

## І НАВАЦЫЙНЫЯ ПАДХОДЫ Ё Вывучэнні І Даследаванні Фразеалагізмаў Беларускай і Англійскай Моў з Выкарыстаннем ІТ-Тэхналогій

Сучасная эпоха прынесла ё лінгвістыку, і ё прыватнасці ё фразеалагічныя даследаванні, не проста новыя інструменты, а цалкам новую філасофію навуковага пазнання. Рэвалюцыйную змену абумовіла інтэграцыя тэхналогій штучнага інтэлекту, у аснове якіх ляжаць складаныя матэматычныя мадэлі і нейронныя сеткі. Гэтыя сродкі пераўтвараюць даследчыка з чалавека, які ўручную збірае і сістэматызуе прыклады, у аналітыка, які кіруе магутнымі алгарытмамі для выяўлення заканамернасцей, недаступных традыцыйнаму назіранню. Калі гаворка ідзе пра параўнальнае вывучэнне фразеалогіі такой славянскай мовы, як беларуская, і такой германскай, як англійская, гэтыя тэхналогіі дазваляюць пранікнуць у самую сутнасць культурна-моўных адрозненняў і падабенстваў на новым лічбавым узроўні.

Асновай гэтай трансфармацыі сталі нейронныя сеткі, архітэктурна якія натхнёна будовай чалавечага мозга. Іх унікальная здольнасць вучыцца на велізарных масівах тэкставых дадзеных без жорсткага праграмавання робіць іх ідэальным інструментам для працы з фразеалогіяй, дзе значэнне цэлага значна перавышае суму значэнняў яго частак. Адным з найбольш важных прынцыпаў з'яўляецца вектарызацыя слоў і выразаў. Кожнае слова або словазлучэнне пераўтвараецца ў шматмерны лічбавы вектар – своеасаблівую каардынату ў сэнсавай прасторы. Фразеалагізмы ё гэтай прасторы займаюць асаблівыя месцы. Іх вектары аказваюцца бліжэй адзін да аднаго, калі выражаюць падобныя паняцці, нягледзячы на літаральны сэнс кампанентаў. Напрыклад, беларускае *быць не ў сваёй талерцы* [1, с. 256] і англійскае *to feel like a fish out of water* [2, с. 137], верагодна, будуць мець блізкія вектары, бо абазначаюць адчуванне дыскамфорту, нягледзячы на розныя вобразы.

Гэтая вектарызацыя адкрывае шлях для аўтаматычнай класіфікацыі і сінаніміі. Нейронная сетка можа беспамылкова аднесці невядомы ёй раней фразеалагізм да пэўнай семантычнай групы: напрыклад, вызначыць, што “летуценнічаць” адносіцца да групы фразеалагізмаў пра мару і няўважлівасць разам з беларускім *будаваць наветраныя замкі* [1, с. 269] і англійскім *to build castles in the air* [2, с. 61]. Амаль цудам выглядае здольнасць гэтых мадэляў выяўляць варыятыўнасць. Пры правядзенні параўнання паміж мовамі такія мадэлі могуць прапанаваць не толькі перакладныя адпаведнікі са слоўнікаў, але і кантэкстуальныя сінанімы, якія больш дакладна перадаюць адценне сэнсу. Гэта асабліва важна для беларускіх ідыём, якія часта маюць глыбокія фальклорныя карані. Іх пераклад на англійскую мову патрабуе пошуку не літаральнага адпаведніка, а культурна-аналагічнага вобразу.

Асобны прарыў звязаны з архітэктурна трансфармераў і мадэлямі тыпу BERT (Bidirectional Encoder Representations from Transformers) [3] – гэта мадэль, якая разбірае тэкст у абодвух напрамках (злева і справа), што дазваляе лепей разумець кантэкст слоў у сказе. Дадзеная ўласцівасць аналізаваць кантэкст з абодвух бакоў дазваляе ім з вялікай дакладнасцю адрозніваць, калі звычайнае словазлучэнне выкарыстоўваецца ў якасці фразеалагізму. Напрыклад, мадэль разумее, што ё выказванні “Ён зноў пайшоў па гэтай лёгкай дарозе” гаворка ідзе пра маральны выбар (фразеалагізм), а не пра фізічную дарогу. Для беларускай мовы, дзе многія ідыёмы ўжываюцца ў вуснай мове і могуць мець нечаканыя кантэксты, такое разуменне неацэнна. Акрамя таго, гэтыя мадэлі могуць прапаноўваць прагнозы: на аснове ўзору яны могуць “дадумаць” прапушчаны кампанент у ідыёме або прапанаваць, які з некалькіх варыянтаў звароту (*праводзіць час ці прагульваць час*) больш імаверны ё пэўным кантэксте.

