

Комп'ютерний переклад

Сьогодні розвиток науки і техніки неможливий без обміну спеціальними даними, що з'являються в різних країнах світу у наукових періодичних виданнях, спеціальних бюлетенях, патентній літературі тощо. При цьому відбувається взаємопроникнення гуманітарних і природничих дисциплін, що обумовлено досягненнями кібернетики, інформатики та їх суттєвим впливом на розвиток інших наук.

Комп'ютеризація інформаційних технологій стимулювала розвиток нового напрямку прикладної лінгвістики – комп'ютерного, до якого входять засоби для управління даними в автоматизованих системах різних типів. У мовознавців з'явився новий засіб діяльності – комп'ютер, зі спеціальними словниками і граматиками, максимально формалізованими, написаними у вигляді алгоритмів (програм).

Сьогодні за допомогою комп'ютерних програм можна здійснювати чимало трудомісткої роботи – укладати словники різних типів: орфографічні, перекладацькі, синонімів, антонімів, омонімів, перекладати стандартизовані тексти, створювати реферати, здійснювати коректорську і частково редакторську роботу, підтримувати масштабні бази даних для найрізноманітніших дослідницьких проектів.

З точки зору лінгвістики, переклад – це відтворення письмового тексту чи усного вислову засобами іншої мови.

Процес перекладу (людиною) охоплює такі етапи:

- 1) зорове чи слухове сприймання матеріалу, усвідомлення його змісту;
- 2) аналіз даних мовою оригіналу і синтез іншою мовою;
- 3) відтворення змісту іншою мовою.

Оскільки переклад – це передавання змісту того, що було висловлено, то перекладаються не слова, граматичні конструкції чи інші засоби мови оригіналу, а думки, зміст оригіналу. Згідно з теорією перекладу, неперекладаваних матеріалів не існує – є складні для перекладу тексти. Під час перекладання можуть виникнути труднощі з відсутністю в цій мові готових відповідників для висловлення того, що вже було висловлено засобами мови оригіналу або ж недостатніми знанням: мови оригіналу; суті предмету; мови, якою перекладають (Рис. 1.).

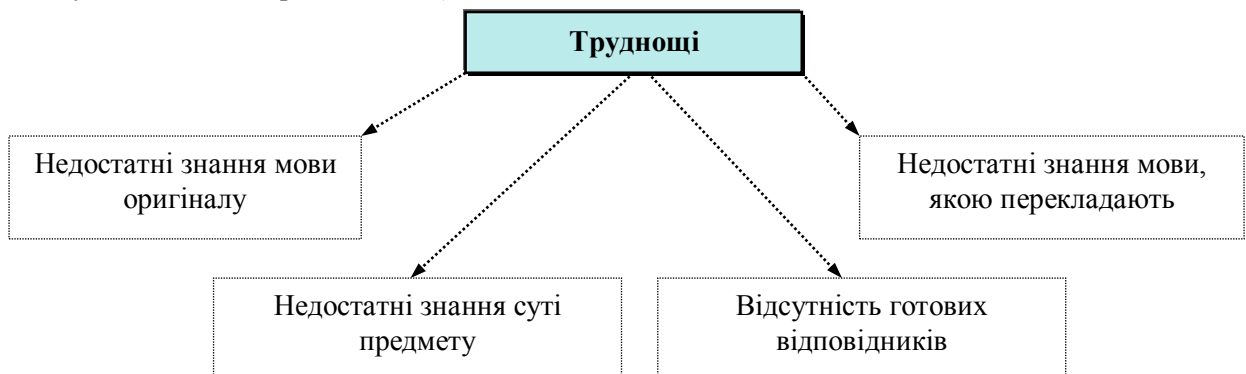


Рис. 1

Взагалі переклад тексту з однієї природної мови на іншу припускає розуміння тексту, і навіть більше – розуміння «картини світу», у якій даний текст з'явився. Майже завжди потрібні знання стосовно розпізнавання ознак тексту, що відносяться до сфери прагматики: зовнішнє оформлення (наприклад, це стаття, вірш чи документ встановленого зразка); галузь знання, до якого текст відноситься (розпізнавання термінології); зв'язаність частин тексту, що не завжди описується за допомогою синтаксичних чи лексико-семантичних критеріїв; і т.д. Для адекватного перекладу перекладач (людина) повинен:

- знати внутрішні структури мов, між якими здійснюється переклад;
- мати ясне уявлення про культуру, історію, мораль, переважні типи мислення народів, що є носіями мови;
- володіти якомога більшим словниковим запасом, більш-менш структурованим за галузями застосування слів (спеціальна термінологія, діалекти, ідіоматика, сленг);
- мати явний чи інтуїтивний тезаурус слів обох мов, тобто вміти запропонувати за даним словом семантичні функції від нього, такі як синонім, антонім, класичний атрибут;
- вміти запропонувати похідні частини мови від даного слова, якщо такі існують (добро – добрий – добріше – подобрів і т.п.).

