

УДК 81'33

А. А. БАРКОВІЧ

МЕТАЛІНГВІСТЫЧНАЯ ІНДЭКСАЦЫЯ Ў КАМП'ЮТАРНА-АПАСРОДКАВАНЫМ ДЫСКУРСЕ

Выяўленне магчымасцей мадэлявання, фіксацыі квалітатыўных і квантытатыўных паказчыкаў маўлення камп'ютарна-апасродкаванай камунікацыі (далей КампАК) – задача важная, не пазбаўленая сваёй навізны. Першыя поспехі ў фармалізацыі мовы з дапамогай камп'ютарных тэхналогій, якія адносяцца яшчэ да другой паловы XX ст., выявілі высокі патэнцыял дадзенага напрамку даследаванняў. Выразнай прыметай сучаснага этапу навукова-тэхнічнага прагрэсу з'яўляецца шырокае выкарыстанне металінгвістычных даных. Высокі попыт на іх стварэнне і развіццё абумоўлены складанасцю камунікацыйных і інфармацыйных сродкаў, неабходнасцю іх суправаджэння верыфікаванымі і структураванымі апісаннямі. Прыкладная лінгвістыка паступова стварае метадалагічную базу для рэалізацыі ўсё больш складаных даследаванняў.

Металінгвістычная індэксацыя. Высокая прэцызійнасць апрацоўкі колькасных паказчыкаў маўлення ўжо сёння ў цэлым паспяхова забяспечваецца сферай КампАК. Моўныя даныя, прадстаўленыя ў электронным выглядзе, валодаюць значным даследчым патэнцыялам у парадыгме прыкладной лінгвістыкі. Прыкладам рэалізацыі такога падыходу можа служыць прымяненне сродкаў прыкладнога характару пры апісанні асаблівасцей апасродкаванага камп'ютарамі маўлення, *металінгвістычнай індэксацыі* – параметрызацыі тэкстаў па паказчыках фармалізацыі, дэрывацыйнасці, семантычнасці, кантэкстуальнасці.

Тэрмін *Meta-Language Indexing* (англ. 'метамоўная індэксацыя') паступова напаяўняецца зместам у англамоўным навуковым дыскурсе. Англійская мова як крыніца мноства лінгвістычных тэрмінаў стварае спрыяльнае метамоўнае асяроддзе для даследавання іншых моў, асабліва ў шматмоўным кантэксце: *We use a meta-language index in our experiments. Instead of distinguishing different languages, all fields are indexed without stemming and stop word removal* (Мы выкарыстоўваем метамоўную індэксацыю ў нашых эксперыментах. Замест дыферэнцыяцыі розных моў, усе палі індэксаваны без «стэміngu» (вылучэнне асновы слова. – А. Б.) і «выдалення стоп-слоў») [1].

Відавочна, што выкарыстанне другасных семіятычных сістэм у практыцы вывучэння маўлення адной мовы таксама можа быць рэалізавана [2]. У такім ракурсе мэтазгодным прадстаўляецца выкарыстанне тэрміна *Metalinguistic Indexing* (англ. ‘металінгвістычная індэксацыя’).

Прэзентацыя тэрмінаў *метамова* і *металінгвістыка* мае добрыя традыцыі ў многіх мовах. З беларускіх крыніц найбольшай аб’ектыўнасцю вызначаюцца пазіцыі А. Я. Міхневіча і В. У. Мартынава. Апошні адзначаў: «Апісваючы якую-небудзь мову, мы заўсёды карыстаемся метамовай лінгвістыкі, г. зн. некаторай спецыялізаванай мовай тэрмінаў, прызначанай спецыяльна для апісання мовы. Без такой метамовы мы не маглі б абысціся, таму што апісаць мову пры дапамозе той жа мовы немагчыма...» [3]. На думку А. Я. Міхневіча, метамова – гэта мова «другога парадку», штучная, навуковая мова, што служыць сродкам апісання натуральнай мовы як прадметнай вобласці [4]. А. Я. Міхневіч вызначае і сутнасць металінгвістыкі: «Паступова ў межах навукі аб натуральных мовах склаўся вельмі важны новы раздзел – металінгвістыка. Металінгвістыка найчасцей арыентавана на адрозненне мовы «першага парадку» (г. зн. натуральнай мовы, што выступае прадметам апісання) і мовы «другога парадку», або метамовы ... Але трэба памятаць, што адрозненне «першая» і «другая» (мовы) – гэта фіксацыя самага простага выпадку суадносін знакавых сістэм, бо калі, напрыклад, з’явіцца апісанне або слоўнік лінгвістычнай тэрміналогіі, то яны ў адносінах да мовы «першага парадку» павінны разглядацца ўжо як створаныя на мове «трэцяга парадку» [5].

