

НЕЙРАСЕТКІ Ў СПРАВЕ ПЕРАКЛАДУ БЕЛАРУСКАЙ БЕЗЭКВІВАЛЕНТНАЙ ЛЕКСІКІ

Дзяржаўская А.В., Дзяржаўская Л.В.

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

Наўроцкая І.В. – старшы выкладчык

Прааналізаваны асаблівасці перакладу безэквівалентнай лексікі з беларускай мовы на рускую і падбору аднаслоўнага адпаведніка рускім словазлучэнням на беларускай мове з дапамогай чатырох нейрасетак (Deepseek, Copilot, Grok, ChatGpt). Дадзенае даследаванне дазваляе ўстанавіць, наколькі карыснымі могуць быць нейрасеткі ў пытаннях вывучэння беларускай мовы.

Адметныя рысы культуры кожнага народа выяўляюцца ў першую чаргу на ўзроўні лексікі, што спрыяе павелічэнню яе выразнасці і пашырэнню слоўніка. Вялікую цікавасць выклікае пласт слоў беларускай мовы, нацыянальная адметнасць якіх выяўляецца ў немагчымасці знайсці ім іншамойны аднаслоўны адпаведнік.

Беларуская і руская мовы з прычыны шматлікіх фактараў моцна ўзаемазвязаныя, і традыцыйны пераклад "слова праз слова" з адной з іх на іншую з'яўляецца самым распаўсюджаным для беларускай лексікі, і, як следства, вельмі важны для носбітаў абедзвюх культур для папаўнення іх моўных рэсурсаў. Аднак такі прынцып не здольны з аднолькавым поспехам дзейнічаць на ўсёй лексічнай прасторы. Нягледзячы на тое, што аднаслоўны эквівалент пры жаданні можна знайсці заўсёды (пераважна шляхам падбору блізказначнага сіноніма), адценні, звязаныя з адметнымі ўмовамі жыцця асобнага народа, беззваротна страчваюцца.

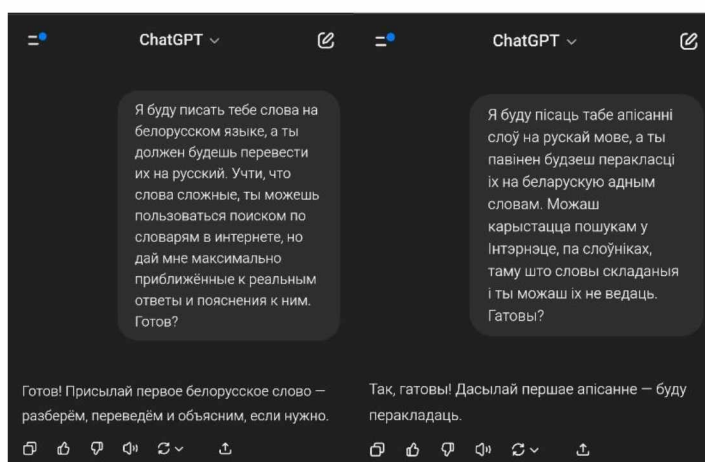
У лексіцы беларускай мовы, як і іншых нацыянальных моў, можна вылучыць шэраг слоў, якія ў рускамоўным дачыненні не маюць аднаслоўнага адпаведніка. Патрэбы перакладчыцкага характару вымушаюць перадаваць семантыку такіх слоў, аб'ём паняцця, што яны абазначаюць, і эмацыянальныя адценні не словамі, а словазлучэннямі рускай мовы. Такія словы вызначаюцца навуковым паняццем "безэквівалентныя", гэта значыць, "якія не маюць адпаведніка", і менавіта аднаслоўнага [1].

