

## ГРАМАТЫЧНЫЯ ПАМЫЛКІ МАШЫННАГА ПЕРАКЛАДУ НА БЕЛАРУСКУЮ МОВУ НАВУКОВЫХ ТЭКСТАЎ ІТ-ТЭМАТЫКІ

Гук М.І.

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

Албут А.А. – старшы выкладчык

Даследуюцца граматычныя памылкі ў беларускай ІТ-тэрміналогіі, якія ўзнікаюць пад уплывам машыннага перакладу. Паказана, што адной з галоўных праблем з'яўляецца ўжыванне дзеепрыметнікаў цяперашняга часу, неўласцівых беларускай літаратурнай мове.

Тэрміналогія галіны алгарытмізацыі і праграмавання ў беларускамоўных інтэрнэт-тэкстах часта з'яўляецца вынікам машыннага перакладу (далей – МП) з рускай мовы. Паколькі беларускамоўны ІТ-кантэнт у інтэрнэце прадстаўлены значна менш, чым рускамоўны, аўтаматычныя сістэмы перакладу, як правіла, працуюць па схеме: англійская → руская → беларуская мовы. Менавіта гэты апасродкаваны шлях з'яўляецца крыніцай сістэматычных граматычных памылак.

Асноўная праблема звязана з дзеепрыметнікамі цяперашняга часу. Згодна з нормамі беларускай літаратурнай мовы, дзеепрыметнікі цяперашняга часу з суфіксамі *-уч-* (*-юч-*), *-ач-* (*-яч-*), *-ем-* (*-ім-*) не ўласцівы беларускай мове і дапускаюцца толькі ў навуковым і афіцыйна-справавым стылях, калі інакш выказацца немагчыма. МП гэтага абмежавання не ўлічвае, бо ў рускай мове такія формы з'яўляюцца нарматыўнымі і шырока ўжываюцца ў тэрміналогіі: *компилируемый язык*, *выполняющийся процесс*, *объявляемая переменная*. Пры аўтаматычным перакладзе на беларускую мову яны ўзнаўляюцца механічна, без уліку граматычнай спецыфікі мовы. Акрамя таго, зваротныя дзеепрыметнікі на *-ся* ў беларускай мове не ўтвараюцца наогул, аднак МП рэгулярна ўзнаўляе і іх.

Параўнальны аналіз тэрмінаў з галіны алгарытмізацыі і праграмавання паказвае наступнае (табліца 1):

Табліца 1 - Параўнанне варыянтаў МП з нарматыўнымі беларускімі адпаведнікамі.

Англійскі варыянт	Варыянт МП	Нарматыўны беларускі варыянт
compiled language	кампілюемая мова X	мова, якая кампілюецца ✓
interpreted language	інтэрпрэтуемая мова X	мова, якая інтэрпрэтуецца ✓
running process	выконваючы працэс X	працэс, які выконваецца ✓
nested loop	угнязджаючы цыкл X	укладзены цыкл ✓
declared variable	аб'яўляемая зменная X	аб'яўленая зменная ✓
increasing function	узростаючая функцыя X	нарастальная функцыя ✓
decreasing function	змяншаючаяся функцыя X	спадальная функцыя ✓
executing program	выконваючаяся праграма X	праграма, якая выконваецца ✓
operating system	кіруючая сістэма X	операцыйная сістэма ✓
branching algorithm	разгаліноўваючыся алгарытм X	алгарытм з разгалінаваннем ✓
returning value	вяртаемае значэнне X	вернутае значэнне ✓

Такім чынам, МП не толькі ўзнаўляе граматычна непажаданыя формы, але і ў шэрагу выпадкаў перашкаджае ўжыванню ўстойлівых беларускіх тэрміналагічных адзінак, замяняючы іх рускімі калькамі ў беларускай графіцы. Негатыўны ўплыў машыннага перакладу на беларускую мову не абмяжоўваецца, аднак, асобнымі граматычнымі памылкамі. Масвае распаўсюджванне аўтаматычна перакладзеных тэкстаў у інтэрнэце прыводзіць да з'явы, якую можна назваць граматычнай эрозіяй: памылковыя формы паступова ўспрымаюцца чытачамі як нарматыўныя, паколькі сустракаюцца значна часцей, чым правільныя. Гэта асабліва небяспечна для тэрміналогіі, бо тэрмін павінен быць сістэмным, гэта значыць, паслядоўна ўжывацца ва ўсіх тэкстах адной галіны.