Кіруючыся рэкамендацыяй А. Я. Міхневіча, звернемся да слоўнікавай рэпрэзентацыі паняццяў *метамова* і *металінгвістыка*: «...найбольш відавочным і пашыраным прыкладам метамовы натуральнай мовы служыць яе тлумачальны слоўнік з яго навуковым, тэрміналагічным і аперацыйным апаратам» [6].

Класічным прыкладам тлумачэння метамовы і металінгвістыкі, іх цеснай узаемазвяззі, з’яўляюцца азначэнні са «Слоўніка лінгвістычных тэрмінаў» В. С. Ахманавай. Тут азначэнне *метамовы* (‘мова «другога парадку», г. зн. такая мова, на якой гавораць аб мове ж (мове-аб’екце); мова, аб’ектам якой з’яўляецца змест і выражэнне іншай мовы’) – адсылае, у сваю чаргу, да паняцця *металінгвістыкі*: «...пар. металінгвістыка (у 2 знач.)» [7].

Семантыка тэрміна *металінгвістыка* ў В. С. Ахманавай адлюстравана дастаткова шырока: «1. (псіхалінгвістыка, экзалінгвістыка). Галіна мовазнаўства, якая вывучае асаблівасці змествага боку мовы ў сувязі з мысленнем і грамадскім жыццём моўнага калектыву як неабходную ўмову пранікнення ў прыроду лінгвістычных адзінак і заканамернасцей іх функцыянавання. У сферу металінгвістыкі ўваходзіць сукупнасць розных мовазнаўчых пытанняў, такіх як успрымальнасць маўлення, вызначэнне асацыятыўных верагоднасцей у розных моўных (адпав. маўленчых) кантэкстах, агульная логіка семантычных дыферэнцыяцый, заканамернасці авалодання мовай

(развіццё маўлення ў дзяцей, двухмоўе і г. д.), адносіны лінгвістычных працэсаў да перцэпцыі, да пазнання і г. д.; пар. *макралінгвістыка, мікралінгвістыка, перадлінгвістыка*. 2. Навука аб мовах «другога парадку» (*метамовах*), для якіх прадметам (матэрыялам, зместам) з'яўляецца натуральная чалавечая мова; пар. *метамова*) [8].

Этымалагічна англійскія лексемы *metalanguage* (англ. 'метамова') і, адпаведна, *Metalinguistics* (англ. 'металінгвістыка') паходзяць з лацінскай мовы. У сваю чаргу, з англійскай мовы дадзеныя лексічныя адзінкі прыйшлі ў рускую, беларускую і іншыя мовы. Але, калі лацінскае паходжанне асновы *lingua* – 'мова' не выклікае сумненняў, то *meta* – гэта не лацінская *meta* ('калона на іпадроме'), а старажытнагрэчаская аснова *μετά* – 'разам з тым, пры дапамозе, сродкамі чаго' [9].

Тэрміналагічны характар лексемы *метамова* – адпаведна і *металінгвістыка* набылі ў першай палове XX ст. На сённяшні дзень у англійскай мове значэнне лексемы *Metalinguistics* раскрываецца, у тым ліку праз адсылку да паняцця *metalanguage*. А значэнне *metalanguage* па даных з розных крыніц – дастаткова аднароднае і кампактнае:

Any language or symbolic system used to discuss, describe, or analyze another language or symbolic system (Любая мова або знакавая сістэма, якая выкарыстоўваецца для абмеркавання, апісання або аналізу іншай мовы або знакавай сістэмы) [10];

A language or system of symbols used to discuss another language or system (Мова або сістэма знакаў, якая выкарыстоўваецца для абмеркавання іншай мовы або сістэмы) [11] і інш.

