УДК 81'33

А. А. БАРКОВІЧ

МЕТАЛІНГВІСТЫЧНАЯ ІНДЭКСАЦЫЯ Ў КАМП'ЮТАРНА-АПАСРОДКАВАНЫМ ДЫСКУРСЕ

(працяг)

Насычанасць камп'ютарна-апасродкаванай камунікацыі (далей КампАК) штучнымі і фармальнымі мовамі стварае своеасаблівы семіятычны фон чалавечай камунікацыі. Гэта выражаецца не толькі ў інтэрнацыяналізацыі і нонканфармізме маўлення, але і ў яго лінгваінфармацыйнасці і статыстычнай «фарматнасці».

Пры невычарпальнасці багацця мовы – у сферы КампАК дэманструецца асаблівая выразнасць і бескампраміснасць маўлення. Дыскурсіўная практыка камунікацыйных асоб ў камп'ютарна-апасродкаваным камунікацыйным асяроддзі дэманструе павышаную варыятыўнасць і разнастайнасць маўлення, з аднаго боку. І з іншага – колькасць інфармацыі ў віртуальнай рэальнасці павялічылася да такога колькаснага аб'ёму, якім нават гіпатэтычна не змагла б здзівіць сфера друкаванага кнігавыдання ці гуказапісу, і працэс працягваецца. КампАК часта арыентуецца практычна на неабмежаванае кола суразмоўцаў, што дазваляе разважаць аб *гіперідэнтычнасці, гіперперсанальнасці, гіперадрасацыі* і г. д. адпаведнай практыкі. Камп'ютарныя сродкі ўвесь час удасканальваюцца і дэманструюць універсальнасць, з якой ніколі ўжо не змогуць супернічаць выдавецтвы папяровых кніг і кампакт-дыскаў. Зразумела, значную частку інтэрнэт-кантэнту прадстаўляюць копіі ўжо выдадзеных традыцыйнымі спосабамі твораў. У КампАК шырока распаўсюджаны і копіі копій, варыянты, кавэр-версіі і г. д. Такі змест інтэрнэту сведчыць, у тым ліку, аб шырокім попыце на электронныя версіі тэкстаў.

Статыстычныя магчымасці камп'ютарна-інфармацыйнай апрацоўкі тэкстаў – пры ўсёй яе павярхоўнасці і недастатковасці – значна пераўзыходзяць магчымасці колькаснай апрацоўкі «матэрыяльнага» кантэнту. Праграмы штучнага інтэлекту па вельмі немудрагелістай статыстычнай мадэлі, але тым не менш паспяхова ў рамках сваёй глыбіні, могуць суправаджацца нескладанай разметкай па фармальных і кадыфікаваных паказчыках вялізныя тэкставыя масівы. Апрабіраваныя методыкі ёсць у прыкладной парадыгме, корпуснай лінгвістыцы.

Адным са стратэгічных пытанняў у дадзеным кантэксце з'яўляецца методыка фарміравання матэрыяльнай базы даследавання і яе адэкватнасць існуючым традыцыйным лінгвістычным патрабаванням і стандартам.

Індэкс семантычнасці. Паказчык *семантычнасці* ўзораў маўлення не вызначае самадастатковыя ўласцівасці адпаведнага дыскурсу. Тым не менш менавіта семантыка, пранізваючы ўсю моўную сістэму, дазваляе атрымаць аб'ектыўнае ўяўленне аб прыярытэтах ужывання асобных лінгвістычных адзінак. У сваю чаргу, адзінкі, якія валодаюць самастойным значэннем, фарміруюць агульную семантычную насычанасць тэксту. Ступень празрыстасці тэкстаў камп'ютарна-апасродкаванага дыскурсу залежыць ад дыферэнцыраваных даных, характэрных маўленню і прама прадстаўленых рэалізацыямі асобнай лемы і сукупнасці лем.

Дзякуючы лінгвістычным корпусам з'явіліся дакладныя даныя аб распаўсюджанасці (частотнасці) класаў слоў, лем, асобных слоў, словазлучэнняў. Напрыклад, нягледзячы на важнасць для моўнай сістэмы самастойных часцін мовы, у рускіх тэкстах найвышэйшымі паказчыкамі частотнасці характарызуюцца службовыя часціны мовы. Як відаць з табл. 1, у прыназоўніка *в* частотнасць будзе ў ≈ 3,55 раза вышэй, чым у дзеяслова *быть*, і ў ≈ 8,1 раза вышэй, чым у назоўніка *год*. Менавіта службовыя часціны мовы занялі першыя пяць радкоў рэйтынга, напрыклад Корпуса газетных текстов (для 11 401 479 словаўжыванняў).

