УДК 81’25

***Поворознюк Р.В.,*** к. філол. н., доц., ***Антонова В.,*** студ.
Інститут філології КНУ імені Тараса Шевченка, м. Київ

# ПРОТИСТОЯННЯ ЛЮДЕЙ ТА МАШИН: ЕВОЛЮЦІЯ МАШИННОГО ПЕРЕКЛАДУ НА ПРИКЛАДІ ПЕРЕДОПЕРАЦІЙНИХ ІНСТРУКЦІЙ, ОПРАЦЬОВАНИХ GOOGLE TRANSLATE

*У статті розглянуто найсучасніші досягнення в сфері машинного перекладу на прикладі роботи онлайн-перекладача Google Translate та проаналізовано основні етапи розвитку машинного перекладу: від статистичної моделі до цифрової нейронної мережі.*

***Ключові слова:*** *машинний переклад, штучний інтелект, онлайн-перекладач, нейронний машинний переклад, статистична модель перекладу, передопераційні інструкції.*

Нерідко можна почути, що останні досягнення в галузі машинного перекладу (МП) вже швидко дозволять штучному інтелекту не тільки конкурувати з перекладачами, але й згодом повністю замінити їх. У Південній Кореї навіть провідні експерти в галузі МП вирішили випробувати свої системи в протистоянні з професійними перекладачами в так званій битві перекладу “Людина” проти “Штучного інтелекту”. Але чи дійсно МП дійсно досяг того рівня, коли професійний перекладач вже не складає конкуренцію?

Термін “машинний переклад” був в ужитку ще до настання сучасної комп’ютерної ери. Деякі з найперших моделей машинного перекладу сприймалися як своєрідний дешифрувальний інструмент під час Другої світової війни та повоєнної епохи. Тому ранні методи машинного перекладу засновувалися на тому, що вихідний документ розуміли як повідомлення, яке потрібно було “дешифрувати”, щоб отримати цільовий текст [6].

На цьому побудовано наступне покоління “мовних технологій”. Підхід, що ґрунтувався на правилах (rules-based approach), розбивав вихідний документ на певний проміжний немовний код, який міг зрозуміти комп’ютер, а потім реконструював цей немовний код для будь-якої потрібної мови. Це призвело до епохи “Babel Fish” (вавилонської рибки), названої так на честь пристрою – рибки, яку поміщали в вухо для одержання миттєвого перекладу – що пішло з романів Дугласа Адамса “Путівник галактикою” [7]. Користувачі цього раннього покоління МП пам’ятають, якими неточними, а іноді й відверто невдалими були кінцеві переклади.

Але успіх був не за горами, і такі спроби проклали шлях до програмного забезпечення для перекладу, що дійсно працювало...

Використання статистичних методів у програмному забезпеченні для перекладу стало наступним кроком уперед, і цей підхід допоміг прославити Google Translate. Приблизно з 2007 року Google та інші великі компанії почали використовувати програмне забезпечення, яке сканувало Інтернет-простір, щоб знайти тексти, вже перекладені людьми. Вони слугували орієнтирами для майбутніх перекладів.

Мета полягала в тому, щоб знайти два різних робочих корпуси, які могли б використовуватися для “тренування” мовного сервісу Google на можливі переклади певних

рядків або фраз. Програмне забезпечення може потім копіювати ці моделі для використання в майбутніх завданнях [3].

Наразі машинний переклад дійсно працює, хоча, судячи з незграбного, а іноді і відверто курйозного перекладу передопераційних інструкцій пацієнтам, так і не скажеш. Наприклад, наступну інформацію можна зустріти на сайті Palisades Eye Surgery Center (США): "You will then be wheeled to a Pre-Operative area outside the operating room" (Вам будет вертеться к дооперационной зоне за пределами операционной). Але, як би там не було… Невже цифрові нейронні мережі працюють як людський мозок?

Найсучасніші модулі автоматичного перекладу працюють на основі системи, відомої як Digital Neural Networks або цифрова нейронна мережа (ЦНМ), що також іноді називається Neural Machine Translation або нейронний машинний переклад (НМП) [8]. Теоретично, вони змодельовані за прикладом людського мозку, але на практиці це просто ще один приклад математичної моделі, принцип роботи якої осмислювався протягом багатьох років.

Хоча зараз графічні процесори (graphics processing unit), як правило, використовуються для обробки найновіших відеоігор, ЦНМ з точки зору якості зробили великий крок уперед. До того ж, ця технологія може бути використана у сфері перекладу. Останнє покоління систем машинного перекладу використовує речення як основу для порівняння з іншими зразками, доступними в Інтернеті.