У тэхніцы (англ. *Technology*) метамова – 'мова камп'ютарнай логікі' і 'мова верыфікацыі іншамоўных выказванняў' 1. *[theorem proving] A language in which proofs are manipulated and tactics are programmed, as opposed to the logic itself (the «object language»)*. *The first ML was the metalanguage for the Edinburgh LCF [Logic for Computable Functions] proof assistant [computer systems that allow a user to do mathematics on a computer]*. (1. [даказванне тэарэмы] Мова, на якой ствараюцца доказы і праграмуюцца алгарытмы, апазіцыйныя ўласна логіцы («аб'ектнай мове»). Першай метамовай [ML – Meta Language] была метамова для дапаможнага модуля «Эдынбургскай праграмы праверкі лагічнасці камп'ютарных вылічэнняў» – камп'ютарнай сістэмы, якая дазваляла рашаць матэматычныя задачы на камп'ютары). 2. *[logic] A language in which to discuss the truth of statements in another language*. (2. [логіка] Мова, з дапамогай якой можна вызначыць сапраўднасць палажэнняў у іншай мове) [12].

Энцыклапедычны артыкул «Брытанікі» азначае метамоу «семантыкі і філасофіі», як мову, якая выкарыстоўваецца для аналізу «аб'ектнай мовы» (мовы, якая апісвае аб'екты свету): «...метамова можа разглядацца як мова аб іншай мове. Такія філосафы, як нямецкі «лагічны пазітывіст» Рудольф Карнап і Альфрэд Тарскі, польскі матэматык, сцвярджалі, што філасофскія

праблемы і філасофскія выказванні могуць быць вырашаны толькі пры апісанні ў межах сінтаксічнай структуры. Логіка семантыкі – гэта тое, што вызначае ісцінасць выказвання хутчэй, чым яго нефармальнасць ці актуальнасць значэння. Карнап лічыў, што «дзякуючы выкарыстанню сімвалічных пазначэнняў *метамовы* і прытрымліваючыся правілаў логікі, можна пазбегнуць метафізічных меркаванняў, якія ў яго сістэме, па азначэнні, з'яўляліся несапраўдным» [13].

Такім чынам, можна пагадзіцца з А. Я. Міхневічам: «У сферы натуральнай мовы адрозніваюцца: а) сама натуральная мова; б) мова яе апісання, або лінгвістыка; в) мова апісання лінгвістыкі, або *металінгвістыка*» [14].

Індэкс фармалізацыі. Выразным паказчыкам аналізу КампАК безумоўна выступае мера адэкватнасці перадачы тэкстаў, як пісьмовых, так і вусных, пры пераадоленні «машыннага бар'еру». Адлюстраванне эмпірычнага матэрыялу ў новым асяроддзі, камп'ютарна-апасродкаваным, не можа не суправаджацца памылкамі. Пры несумненнай запатрабаванасці эталонных узораў дыскурсіўна бездакорных тэкстаў большасць з іх належыць да перыяду дакамп'ютарнай і дасеткавай «эпохі» культурнага жыцця чалавецтва. І сёння, і яшчэ вельмі доўга прызнанымі ўзорамі тэкстаў, створаных людзьмі, з'яўляюцца і будуць з'яўляцца літаратурныя творы вядомых пісьменнікаў. Дарэчы, многія з гэтых тэкстаў валодаюць і вельмі немалаважнай дадатковай перавагай у выглядзе абмежаванасці тэрмінаў рэгулявання аўтарскага права (50–70 гадоў).

Такім чынам, ёсць падставы для суправаджэння прызначанага для навуковага выкарыстання эмпірычнага матэрыялу лічбавымі паказчыкамі фармальнай тоеснасці электроннай копіі арыгінальнай першаснай версіі. Такія паказчыкі запатрабаваныя і рэlevantныя ў прыкладной лінгвістыцы практычна на кожным кроку.