Табліца 1. **20 самых частотных лем рускай мовы [1]**

№	Слова	Інфарм.	Асабова-публіц.	Інф.-публіц.	Мастацкі	Мастацка-публіц.	Рэклама	Афіц.-дзелавы	Разм.-пісьм.	Астатнія жанры	Усе жанры
1.	В	40104	72705	223130	7760	24912	4852	4475	229	28511	406678
2.	И	26368	69056	185954	10287	24957	3207	4105	527	25185	349646
3.	На	17612	30230	94984	4255	11497	2008	2190	124	12445	175345
4.	Не	9305	34316	95905	5449	13457	818	1856	345	13254	174705
5.	С	10917	22892	69794	3140	8945	1618	1672	359	9549	128886
6.	Этот	7872	23143	68714	2695	7602	722	1275	88	9166	121277
7.	Быть	8273	21125	63479	3065	8365	684	1068	130	8651	114840
8.	Что	6829	20870	66187	2969	7509	460	1035	92	8373	114324
9.	Тот	5092	17878	51373	68	5618	431	823	60	6714	88057
10.	А	4426	14587	40683	2840	6044	767	772	200	6516	76835
11.	По	8691	12943	41833	1275	4131	874	966	33	5586	76332
12.	Весь	4540	14711	41630	2338	5805	557	761	145	5762	76249
13.	Как	3908	12176	34148	2179	4831	305	543	57	4357	62504
14.	К	4288	10962	30331	1486	3737	439	620	168	3937	55968
15.	О	4329	10653	30735	1012	3033	386	617	77	3888	54730
16.	Из	5172	9692	28336	1116	3594	495	471	398	3937	53211
17.	Но	2398	10201	29354	1581	4073	249	528	60	4034	52478
18.	Год	5226	9877	27120	583	2909	389	509	23	3624	50260
19.	Свой	3299	9189	25888	1084	3481	309	332	57	3163	46802
20.	За	4132	8558	25000	1162	3158	362	491	31	3481	46375

Падобная, па даных *BNC* (Брытанскага нацыянальнага корпуса), сітуацыя назіраецца і ў англійскай мове. Прыназоўнікі *in, to, for, on, with, at*, і *by* (англ. – *preposition*), саюзы *and* і *that* (англ. – *conjunction*) і часціца *to* (англ. – *particle, infinitive-marker*), нягледзячы на невялікую колькасць у сістэме мовы службовых часцін мовы, прадстаўлены на 11 з 20 пазіцый рэйтынга ў табл. 2.

Табліца 2. **20 самых частотных лем *BNC* [2]**

Рэйтынг	Частотнасць	Лема	Часціна мовы
1	6187267	The	Determiner
2	4239632	Be	Verb
3	3093444	Of	Preposition
4	2687863	And	Conjunction
5	2186369	A	Determiner
6	1924315	In	Preposition
7	1620850	To	Infinitive-marker
8	1375636	Have	Verb
9	1090186	It	Pronoun
10	1039323	To	Preposition
11	887877	For	Preposition
12	884599	I	Pronoun
13	760399	That	Conjunction
14	695498	You	Pronoun
15	681255	He	Pronoun
16	680739	On	Preposition
17	675027	With	Preposition
18	559596	Do	Verb
19	534162	At	Preposition
20	517171	By	Preposition

Адным з паказчыкаў, якія могуць павысіць лінгвістычную пераканаўчасць вынікаў, атрыманых у асяроддзі і сродкамі прыкладной кампетэнцыі, з'яўляецца лічбавае прадстаўленне насычанасці тэкстаў рознаўзроўневай семантыкай. Такую інфармацыю адносна лексічнага ўзроўню семантыкі камп'ютарызаванага дыскурсу ў корпусным кантэксце можна атрымліваць, напрыклад, з дапамогай нескладаных статыстычных мадэляў.