Сітуацыю ўскладняе тое, што беларускамоўны сегмент інтэрнэту ў галіне ІТ невялікі. Карыстальнік, які шукае інфармацыю пра праграмаванне па-беларуску, з вялікай імавернасцю трапіць на машынна перакладзены тэкст і не будзе мець магчымасці параўнаць яго з нарматыўным узорам. У выніку МП не дапамагае развіваць беларускую тэрміналогію ў галіне праграмавання, а наадварот – замацоўвае памылковыя формы як нібы нарматыўныя.

Для пацвярджэння вышэйсказанага быў праведзены эксперымент. Зыходны тэкст на рускай мове з базай лексікай алгарытмізацыі і праграмавання быў перакладзены шасцю сістэмамі МП: DeepSeek, Kimi, Qwen, ChatGPT, Яндэкс Перакладчык і Google Translate. Для эксперыменту быў выкарыстаны наступны тэкст на рускай мове: «Компилируемый язык программирования

преобразует исходный код в исполняемый файл с помощью компилятора. В отличие от интерпретируемых языков, выполняемых непосредственно во время запуска, компилируемые программы обычно работают быстрее. Основными элементами любого алгоритма являются объявляемые переменные, ветвящиеся условия и повторяющиеся циклы. Возрастающая или убывающая функция может быть задана с помощью рекурсивно вызываемой подпрограммы. Операционная система управляет выполняющимися процессами и распределяет ресурсы между запущенными приложениями».

Параўнальны аналіз перакладаў паказвае наступнае (табліца 2):

Табліца 2 - Колькасць памылак у перакладах шасці сістэм МП.

Сістэма МП	Колькасць памылак	Асноўныя тыпы памылак
Kimi	2	дзеепрыметнікі цяперашняга часу (-ем-)
DeepSeek	3	дзеепрыметнікі цяперашняга часу (-уч-, -ач-, -ем-)
Qwen	3	дзеепрыметнікі цяперашняга часу (-ем-, -уч-)
ChatGPT	3	дзеепрыметнікі цяперашняга часу (-уч-), транслітарацыя з рускай мовы
Google	4	памылкі ў часе, семантычныя памылкі, дзеепрыметнікі цяперашняга часу(-уч-, -ач-)
Яндэкс	5+	граматычны разлад, памылкі ў часе, несупадзенне форм

Вынікі эксперымента паказваюць, што ніводная з шасці сістэм МП не справілася з перакладам поўнасцю. Найлепшы вынік паказала Kimi (2 памылкі), аднак і яна дапусціла дзеепрыметнікі цяперашняга часу. Большасць сістэм зрабіла па 3–4 памылкі, найгоршы вынік – у Яндэкс Перакладчыка (5 і больш памылак). Галоўная агульная праблема для ўсіх сістэм – перадача рускіх дзеепрыметнікаў цяперашняга часу, якіх у такой форме ў беларускай мове не існуе. Варта адзначыць, што нейрасеткавыя сістэмы (Kimi, DeepSeek, Qwen, ChatGPT) у цэлым паказалі крыху лепшы вынік у параўнанні з класічнымі перакладчыкамі (Google Translate, Яндэкс), аднак праблема з дзеепрыметнікавымі канструкцыямі застаецца агульнай для ўсіх тыпаў сістэм. Такім чынам, усе разгледжаныя сістэмы МП практычна аднолькава не ўлічваюць граматычную спецыфіку беларускай мовы ў галіне тэрміналагічных дзеепрыметнікавых канструкцый.