МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІВСЬКА ПОЛІТЕХНІКА»

Інституту комп' ютерних наук

та інформаційних технологій

Кафедра прикладної лінгвістики

РЕФЕРАТ

НА ТЕМУ

МАШИННИЙ ПЕРЕКЛАД

Львів – 2009р.

Аnnotation

A topic of machine translation is exposed in this report. The aim of this paper is to investigate the history, the problem, the present situation and some aspects of machine translation. Firstly is discussed what machine translation is, how it works and the process of translation. Than the history is shown: when the first idea of MT was accepted and to whom it belongs, the progess and how it was developing. Next, the theory of MT and direction of development are discussed, and some examples of software programs for translating natural language are pointed. In conclusion, advantages and disadvantages are summed up and the place of machine translation nowadays is shown.

1. Вступ

Машинний переклад - виконувана на комп'ютері дія, яка перетворює текст з однієї природньої мови в еквівалентний за змістом текст на іншій мові, а також результат такої дії. Сучасний машинний, або автоматичний переклад здійснюється за допомогою людини: перед-редактора, який тим або іншим чином заздалегідь обробляє текст для перекладу, інтер-редактора, який бере участь в процесі перекладу, або пост-редактора, який виправляє помилки і недоліки в перекладеному комп’ютером тексті.

Для здійснення машинного перекладу в комп'ютер вводиться спеціальна програма, що реалізує алгоритм перекладу, тобто послідовність однозначно і строго визначених певних дій над текстом для знаходження відповідностей для перекладу в даній парі мов при заданому напрямі перекладу (з однієї конкретної мови на іншу).

Програми перекладу (системи машинного перекладу) з'явилися у відповідь на потреби користувачів в оперативному перекладі різної комерційної, технічної або INTERNET-інформації, яка подана в електронному вигляді.

комп’ютер машинний переклад

2. Історія машинного перекладу

Історія машинного перекладу як науково-прикладного напряму почалася в кінці 40-х років минулого століття. У березні 1947 р. Уоррен Уївер, директор відділення природничих наук Рокфеллерівського фонду, в листуванні з Ендрю Бутом і Норбертом Вінером вперше сформулював концепцію машинного перекладу. Уївер писав : «У мене перед очима текст, написаний по-російськи, але я збираюся зробити вигляд, що насправді він написаний по-англійськи і закодований за допомогою певних дивних знаків. Все що мені потрібне — це зламати код щоб витягувати інформацію, ув’язнену в тексті ».

Ідеї Уївера лягли в основу підходу до МП, заснованого на концепції interlingva: стадія передачі інформації розділена на два етапи; на першому етапі текст перекладається на мову-посередник (створену на базі спрощеної англійської мови), а потім результат цього перекладу представляється засобами вихідної мови.

У 1947 році Е.Бут і Д. Бриттен розробили докладний "код" для послівного машинного перекладу. А в 1948 році Р. Річенс запропонував правила розбивки словоформ на основу й закінчення.

В ті роки комп’ютери сильно відрізнялися від сучасних. Це були дуже великі і дорогі машини, які займали цілі кімнати і вимагали для свого обслуговування великий штат інженерів, операторів і програмістів. В основному ці комп’ютери використовувалися для здійснення математичних розрахунків для потреб військових установ, а також математичних і фізичних факультетів, які були пов'язані з військовою сферою. Тому на ранніх етапах розробка МП активно підтримувалася військовими, при цьому в США основна увага приділялася російсько-англійському направленню, а в СРСР — англо-російському.

У 1952 році відбулася перша конференція в Масачусетському технологічному університеті, а в 1954 році в Нью-Йорку була представлена перша система МП — IBM Mark II, розроблена компанією IBM спільно з Джоржтаунським університетом (ця подія увійшла в історію як експеримент Джорджтауна). Була представлена програма, яка дуже обмежена в своїх можливостях (вона мала словник в 250 одиниць і 6 граматичних правил), що здійснювала переклад з російської мови на англійську. Здавалося, що створити системи якісного автоматичного перекладу цілком можливо в межах декількох років (при цьому акцент робився на розвитку повністю автоматичних систем, що забезпечують високоякісні переклади; участь людини на етапі постредагування розцінювалася як тимчасовий компроміс). Професійні перекладачі серйозно непокоїлись незабаром залишитися без роботи...