Пры аўтаматычнай апрацоўцы моўнага матэрыялу робіцца нямала памылак. Суадносная колькасць такіх памылак, каэфіцыент *шуму*, залежыць ад многіх фактараў. У дачыненні да пэўнай тэкставай сукупнасці, прадстаўленай сродкамі праграмнай апрацоўкі даных, маюцца важкія аргументы для ўключэння ў сукупнасць экстралінгвістычных даных і адпаведнай разметкі: **індэкса фармалізацыі** (англ. *Formalisation Index*), ці I_F , які адлюстроўвае суадносіны (вынік дзялення) прадстаўленых у тэксце і распазнаных пэўнымі праграмнымі сродкамі моўных адзінак таго ці іншага ўзроўню. Калі тэкст складаецца са ста слоў, напрыклад, – а праграма па распазнанні тэксту бездакорна выявіла 90 – індэкс фармалізацыі будзе роўны $\approx 1,11$, і чым больш лічба будзе набліжацца да значэння $1,0$, тым лепш. Важнасць вызначэння такіх паказчыкаў абумоўлена неабходнасцю ўстаўлення *парога кампетэнтнасці* праграмных сродкаў. Для рацыянальнай арганізацыі і параметрызацыі тэкстаў даследчыку патрэбна ведаць: дзе павінны адбыцца пераход ад аўтаматычнай да «аўтаматызаванай» апрацоўкі (з удзелам чалавека-рэдактара) даных. Такі шлях дазваляе аптымізаваць ўдасканаленне саміх праграмных сродкаў як працэс.

Дакладныя лічбы адпаведнага індэкса будуць дасягальныя пры параўнанні аўтэнтычнага варыянту тэксту з яго электроннай копіяй. Статыстычныя звесткі з розных крыніц сведчаць аб сярэдняй прыбліжанасці паказчыка памылак праграм да 5%, што будзе адпавядаць значэнню індэкса фармалізацыі $I_F \approx 1,05$ для англійскай мовы. Намнога больш сінтэтычныя беларуская ці руская мова, пры ўсіх роўных іншых умовах, зразумела, будуць мець значна горшы паказчык *шуму*.

У прынцыпе памылкі корпусных менеджараў, машынных перакладчыкаў, праграм графічнага распазнання тэксту добра бачныя і могуць быць растлумачаны. Наколькі глыбокія звесткі аб структуры мовы даступны штучнаму інтэлекту – добра бачна па змесце корпусных канкардансаў, якія насычаны памылкамі, што пацвярджае неглыбокае «веданне» праграмамі граматыкі і семантыкі. Магчыма, *сістэмныя памылкі* праграм – прыцыпова невырашальныя да гіпатэтычнага моманту камп'ютарнай апрацоўкі ўсёй моўнай сістэмы. На практыцы нават карэктная экстрапаляцыя графічнага боку тэкстаў у электронным выглядзе пакуль сведчыць не на карысць дасканаласці праграмнай стратэгіі, а, хутчэй, аб удзеле чалавека ў папярэдняй падрыхтоўцы і постапрацоўцы тэксту.

Метадалагічны шлях да разумення і вырашэння задач на стыку лінгвістыкі і матэматыкі знаходзіцца ў межах своеасаблівага кампрамісу абстрактнай інварыянтнай логікі і яе сістэмнай рэалізацыі ў гуманітарнай сферы. Такія магчымасці прадстаўляе, напрыклад, матэматычная «Тэорыя недакладных мностваў (падмностваў)», прадстаўленая прафесарам Каліфарнійскага ўніверсітэта (*University of California, Berkeley*) Лотфі Задэ (*Lotfi A. Zadeh*, ці *Lotfi Aliaskerzadeh*). Незвычайнай папулярнасцю гэта тэорыя абавязана свайму зместу, які не супярэчыць «гуманітарнаму» (лінгвістычнаму) профілю ў сістэме пошуку шляхоў да ўдасканалення ўзаемадзеяння штучнага інтэлекту і жывой мовы [15].

Індэкс дэрывацыйнасці. Пры спробах стварыць праграмныя сродкі апрацоўкі тэкстаў, між іншым, становіцца асабліва відавочнай звыклая неадпаведнасць класічных граматычных канонаў практыцы словаўжывання ў працэсе камунікацыі. Так адбываецца не толькі ў КампАК, але і пры камунікацыі, зарыентаванай у спрыяльных сацыяльных ці прагматычных абставінах на высокі ўзровень адпаведнасці норме. Характэрнымі сведчаннямі недастатковай аб'ектыўнасці і, калі называць рэчы сваімі імёнамі, штучнасці граматык з'яўляюцца пошукі актуальнага фармату *машыннага перакладу, заснаванага на правілах*, нестандартная для тэорыі мовы прэзентацыя граматычных мадэляў у канкардансерах і г. д.