Переклад поділяють на такі види:

- *За формою:*
 - ✓ усний;
 - ✓ письмовий;
 - ✓ **автоматизований (машинний).**
- *За способом:*
 - ✓ буквальний;
 - ✓ адекватний.
- *За змістом:*

- ✓ суспільно-політичний;
- ✓ художній;
- ✓ науково-технічний (технічний).

Проблема автоматичного перекладу з'явилася майже одночасно з появою перших ЕОМ і донині залишається актуальною. Інтерес до машинного перекладу постійно зростає, практично прямо пропорційно росту обсягів повідомлень і всеможливих матеріалів, які доводиться читати іноземними мовами. Різні описи, інструкції, електронна пошта, Інтернет-сайти, – все це потрібно перекладати якісно та швидко. Саме для вирішення цих завдань існують програми-перекладачі. Тут доречно відзначити, що часто виникає плутанина, і перекладачем називають будь-яку програму для перекладу слів. Разом з тим за допомогою системи машинного перекладу перекладається весь текст, на відміну від електронних словників, за якими отримується тільки переклад окремих слів.

Потреби в автоматичному перекладі щорічно збільшуються, тому актуальним сьогодні є пошук раціональних шляхів вирішення проблеми швидкого перекладу значних за обсягом матеріалів. Цю проблему часто розв'язують за допомогою автоматизованого (комп'ютеризованого, напівавтоматизованого) перекладу.

Ідея автоматизованого перекладу виникла ще в 1924 р., а 1933 року радянському інженерові П.Смирнову-Троянському було видано патент на машину для перекладання, за принципом зіставлення механічним способом відповідників з різних мов. Комп'ютеризований переклад у сучасному розумінні цього терміна вперше було здійснено 1954 року в Джорджтаунському університеті. Сьогодні створено багато експериментальних і практичних систем автоматичного перекладу, напр., системи SYSTRAN, LOGOS, ALPS, METAL, GETA, EUROTRA тощо, до яких входить понад 15 версій для різних пар мов.

У процесі перекладання за допомогою комп'ютера на різних мовних рівнях: розпізнаються графічні образи, проводиться морфологічний аналіз і переклад слів і словосполучень, аналізується синтаксис тексту (словосполучення і речення), проводяться семантичні (сміслові) перетворення, що забезпечує змістовну відповідність введеного й отриманого речення або тексту. Без розуміння призначення слова в тексті, без аналізу контексту (змісту) певного слова та аналізу речення не може бути ніякого перекладу. Перекладання тексту з однієї мови на іншу за допомогою комп'ютера є не простим завданням, оскільки вимагає не заміни слів однієї мови словами іншої, а відтворення змісту у повному обсязі, з усіма відтінками.

Сучасні програми комп'ютерного перекладу можна використовувати для комп'ютерної підтримки перекладів різноманітних матеріалів, проте перекладений текст слід перевірити, звернувши особливу увагу на переклад власних назв, термінів, мовних реалій (*гордіїв вузол*), слів у непрямому значенні та багатозначних, паронімів (порівняти російською і українською –*луна*), омонімів (*английський смог* може бути перекладено, напр., як *англійський зміг*), граматичних форм (*человек мог – людина міг*).

Про переклад за допомогою окремих програм уже написано чимало. Однак порівняємо машинні перекладачі щодо технологій. Сьогодні у світі існують дві технології машинного перекладу: традиційна, на основі правил (*rule-based machine translation*) і статистична (*statistical-based machine translation*). Необхідно визначити переваги та недоліки цих технологій:

Технологія статистичного перекладу. Слово «статистичний» у назві технології відразу наводить на думку про те, що для отримання перекладу в системі застосовуються математичні методи. Точніше, принцип роботи заснований на статистичному обчисленні ймовірності того, що фрази з вихідного тексту співпадут з фразами, що зберігаються в базі системи перекладу.