На перший погляд видається непоганим такий алгоритм дій та функціонал. Які ж тоді існують перепони на шляху до "ідеального" машинного перекладу?

Програмне забезпечення з перекладу "на базі людського мозку" звучить як щось із наукової фантастики, і, судячи з перших досліджень у цій галузі, майже так воно і є! Однак, існують проблеми, які впливають на якість перекладу, який здійснюється штучним інтелектом (AI). До них відносяться труднощі, пов'язані з:

1. Програмуванням систем, які були б здатні розпізнавати винятки з правил граматики.

2. Створенням системи, яка б розуміла контекст слова.

3. Розміщенням текстів у вільному доступі на необхідні пари мов [1] (для тренування системи).

Проблема з винятками не викликає питань, тож пропонуємо розглянути два останні пункти більш докладно.

Аспекти, пов'язані з контекстом, залишаються, мабуть, найбільш проблемними для майбутніх поколінь програмного забезпечення з автоматичного перекладу.

"Машини" просто не розуміють мови так, як людина. Звідки, наприклад, системі знати, про що саме йшлося в документі, який включав словосполучення "French teacher" – це стосувалося походження викладача чи сфери його викладання? Наразі AI має певні прогалини знань, тому поки що система не "сприймає" подібний контекст.

Одна справа, якщо контекст слугує уточнювальним елементом, деталізує явище чи суб'єкт, про який йшлося раніше (як у вище наведеному прикладі), а якщо він є сенсоутворювальним? Якщо від контексту залежатиме здоров'я та життя людини, як у передопераційних інструкціях пацієнтам? Адже навіть ті медичні захворювання, які мають універсальний характер, можуть набувати різних назв, спантеличуючи непідготовлених

читачів. Наприклад, Т.Дж. Пекхам, опитавши 180 пацієнтів з ортопедичними порушеннями, відзначав, що 80% респондентів вважали broken bone і fractured bone різними видами переломів [4], відповідно поради щодо процедури підготовки до операції можуть бути різними.

Друга проблема, що стосується навіть найновіших статистичних систем нейронного машинного перекладу (НМП), полягає в тому, що ці системи залежать від блоку даних, вже перекладених людиною. Навіть у доступних зараз трильйонах веб-сторінок деякі мовні пари просто не трапляються так часто.

Що стосується європейських мов, то це питання є менш актуальним, адже Європейський Союз має 24 офіційні мови і зазвичай створює блоки даних кількома мовами, до ресурсів яких можуть звернутися системи нейронного машинного перекладу. Для інших мовних пар, наприклад, фарсі та іспанської, або урду та італійської мов, кількість наявних текстів є значно меншою, що ускладнює роботу програмного забезпечення [6].

До того ж, у випадку з приватними компаніями, що прагнуть заощадити на професійному перекладі, набагато простіше скористатися машинним перекладом нижчої якості. Більше того, існує помилкова думка, що всі вихідці з колишніх республік СРСР вільно володіють російською мовою, а тому економічно невигідно розробляти різномовні версії (включно з українською).

Тому зараз можна зустріти таке явище, як "вбудований" в сайт перекладач. Наприклад, багато приватних закордонних клінік, як-от Colon & Rectal Surgery Associates (Міннеаполіс, шт. Міннесота, США) або St. Anthony Regional Hospital, and Nursing Home (Керолл, шт. Айова, США) , що пропонують широкий спектр послуг клієнтам – від пластичної хірургії до стоматологічних операцій – розміщують передопераційні інструкції (pre-operative (pre-op) instructions) з метою ознайомлення з низками заходів та правил, які слід виконати перед тим, як лягати на операційний стіл. Здавалося б, ступінь відповідальності максимально високий, і все ж, замість того, щоб надати якісний переклад, можливі клієнти-носії інших мов (відмінної від англійської) змушені читати коструватий переклад від Google Translate: "Q: Will it hurt to move my bowels? A: There should be no pain." (Q: Чи буде боляче поворухнути кишечник? A: Там не повинно бути ніякого болю), або: "You will be contacted by St. Anthony's pre-admission nurses after you and your surgeon have decided on surgery and scheduled the date for your procedure" (Вам буде зв'язатися по догоспитальной медсестрі Святого Антонія після ви і ваш хірург вирішили по хірургії і призначили дату процедури).

