

УДК 37.022:355.231(477)

ОСОБЛИВОСТІ ВИКОРИСТАННЯ МАШИННОГО ПЕРЕКЛАДУ В УМОВАХ ПРОФЕСІЙНОЇ ПІДГОТОВКИ МАЙБУТНІХ ПЕРЕКЛАДАЧІВ

Янковець Андрій Володимирович
м.Хмельницький

Розглядається питання особливостей використання машинного перекладу в умовах професійної підготовки майбутніх перекладачів. На думку автора, системи машинного перекладу допомагають скоротити часові витрати, якщо текст досить об'ємний і містить одноманітну термінологію, але можуть застосовуватися лише там, де використовується максимально стандартизована мова з простою граматикою й порівняно невеликим запасом слів.

Ключові слова: машинний переклад, професійна підготовка, майбутній перекладач, система машинного перекладу.

Постановка проблеми. Прогресивний розвиток інформаційних і комп'ютерних технологій та безперервне розширення меж інформаційного простору відкривають перед майбутніми фахівцями безмежний океан можливостей, про які людство навіть і не мріяло сімдесят або лише двадцять років тому.

З огляду на об'єктивні еволюційні процеси та зміни у суспільстві, підготовку майбутніх перекладачів слід розглядати як пріоритетне завдання, що зумовлюється розширенням міжнародних зв'язків України, розвитком усіх сфер суспільного життя.

Питання особливостей використання машинного перекладу не залишаються осторонь як з боку науковців, так і виробників програмного забезпечення. Існують цілі асоціації машинного перекладу, які безпосередньо переймаються проблемами цього напрямку. На їхніх сайтах можна знайти відомості про щорічні конференції з проблем машинного перекладу, інформацію про новітні технології перекладу, матеріали про наукові організації, що працюють в галузі машинного перекладу.

За останні 10-15 років характер роботи перекладача й вимоги до його підготовки істотно змінилися. У першу чергу зміни торкнулися письмового перекладу науково-технічної, офіційної й ділової документації. Сьогодні, як правило, уже недостатньо просто перекласти текст, користуючись комп'ютером як друкарською машинкою. Від перекладача очікують, що оформлення готового документа відповідатиме зовнішньому вигляду оригіналу настільки точно, наскільки це можливо, і при цьому має задовольняти прийняті у даній країні стандарти.

У наші дні, з огляду на бурхливий розвиток інформаційних технологій і неухильний ріст обсягу інформації, перекладацька діяльність стає все більш затребуваною. Сучасному майбутньому перекладачеві доводиться не тільки переглядати професійну та іншу літературу з метою витягу інформації, але й обробляти її, тобто, практично, займатися перекладом.

Як наслідок технічного прогресу – підвищення вимог і до підготовки майбутніх перекладачів, зокрема, до особливостей використання машинного перекладу в умовах їхньої професійної підготовки.

Тому навчання майбутніх перекладачів використанню машинного перекладу, умінню грамотно й чітко викладати інформацію, отриману з першоджерела, набуває все більшого значення, стає досить актуальним.

Аналіз попередніх досліджень та публікацій. Вперше думку про можливість машинного перекладу висловив Чарльз Беббідж (1791-1871), який розробив у 1836-1848 рр. проект цифрової аналітичної машини – механічного прототипу електронних цифрових обчислювальних машин, що з'явилися через 100 років [1].

У 18 і 19 століттях було розроблено багато універсальних мов, у тому числі й Есперанто. Відомо, що ідея універсальної мови для машинного перекладу, ніяк не проявилася на початкових етапах розвитку цієї технології. Замість неї розглядалися тільки пари мов. Однак, протягом 1950-х і 60-х років, дослідники в Кембриджі на чолі з Маргарет Мастерман, у Ленінграді на чолі з Миколою Андрєєвим і в Мілані Сільвіо Чеккато почали роботу в цій галузі [1].

У 1970-х і 1980-х роках були зроблені певні успіхи в цій галузі й був побудований ряд систем машинного перекладу.