Для мовы пераважна сінтэтычнага складу, такой як беларуская, намінацыйныя характарыстыкі вельмі цесна звязаны з дэрывацыйнымі. Менавіта тандэм намінацыйна-дэрывацыйных тэндэнцый вызначае напрамак развіцця мовы. Адметныя рысы стану моўнай сістэмы і фактары, якія фарміруюць паказчыкі яе напаўнення, шмат у чым вызначаюць логіку распрацоўкі

катэгорый. Камп'ютарна-апасродкаваны сегмент агульнамоўнай прасторы характарызуецца ярка выражаным фактарам надзвычайных статыстычна-і інфармацыйна-сумяшчальных магчымасцей сферы КампАК.

Лічбавы характар даных апрацоўкі тэкстаў, складаючых аб'ектыўны бок сферы, дазваляе рабіць цэлы шэраг высноў высокай дакладнасці адносна ў першую чаргу фармальных рыс тэкстаў. Прыкладам вырашэння ў такім ключы прыватнай лінгвістычнай задачы можа служыць кваліфікацыя зместу тэкстаў у электронным выглядзе ў частцы ўстанаўлення ступені лінгвістычнай, у тым ліку граматычнай, варыятыўнасці фрагмента маўлення. Алгарытм такога даследавання патрабуе распрацаванасці граматычнага кодэксу мовы ў інтэрпрэтацы адпаведнага праграмнага забеспячэння. У дадзеным кантэксце тыповая навуковая задача – напрыклад, даследаванне такога спецыфічнага аб'екта моўнай рэальнасці, як ідыялект пісьменніка.

Даследаванні прыкладнога характару ёсць сэнс праводзіць на базе дастаткова прадстаўнічых масіваў тэкстаў. Сведчанні запатрабаванасці і выканальнасці такіх даследаванняў прысутнічаюць у працах аўтарытэтных замежных спецыялістаў. Разнастайнасць задзейнічанага ў корпусных рэсурсах слоўнікавага складу ўжо паспяхова вызначалася з дапамогай «суадносін словаформа-токен» (англ. *Type-Token Ratio*, або *TTR*): «Мера слоўнікавай разнастайнасці ў корпусе ў лічбавым выражэнні роўная суме словаформаў, падзеленай на суму токенаў. Чым бліжэй суадносіны да 1 (або 100 %), тым шырэй варыятыўнасць слоўніка. Гэтая статыстыка не дазваляе непасрэдна параўноўваць паміж сабой корпусы розных памераў» [16].

Выніковыя эксперыменты праводзіліся на матэрыяле англійскай мовы, матэрыял іншамоўных тэкставых масіваў таксама дасягальны для падобнай апрацоўкі. Больш за тое, для значна больш сінтэтычнай беларускай мовы разгорнутыя па методыцы вылічэння *індэкса*, а не каэфіцыента даследаванні несумненна якраз больш рэлевантныя: аб'ектыўным паказчыкам для разгалінаванай сістэмы словазмянення славянскай мовы з'яўляецца арыентацыя менавіта на ўзровень *лемы*. Вывучэнне адноснай – без жорсткай абумоўленасці аб'ёмам тэксту ці гіпертэксту – адпаведнай характарыстыкі актуальна для даследаванняў мовы ў прыкладным рэчышчы.

Тэрмін *лема* часта выкарыстоўваецца паралельна з іншым тэрмінам – *лексемай*. Іх карэктнае ўжываннем лёгка дыферэнцыраваць, калі разглядаць паняцці ў аспекце дыхатаміі «мова / маўленне». Такім чынам *лема*, як атрыбут маўленчай рэалізацыі і 'сукупнасць формаў слова' відавочна адзначаецца ад *лексемы*, як абстрактнай характарыстыкі моўнай сістэмы і 'паняцця'. Асобна трэба сказаць аб карэктным ужыванні тэрміна *словаўжыванне*, які набыў шматзначнасць і разнапланавасць. Традыцыйнае разуменне пад *словаўжываннем* 'ужывання лексемы ў розных значэннях' у прыкладной лінгвістыцы валодае больш шырокім аб'ёмам семантыкі – 'ужывання слова як у розных значэннях, так і ў розных формах і кантэкстах'.