На початку 50-х років дослідницькі групи в США і в Європі працювали в області МП. Результати дуже скоро розчарували інвесторів. Однією з головних причин невисокої якості МП в ті роки були обмежені можливості апаратних засобів: малий об'єм пам'яті, неможливість повноцінного використання мов програмування на високому рівні, відсутність теоретичної бази, необхідної для вирішення лінгвістичних проблем (переклад текстів відбувався без синтаксичної цілісності). У 1959 році філософ Бар-Хіллел виступив з твердженням, що це не реально досягнути високоякісного автоматичного перекладу. Як приклад він привів проблему знаходження правильного перекладу для слова pen в наступному контексті: John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy. (Джон шукав свою іграшкову коробку. Нарешті він її знайшов. Коробка була в манежі. Джон був дуже щасливий.). Pen в даному випадку повинно перекладатись не як "ручка" , а як "дитячий манеж". Вибір того чи іншого перекладу в цьому випадку зумовлений знанням позамовної дійсності, а це знання дуже обширне, щоб вводити його в комп'ютер. Проте Бар-Хіллел не заперечував ідею МП як таку, вважав перспективним розроблення таких машинних систем, але в перекладі повинна безпосередньо брати участь людина.

Цей виступ негативно відбився на розвитку МП в США. У 1966 р. спеціально створена Національною Академією наук комісія ALPAC (Automatic Language Processing Advisory Committee), грунтуючись у тому числі і на висновках Бар-Хіллела, прийшла до висновку, що машинний переклад неприбутковий. Доповідь суттєво загальмувала розвиток МП у цілому.

З розвитком обчислювальної техніки в кінці 70-х років (поява мікрокомп'ютерів, розвиток мереж, збільшення ресурсів пам'яті) машинний переклад увійшов до епохи "Ренесансу". Але тепер досліджувалась участь людини на різних стадіях перекладу. Системи МП із "ворога" і " конкурента" професійного перекладача перетворюються на незамінного помічника, який сприяє заощадженню часу і людських ресурсів.

Про відродження МП у 70-80 роках свідчить ряд таких фактів. Комісія Європейських общин купує англо-французьку версію Systran, а також систему перекладу з російської на англійську (остання розвивалася після доповіді ALPAC і продовжувала використовуватися ВВС США і НАСА); крім того, CEC замовляє розробку франко-англійськой і італійсько англійської версій. У той час завдяки CEC були закладені основи проекту EUROTRA, заснованого на розробках груп SUSY і GETA. Одночасно відбувається швидке розширення діяльності по створенню систем МП в Японії (головним чином заснованих на технології interligva, розробленою Уївером в кінці 40-х рр.); у США Панамериканська організація охорони здоров'я замовляє розробку іспано-англійську версію (система SPANAM); ВВС США фінансують розробку системи МП в Лінгвістичному дослідницькому центрі при Техаському університеті в Остіні; група TAUM в Канаді досягає помітних успіхів в розробці своєї системи METEO (яка використовувалася в основному для перекладу метеорологічних досліджень). Цілий ряд проектів, започаткованих в 70-80 рр. згодом розвинулися в повноцінні комерційні системи.

За період 1978-93 рр в США на дослідження в області МП витрачено 20 мільйонів доларів, в Європі — 70 мільйонів, в Японії — 200 мільйонів.

Однією з нових розробок стала технологія ТМ (translation memory), що працює за принципом накопичення: в процесі перекладу зберігається початковий сегмент (пропозиція) і його переклад, внаслідок чого утворюється лінгвістична база даних; якщо ідентичний або подібний початковому сегмент виявляється в тексті, що знов перекладається, він відображується разом з перекладом і вказівкою збігу у відсотках. Потім перекладач приймає рішення (редагувати, відхилити або прийняти переклад), результат якого зберігається системою. А зрештою "не потрібно двічі перекладати одну і ту ж пропозицію!". В даний час розробником відомої комерційної системи, заснованої на технології ТМ, є система TRADOS (заснована в 1984 р.).