Фактично історія машинного перекладу починається з „Джорджтаунського експерименту“. У січні 1954 р. відбулася перша публічна демонстрація машинного перекладу з російської на англійську, здійсненого на машині IBM-701.

Інший напрямок робіт виник у Відділенні прикладної математики Математичного інституту Академії Наук (МІАН) СРСР за ініціативою А. А. Ляпунова. До робіт з машинного перекладу математичних текстів з французької мови на російську він залучив О. С. Кулагіну, аспірантку МІАН, своїх учениць Т. Д. Вентцель і Н. Н. Рікко [2].

З кінця 1955 р. у цих роботах брала участь Т. Н. Молошная, яка пізніше приступила до самостійної роботи над алгоритмом англо-російського перекладу.

А. А. Ляпунов та О. С. Кулагіна свої уявлення про використання обчислювальних машин для перекладу з однієї мови на іншу опублікували в журналі „Природа“, 1955, № 8. Перші програми машинного перекладу, розроблені цим колективом, були реалізовані на машині „Стріла“ [2].

Перше покоління систем машинного перекладу базувалося на алгоритмах послідовного перекладу „слово за словом“, „фраза за фразою“. Можливості таких систем визначалися доступними розмірами словників, прямо залежали від обсягу пам'яті комп'ютера. Переклад тексту здійснювався окремими пропозиціями, смислові зв'язки між ними ніяк не враховувалися. Такі системи називають системами прямого перекладу. На зміну їм з часом прийшли системи наступних поколінь, в яких переклад від мови до мови здійснювався на рівні синтаксичних структур. В алгоритмах перекладу використовувався набір операцій, що дозволяє шляхом аналізу перекладного пропозиції побудувати його синтаксичну структуру за правилами граматики мови вхідного пропозиції (так само, як навчають дітей мові в середній школі), а потім перетворити її в синтаксичну структуру вихідного пропозиції і синтезувати вихідну пропозицію, підставляючи потрібні слова зі словника. Такі системи називаються Т-системами (Т - від англійського слова „transfer - перетворення“).

З 1960 р. почалася підготовка кадрів в галузі автоматичної переробки текстів на філологічному факультеті Московського Державного Університету, в Ленінградському та Новосибірському університетах.

Мета статті. Дана стаття має на меті розглянути особливості використання машинного перекладу (МП) в умовах професійної підготовки майбутніх перекладачів.

Виклад основного матеріалу. Науково-технічний прогрес спричинив зростання значних об'ємів науково-технічної інформації: книг, статей, технічної документації. Переклад за допомогою людських ресурсів просто не може впоратися з цим потоком. Використання МП у різних відомствах і організаціях виявилось досить ефективним через високу стандартизованість і достатньо низькі вимоги до якості перекладу, що й дозволяло використовувати в роботі навіть системи першого покоління.

Інтерлінгвістичний машинний переклад є одним із класичних підходів до машинного перекладу. Вхідний текст трансформується в абстрактне представлення, яке не залежить від мови (на відміну від трансфертного перекладу).

Перекладний текст утворюється на основі цього представлення. Основною перевагою такого підходу є спосіб додавання нової мови в систему. Можна довести математично, що в рам-

ках цього підходу, створення кожного нового інтерпретатора мови для такої системи буде спрощувати її, у порівнянні, наприклад, із системою трансфертного перекладу.

Крім того, у рамках такого підходу є наступні можливості:

- реалізувати „переказ тексту“, перефразовування вхідного тексту в рамках однієї мови;
- відносно просто реалізувати переклад у дуже відмінних мов, таких як, наприклад українська й арабська.

Однак, дотепер не існує реалізації такого підходу, які б коректно працювали хоча б для двох мов. Багато експертів висловлюють сумніви в можливості такої реалізації. Найбільша складність для створення подібних систем полягає в проектуванні міжмовного представлення. Воно повинно бути одночасно абстрактним і незалежним від конкретних мов, але в той же час воно повинно відображати особливості будь-якої існуючої мови.