Ці проблеми стали ще більш явними після так званої битви перекладу "Людина" проти "Штучного інтелекту" в Кореї 21 лютого 2014 року.

Штучний інтелект представляли три системи перекладу. Одну надала компанія Google Inc. (Translate), іншу (Papago) забезпечила одна з найкращих компаній Інтернет-провайдерів в Південній Кореї Naver Inc., і останньою стала система від компанії Systran International [5]. Всі системи, запропоновані технологічними гігантами, спиралися на найновішу технологію НМП.

Проти них виступали чотири професійні перекладачі, кожен з яких мав щонайменш 5-річний досвід роботи.

Загалом, конкурсанти мали перекласти 8 статей: чотири з них написані англійською мовою, чотири – корейською. Кожному перекладачеві-людині було подано 1 статтю англійською мовою для перекладу на корейську та 1 статтю корейською мовою для перекладу на англійську. Кожна система перекладу від AI повинна була опрацювати всі статті.

Перекладачі мали 50 хвилин, аби впоратись із завданням. Учасники, що представляли AI, завершили переклад 8 статей протягом 10 хвилин.

Два незалежні професійні перекладачі, обрані Корейською міжнародною асоціацією з письмового та усного перекладу (International Interpretation Translation Association) та кіберуніверситетом Сон (Sejong Cyber University), оцінювали результати. І що ж вони показали?

За 30-ти бальною шкалою, команда AI набрала 10-15 балів, а команда перекладачів-людей 25! Так тримати!

Як ми вже мали змогу переконатися, навіть за участі всім і кожному відомого Google Translate переклад був далеким від "ідеального". Повертаючись до тих же передопераційних інструкцій, є компанії (приватні лікарні, клініки та центри догляду за здоров'ям в цьому випадку), які прагнуть зацікавити своїми послугами якомога більшу кількість клієнтів, звертаючись до сервісу перекладача Google. Так, на сайтах подібних до таких, що мають Maine Medical Center (Скарборо, штат Мен, США), Medizin & Ästhetik (Мюнхен, Германія), Ελληνικό Κρανιοπροσωπικό Κέντρο (передмістя Халандрі, Греція), розміщується інформація багатьма мовами, і шанс того, що зі списку перерахованих всі матимуть якісний переклад, надзвичайно малі. Наприклад, "No fish oil, Flax seed, vitamin E, or supplemental Garlic for 2 weeks prior to surgery" (Нет рыбий жир, льняное семя, витамин Е, или дополнительный чеснок в течение 2-х недель до операции), або: "If you develop a cold or infection of any kind, elective surgery will have to be postponed" (Простуда или другие инфекции ведут к переносу необязательных операций), або: "If you need blood for the surgery, something that will have been discussed in a previous appointment with your surgeon, please confirm that the necessary amount has been given in time" (Если вам потребуется кровь в течение хирургии, обсудите этот вопрос с вашем хирургом на предыдушей встрече, пожалуйста, убедитесь, что необходимое количество удалось собрать).

Однак, справедливим буде зазначити, що попри поразку в битві, Google Translate успішно пройшов певні етапи розвитку. Розглянемо детальніше хронологію подій.

Онлайн-перекладач Google Translate стає доступним у квітні 2006 року [2]. Він не застосовує граматичних правил, оскільки його алгоритм заснований на статистичному аналізі, а не на традиційному аналізі правил з підручника.

За словами Франца Йозефа Оча, придатна для використання база для розробки системи статистичного машинного перекладу для нової пари мов з нуля повинна складатися з двомовного корпусу текстів на більш ніж 150-200 мільйонів слів і двох одномовних корпусів кожної з мов у парі на більш ніж мільярд слів. Статистичні моделі на основі цих даних потім використовуються для перекладу в межах цієї мовної пари.

Якщо такої бази даних немає, то перекладач Google не перекладає з однієї мови на іншу (L1 → L2), замість цього він перекладає спочатку на англійську, а потім на цільову мову (L1 → EN → L2) [7].

Коли Google Translate генерує переклад, він шукає мовні шаблони в сотнях мільйонів документів, щоб обрати найкращий варіант перекладу. Виявляючи такі шаблони у вже перекладених людиною документах, Google Translate, буквально, пробує вгадати, як і що слід перекласти.

До жовтня 2007 року для усіх інших мов, окрім арабської, китайської та російської, перекладач Google використовував програмне забезпечення SYSTRAN; після – вже спирався на власну технологію, засновану на статистичному машинному перекладі.