Паколькі сам па сабе падбор перакладу нават традыцыйным спосабам — справа не заўсёды дакладная і шматварыятыўная, пераклад безэквівалентнай лексікі можа стаць сапраўднай праблемай для чалавека, які пачынае вывучаць беларускую мову. Калі звыклыя анлайн-перакладчыкі і слоўнікі не спраўляюцца з прадстаўленай задачай нават пры наяўнасці кантэксту слоў, якія трэба перакласці, на дапамогу прыходзяць нейрасеткі — шматфункцыянальны інструмент, закліканы ў тым ліку садзейнічаць у пытаннях адукацыі. Таму важна ведаць, якімі з іх лепш карыстацца ў пэўных мэтах, каб дасягнуць найбольшай эфектыўнасці.

Для аналізу дакладнасці перакладу безэквівалентнай беларускай лексікі пры дапамозе нейронных сетак намі былі выкарыстаны наступныя платформы: DeepSeek, Copilot, Chat GPT, Grok. Менавіта гэтыя нейрасеткі ляглі ў аснову даследавання па некалькіх прычынах. Па-першае, яны вельмі папулярныя і часта выкарыстоўваюцца ў паўсядзённым жыцці, і, адпаведна, вынікі будуць актуальнымі для практычнага прымянення. Па-другое, усе гэтыя платформы былі спраэктаваны рознымі кампаніямі. DeepSeek — кітайскай арганізацыяй, якая спецыялізуецца на

стварэнні вялікіх моўных мадэляў і іншых тэхналогій штучнага інтэлекту. Copilot — прадукт Microsoft, аднаго з сусветных лідараў у галіне распрацоўкі праграмнага забеспячэння і воблачных тэхналогій. Chat GPT створаны амерыканскай кампаніяй OpenAI, вядомай сваімі перадавымі даследаваннямі ў галіне штучнага інтэлекту. Grok належыць кампаніі Ілана Маска, якая займаецца ўвасабленнем мадэляў штучнага інтэлекту з акцэнтам на максімальна прайздзівя і дакладныя адказы. Па-трэцяе, DeepSeek і Grok з'яўляюцца больш маладымі нейрасеткамі ў параўнанні з дзвюма іншымі, а значыць, яны могуць валодаць палепшанымі магчымасцямі і высокай канкурэнтаздольнасцю ў вобласці штучнага інтэлекту. Такі выбар нейрасеткавых сэрвісаў дазваляе выявіць моцныя і слабыя бакі кожнай платформы і атрымаць комплекснае ўяўленне аб магчымасцях нейронных сетак ў перакладзе безэквівалентнай лексікі.

Пачаць было неабходна з промпта — запыту нейрасетцы, каб яна выдавала карэктны вынік без просьбаў перарабіць. Мы сфармулявалі два паведамленні для двух выпадкаў: перакладу з беларускай на рускую і падбору аднаслоўнага адпаведніка на беларускай мове для рускага словазлучэння. І першае, і другое былі напісаны падобным чынам, але на розных мовах і з удакладненнем, што падабраць пераклад з рускай мовы на беларускую трэба менавіта аднаслоўны: прыклады праілюстраваны на малюнку 1. У абодвух выпадках нейрасеткі адказвалі на мове запыту.