Словаўжыванні ў прыкладным разуменні маюцца на ўвазе, напрыклад, у корпусных даследаваннях. Так, у праекце «*InterCorp: projekt paralelnich korpusů Filozofické fakulty Univerzity Karlovy v Praze*» заяўлена наяўнасць каля 7 588 000 рускіх, 1 493 000 – украінскіх і 1 308 000 – беларускіх словаўжыванняў, або токенаў (англ. *token* ‘словаўжыванне’). Зрэшты, у дадзеным праекце і іншыя славянскія мовы прадстаўлены дастаткова шырока: 26 879 000 – балгарскіх, 12 625 000 – харвацкіх, 2 664 000 – македонскіх, 47 640 000 – польскіх, 40 108 000 – славацкіх, 33 741 000 – славенскіх, 6972 – сербскіх і, натуральна, 99 547 000 – чэшскіх адзінак з агульнай колькасці ў 867 287 000 словаўжыванняў [17]. Інфармацыя вельмі красамоўная: свайго роду рэйтынг, які шмат у чым адлюстроўвае ўзровень развіцця прыкладной лінгвістыкі ў кожнай асобна ўзятай славянскай краіне.

Шматмоўны корпусны праект *ParaSol: A Parallel Corpus of Slavic and other languages* быў створаны Рупрэхтам фон Вальдэнфельсам (*Ruprecht von Waldenfels*) і Роландам Маерам (*Roland Meyer*) у Інстытуце славянскіх моў і літаратур Бернскага ўніверсітэта (*Institute of Slavic Languages and Literatures, University of Bern*) і Інстытуце славістыкі Регенсбургскага ўніверсітэта (*Institute of Slavistics, Regensburg University*) у Швейцарыі [18]. На фоне 18 085 532 словаўжыванняў у рубрыцы «Іншыя мовы» (англ. *other languages*) паказчык у 7 646 832 токенаў прадстаўленасці славянскіх моў вельмі высокі. У гэтым рэсурсе нараўне з іншымі славянскімі мовамі прадстаўлены рускія тэксты ў выглядзе 3 637 357 токенаў і 78 997 лем; украінскія тэксты – 1 017 057 токенаў і 33 563 лемы; беларускія тэксты – 482 467 токенаў і 24 131 лемы [19].

Для ацэнкі *парадыгматычнай* разнастайнасці лексічнага складу корпуса мэтазгодна параўнанне колькасці *словаўжыванняў* корпуса і *лем*, да якіх яны належаць. Такія суадносіны ў лічбавым выражэнні, вызначаюцца як *індэкс дэрывацыйнасці* (англ. *Derivability Index*), ці I_D – паказчык, які фарміруецца ў выніку дзялення колькасці словаўжыванняў тэксту ці сукупнасці тэкстаў на колькасць словаўжыванняў лемы. У корпусе *ParaSol: A Parallel Corpus of Slavic and other languages* для рускай мовы індэкс I_D будзе роўны $\approx 46,04$ (3 637 357 токенаў / 78 997 лем), украінскай – $\approx 30,3$ (1 017 057 токенаў / 33 563 лемы), беларускай – $\approx 19,99$ (482 467 токенаў / 24 131 лема) адпаведна.

Несумненна, што дэрывацыйная дынаміка дыскурсіўнай практыкі будзе залежыць ад тэкставых асаблівасцей, але шэраг даследаванняў дадзенай накіраванасці дазваляе вызначаць некаторыя рэлевантныя ў агульнамоўным кантэксце характарыстыкі. У нашым прыкладзе тэксты корпуса *ParaSol* у адпаведнасці з патрабаваннем рэпрэзентатыўнасці корпуснага дыскурсу – падобнай і параўнальнай метатэкставай якасці. Значыць дэрывацыйная актыўнасць рускага маўлення – вышэйшая за аналагічныя паказчыкі ўкраінскага і беларускага маўлення, пры ўсім падабенстве ўсходнеславянскіх моў і граматык.

