behavior [Electronic resource]. URL: www.pnas.org/content/early/2013/03/06/1218772110.full.pdf. Date of access: 29.03.2013.
- Monoconc: Text Searching Software [Electronic resource]. URL: <http://www.monoconc.com>. Date of access: 29.03.2013.
- Nooj [Electronic resource]. URL: <http://www.nooj4nlp.net/pages/nooj.htm>. Date of access: 29.03.2013.
- Russian is now the second most used language on the web [Electronic resource]. URL: http://w3techs.com/blog/entry/russian_is_now_the_second_most_used_language_on_the_web. Date of access: 29.03.2013.
- Slounik.org: беларускія слоўнікі і энцыклапедыі [Электронны рэсурс]. URL: <http://www.slounik.org>. Дата доступу: 29.03.2013.
- Tatoeba project [Electronic resource]. URL: <http://tatoeba.org/rus>. Date of access: 29.03.2013.
- WordSmith Tools [Electronic resource]. URL: <http://www.lexically.net/wordsmith>. Date of access: 29.03.2013.

Паступіў у рэдакцыю 28.06.13.

Аляксандр Аркадзевіч Барковіч – кандыдат філалагічных навук, дацэнт кафедры прыкладной лінгвістыкі.