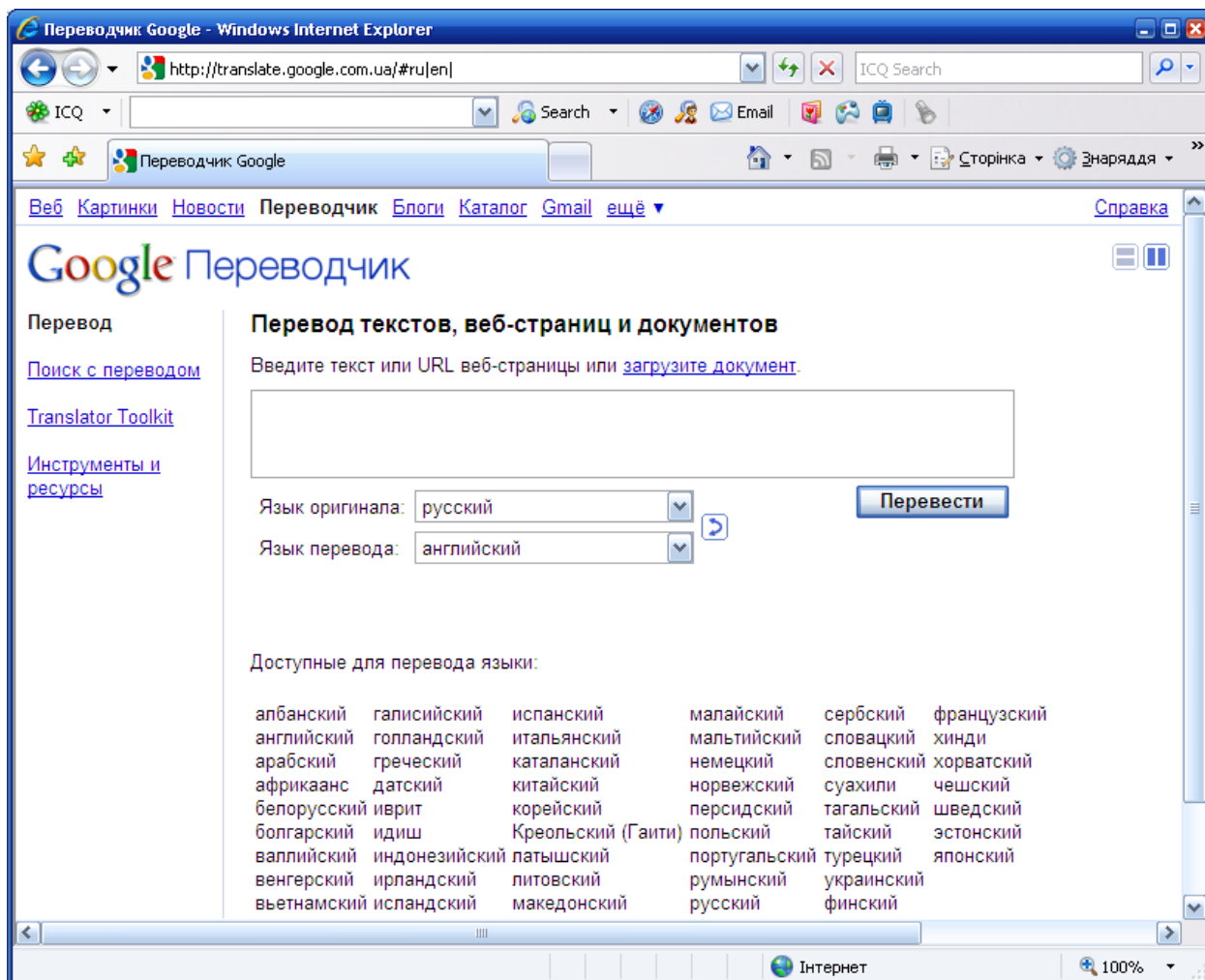


Рис. 2.

