



СЕГМЕНТАЦІЯ ТЕКСТУ ПЕРЕКЛАДУ У COMPUTER-AIDED TRANSLATION СИСТЕМАХ

Ільїнський Б. В.

*Національний технічний університет
«Харківський політехнічний інститут»
м. Харків, вул. Пушкінська, 79/2, тел. 707-63-60.
e-mail: bogdan.ilinski@gmail.com*

За останні роки характер роботи перекладача і вимоги до результатів його роботи істотно змінилися. У першу чергу зміни торкнулися перекладу науково-технічної, офіційної та ділової документації. У наш час вже недостатньо просто перекласти текст, користуючись комп'ютером як друкарською машинкою. Замовник очікує від перекладача, що оформлення готового документа буде відповідати зовнішньому вигляду оригіналу настільки точно, наскільки це можливо, при цьому задовольняти прийнятим у даній країні стандартам. Терміни виконання таких замовлень зменшуються, тому перекладачеві, який не користується допоміжними засобами автоматичної обробки мовної інформації, важко вкладатися в установлені часові рамки. Цих жорстких, часто суперечливих умов можливо дотримуватись лише у тому випадку, якщо перекладач не тільки досконало володіє рідною та іноземною мовами та глибоко вивчив обрану ним предметну область, але й впевнено орієнтується в сучасних комп'ютерних технологіях [3].

Обсяг перекладів, пов'язаних з інформаційними технологіями, також стрімко зростає, причому перекладацьким та комп'ютерним компаніям доводиться мати справу не лише з підготовкою документації, але і з локалізацією програмного забезпечення, тобто з перекладом ресурсів, що містяться в exe- і dll-файлах. А також з наступним тестуванням програмного забезпечення.

Ключовою для перекладача текстів технічної направленості є технологія Translation Memory (ТМ – пам'ять перекладів), що широко використовується в системах автоматизованого перекладу (Computer-Aided Translation або CAT). Крім прискорення процесу перекладу повторюваних фрагментів і змін, внесених до вже перекладених текстів (наприклад, нових версій програмних продуктів або змін у законодавстві), системи пам'ять перекладів також забезпечують однаковість перекладу термінології в однакових фрагментах, що особливо важливо при технічному перекладі [4].

Найбільш розвиненими та розповсюдженими CAT-системи вважаються наступні: SDL Традос, Deja Vu, Wordfast, (що не є повноцінною CAT-системою, а представляється у якості набудови до MS Word). Все ці системи автоматизованого перекладу ґрунтуються на пам'яті перекладів, тобто базі даних (БД), що містить набір фраз із раніше перекладених текстів [1]. У кожній конкретній системі пам'яті перекладів дані зберігаються у своєму власному форматі (текстовий формат у Wordfast, база даних Access в Deja Vu і т.ін), але



існує міжнародний стандарт TMX (англ. Translation Memory eXchange format), що заснований на XML і з яким можуть працювати практично всі системи пам'яті перекладів.

Одним з основних принципів роботи CAT-систем є сегментація тексту тобто лінійне членування мовного потоку на складові відрізки – сегменти, співвідносні з певними одиницями мови: значущими реченнями, словосполученнями, словами, морфемами. Один запис у БД автоматизованих систем перекладу відповідає «одиниці перекладу», за яку зазвичай приймається одне речення. Якщо чергове речення вихідного тексту в точності збігається з реченням, що зберігається в базі, воно може бути автоматично підставлено в переклад. Якщо нове речення чимось відрізняється від того, що зберігається в базі, відповідність буде неповна та перекладачу надається можливість внести зміни у запропонований переклад.

Практично всі системи автоматизованого перекладу членують текст на речення, бо саме речення вважається найбільш оптимальною складовою для перекладу тексту. Але речення, на наш погляд, є досить крупною одиницею тексту. Повний збіг подібних сегментів можливий лише у формальних елементах оформлення документа (шапках, колонтитулах і т.п.) [2]. Тому метою нашої роботи стала розробка програмного продукту, який би сегментував текст, беручи за одиницю членування словосполучення, що з одного боку, безумовно є меншою за складом одиницею, ніж речення, з іншого – несе у собі певне граматичне та семантичне значення.

У якості матеріалу дослідження з усіх видів словосполучень розглядаються дієслівні словосполучення у перекладі з англійської мови на російську.

Одним з найбільш розповсюджених типів дієслівних словосполучень в англійській мові є Verb Patterns (VP), які будуються за наступними моделями [5]:

- Дієслово + інфінітив (V + to + V);
- Дієслово + ing-форма (V + V + ing);
- Дієслово + інфінітив або ing-форма;
- Дієслово + об'єкт + інфінітив (без to);
- Дієслово + об'єкт + ing-форма(герундій).

Результатом дослідження є алгоритм роботи програмного забезпечення, що виконує членування тексту на словосполучення, та пропонує варіанти перекладу виділених словосполучень. У перспективі подальшого дослідження буде розглянуто інші типи словосполучень та мовні пари.

Список літератури

1. *John Hutchins*. The origins of the translator's workstation. – NY: Kluwer Academic Publishers, 1998. – 287 – 307 p.
2. *Алимов В.В.* Теория перевода. Перевод в сфере профессиональной коммуникации. – М.: Едиториал УРСС, 2005. – 80 с.
3. *Бархударов Л.С.* Язык и перевод. Вопросы общей и частной теории перевода. – М.: Международные отношения, 1975. – 42 с.
4. *Гарбовский Н.К.* Теория перевода. – М.: Изд-во Моск. Ун-та, 2004. – 56 с.
5. *Казакова Т.А.* Практические основы перевода. English-Russian. – СПб.: Лениздат, 2003. – 35 с.