90-ті роки принесли з собою бурхливий розвиток ринку ПК (від настільних до кишенькових) і інформаційних технологій, широке використання мережі Інтернет (яка стає усе більш інтернаціональною). Все це зробило можливим подальший розвиток систем МП. Переваги використання машинного перекладу особливо очевидні при одержанні онлайнового перекладу.

3. Теорія машинного перекладу

Загальна теорія перекладу основується на даних, які отримані при спостереженні об’єкту в різних його проявах. Ці прояви можуть бути випадковими, нетиповими, і тоді вони не враховуються. Ці прояви можуть повторюватися, бути типовими, і тоді теорія вважає їх закономірними загалом або для того чи іншого виду перекладу зокрема. Довший час не викликала сумніву аксіома: міжмовні перетворення можуть здійснюватися лише людиною. Науково-технічний прогрес у ХХ столітті вніс в це положення суттєве уточнення: міжмовні перетворення може здійснювати як людина, так і машина. Так з’явилася перша основа для класифікації перекладу: механізми міжмовних перетворень. На цій основі було започатковано теорію машинного перекладу.

Теорія машинного перекладу пройшла бурхливий період розвитку. За свій короткий вік вона зробила в загальну теорію перекладу не менший вклад, ніж існуючі теорії перекладу, здійснюваного людиною. Не можна, звичайно, забувати, що машинний переклад ввібрав у себе всі досягнення «домашинного» перекладу. Вирішальне значення в успіхах теорії машинного перекладу мало те, що його автори змушені були моделювати діяльність перекладача і забезпечити функціонування механізмів перекладу. Саме в машинному перекладі була виявлена необхідність окремо розглядати аналіз вихідного тесту і синтез тексту перекладу. На жаль, на відміну від перекладу, здійснюваного людиною, машинний переклад практичними результатами похвалитися не може. Але, можливо, це тільки наслідок недостатніх технічних можливостей, і отримання високоякісного машинного перекладу – це питання часу.

Теорія машинного перекладу поки що існує як дещо єдине ціле. Це пов’язано із стабільністю умов процесу машинного перекладу і невисокою компетентністю машин. Інша справа людина. Їй не тільки доводиться працювати в різних умовах, їй доводиться в залежності від ситуації перекладати письмово або усно, сприймати оригінал очима або на слух. Важко не помітити різницю, що існує між письмовим перекладом, не обмеженим по часу, що дозволяє користуватися словниками і різноманітними довідниками, і усним перекладом, який виконується в екстремальних умовах функціонування розумових механізмів перекладача. Різниця, яка існує між письмовим і усним перекладом, пов’язана з умовами функціонування розумових механізмів, котрі, таким чином, є основою для класифікації перекладу, що здійснюється людиною.

4. Напрями розвитку машинного перекладу

Розрізняють два магістральні напрямки створення та застосування машинного перекладу. В першому випадку система машинного перекладу функціонує на великій ЕОМ і представляє “сирий”, чорновий переклад, який згодом редагують кваліфіковані перекладачі. Як правило, така методика використовується у великих організаціях, які змушені готувати документи на різних мовах. Деколи досить успішно використовується попереднє редагування вихідних текстів. Деякі фірми вводять у себе так звані “контрольовані природні мови”: коли кожен працівник фірми, що готує документацію, повинен її готувати саме з дотриманням вимог цієї обмеженої мови (наприклад, вимога відсутності складних синтаксичних конструкцій). Використання контрольованої природної мови спрощує роботу машинного перекладу і зменшує обсяги постредагування, яке дорого коштує через необхідність залучення спеціалістів високої кваліфікації. Другим магістральним напрямком машинного перекладу є використання систем, орієнтованих на персональні комп’ютери. Такі системи вперше появилися ще на початку 1980-х років (наприклад, MicroCAT фірми Weidner). Найбільший успіх в застосуванні цих систем перекладу припав на 90-і роки.