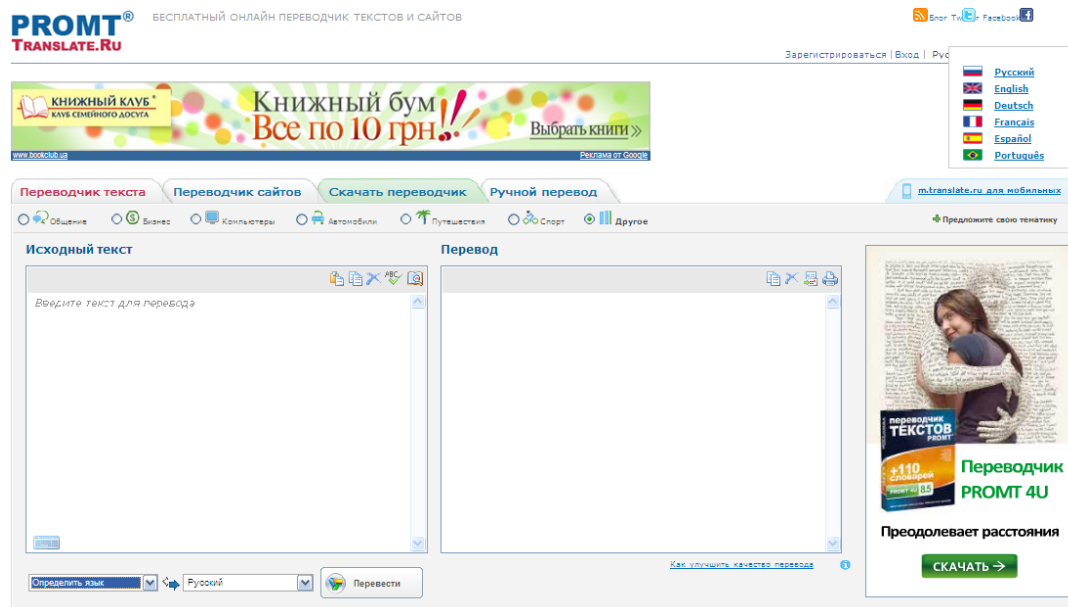


Рис. 3

Для правильной работы таких программ необходима великая база текстов вихідною мовою з їх перекладами (база паралельних текстів), яке за допомогою спеціальної програми аналізується за певними алгоритмами. Потім визначається найбільш імовірний набір слів вихідної мови, який вважається найбільш відповідним перекладові вхідного тексту.

Перевагою статистичної технології є те, що немає потреби в спеціальних лінгвістичних алгоритмах. Як заявляють розробники, при наявності набору статистичних алгоритмів і достатньої бази паралельних текстів можна зробити систему перекладу з будь-якої мови.

Недоліком статистичного перекладу є те, що в ньому не враховуються граматичні правила. Іноді при використанні «статистичного перекладача» замість зв'язного тексту можуть бути отримані ніяк не погоджені слова.

Друга проблема полягає в тому, що при використанні такої системи необхідна дуже велика кількість паралельних текстів. По суті, чим більша база паралельних текстів, тим більша ймовірність знайти необхідне співпадання. Очевидно, що практично єдиним (зате невичерпним) джерелом такої бази служить Інтернет. Саме тому в порівняльних тестах перекладу з арабського й китайського на англійський, проведених Національним інститутом науки й технологій США в 2005 році, кращим був визнаний онлайн-перекладач Google Translate (Рис. 2).

Google Translate – це сервіс компанії [Google](#), за допомогою якого можна автоматично перекладати слова, фрази та web-сторінки з однієї мови на іншу. В системі Google використовується власне програмне забезпечення для перекладу на основі [статистичного машинного перекладу](#). З вересня 2008 підтримуються й переклади українською мовою. Користувач вводить текст, поданий мовою оригіналу, та вказує мову, якою цей текст потрібно подати.

Під перекладом за правилами розуміють переклад з використанням лінгвістичних правил вхідної та вихідних мов. Ця технологія існує з початку 50-х років минулого століття й по праву може називатися традиційною. На зорі комп'ютерної ери не могло бути мови не тільки про бази паралельних текстів, але й про машинні ресурси, достатні для їх опрацювання. Тому протягом кількох десятиріч розвивалися тільки технології перекладу, засновані на лінгвістичних алгоритмах. На такій основі створені системи PROMT, Systran, Linguatex та ін. Розглянемо систему перекладу з іноземних мов на українську виробництва російської компанії PROMT (Рис. 3).