З іншого боку, у рамках штучного інтелекту, завдання виділення змісту тексту на даний момент дотепер не вирішене.

У цьому методі перекладу міжмовне представлення можна розглядати як спосіб опису аналізу тексту мовою оригіналу. При цьому, у представленні зберігаються морфологічні та синтаксичні характеристики тексту. Передбачається, що в такий спосіб можна передати „зміст“ при створенні перекладного тексту.

При цьому іноді використовується два міжмовних представлення. Одне з них більш відбиває характеристики вхідної мови. Інше – мови перекладу. Переклад у цьому випадку проводиться у два етапи.

У деяких випадках використовуються два та більше представлень одного рівня (однаково близькі до обох мов), але такі, що відрізняються за тематикою. Це необхідно для підвищення якості перекладу специфічних текстів.

Такий підхід не новий для лінгвістики. Він заснований на ідеї близькості мов. Для поліпшення якості перекладу, природня мова використовується в якості моста між двома іншими мовами. Наприклад, при перекладі з української на англійську, іноді використовується російська мова.

Для використання системи інтерлінгвістичного машинного перекладу необхідні:

- словники для аналізу й генерації текстів;
- опис граматики мов;
- база знань понять (для створення міжмовного представлення);
- правила проєкції понять для мов і представлення [3].

Переклад заснований на прикладах – один з підходів до машинного перекладу, при якому використовується двомовний корпус тексту. Цей корпус тексту під час перекладу використовується як база знань.

Для побудови системи машинного перекладу, яка заснована на прикладах потрібен мовний корпус, складений з пар речень. Мовні пари – це тексти, що містять речення однією мовою й відповідні їм речення іншою, які можуть бути як варіантами написання двох речень майбутнім перекладачем – носієм двох мов, так і набором речень та їх перекладів, виконаних курсантом.

Переклад, заснований на прикладах, найкраще підходить для таких явищ як фразові дієслова. Значення фразових дієслів сильно залежить від контексту. Фразові дієслова дуже часто зустрічаються в розмовній англійській мові. Вони складаються з дієслова із прийменником або прислівником. Зміст такого виразу неможливо одержати зі змістів складових частин. Класичні методи перекладу в цьому випадку застосувати неможливо.

Цей метод перекладу можна використовувати для визначення контексту речень.

Статистичний машинний переклад – це метод машинного перекладу, який використовує порівняння великих об'ємів мовних пар, так, як і машинний переклад, заснований на прикладах [3]. Статистичний машинний переклад має властивість „самонавчання“. Чим більше в розпорядженні є мовних пар і чим

точніше вони відповідають одна одній, тим кращим є результат статистичного машинного перекладу. Статистичний машинний переклад заснований на пошуку найбільш імовірного перекладу речення з використанням даних із двомовних корпусів текстів. У результаті при виконанні перекладу комп'ютер не оперує лінгвістичними алгоритмами, а обчислює ймовірність застосування того або іншого слова або виразу. Слово або послідовність слів, що мають оптимальну ймовірність, вважаються найбільш відповідними перекладу вхідного тексту й підставляються комп'ютером в текст, який одержується у результаті цього ж перекладу.

При статистичному машинному перекладі ставиться завдання не перекладу тексту, а завдання його розшифрування.

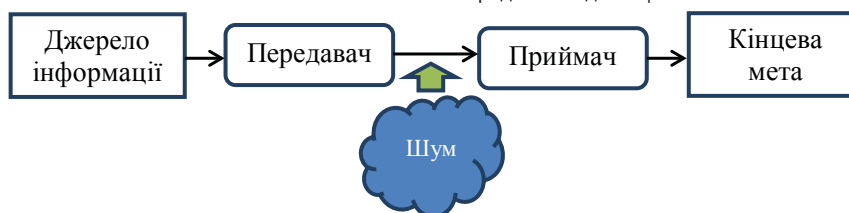


Рис. 1. Модель Шеннона

Через шум отримана приймачем інформація в загальному випадку не збігається з інформацією, відправленою передавачем. Однак, згідно Шеннону, створюючи надлишкову інформацію, вихідні дані можна відновити з якою завгодно високою ймовірністю. Для виявлення помилок використовуються контрольні суми, для їхнього виправлення – спеціальні коригувальні коди (за умови, що ступінь шуму не перевершує певної границі).

