

## **МАГЧЫМАСЦІ БЕЛАРУСКАМОЎНЫХ КОРПУСАЎ: ГРАМАТЫЧНАЯ НАРМАТЫЎНАСЦЬ**

Моўная сістэма, унармаваная і даволі ўстойлівая, усё ж такі паступова змяняецца. Адступленні ад норм літаратурнай мовы, якія характарызуюцца паўтаральнасцю і з цягам часу набываюць устойлівасць, у канчатковым выніку прыводзіць да рэформы правапісу. Гэта складаная з'ява, якая патрабуе грунтоўных мовазнаўчых і сацыялінгвістычных даследаванняў.

Большасць лінгвістаў разглядаюць моўную сістэму як статычную, маларухомую з'яву, якой уласцівы шэраг норм (арфаэпічных, акцэнталагічных, словаўтваральных, марфалагічных, лексічных, арфаграфічных, сінтаксічных, пунктуацыйных, стылістычных), што рэгулююць ужыванне моўных адзінак. У сваю чаргу, маўленне «актыўнае, лінейнае, імкнецца да аб'яднання слоў у маўленчай плыні. У адрозненне ад мовы яно менш кансерватыўнае, больш дынамічнае, рухомае» [5, с. 12]. Гэта значыць, што ў маўленні можа назірацца адступленне ад норм. Агульнавядома, што любая мова імкнецца да эканоміі моўных сродкаў, у выніку чаго адбываецца ўніфікацыя норм. Спрашчэнне сістэмы правілаў літаратурнай мовы перш за ўсё праяўляецца ў вусным маўленні, якое больш гнуткае, прыставае да змен. Паступова некаторыя з'явы пераносяцца ў пісьмовае маўленне і замацоўваюцца ў ім.

Корпусная лінгвістыка дазваляе ў значнай ступені спрасціць працэс выяўлення змен у мове. Лінгвістычны корпус (моўны корпус, корпус тэкстаў) — «вялікі, прадстаўлены ў электронным выглядзе, уніфікаваны, структураваны, размечаны, філалагічна кампетэнтны масіў моўных

дадзеных, прызначаны для вырашэння канкрэтных лінгвістычных задач» [3, с. 55].

Як вядома, «*Беларускі N-корпус* — першая спроба стварэння публічнага агульнага корпусу беларускай мовы» [2]. Аб'ём корпуса складае прыкладна 30 млн. словаўжыванняў. У склад корпуса ўваходзяць тэксты з мастацкай (вершаваныя і праязічныя) і публіцыстычнай літаратуры. Наяўнасць пашпартызацыі дазваляе карыстальніку атрымаць звесткі пра аўтара, назву твора, а таксама пра год напісання.

Такім чынам, прысутнасць экстралінгвістычнай разметкі (указанне звестак пра аўтара, жанр, дату напісання твора) дазваляе прасачыць ужыванне той ці іншай словаформы, слова, спалучэння слоў у розныя перыяды; выявіць паўтаральнасць аналізуемай з'явы; ахарактарызаваць спецыфіку яе ўжывання.

Наяўнасць марфалагічнай разметкі дае магчымасць даследавання мовы ў граматычным аспекце. Граматыкай называецца «раздзел мовазнаўства, змест якога складае вучэнне аб заканамернасцях формаўтварэння і словазмянення, тыпах словазлучэнняў і сказаў у пэўнай мове або групе роднасных моў», а таксама «граматычная будова мовы, сістэма аб'ектыўна дзеючых у мове правіл словазмянення, сувязі слоў у спалучэнні і сказы» [8, с. 38]. Так, пошук у корпусе магчымы па шматлікіх часцінамоўных катэгорыях. У раздзеле «Граматыка» прадстаўлены наступныя часціны мовы: назоўнік, лічэбнік, прыметнік, дзеяслоў (у склад якога ўваходзіць і дзеепрыслоўе), асобна вылучаны дзеепрыметнік, прыслоўе, злучнік, прыназоўнік, выклічнік, пабочнае слова, прэдыкатыў. Ёсць магчымасць ажыццяўляць пошук кантэкстаў з пэўнымі знакамі прыпынку — гэта значыць аналізаванаць пунктуацыю.