Спочатку за допомогою системи досліджується вхідний текст: виконується морфологічний аналіз слів у пропозиції (визначається рід, число, особа й інші морфологічні характеристики), потім синтаксичний аналіз (визначаються члени речення: підмет, присудок, додаток) і на заключному етапі виконується синтез речення у перекладеному тексті.

Перевагою такої системи є те, немає потреби в доступі до баз паралельних текстів. Крім того, систему можна налаштувати, що підвищує якість перекладу спеціалізованих текстів.

Що стосується програмних засобів для перекладу, які використовуються в Україні, то до найбільш поширених відноситься програма **Pragma**. Більшість урядових установ в Україні, великі індустріальні та фінансові компанії, загальноосвітні школи, а також малий бізнес та приватні особи як в Україні, так і в усьому світі є користувачами цієї програми.

Багаторічний досвід досліджень і розробок в галузі машинного перекладу втілено в останньому проекті Pragma версії 6.x.

Оцінити якість перекладу можна за матеріалами web-сайту (<http://www.trident.com.ua/uk/translation/on-line/>) на закладинці *Переклад*, відповідно обравши *Онлайн перекладач*. В десктоп-програмі Pragma 6.x використовується той самий алгоритм перекладу, але програма є платною.

The image shows a screenshot of the Trident Software website. At the top, there is a navigation menu with links: ГОЛОВНА, ПРОДУКТИ, ПЕРЕКЛАД, ПІДТРИМКА, ЗАМОВЛЕННЯ, ПАРТНЕРАМ, КОМПАНІЯ. Below the menu is a search bar with the text 'Пошук' and flags for USA, UKR, and RUS. The main content area is divided into two columns. The left column has a heading 'ВХІД' and a message: 'Для доступу до закритого розділу сайту вам треба авторизуватися.' Below this is a login form with fields for 'Ім'я користувача' and 'Пароль', a 'Запам'ятати мене' checkbox, and a 'Вхід' button. There are also links for 'Забули пароль?' and 'Забули ім'я?'. The right column has a heading 'ВХІД / РЕЄСТРАЦІЯ' and a registration form with fields for 'Ім'я користувача' and 'Пароль', a 'Запам'ятати мене' checkbox, and a 'Вхід' button. Below this are links for 'Забули свій пароль?', 'Забули своє ім'я користувача?', and 'Зареєструватися'. There is also a 'НОВИНКА' section with a heading 'Pragma 6.x!' and a link to 'Як вивчити іноземний?'. At the bottom, there are several advertisements: 'Ads by Google', 'Ukrainian Translator', 'Russian', 'Talk English', 'English Lesson', and 'Ads by Google'. The footer contains the text: 'Авторські права © 2009 ТОВ Трайдент Софтвар. Усі права захищені.'