У 2012 році дуже популярною в соцмережах і на різних форумах була розвага тестувати Google Translate і ділитися з усіма перлами, як він видавав. Наприклад, просте речення "My cat has given birth to four kittens: two brown, one white and one black" (Моя кішка народила чотирьох кошенят: двох коричневого кольору, одного білого і одного чорного) перетворювалося на "Мій кіт народив чотирьох кошенят: два коричневі кольори, одне біле і одного афроамериканця". Або: до 4 березня 2012 року (дата виборів президента РФ) фраза "Доброго ранку, Володимире Володимировичу" перекладалась як "Good morning, Mr President", після – як "Good morning, Volodymyr Volodymyrovych". Так відбувалося через те, що статистична система перекладу Google Translate завантажувала дані з Інтернету, упорядковувала ці дані і видавала найкращу, на думку самої системи, комбінацію в будь-якій мовній парі. Недивно, що такий принцип давав збої – ще донедавна словосполучення "революція гідності" перекладалося як "політична криза на Україні".

У січні 2015 року Google представив оновлений додаток Google Translate для iOS і Android. У версії сервісу став можливим миттєвий переклад написів з англійської на російську, французьку, німецьку, італійську, португальську та іспанську мови, а також з цих мов на англійську. Досить просто навести камеру на потрібну вивіску або текст, і переклад відразу ж відобразиться на екрані навіть без інтернет-підключення.

У вересні 2016 р. Google Translate оголосив про масштабне оновлення сервісу: в основу роботи перекладача будуть покладені нейромережі. Розробники стверджують, що нейромережі значно покращують якість перекладу, оскільки "машини" можуть аналізувати не окремі слова і фрази, а повноцінні речення і контекст [8].

Станом на 7 березня 2017 року такий принцип роботи поширюється на дев'ять мовних пар (а саме на: англійську, французьку, німецьку, іспанську, португальську, китайську, японську, корейську і турецьку мови), з якими працює Google Translate (всього перекладач підтримує більше 100 мов і близько 10 тис. мовних пар, щодня перекладаючи приблизно 140 млрд слів).

На думку експертів, метод перекладу Google, який отримав назву Zero-shot Translation ["Прицільний переклад"], не зрівняється з "людським" перекладом, хоч і демонструє хороший потенціал. Сподіваємося, що незабаром ті самі передопераційні інструкції будуть перекладатися краще, щоб не нагадувати фільми жахів ("Если вы получили комплект глаз и / или рецепт глазных капель от врача, пожалуйста, принесите их с собой") або якийсь невдалий анекдот ("Здесь вам будет предложено, чтобы вымыть лицо со специальным моющим средством и покрыть волосы с крышкой").

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. *Franz Josef Och*. Statistical Machine Translation: Foundations and Recent Advances. Tutorial at MT Summit 2005, Phuket, Thailand. – Google, Inc. – Access mode : http://www.mt-archive.info/MTS-2005-Och.pdf. 2. *Google Research Blog* / Statistical machine translation live. – Access mode : https://research.googleblog.com/2006/04/statistical-machine-translation-live.html. 3. *MT on and for the Web* / Christian Boitet, Hervé Blanchon, Mark Seligman, Valérie Bellynck; edt by Christian Boitet. – Access mode : http://www.academia.edu/4294657/MT_on_and_for_the_Web. 4. *Peckham, T.J.* Doctor, have I got a fracture or a break? – Access mode : http://www.injuryjournal.com/article/0020-1383(94)90065-5/abstract. 5. *The Jeju Weekly* / Battle of Korean translation apps: Google vs. Papago. – Access mode : http://www.jejuweekly.com/news/articleView.html?idxno=5596. 6. *Tim Adams.* Can Google break the computer language barrier? / Google The Observer / The Guardian. – Access mode : https://www.theguardian.com/technology/2010/dec/19/google-translate-computers-languages. 7. *Wikipedia, the free encyclopedia* / Google Translate. – Access mode : http://en.wikipedia.nom.si/wiki/Google_Translate. 8. *Бенюмов К.* "Как думаете, какой запрос самый распространенный?" Глава Google Translate Барак Туровски – о том, как сервис переходит на нейросети / К. Бенюмов. – Режим доступа : https://meduza.io/feature/2017/03/07/kak-dumaete-kakoy-zapros-samyy-rasprostranennyy.