Аднак самай інавацыйнай з'явай апошніх гадоў сталі генератыўныя вялікія моўныя мадэлі, што ляжаць у аснове сістэм, падобных да DeepSeek або ChatGPT. Яны ўжо не проста аналізуюць мову – яны яе мадэлююць. У дачыненні да фразеалогіі гэта азначае, што такую мадэль можна выкарыстоўваць як універсальны інструмент для гіпатэзавання. Ёй можна прапанаваць сфармуляваць гіпотэзу аб семантычным полі, якое аб'ядноўвае беларускія фразеалагізмы з назвамі жывёл, а затым прапанаваць меркаванья англійскія паралелі, заснаваныя па падобнай логіцы, а не на прамым перакладзе. Яна можа генераваць тлумачэнні складаных зваротаў, складаць прыклады іх ужывання ў розных стылях, што рэальна паскарае працэс стварэння навучальных і перакладчыцкіх рэсурсаў для беларускай мовы. Але ў гэтай магутнасці схаваны і галоўны падвох: гэтыя мадэлі схільныя да стварэння ілюзорных звестак. Яны могуць выдумаць неіснуючы фразеалагізм з этнаграфічнай легендай аб яго паходжанні. Таму іх ролю нельга пераацаніць: гэта не аўтарытэтныя крыніцы, а дапаможнікі-гіпатэзатары, кожнае слова якіх патрабуе перакрываючай праверкі з дапамогай традыцыйных метадаў і надзейных крыніц.

Роля нейронных сетак у развіцці беларускай мовы аказваецца ўсё больш значнай у эпоху тэхналагічнага прагрэсу, бо гэтыя мадэлі дазваляюць значна павысіць якасць аўтаматычнай апрацоўкі тэксту, сур'езна пашыраюць магчымасці распазнавання, перакладу і генерацыі мовы. У прыватнасці, нейронныя сеткі выкарыстоўваюцца для стварэння сістэм аўтаматычнага перакладу, аўтаматычнага распазнавання мовы, апрацоўкі тэксту і вызначэння моўных асаблівасцей, што важна для падтрымкі і папулярызавання беларускай мовы ў сучасным інфармацыйным асяроддзі. Дарэчы, развіццё гэтых тэхналогій дазваляе спрасціць доступ да беларускай мовы для шырокай аўдыторыі, пашырыць магчымасці адукацыйных рэсурсаў, інтэграваць беларускую мову ў інтэрнэт-асяроддзі, сацыяльныя сеткі і іншыя сучасныя каналы. Акрамя таго, нейронныя сеткі могуць дапамагчы ў аднаўленні гістарычных тэкстаў, паляпшэнні аўдыя- і відэаконтэнту на беларускай мове, што ў цэлым спрыяе захаванню і развіццю нацыянальнай культуры і мовазнаўства.

Прамы вынік такой навуковай рэвалюцыі – стварэнне практычных інструментаў, даступных не толькі вузкаму колу спецыялістаў. На аснове апісаных нейрасеткавых мадэляў ужо сёння ствараюцца інтэлектуальныя сістэмы, якія могуць выконваць рэальныя задачы: ад аўтаматычнай разметкі вялікіх тэкставых архіваў да стварэння інтэрактыўных адукацыйных платформаў. Для беларускай мовы, дзе аб'ём лічбавых лінгвістычных рэсурсаў яшчэ адстае ад сусветнага ўзроўню, гэта асабліва важна. Можна скапіраваць уласныя, лакалізаваныя версіі такіх інструментаў, якія будуцца з улікам спецыфікі беларускай фразеалогіі, дыялектных асаблівасцей і культурнага кантэксту. Тэхналогіі дазваляюць не толькі вывучаць мову, але і актыўна фарміраваць яе лічбавае асяроддзе, захаваць яе багацце для будучых пакаленняў у новым інтэрактыўным фармаце. Гэта шлях ад пасіўнай фіксацыі да актыўнай лічбавай рэвіталізацыі моўнай спадчыны.

Такім чынам, уплыў ІТ-тэхналогій, у аснове якіх ляжаць нейронныя сеткі, на даследаванне фразеалогіі з'яўляецца пераваротным. Ён пераносіць цэнтр цяжару з апісання і класіфікацыі на разуменне і мадэляванне. Для беларускай мовы фразеалагічны фонд, які асабліва багаты і мае важнае культурнае значэнне, гэта адкрывае ўнікальны шанец. У выніку з'яўляецца магчымасць не толькі сістэматызаваць спадчыну з неверагоднай хуткасцю і дакладнасцю, але і па-новаму ўбачыць яе месца ў сусветнай мазаіцы моўных свядомасцей. Будучыня беларускай фразеалогіі ляжыць у плённай калабарацыі паміж лінгвістамі, якія разумеюць тонкасці матэрыялу, і спецыялістамі па машынным навучанні, здольнымі перакласці гэтыя тонкасці на мову алгарытмаў. Толькі такім шляхам можа быць цалкам раскрыты патэнцыял, які нейронныя сеткі ўносяць у разуменне сапраўднай духоўнай скарбніцы нацыі, захаванай у сталых словазлучэннях яе мовы.

## Літаратура

1. Лепешаў, І. Я. Этымалагічны слоўнік фразеалагізмаў / І. Я. Лепешаў. – Мінск: БелЭн, 2004 – 448 с.
2. Oxford Dictionary of Idioms / Oxford University Press. – Oxford University Press, 2020. – 416 с.
3. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL) // Штоквартальны часопіс. – 2019. – Т. 8, № 4. – с. 45-59. – Рэжым доступу: <https://aclanthology.org/N19-1423>. – Дата доступу: 06.02.2026.