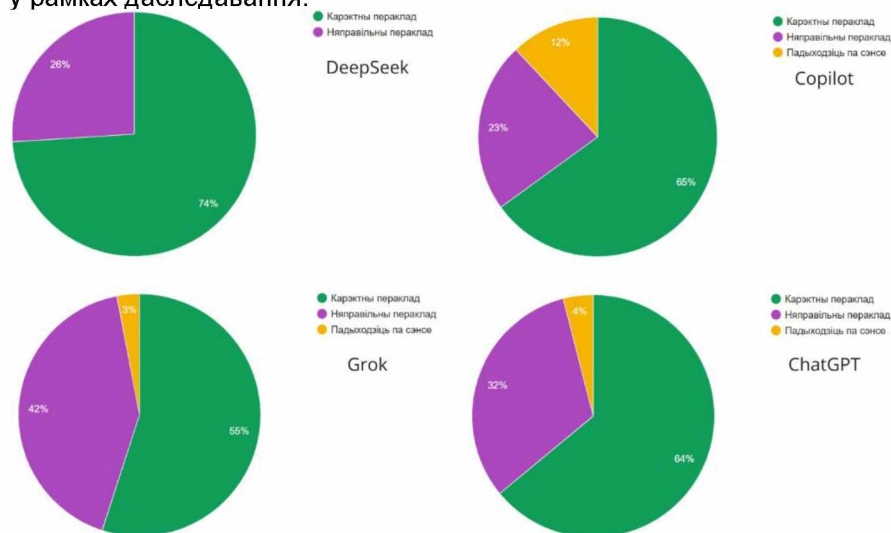
Малюнак 1 – Промпты для перакладу

Потым мы задумаліся, як ацэньваць пераклад. Усё зразумела, калі ён цалкам карэктны, але, як згадвалася раней, гэта ўсё роўна працэс часам недакладны, звязаны з дапушчэннямі і набліжэннямі. А калі яшчэ і прывязаць да гэтага працэсу моўныя мадэлі, як мы зрабілі, вынік і зусім можа стаць нечаканым: многія з нейрасеткавых сэрвісаў аказаліся схільнымі прыдумляць новыя словы. Таксама мы ўвялі катэгорыі “Падыходзіць па сэнсе, але не тое, што мы чакалі”, “Абсалютна няправільны пераклад” і “Пераклад правільны, але дапушчана арфаграфічная памылка” — апошнія таксама аказалася не рэдкасцю. На малюнку 2 прадстаўлены фрагмент складзенай намі табліцы з вынікамі. Яны сапраўды атрымаліся разнастайнымі, так што нельга адназначна рэкамендаваць да выкарыстання адну канкрэтную нейрасетку, паколькі ў вялікай выбарцы з больш чым трыццаці безэквівалентных слоў памылкі і недахопы былі ва ўсіх. Аднак мы можам назваць тыя нейрасеткі, якія справіліся з заданнем лепш за астатнія.

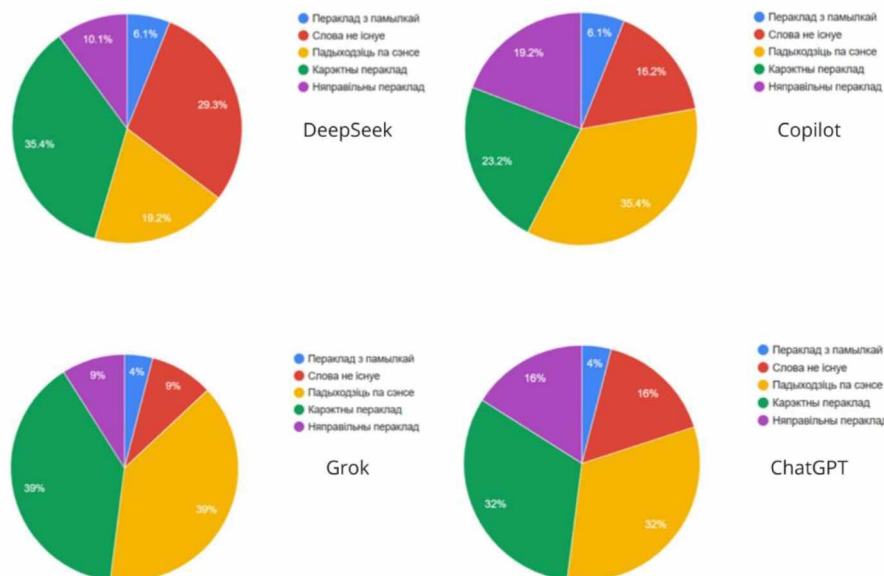
Беларуская мова	Русский язык	DeepSeek пераклад на рускую мову	DeepSeek пераклад на беларускую мову	Copilot пераклад на рускую мову	Copilot пераклад на беларускую мову	Grok пераклад на рускую мову	Grok пераклад на беларускую мову	Chat GPT пераклад на рускую мову	Chat GPT пераклад на беларускую мову
Талака	Коллективная помощь, работа сообща	Коллективная помощь, толока	талака	совместная работа, взаимопомощь	тапока	толока, совместная помощь, коллективная работа	талака	совместная работа, помощь, коллективный труд	Талока
Тэлефанаваць	Звонить по телефону	Звонить по телефону	тэлефанаваць	звонить	тэлефанаваць	Звонить по телефону	тэлефанаваць	звонить по телефону	Тэлефанаваць
Усцяж	На всём пространстве, во всех направлениях	вдоль, по всей длине, на протяжении	уздоўж	вдоль, постоянно	уздоўж	во всё время, непрерывно	усюды	все время, постоянно, непрерывно	Паўсюль
Фатэль	Мягкое кресло	полотенце	фатэль	кресло	фацэль	кресло	фатэль	кресло, кресло-качалка	Фатэль
Хмурынец	Становится пасмурным, сумрачным	хмуриться, становится мрачным	хмурыцца	хмуриться, пасмурнеть	хмурыцца	хмуриться, становится мрачным	хмурыцца	хмуриться, хмурить лицо, принимать недовольное или печальное выражение лица	Хмурынец
Цымус	Самое важное, самое существенное, суть чего-либо	суть, изюминка, главная идея	сутнасць	суть, изюминка, что-то особенное	аснова	изюминка, суть, самое вкусное или интересное	сутнасць	характерная черта, особенность, изюминка	Сутнасць
Часцець	Становится чаще, повторяется чаще	становиться чаще, повторяться чаще	часціцца	чистить, очищать	часцейшыць	угощать, потчевать	частаць	пытать, расспрашивать	Часцець

