

# БЕЛАРУСКІ N-КОРПУС І ПРАБЛЕМА КУЛЬТУРЫ МАЎЛЕННЯ

*Каваленка Я.І., Солтан М.В.*

*Беларускі дзяржаўны ўніверсітэт інфарматыкі і радыёэлектронікі  
г. Мінск, Рэспубліка Беларусь*

*Албут А.А. – ст. выкладчык*

Сёння цяжка ўявіць чалавека, які можа абыходзіцца без лінгвістычных ведаў, якія трэба ўвесь час удасканальваць, абнаўляць. У гэтым можа дапамагчы праца з корпусам тэкстаў мовы, на базе якога ў ідэале сёння складаюцца даведнікі па культуры маўлення. Разгледжана гісторыя распрацоўкі Беларускага нацыянальнага корпуса. Важна адначыць вялікую колькасць уключаных у яго публіцыстычных тэкстаў, паколькі гэты стыль мовы дынамічны і найперш упывае на змены ў моўнай норме. Важна папулярызаваць дадзеную распрацоўку беларускіх навукоўцаў, каб ёй актыўна карысталіся, пашыралі.

Мовазнаўства, або лінгвістыка, — навука аб будове, функцыянаванні і развіцці моў свету. Мова не дадзена лінгвісту ў прамым назіранні, толькі тыя факты маўлення, што непасрэдна назіраюцца, ці моўныя з’явы, ці маўленчыя акты носьбітаў жывой мовы. Таму і з’явілася корпусная лінгвістыка – частка лінгвістыкі, якая займаецца непасрэдна стварэннем тэкставых карпусоў. З развіццём інфармацыйных тэхналогій амаль кожная мова атрымала свой нацыянальны корпус, не стала выключэннем і наша – беларуская.

Развіццё беларускага корпуса мае даволі працяглую гісторыю. Першыя спробы стварыць корпус у канцы 1970-х гг. былі не вельмі ўдалымі, але ўжо у 2006 г. у рамках дзяржаўнай праграмы «Электронная

Беларусь» быў створаны паралельны руска-беларускі корпус тэкстаў аб'ёмам каля 1 млн. словаўжыванняў. У 2005 г. скончыўся тэрмін выканання дзяржаўнай праграмы, а ў 2006 г. пачалася новая дзяржаўная праграма на 2006–2010 гг. Над корпусам працавалі ў Мінскім дзяржаўным лінгвістычным універсітэце пад кіраўніцтвам прафесара Аляксандра Зубава, за аснову Корпуса беларускай мовы былі ўзяты напрацоўкі папярэдняга этапу. На гэтым этапе адпрацоўвалася сістэма граматычнай і структурнай разметкі, пашыралася граматычная база. Ужо на гэтым этапе быў створаны аднамоўны беларускі корпус.

Але вышэйапісаная беларускія корпусныя даследаванні не выходзілі на інтэгральны ўзровень нацыянальнага корпуса. Толькі з 2015 г. пачалася праца над стварэннем уласна Нацыянальнага корпуса. Тады быў зацверджаны асобны праект на 2015–2016 гг. «Стварэнне Нацыянальнага корпуса і лексіка-граматычнай базы беларускай мовы і іх праграмнае суправаджэнне», што выконваўся ў Інстытуце мовазнаўства імя Якуба Коласа НАН Беларусі. У выніку выканання праекта створаны электронны корпус тэкстаў аб'ёмам 251 млн. словаўжыванняў са структурнай і граматычнай разметкай. Але на сёняшні дзень дапрацоўкай корпуса займаюцца амаль толькі энтузіясты [1].

Разгледзім, як выглядае беларускі электронны корпус [2] і што можна з ім рабіць. Найперш нас сустракае галоўная старонка, дзе можна больш дакладна пазнаёміцца з гісторыяй праекта. Далей ідзе пошук тэкстаў ці слоў у корпусе. З дапамогай гэтай функцыі можна ўбачыць, ці ўжываюць слова, якое адсутнічае у слоўніках, у мове або мае варыянты. Напрыклад, разбярэм такія словы: *рэптыліі* ('паўзуны'), *сысуны* ('млекакормячыя') і *дадзеныя* ('даняны'). Калі пашукаць *рэптыліі* ў корпусе, можна заўважыць, што гэтыя слова ўжывалася ў перыёдыцы і не ўжывалася ў звычайных тэкстах. Калі ўзяць слова *сысуны*, таксама атрымаем спасылкі толькі на перыёдыку, але выдач будзе менш і крыніцы будуць яўна арыентаваныя на тарашкевіцу (цікава, што адзіная спасылка на газету «Звязда» (грамадска-палітычнае выданне са строгімі правіламі нармаванкі) утрымлівае аманімічнае прозвішча да нашага запыту). Але лексема *дадзеныя* ўжываецца як у шматлікіх часопісах і да т.п., так і ў розных іншых па стылі тэкстах (спасылак было шмат, бо памятаем не толькі пра субстантыў, але і звычайны дзеепрыметнік). Таму можна зрабіць выснову, што *рэптыліі* выкарыстоўваюць часцей за *паўзуноў*. *Сысуны* ўжываюцца, таксама як і *млекакормячыя*, але з папраўкай на стандарт сучаснай беларускай літаратурнай мовы. А слова *дадзеныя* ўжываецца часцей у звычайных тэкстах, насуперак слову *даняны* – часцейшаму ў публіцыстычных і афіцыйных (навуковых, афіцыйна-справавых).

На сайце Беларускага N-корпуса змешчаны раздзел з граматычнай базай, у якой мы можам пашукаць нейкае слова, даведацца, ці ёсць такое ў слоўніку, разгледзець яго марфалагічныя асаблівасці (парадыгму). Таксама ўвазе наведвальніка прапанавана старонка, дзе можна загрузіць на свой камп'ютар праграму для праверкі беларускай арфаграфіі для розных тэкставых рэдактараў.

Такім чынам, Беларускі N-корпус – цікавы праект, які можа быць карысным і запатрабаваным не толькі сярод лінгвістаў, але сярод навучэнцаў, тых, хто вывучае беларускую мову, працуе на ёй. Мы пераканаліся, што корпус можна выкарыстоўваць у якасці спосабу праверкі правільнасці маўлення. Мова не існуе закансервавана, таму наяўнасць у маўленні варыянтаў – прадказальная з'ява. Выбіраючы тое ці іншае слова, аналізуючы кантэксты, лёгка заўважыць дыскурсіўную абумоўленасць моўнай нормы.

**Спіс выкарыстаных крыніц:**

1. Кошчанка, У. Нацыянальны корпус беларускай мовы ў кантэксце корпуснай лінгвістыкі / У. Кошчанка [Электронны рэсурс]. – Рэжым доступу : [https://bnkorus.info/articles/Artykul\\_Koschanki\\_2018.html](https://bnkorus.info/articles/Artykul_Koschanki_2018.html). – Дата доступу : 17.04.2020.
2. Беларускі N-корпус [Электронны рэсурс]. – Рэжым доступу : <https://bnkorus.info/index.html#>. – Дата доступу : 17.04.2020.