Рис. 4

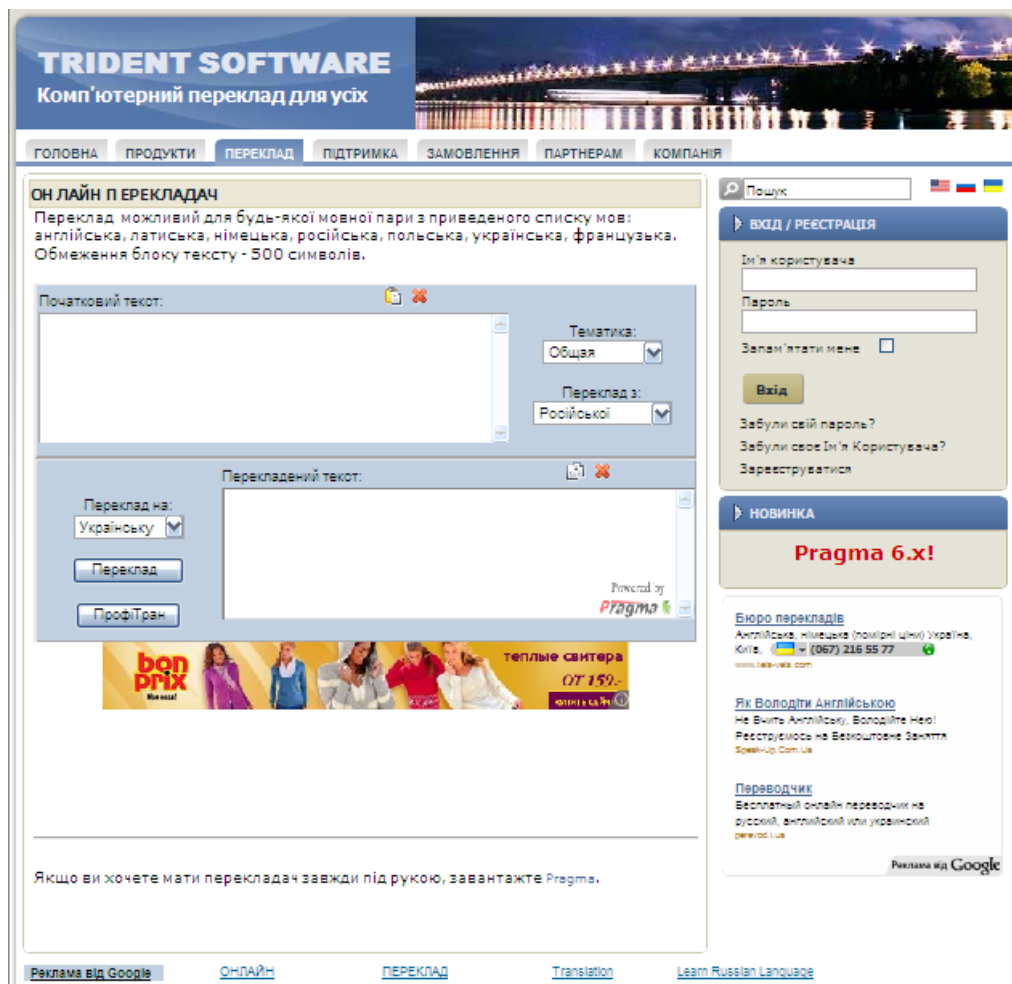


Рис. 5

Програму-перекладач можна встановити на веб-сайті, і до того ж безкоштовно. Для цього треба зареєструватися на сайті (<http://www.trident.com.ua>) вибравши закладку *Партнерам*→*Розробникам* (Рис. 4).

На даний момент за її допомогою можна здійснювати багатомовні машинні переклади з англійської, латиської, німецької, польської, російської, української, французької та казахської мов. Переклад з будь-якої мови на будь-яку це – 56 напрямів перекладу з обмеження блоку тексту – 500 символів (Рис. 5).

Недоліком програм, в яких використовується технологія *переклад за правилами*, можна назвати те, що для підвищення якості перекладу потрібно дуже ресурсоемна робота – адже для цього потрібно не тільки розробляти нові алгоритми, але й поліпшувати раніше створені.

Кожна із названих технологій має як свої переваги, так і недоліки. Для кращого перекладу можливо доцільне поєднання цих технологій. Разом з тим на сьогодні найкращими поки що залишаються напівавтоматичні переклади спеціально підготовленими фахівцями, які досконало володіють мовами, предметними знаннями у відповідній галузі, а також сучасними інформаційно-комунікаційними технологіями.

Література

1. Бондаренко В.В., Дубічинський В.В., Кухаренко В.М. Переклад науково-технічної літератури. – Харків, 2001.
2. Войналович О., Моргунюк В. Російсько-український словник наукової і технічної мови (термінологія процесових понять). – К., 1997.
3. Онлайн перекладач. Перекладач і онлайн переклад від Трайдент Софтвр. [Електронний ресурс]. – Режим доступу: <http://www.trident.com.ua/uk/translation/on-line/>
4. Прикладна лінгвістика, комп'ютерний переклад [Електронний ресурс]. – Режим доступу: <http://damar.ucoz.ru/publ/9-1-0-238>.
5. Російсько-український словник / За ред. В.В. Жайворонка. – К., 2003.
6. Федоров А.В. Основы общей теории перевода. – Москва, 1983.
7. Google Переводчик – Википедия. [Електронний ресурс]. – Режим доступу: http://ru.wikipedia.org/wiki/Google_Переводчик.
8. Online-переводчик текста компании ПРОМТ: английский, русский, немецкий, французский, испанский, итальянский и португальский языки. [Електронний ресурс]. – Режим доступу: <http://www.translate.ru/>.
9. Statistical machine translation – Wikipedia, the free encyclopedia. [Електронний ресурс]. – Режим доступу: http://en.wikipedia.org/wiki/Statistical_machine_translation.