***Поворознюк Р.В., к. филол. н., доц., Антонова В.В., студ.,***
***Институт филологии КНУ имени Тараса Шевченко, Киев***


## ПРОТИВОСТОЯНИЕ ЛЮДЕЙ И МАШИН: ЭВОЛЮЦИЯ МАШИННОГО ПЕРЕВОДА НА ПРИМЕРЕ ПЕРЕДОПЕРАЦИОННЫХ ИНСТРУКЦИЙ, ПЕРЕВЕДЁННЫХ В GOOGLE TRANSLATE


*В статье рассмотрены современные достижения в сфере машинного перевода на примере дооперационных инструкций, переведённых Google Translate, и проанализированы основные этапы развития машинного перевода от статистической модели к цифровой нейронной сети.*



***Ключевые слова**: машинный перевод, искусственный интеллект, онлайн-переводчик, нейронный машинный перевод, статистическая модель перевода, дооперационные инструкции.*



***Povorozniuk R., PhD., Ass. Prof., Antonova V., stud.***
***Taras Shevchenko National University of Kyiv***


## PEOPLE VS. MACHINES: THE EVOLUTION OF MACHINE TRANSITION ON THE EXAMPLE OF PRE-OPERATIVE INSTRUCTIONS PROCESSED IN GOOGLE TRANSLATE


*The article considers the most advanced achievements in the field of machine translation on the example of pre-operative (pre-op) instructions processed by Google Translate and analyzes the main stages of the machine translation development: from the statistical model to the digital neural network.*



***Key words**: machine translation, artificial intelligence, online translator, neural machine translation, statistical translation model, pre-operative (pre-op) instructions.*

УДК 811.161.2’243:378.147

***Самусенко О.М.,*** к. філол. н., доц.
Інститут філології КНУ імені Тараса Шевченка, м. Київ

# ЛЕКСИЧНІ МІНІМУМИ З УКРАЇНСЬКОЇ МОВИ ЯК ІНОЗЕМНОЇ: ОСНОВНІ ПІДХОДИ ДО УКЛАДАННЯ

*Стаття присвячена методиці викладання української мови як іноземної. У праці розглянуто питання укладання лексичних мінімумів з української мови як іноземної. У розвідці наведено європейські рекомендації щодо володіння лексикою іноземної мови на різних рівнях та проаналізовано зразки вітчизняних розробок у цій галузі. Основної уваги надано питанням обсягу словникового мінімуму для кожного рівня, принципам та критеріям відбору лексичних одиниць.*

*Ключові слова: українська мова як іноземна, методика викладання, лексичний мінімум, іноземні студенти.*

На сучасному етапі формування наукового та методичного забезпечення вивчення української мови іноземцями постало питання укладання лексичних мінімумів з української мови як іноземної для різних рівнів. Незважаючи на те, що проблемі створення лексичних мінімумів у навчанні мов як іноземних присвячена велика кількість наукових та науково-методичних праць, як зауважують дослідники, останніми роками помітне “переважання емпіричних напрацювань над теоретичними”, і виникла потреба “нової концепції лексичного мінімуму” [5, с. 86-87]. Ця концепція лексичних мінімумів має ґрунтуватися, на нашу думку, як на відомих лінгводидактичних здобутках, так і на досягненнях сучасних інформаційно-комунікативних технологій, застосування яких може стати у великій нагоді для таких робіт.

З одного боку, незаперечним є той факт, що для ефективного вивчення іноземної мови необхідно засвоєння всіх її систем, зокрема й лексичної. З іншого, укладання лексичних мінімумів є необхідним і доцільним й тому, що вони адресовані широкому колу реципієнтів: студентам та усім охочим вивчати іноземну мову і складати іспити з мови; викладачам та тестерам, що готують тренувальні, сертифікаційні тести та екзаменаційні матеріали, а також проводять тестування; фахівцям, що викладають мову як іноземну; авторам підручників та посібників з української мови як іноземної. Лексичні мінімуми зорієнтують викладачів, іноземних студентів та всіх охочих під час вивчення української мови та в процесі підготовки до складання іспитів з певного рівня володіння мовою.

Отже, **актуальність** пропонованої праці обумовлена необхідністю формулювання нових засад для укладання лексичних мінімумів та визначається потребою створення лексичних мінімумів з української мови як іноземної для різних рівнів володіння мовою, що є одним із компонентів комплексного забезпечення навчально-методичною бази викладання української мови в іноземній аудиторії. **Метою** статті є вивчення основних підходів до укладання лексичних мінімумів з української мови як іноземної. Досягнення мети передбачає розв’язання таких **завдань**: проаналізувати вітчизняні надбання зі створення лексичних мінімумів з української мови для іноземців; опрацювати