Індэкс семантычнасці (англ. – *Semanticity Index*), або $\boldsymbol{I_S}$ (колькасць словаўжыванняў тэксту / корпуса (тэкстаў), падзеленая на колькасць словаўжыванняў асобнай лемы) дазваляе ўстанаўліваць ступень насычанасці тэкстаў асобнымі лемамі і іх групамі, семантычнымі катэгорыямі. Напрыклад, сярод 20 самых частотных слоў англійскай мовы – толькі 3 дзеясловы: *be* – 'быць', *have* – 'мець' і *do* – 'рабіць' [3]. Адпаведна, індэкс семантычнасці ў першай «дваццатцы» найбольш частотных слоў BNC (для 110 691 482 словаўжыванняў) будзе вагацца ад $\boldsymbol{I_S} \approx 17{,}89$ для *the* (частотнасць 6 187 267) – да $\boldsymbol{I_S} \approx 214{,}03$ для *by* (517 171).

Выкананне падобных падлікаў цалкам рэлевантна і для беларускай мовы – пасля стварэння рэпрэзентатыўных тэкставых масіваў у электронным фармаце і доступу да зыходных кодаў разметкі, калі адпаведная спецыфікацыя будзе адсутнічаць па змоўчванні.

Адносна ўжывальнасці беларускіх адзінак статыстычныя звесткі пацвярджаюць агульную тэндэнцыю па міжчасцінамоўных адносінах у тэксце. Згодна з данымі У. А. Кошчанкі, у складзе *Беларускага N-корпуса* з пераканаўчай стабільнасцю пераважаюць службовыя часткі мовы (табл. 3). Так, чатыры першыя радкі, прадстаўленыя *у, і, на, з* – належаць прыназоўнікам і злучніку, а тры першыя, дарэчы, адпавядаюць і рускамоўным варыянтам (гл. табл. 1).

Табліца 3. **20 самых частотных лем *Беларускага N-корпуса* [4]**

Рэйтынг	Частотнасць	Лема	Часціна мовы
1	2994603	У	Прыназоўнік
2	1989721	І	Злучнік
3	1499682	На	Прыназоўнік
4	1161539	З	Прыназоўнік
5	983568	Што	Злучнік*
6	909338	Гэты	Займеннік
7	894314	Не	Часціца
8	589441	Быць	Дзеяслоў
9	559775	А	Злучнік
10	528638	Да	Прыназоўнік
11	491321	Той	Займеннік
12	481396	За	Прыназоўнік
13	464567	Які	Займеннік*
14	438824	Як	Прыслоўе*
15	435681	Па	Прыназоўнік
16	410385	Я	Займеннік
17	328822	Мы	Займеннік
18	294385	Год	Назоўнік
19	293174	Але	Злучнік
20	292791	Ён	Займеннік

* Дакладная часцінамоўная прыналежнасць не вызначана.

Прыведзеныя даныя адносна часцінамоўнага складу *Беларускага N-корпуса* патрабуюць далейшай апрацоўкі па прычыне «нязнятай аманіміі» ў корпусе. Напрыклад, вызначыць прапарцыянальныя суадносіны злучніка *што* і займенніка *што* на сённяшні дзень можна толькі прыблізна. Кантэксты корпуснага канкардансу сведчаць аб адносна невялікай колькаснай перавазе злучніка *што* над займеннікам *што*:

*Ён пабачыў, **што** я хачу падняцца, а не падымуся*[*];

*Мы пасталі: думалі спачатку – немцы, зірнулі – пазналі, **што** гэта ідуць партызаны*;

*Тое, **што** адбывалася з Вялікімі Прусамі, дзве гэтыя жанчыны бачылі, перажылі кожная па-свойму*;

*Пытанне: – А пра **што** яны ўвечары гаварылі з людзьмі?* [5].

Усе прыведзеныя прыклады корпусны менеджар *Беларускага N-корпуса* суправадзіў, на жаль, не вельмі змястоўным і недыферынцыраваным у адносінах часцінамоўнай прыналежнасці адзінак метаапісаннем:

Lemma: што

CSX: Злучнік, падпарадкавальны,

SALXXAS: Займеннік, як у прыметніка, пытальна-адносны, вінавальны, адзіночны

SALXXNS: Займеннік, як у прыметніка, пытальна-адносны, назоўны, адзіночны [6].

Падобным чынам арганізавана метамоўнае суправаджэнне тэкстаў у большасці лінгвістычных корпусаў, пры гэтым, напрыклад, у шырокавядомым «Нацыянальным корпусе рускай мовы» аманімія знята толькі для параўнальна нязначнага аб'ёму (у суме каля 6 000 000 словаўжыванняў) тэкстаў – праца выканана супрацоўнікамі праекта «ўручную». Такім чынам, аманімія камп'ютарна-апасродкаванага дыскурсу пакуль можа быць безпамылкова інтэрпрэтавана толькі ў працэсе постапрацоўкі тэкстаў чалавекам.