Несумненна, вывучэнне тэкстаў КампАК вызначаецца шматбаковым і шматаспектным характарам. Лінгвістычна-рэlevantныя характарыстыкі тэкстаў сферы КампАК – комплексныя і складаныя, што абумоўлена лінгваінфармацыйным фармаатам, напрыклад інтэрнэту. Істотную ролю адыгрываюць інтэр- і мультыдысцыплінарныя ўласцівасці функцыянавання камп’ютарна-апасродкаванага дыскурсу.

(працяг будзе)

Літаратура

1. Jadidinejad, A. H. Query Wikification: Mining Structured Queries From Unstructured Information Needs using Wikipedia-based Semantic Analysis / A. H. Jadidinejad, F. Mahmoudi // In Proceeding of the Cross Language Evaluation Forum (CLEF 2009 Working Notes). – [Technical Report]. – Corfu, Greece, 2009 [Electronic resource]. – Mode of access: <http://ceur-ws.org/Vol-1175/CLEF2009wn-adhoc-HosseinEt2009.pdf>. – Date of access: 29.09.2014.
2. Барковіч, А. А. Корпус текстов как вторичная семиотическая система / А. А. Барковіч // Актуальныя праблемы паланістыкі. – 2006. – С. 63–69.
3. Мартынов, В. В. Кибернетика. Семиотика. Лингвистика / В. В. Мартынов. – Минск: Наука и техника, 1966. – С. 57.
4. Міхневіч, А. Я. Метонімы ў беларускай мове / А. Я. Міхневіч // Працы кафедры сучаснай беларускай мовы / пад рэд. А. Я. Міхневіча. – Мінск: Выд-ва БДУ, 2007. – Вып. 6. – С. 13.
5. Міхневіч, А. Я. Метонімы ў беларускай мове. – С. 13.
6. Міхневіч, А. Я. Метонімы ў беларускай мове. – С. 13.
7. Ахманова, О. С. Словарь лингвистических терминов / О. С. Ахманова. – М.: Сов. энцикл., 1966. – 607 с.
8. Ахманова, О. С. Словарь лингвистических терминов.
9. Online Etymology Dictionary [Electronic resource]. – Mode of access: <http://www.etymonline.com>. – Date of access: 29.09.2014.
10. Dictionary. com [Electronic resource]. – Mode of access: <http://dictionary.reference.com>. – Date of access: 29.09.2014.
11. Collins Dictionaries [Electronic resource]. – Mode of access: <http://www.collinsdictionary.com>. – Date of access: 29.09.2014.
12. FOLDOC – Computing Dictionary [Electronic resource]. – Mode of access: <http://foldoc.Org>. – Date of access: 29.09.2014.
13. Encyclopedia Britannica [Electronic resource]. – Mode of access: <http://www.britannica.Com>. – Date of access: 29.09.2014.
14. Міхневіч, А. Я. Метонімы ў беларускай мове. – С. 14.
15. Кофман, А. Введение в теорию нечетких множеств / А. Кофман; пер. с фр. – М.: Радио и связь, 1982. – 432 с.
16. McEnery, A. Corpus Linguistics: Method, Theory and Practice / A. McEnery, A. Hardie. – Cambridge: Cambridge University Press, 2012. – P. 253.
17. InterCorp: projekt paralelnich korpusů Filozofické fakulty Univerzity Karlovy v Praze [Electronic resource]. – Mode of access: <http://www.korpus.cz/intercorp>. – Date of access: 29.09.2014.
18. ParaSol: A Parallel Corpus of Slavic and other languages [Electronic resource]. – Mode of access: <http://parasol.unibe.ch>. – Date of access: 29.10.2014.
19. ParaSol: A Parallel Corpus of Slavic and other languages.

Summary

Language data represented in electronic form have significant research potential in the paradigm of Applied Linguistics. Saturation of Computer-Mediated Communication (CMC) and artificial formal languages creates a kind of semiotic background of human communication. Metalinguistic Indexing is the parameterization of texts in terms of formalization, derivation, semantics, contextuality. One of the strategic issues in this context is the method of forming the material base of research and the adequacy of existing traditional linguistic requirements and standards. Linguistically-relevant characteristics of the text sphere of CMC are complex and have lingvo-inforvational format, for example, in the Internet. Inter- and multidisciplinary features of functioning of Computer-Mediated Discourse play significant role.

Беларускі дзяржаўны ўніверсітэт

03.02.2015 г.