Варто відзначити, що будь-яка інформація до певної міри надлишкова. Людська мова надлишкова – щоб уловити зміст речення, найчастіше необов'язково чути його повністю. Аналогічно, письмова мова, теж надлишкова, і при перекладі цим можна скористатися. Якщо речення в цілому зрозуміло, але є кілька незнайомих слів, то звичайно не важко догадатися про їхнє значення [5].

Таким чином, для перекладу тексту необхідно знайти спосіб декодування, що використовує природну надмірність, у зв'язку із чим декодування повинне бути ймовірнісним.

Завдання такого декодування полягає в тому, щоб, при даному повідомленні, знайти вхідне повідомлення, якому відповідає найбільша ймовірність. Для цього ж необхідно для будь-яких двох повідомлень уміти знаходити умовну ймовірність того, що перекладене повідомлення, пройшовши через канал із шумом, перетвориться у вхідне повідомлення.

У цьому випадку потрібна модель джерела (модель мови) і модель каналу (модель перекладу). Модель мови дає оцінку ймовірності фразам перекладної мови, а модель перекладу оцінює ймовірність вихідної фрази за умови фрази перекладною мовою.

Якщо нам потрібно перекласти фразу з української на англійську, то ми повинні знати, що саме зазвичай говорять англійською мовою і як англійські фрази спотворюються до стану української мови. Сам по собі переклад перетворюється в процес пошуку такої англійської фрази, яка максимізувала б добуток безумовної ймовірності англійської фрази й ймовірності української фрази (оригіналу) за умови даної англійської фрази.

У системах статистичного перекладу, у якості моделі мови використовуються варіанти n-грамної моделі (наприклад, у перекладачі Google використовується 5-грамна модель). Згідно із цією моделлю, правильність вибору того або іншого слова залежить тільки від попередніх (n-1) слів.

Найпростішою статистичною моделлю перекладу є модель дослівного перекладу. У цій моделі, відомої як Модель IBM №1, передбачається, що для перекладу речення з однієї мови на

Ми припускаємо, що стаття, яка написана англійською мовою, насправді є статтею написаною англійською, але текст зашифрований (або перекучений шумом). При такому підході стає зрозуміло чому, чим більш віддалені мови, тим краще працює статистичний метод, у порівнянні із класичними підходами. Математична модель статистичного машинного перекладу відома як модель Шеннона [5].

Модель складається з п'яти елементів: джерела інформації, передавача, каналу передачі, приймача й кінцевої мети, розташованих лінійно (рис.1). Передавач кодує інформацію, отриману від джерела, і передає її на канал. По каналу передачі, на який діє шум – перешкоди будь-якого роду, що спотворюють інформацію, дані надходять у приймач, де вони декодується й передаються до кінцевої мети.

іншу досить перекласти всі слова (створити „мішок слів“), а розміщення їх у правильному порядку забезпечить модель мови. Єдиним масивом даних, яким оперує Модель IBM №1, є таблиця ймовірностей парних перекладних відповідностей слів двох мов. Звичайно використовуються більш складні моделі перекладу. Багато з них є комерційними таємницями компаній розробників.

Робота статистичних систем, так само як і систем заснованих на прикладах відбувається у двох режимах: навчання й експлуатації.

У режимі навчання проглядаються паралельні корпуси тексту й обчислюються ймовірності перекладних відповідностей. Будується модель мови перекладу. Відразу визначаються ймовірності кожного n-грама.

У режимі експлуатації, для фрази з вихідного тексту шукається фраза перекладного тексту, так, щоб максимізувати добуток ймовірностей.