Тым не менш індэкс семантычнасці, прадстаўляючы базавую інфармацыю аб пэўным параметры семантычнай ідэнтычнасці тэксту, дазваляе праводзіць дастаткова шырокія абагульненні. Практыка сведчыць, што праграмнымі сродкамі можна забяспечыць выкананне цэлага шэрагу алгарытмаў лінгвістычнай апрацоўкі тэкстаў. Распрацоўка вузкаспецыяльных першапачаткова задач у прыкладным рэчышчы часта вызначаецца перспектыўнасцю, фарматнасцю і магчымасцямі ўключэння ў комплексныя даследаванні.

Індэкс кантэкстуальнасці. Прадстаўленне лексічных адзінак у лічбавым фармаце, іх максімальная «інфарматызацыя» – праграма-мінімум для мадэлявання мовы віртуальнай рэальнасці. Пры тым, што мова (і маўленне) «...выкарыстоўвае толькі невялікую частку ад велізарнай колькасці тэарэтычна магчымых камбінацый, якую магло б даць адвольнае злучэнне мінімальных асноўных элементаў...» – калекцыя лексічных адзінак жывой мовы, якая ўжо створана, не можа быць поўнасцю ахоплена нават тэарэтычна [7].

Тым не менш сістэматызацыя КампАК на падставе лексічнай семантыкі безумоўна неабходна на практыцы для аптымізацыі прыкладных даследаванняў. Маўленне стварае тканіну неперарыўнасці дыскурсіўнай практыкі: «...семантычная прастора мовы аказваецца ... неперарыўнай» [8].

---

[*] Тут і далей прыклады па: Алесь Адамовіч, Янка Брыль, Уладзімір Калеснік. Я з вогненнай вёскі (1975) [Электронны рэсурс]. – Рэжым доступу: http://bnkorpus.info. – Дата доступу: 29.10.2014.

Маўленне і мова ўзаемна дэтэрмінаваныя, і маўленне ў значнай ступені вызначае моўную канстытуцыю: «Можна было б сказаць, перафразуючы класічнае выслоўе (лац. *Nihil est in intellectu, quod non prius fuerit in sensu* – ‘няма нічога ў свядомасці, чаго б не было ў адчуваннях’): *nihil est in lingua quod non prius fuerit in oratione* «няма нічога ў мове, чаго не было б раней у маўленні» [9].

У дадзеным кантэксце значным патэнцыялам валодае методыка прысваення тэкставым рэалізацыям лексемы лічбавай формы, прадстаўлення адносін яе кантэкстуальнасці / узуальнасці ў тэксце – ***індэкса кантэкстуальнасці*** (англ. – *Contextuality Index*), ці $\boldsymbol{I_C}$.

Алгарытм фарміравання такой анатацыі тэкстаў улічвае існаванне тлумачальных слоўнікаў, практыку дыферэнцыяцыі сукупнасці лексічных значэнняў у слоўнікавым артыкуле па прынцыпе парадкавага прыярытэту – спачатку найбольш распаўсюджаныя, потым усё больш рэдкія значэнні лексемы. Выкарыстанае ў пэўным кантэксце адценне значэння лексемы, што адпавядае дакладнай лічбе са спісу прэзентацыі семем у тлумачальным слоўніку, суадносіцца з агульнай колькасцю такіх вядомых і вызначаных у слоўніку семем – з дапамогай функцыі дзялення.

Разгледзім прыклад. Лексема *бегчы* ў ТСБМ прадстаўлена ў 9 значэннях і ў тэксце можа прысутнічаць, з улікам вядомага кантэксту, у 1-м значэнні – ‘хутка рухацца, перамяшчацца, моцна адштурхоўваючыся ад зямлі нагамі // імкліва накіроўвацца куды-н., рухацца ў якім-н. напрамку // спяшаючыся ісці куды-н.’: Якая гэта радасць, якое шчасце ісці і бегчы па толькі табе вядомых таемных лясных сцяжынках, дзе ўсё заслана снегам або залатым восеньскім лісцем, уздымаць з логавішчаў звера, крычаць яму ўслед, чуць, як спуджана ўцякае ён, прадзіраючыся праз ельнікі і хмызнякі (Леанід Дайнека. Меч князя Вячкі (1987)) [10]. Нескладаныя маніпуляцыі, якія нават на велізарных аб’ёмах могуць імгненна выконвацца праграмнымі сродкамі, дазваляюць даведацца, што дадзенае слова мае індэкс кантэкстуальнасці $\boldsymbol{I_C} \approx 0{,}11$ (1/9).