Яшчэ адным беларускамоўным корпусным рэсурсам з'яўляецца *Паралельны беларуска-рускі і руска-беларускі корпус* у складзе *Нацыянальнага корпуса рускай мовы*. Ён змяшчае «арыгінальныя тэксты і суадносныя з імі (роўныя сказ у сказ) пераклады гэтых тэкстаў» [6]. На сённяшні дзень аб'ём *Паралельнага (беларускага) корпуса* складае 350 308 сказаў, 3 487 846 слоў. У ім змешчаны пераклады мастацкіх тэкстаў з беларускай мовы на рускую і наадварот. У асноўным выкарыстанне паралельных карпусоў арыентавана на падыходзіць «да арыгінальных і перакладзеных тэкстаў з сапраўды навуковага пункту погляду: выкарыстоўваць статыстычныя метады для выяўлення тых або іншых заканамернасцяў, абапіраючыся пры гэтым на аб'ектыўныя дадзеныя» [4]. Аднак *Паралельны (беларускі) корпус* таксама з'яўляецца прыдатным не толькі для параўнальнага аналізу беларускай і рускай моў, але і для даследавання праблем уласна беларускага мовазнаўства.

Навуковыя тэксты прадстаўлены ў *Corpus Albaruthenicum* [1]. Гэты рэсурс невялікі па аб'ёме (74 тэксты, 350027 слоў), таму складана сцвярджаць пра рэлевантнасць лінгвістычных дадзеных, атрыманых з яго дапамогай. Такая колькасць моўных адзінак абмяжоўвае магчымасці даследчыка, але той факт, што ўсе тэксты *Corpus Albaruthenicum* адносяцца

да аднаго функцыянальна-стылістычнага тыпу, дазваляе аналізаваць спецыфіку ўжывання моўных адзінак розных узроўняў менавіта ў навуковым стылі.

Лінгвістычны корпус дае магчымасць вылучыць выпадкі пашырэння нормы. Так, напрыклад, на сённяшні дзень адсутнічае аднастайнасць у выбары канчаткаў назоўнікаў множнага ліку роднага склону. У залежнасці ад асновы назоўнікі ў множным ліку могуць набываць канчаткі *-аў (-яў)*, *-оў (-ёў)*, *-ей (-эй)* і *нулявы*. Неабходна заўважыць, што многія з такіх назоўнікаў ужываюцца з канчаткам *-аў*, які мае тэндэнцыю да пашырэння. Для прыкладу разгледзім лексему «*слова*», аснова якой заканчваецца на адзін зычны, і якая, адпаведна, у родным склоне мае нулявы канчатак, што і пацвярджае Беларускі арфаграфічны слоўнік, 2010 г. пад рэд. А.А. Лукашанца. Пры гэтым часта сустракаецца і форма «*словаў*». Каб прасачыць частотнасць ужывання кожнай з гэтых форм скарыстаемся магчымасцямі лінгвістычных корпусаў. Дадзеныя пошуку прадстаўлены ў табліцы:

Табліца

Форма слова	Колькасць уваходжанняў		
	Беларускі-N-корпус	Паралельны корпус (беларускі)	Corpus Albaruthenicum
слоў	больш за 100 (больш за 53%)	287 (95%)	12 (86%)
словаў	89 (каля 45 %)	12 (5%)	2 (14%)

З дадзеных, прыведзеных у табліцы вынікае, што ўжыванне нарматыўнага варыянту пераважае. Але спецыфіка кожнага з выкарыстаных корпусаў дазваляе прааналізаваць асаблівасці ўжывання тых або іншых форм у залежнасці ад відаў, тыпаў і жанраў тэкстаў. Напрыклад, у *Corpus Albaruthenicum* змяшчаюцца тэксты з розных галін навукі. Як вядома, навуковы стыль адметны сваёй унармаванасцю, што і тлумачыць адзінкавыя выпадкі адхілення ад нормы. Большасць тэкстаў *Паралельнага корпуса* можна аднесці да мастацкай літаратуры, дзе таксама адлюстравана нарматыўнасць мовы. З прычыны таго, што ў *Беларускім-N-корпусе* змешчана вялікая колькасць публіцыстычных тэкстаў, гэты рэсурс найбольш ярка дэманструе пашырэнне нормы, у прыватнасці, імкненне носбітаў мовы да ўніфікацыі канчаткаў назоўнікаў множнага ліку роднага склону.

Такім чынам, магчымасці лінгвістычных корпусаў дазваляюць прасачыць змены ў мове, выявіць спецыфіку ўжывання як асобных моўных адзінак, так і іх спалучэнняў. Адлюстраванне адхіленняў ад нормы дае магчымасць прасачыць паўтаральнасць назіраемай з'явы, частотнасць і ўстойлівасць яе ўжывання. Пералічаныя прыметы могуць паўплываць на пашырэнне або звужэнне літаратурнай нормы.

#### ЛІТАРАТУРА

1. Corpus Albaruthenicum [Электронны рэсурс]. — Режим доступа: URL: <http://grid.bntu.by/corpus>. — Дата доступа: 11.03.2015.