Малюнак 2 – Вынікі для перакладу з беларускай мовы на рускую

На малюнках 3 і 4 вынікі даследавання прадстаўлены ў выглядзе дыяграм. З іх відаць, што самай эфектыўнай нейрасеткай для перакладу беларускіх слоў на рускую мову можна лічыць Deepseek, а таксама з яго вынікамі можа супернічаць Copilot, які, тым не менш, прыўнёс у пераклад шмат недакладнасцей. Больш складана нейрасеткам аказалася падбіраць беларускі аднаслоўны адпаведнік для словазлучэнняў на рускай мове, і тут перамогу атрымаў Grok: акрамя таго, што ў яго самая вялікая колькасць цалкам карэктных адказаў, ён таксама значна часцей за іншыя нейрасеткі апынаўся бліжэй за ўсіх да правільнага перакладу, часам называючы не тыя словы, якія мы ад яго чакалі ў рамках даследавання.



Малюнак 3 – Вынікі для перакладу з беларускай мовы на рускую



Малюнак 4 – Вынікі для перакладу з рускай мовы на беларускую

Зыходзячы з усёй пададзенай вышэй інфармацыі, можна зрабіць выснову аб тым, што, хоць нейрасеткі не валодаюць беларускай мовай ідэальна, магчыма таму, што нячаста з ёй працуюць у параўнанні з іншымі мовамі. Аднак мы можам зафіксаваць становую дынаміку ў развіцці нейрасеткавых сэрвісаў. Так, Deepseek і Grok, нягледзячы на іх навізну, паказалі найлепшыя вынікі. Гэта сведчыць пра тое, што ў найбліжэйшым часе штучны інтэлект можна будзе сапраўды якасна выкарыстоўваць у галіне працы з беларускай лексікай. І ўжо цяпер, правяраючы па друкаваных крыніцах прадстаўленую сэрвісамі інфармацыю, можна зрабіць выснову, што з дапамогай

нейрасетак магчыма рабіць пераклад на нашу родную мову, што аднак не выключае рэдактарскай работы перакладчыка-чалавека над атрыманым праз нейрасеткі перакладам.

Спіс выкарыстаных крыніц:

1. Шкраба І.Р.Самабытнае слова: слоўнік беларускай безэквівалентнай лексікі (у рускамоўным дачыненні) /І. Р. Шкраба. — Мінск: Беларуская Энцыклапедыя імя Петруся Броўкі, 1994. — 207 с.