Дадзены індэкс можа быць сумай некалькі такіх дробаў у выпадку неадназначнасці / неаднаразовасці кантэкстнай прадстаўленасці лексемы. Цікавыя вынікі дадуць нават арыфметычныя сумы такіх індэксаў тэксту ў абсалютным ці адносным – прапарцыянальным дакладнай колькасці слоў тэксту – выглядзе. Выкарыстанне прамой, а не зваротнай прапарцыянальнасці дробу дазваляе скараціць парадак разрадаў лічбаў для вялікіх па аб’ёме тэкстаў.

Аб магчымасцях і неабходнасці кантэкстуальнай распрацоўкі зместу КампАК сведчыць пільная ўвага амерыканскіх і еўрапейскіх навукоўцаў да індэксацыі кантэксту, якая проста звышпрадуктыўная: даследаванні накшталт *Contextual Indexing and Joining: Supporting Efficient, Scalable Entity Search* (Кантэкстуальная індэксацыя і інтэграцыя: падтрымка каэфіцыентных, маштабных статыстычных даследаванняў) Ч. Тао і К. Кевіна; *Applying*

Random Indexing to Structured Data to Find Contextually SimilarWords (Прымяненне выпадковага індэксіравання да структурыраваных даных для знаходжання кантэкстуальна падобных слоў) Д. Дамлянавіч, У. Крушвіца, М–Д. Альбакура, Ё. Петрака, М. Лупу; *Latent Contextual Indexing of Annotated Documents* (Латэнтная кантэкстуальная індэксацыя анатаваных дакументаў) М. Зэнгштока, М. Герца і інш. – ствараюць самастойны вектар даследчай практыкі [11–13]. Не ўсе публікацыі застануцца запатрабаванымі і актуальнымі надоўга, але можна канстатаваць, што напрамак сфарміраваны.

Паняцці *кантэкстуалізацыі* (англ. – *contextualization*) як працэсу, *кантэкстуалізацыйных умоў* (англ. – *contextualization conventions*) як асяроддзя і *кантэкстуалізацыйных ключоў* (англ. – *contextualization cues*) як паказчыкаў – суадносяцца з важнейшымі тэкставымі і дыскурсіўнымі характарыстыкамі, хаця іх выкарыстанне Дж. Гамперцам, ніяк не асацыіравалася са сферай прыкладной лінгвістыкі [14]. Але паступальнае метадалагічнае афармленне дыскурсіўнай практыкі паслядоўна прыводзіць да прыкладной лінгвістычнай традыцыі. Супрацьпастаўленне паняццяў *кантэкстуалізм*, які валодае маўленчай пэўнасцю, і *ўзуалізм*, зарыентаванага на абстрактна-слоўнікавыя ўласцівасці семантыкі слоў, дазваляе пашырыць методыку інтэрдысцыплінарных лінгвістычных даследаванняў [15].

Такім чынам, адным з магчымых шляхоў рэалізацыі прыватных лінгвістычных «прылажэнняў» з'яўляецца апісанне і мадэляванне працэсу фарміравання кантэкстуальна залежных функцыянальных характарыстык моўных адзінак. Традыцыйна даследаванне «элементнай базы» камп'ютарна-апасродкаванага дыскурсу засяроджваецца на лексічным матэрыяле. Уважлівы аналіз адпаведнага ўзроўню маўленчай праекцыі інтэрнэт-мовы як мінімум спрыяльны ў кантэксце захавання пераемнасці лінгвістычных традыцый. Хаця і марфемныя, і сінтаксічныя, і тэкставыя (напрыклад, для прэцэдэнтных тэкстаў) даследаванні могуць выконвацца па аналагічных мадэлях.