Висновки. Отже, сьогодні якість машинного перекладу залишає бажати кращого, а наявність такого методу перекладу поки потрібно сприймати як предмет наукових досліджень. В умовах професійної підготовки майбутніх перекладачів у більшості випадків використання систем машинного перекладу не виправдане, оскільки:

- системи МП не дають прийнятної якості вихідного тексту. Більш високої якості можна досягти за допомогою попереднього настроювання системи (продукти серії PROMT надають користувачеві безліч можливостей для цього), що зовсім неприйнятно при невеликих обсягах перекладного тексту, і/або шляхом наступного редагування, а це тільки уповільнює роботу, якщо перекладач використовує сліпий метод друку;
- системи МП не гарантують дотримання єдності термінології, особливо при роботі колективу перекладачів над великим проектом. Вірніше, можуть гарантувати за умови дуже уважного поводження з користувацькими словниками, а на це не завжди варто розраховувати.

Однак у деяких випадках використання систем МП все ж таки допомагає скоротити тимчасові витрати. Це відбувається, якщо текст досить об'ємний і містить одноманітну термінологію, що дозволяє порівняно швидко настроїти під нього систему МП. Тоді редагування тексту не займе занадто багато часу. Однак у цьому випадку слід особливо уважно поставитися до стилю тексту перекладу. Машинний переклад формальний, тому ви-

сока ймовірність калькування синтаксичних структур мови оригіналу, що характерно для перекладу взагалі, а тому цілком може бути пропущене при редагуванні.

Загалом кажучи, системи МП цілком можуть застосовуватися там, де використовується максимально стандартизована мова з простою граматикою й порівняно невеликим запасом слів. Досить успішним проектом системи МП вважається німецька програма *Meteo*, що виконує переклад метеопрогнозів із французької мови на англійську й навпаки. Для полегшення роботи перекладачів і технічних письменників компанією *Boeing* у свій час був розроблений стандарт мови для написання технічної документації, який відомий як *Boeing English*.

Перспективи подальших розвідок за цим напрямом. Таким чином, говорячи про найбільш перспективні шляхи розвитку систем автоматизації перекладу, випливає, імовірно, зосередитися на тому, що здійснено на даний момент, тобто на створенні більш ефективних електронних словників з як можна більш ефективним механізмом пошуку й індексації, з як можна більш інтегрованою системою словникових статей. Якщо ж брати до уваги розвиток систем машинного перекладу, то найбільш перспективним напрямком тут виявляється вдосконалення підсистем граматичного аналізу й синтезу, а також збільшення обсягу контекстуального охоплення тексту й удосконалювання семантичних ланцюжків з метою більш точного добору значень слів.

Література і джерела

1. Бацевич Ф. С. Основи комунікативної лінгвістики / Ф. С. Бацевич. - К. : Вид. центр „Академія”, 2004. – С. 304-342
2. Слепович В. С. Курс перевода (английский – русский): Translation Course / В. С. Слепович. – Мн. : Тетра Системс, 2001. – 278 с.
3. Коваленко А. Я. Общий курс научно-технического перевода: Пособие по переводу с английского языка на русский / А.Я. Коваленко. – Киев, 2004. – 356 с.
4. Виноградов В. С. Введение в переводоведение / В. С. Виноградов. – М. : Изд-во ИОСО РАО, 2001. – С.222
5. Shannon C. A mathematical theory of communication / C. Shannon // Bell System Technical Journal. – 27(3) – 1948. – PP.379-423

Рассматриваются вопросы особенностей использования машинного перевода в условиях профессиональной подготовки будущих переводчиков. По мнению автора, системы машинного перевода помогают сократить временные затраты, если текст довольно объемный и содержит однообразную терминологию, но могут применяться лишь там, где используется максимально стандартизированный язык с простой грамматикой и сравнительно небольшим запасом слов.

Ключевые слова: машинный перевод, профессиональная подготовка, будущий переводчик, система машинного перевода.

The issue of special usage of machine translation in context of professional training of future translators has been considered. According to the author machine translation systems help to reduce time if a text is long enough and contains uniform terminology but can be used only in cases when the language is at most standardized with simple grammar and comparatively inconsiderable vocabulary.

Key words: machine translation, professional training, future translator, machine translation system.