Зараз інтенсивно проводяться дослідження в галузі перекладу усної мови. Компанія IBM, яка вважається лідером в цій області, випустила програмний пакет ViaVoice 4.1, яка дозволяє комп’ютеру сприймати до 140 слів на хвилину неперервної диктовки. Попередні версії цієї програми передбачали лише дискретний спосіб введення мови. Для налагодження на персональні особливості мовлення певної людини при початковій установці програми в залежності від якості вимови і конкретного діалекту потрібно повторити від 104 до 256 речень, що попередньо задаються. Але розпізнавання мови не означає розуміння її змісту. Слід розрізняти розпізнавання машинної мови і перетворення її в текст або ж її використання у вигляді команд, і справжнє розуміння її змісту, як це робить людина. Останнє вимагає знання комп’ютером всього обсягу стилістичних та семантичних конструкцій, правил використання слів та висловів, при чому останні повинні задаватися не жорстко – адже люди говорять, що завгодно і як завгодно, не звертаючи уваги наскільки це грамотно і літературно. Іноді навіть те, що говориться, не відповідає тому, що малося на увазі. Так що навчити машину розуміти людей – задача незрівнянно складніша і віддалена в часі.

5. Приклади систем машиного перекладу

«Promt 98» — професійна 32-розрядна система для автоматичного перекладу документів, що успадкувала одночасна лінгвістичні й інтерфейсні розробки систем машинного перекладу сімейства Stylus. Ця система складається з кількох модулів, пов'язаних між собою так називаним інтегратором. У її склад входить модуль FileTranslator для пакетної обробки великої кількості документів. Для цього необхідно вказати системі файли, який потрібно перевести, і запустити процес перекладу. Можна формувати чергу перекладу чи безпосередньо в модулі File Translator, чи відправляти на переклад документи, відкриті в системі "Промт". Модуль WebView — броузер, що дозволяє одержати синхронний переклад HTML-сторінок при роботі в Internet. Модуль QTrans використовується для перекладу невеликих неформатованих текстів з буфера обміну чи введених із клавіатури. Promt 98 дозволяє виконати переклад для трьох мовних пар у двох напрямках: англійський – російський, німецький – російський і французький – російський. Система підтримує формати: .txt, .doc, .rtf, .wri, .htm; має інтуїтивно зрозумілий інтерфейс. Promt 98 вмонтована в Word і Excel, вона використовує системи перевірки правопису Lingvo Corrector, "Пропис", "Орфо" і ін.; надає можливість користувачу вибрати варіант перекладу серед кількох пропонованих.

Система "Плай" — універсальний російсько-український перекладач. Забезпечує миттєвий переклад будь-яких редакторських виправлень у вихідному тексті, має інтелектуальну технологію введення нових слів і словосполучень у словники користувача, забезпечує повне збереження структури документа, створення списку зарезервованих слів, а також представляє словники синонімів українських і російських слів, російсько український електронний словник і вмонтовані лінгвістичні модулі системи перевірки правопису і граматики "Рута".

Система Pars має такі особливості, як одночасна робота багатьох користувачів; проста процедура поповнення словників, що дає можливість настроїти систему на потрібні тематики. Система надає користувачам наступні можливості: задавати словники потрібної тематики і встановлювати їхню пріоритетність (це буває зручно, коли зустрічається слово, що має кілька перекладів, система вибирає те значення, що знаходиться в словнику з найбільшим пріоритетом); автоматично транслювати власні імена: наприклад, англійські чи прізвища назви міст можна просто записати російськими буквами; вибирати той варіант перекладу, що найбільше підходить по тексту з пропонованих системою варіантів; робота безпосередньо в WinWord зі збереженням формату вхідного тексту.

6. Висновок

Отже, машинний переклад значно дешевший і швидший від традиційного, хоч і поступається йому по якості. Ним користуються в тих випадках, коли важливіше зрозуміти зміст документу, ніж перекласти текст відповідно до літературних критеріїв. На мою думку, машинний переклад стане важливим інструментом для розвитку міждержавної торгівлі, тому що він спроможний значно спростити і прискорити одержання інформації про товари, що випускаються в інших країнах. Останнім часом в цій галузі досягнуто значних успіхів. У багатьох випадках навіть не дуже якісний переклад краще, чим відсутність якого-небудь. Там, де відсутність часу й доступність - більш важливі умови, ніж абсолютна стилістична точність, МП підтверджує свою неоціненну рентабельність.

Список використаної літератури:

1. Суханова М. Кто лучше переводит" // Мир ПК. – 1997. – N1.
2. Кулагина О. С. Исследования по машинному переводу. – М. – 1979.