Паказчыкі *металінгвістычнай індэксацыі* з'яўляюцца карыснымі для выяўлення металінгвістычнай спецыфікі тэкстаў у камп'ютарна-апасродкаваным фармаце, адлюстроўваючы значымыя параметры семантыкі. Рэпрэзентацыя, як у асобным тэксце, так і ў вялікім корпусе істотных для моўнай сістэмы парадыгматычных адносін можа быць аб'ектыўна ахарактарызавана з дапамогай *індэкса семантычнасці*. Імплементацыя семантычнай катэгарыяльнасці лексемы абумоўлена яе кантэкстнай (маўленчай) і лексікаграфічнай (моўнай) лакалізацыямі, што можа быць апісана з дапамогай *індэкса кантэкстуальнасці*. Даныя аб адпаведных уласцівасцях моўных адзінак дазваляюць мэтанакіравана і паслядоўна аналізаваць фарміраванне і рэалізацыю семантычнага патэнцыялу як лексічнага, так і металексічнага ўзроўню.

Інтэрпрэтацыя дыскурсіўных асаблівасцей тэксту ў металінгвістычным фармаце мае істотнае значэнне для развіцця сучаснай маўленчай практыкі і яе навуковага суправаджэння. Пакуль прымяненне прапанаваных методык да беларускіх тэкстаў ускладняецца дэфіцытам як беларускіх тэкстаў,

так і слоўнікаў у электронным фармаце і інтэрнэт-доступе, недастаткова глыбокай анатацыяй ужо створаных рэсурсаў. Несумненныя перавагі лінгваінфармацыйнага існавання, функцыянавання і ўдасканалення лексікаграфічных і тэкставых масіваў (камп'ютарна-апасродкаванага дыскурсу, найперш) абумоўліваюць высокую надзённасць вырашэння праблем камп'ютарнага забеспячэння беларускамоўнага дыскурсу.

Літаратура

1. МГУ–ЛОКЛЛ: Компьютерный корпус текстов русских газет конца XX века [Электронны рэсурс]. – Рэжым доступу: http://www.philol.msu.ru/~lex/corpus/corp_descr.html. – Дата доступу: 29.10.2013.

2. British National Corpus [Electronic resourse]. – Mode of access: URL: http://www.natcorp.ox.ac.uk. – Date of access: 29.09.2014.

3. British National Corpus.

4. Беларускі N-корпус [Электронны рэсурс]. – Рэжым доступу: http://bnkorpus.info. – Дата доступу: 29.10.2014.

5. Беларускі N-корпус.

6. Беларускі N-корпус.

7. Бенвенист, Э. Общая лингвистика / Э. Бенвенист; пер. с фр.; под ред., с вступит. ст. и коммент. Ю. С. Степанова. – М.: Прогресс, 1974. – С. 24.

8. Апресян, Ю. Д. Избранные труды: в 2 т. / Ю. Д. Апресян. – М.: Языки русской культуры, 1995. – Т. 1: Лексическая семантика. Синонимические средства языка. – С. 252.

9. Бенвенист, Э. Общая лингвистика. – С. 140.

10. Национальный корпус русского языка [Электронны рэсурс]. – Рэжым доступу: http://www.ruscorpora.ru. – Дата доступу: 29.09.2014.

11. Cheng, T. Contextual Indexing and Joining: Supporting Efficient, Scalable Entity Search [Electronic resource] / T. Cheng, K. Chang // UIUC Technical Report: UIUCDCS-R-2007-2911, Oct. 2007. – Mode of access: https://www.ideals.illinois.edu/handle/2142/11405. – Date of access: 14.09.2014.

12. Damljanović, D. Applying Random Indexing to Structured Data to Find Contextually SimilarWords / D. Damljanović [et al.] // Proceedings of the LREC, May 2012. – Istanbul, Turkey, 2012. – P. 2023–2030.

13. Sengstock, C. Latent Contextual Indexing of Annotated Documents (poster paper) / C. Sengstock, M. Gertz // Proceedings of the 21th International Conf. on World Wide Web, WWW, Apr. 16–20, 2012. – Lyon, France, 2012. – P. 593–594.

14. Gumperz, J. Discourse strategies / J. Gumperz. – Cambridge: University Press, 1982. – 225 p.

15. Баркович, А. А. Интернет-дискурс: компьютерно-опосредованная коммуникация: учеб. пособие / А. А. Баркович. – М.: Флинта: Наука, 2015. – 288 с.

Summary

The specificity of the meta descriptions of modern discourse is largely due to the high demand for the tools of interpretation and representation data of the multilevel semantics – from separated language units to complicated fragment of speech practice. The metalinguistic indexing algorithms allow to identify and aggregate present in any text metalinguistic potential. Indexes of semantic and contextuality are formed on the basis of linguistically relevant information which can be analyzed in a wide linguoinformational context.

Беларускі дзяржаўны ўніверсітэт *03.02.2015 г.*