

Міністерство освіти і науки України  
Національний педагогічний університет імені М.П. Драгоманова

СТАХМИЧ ЮЛІЯ СТАНІСЛАВІВНА

УДК (811.111'322+811.161.2'322+811.161.1'322):004.934

КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ ПРИРОДНОЇ МОВИ  
(на матеріалі систем програмного забезпечення перекладу)

10.02.15 – загальне мовознавство

Автореферат дисертації на здобуття наукового ступеня  
кандидата філологічних наук

Київ – 2015

Дисертацією є рукопис

Робота виконана у Національному педагогічному університеті імені М.П. Драгоманова, Міністерство освіти і науки України

**Науковий керівник** доктор філологічних наук, професор  
**Денисова Світлана Павлівна**,  
Київський національний лінгвістичний університет,  
професор кафедри загального і порівняльного  
мовознавства та новогрецької філології

**Офіційні опоненти:** доктор філологічних наук, професор  
**Компанцева Лариса Феліксівна**,  
Національна академія Служби безпеки України,  
завідувач кафедри теорії та практики перекладу

кандидат філологічних наук, доцент  
**Вусик Ганна Леонідівна**,  
Бердянський державний педагогічний університет,  
доцент кафедри загального мовознавства  
та слов'янської філології

Захист відбудеться “03” березня 2015 р. о 13.00 годині на засіданні спеціалізованої вченої ради К 26.053.15 у Національному педагогічному університеті імені М.П. Драгоманова за адресою: 01601, м. Київ-30, вул. Пирогова, 9

З дисертацією можна ознайомитись у бібліотеці Національного педагогічного університету імені М.П. Драгоманова за адресою: 01601, м. Київ-30, вул. Пирогова, 9

Автореферат розісланий “02” лютого 2015 р.

Учений секретар  
спеціалізованої вченої ради

Т. С. Толчеєва

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Дисертацію присвячено вивченню комп'ютерного моделювання природних мов за допомогою систем програмного забезпечення англо-українського й англо-російського перекладу. У роботі проаналізовано сучасні системи програмного забезпечення перекладу, розроблено методику оцінки якості їх лінгвістичної складової, виявлено типові мовні помилки у текстах машинних перекладів, визначено можливі причини їхнього виникнення та способи усунення.

У сучасному наукознавстві природна мова є міждисциплінарним об'єктом. Автоматичну обробку природної мови досліджували з позицій *кібернетики* (А. Тьюринг), *програмування* (А. Бут, Ю. Моторін), *математики* (У. Вівер, Д. Панов), *філософії* (І. Бар-Хіллель) тощо.

У лінгвістиці природна мова активно інтерпретується представниками сфери прикладного мовознавства, зокрема комп'ютерної лінгвістики. У цьому аспекті у центрі уваги вчених перебуває аналіз структури і функціонування мовних одиниць як результатів інтелектуальної діяльності людини (А. Баранов, Є. Карпіловська, М. Кочерган), а також їх моделювання у системах комп'ютерної обробки мови (Ю. Марчук, Л. Нелюбін, З. Партико, Р. Піотровський).

Одним зі способів комп'ютерного моделювання природної мови є системи програмного забезпечення перекладу – комп'ютерні програми, які використовуються для перекладу інформації з однієї природної мови іншою. Системи програмного забезпечення перекладу є інженерно-лінгвістичними моделями – комплексом, який охоплює лінгвістичну модель мови й відповідне програмне забезпечення.

У мовознавстві під системами машинного перекладу розуміють такий тип програмного забезпечення, який здійснює обробку та переклад тексту з однієї природної мови іншою з максимальним збереженням змісту та структури оригіналу (Т. Коваль, М. Мамедова, Л. Нелюбін, J. Hutchins). Проте робота алгоритму системи перекладу не підлягає прямому лінгвістичному спостереженню, тому у прикладному мовознавстві предметом аналізу стають машинні переклади текстів.

Машинний переклад залишається одним із основних напрямів теорії штучного інтелекту (В. Бурдаєв, Г. Гайна, О. Гладченко, В. Доля, І. Зубова, О. Зубов, Є. Карпіловська, Ю. Рождественський, О. Спірін). Сучасні системи програмного забезпечення перекладу здійснюють переклад тексту з урахуванням морфологічних, синтаксичних та семантичних даних. Утім, якість перекладу не є високою, що зумовлено недосконалістю лінгвістичного та програмного забезпечення систем унаслідок неможливості повної формалізації мови та процесів мислення людини.

Дослідження якості машинного перекладу базується на різних методах і способах її оцінки. Так, представлено *суб'єктивну оцінку людини й анкетування інформантів* (А. Бірюков, Ю. Марчук, D. Arnold, M. Glenn, J. Lopez, D. Vilar), *зворотний переклад* (N. Bach, M. Dillinger, P. Koehn, S. Soderland), *зіставлення тексту перекладу з текстом оригіналу* (Л. Бархударов, Ю. Ванников, Б. Виноградов, В. Комісаров, О. Швейцер), *застосування автоматичних метрик оцінки якості машинних перекладів* (В. Яковина, J. Olive, M. Snover, L. Specia, C. Tillmann) та *систем кількісної оцінки якості перекладу* (G. Christian, A. Gerasimov, G. Saldanha, A

. Zerfass).

Проте, на сьогодні не розроблено єдиної методики оцінки якості систем програмного забезпечення перекладу, яка б пояснювала лінгвістичну природу помилок у машинних перекладах та можливі причини їхнього виникнення, що становить окреме лінгвістичне завдання.

**Актуальність** дисертаційної роботи зумовлена загальними тенденціями теоретичної і прикладної лінгвістики до вивчення структур представлення знань людини, операцій мислення та свідомості з метою їх моделювання у створених формальних системах. Системи програмного забезпечення перекладу є одним із напрямів розвитку теорії штучного інтелекту, центральним завданням якої є побудова автомата, здатного виконувати функції, притаманні людському інтелекту, до яких належить переклад як особливий вид когнітивно-креативної, а відтак, і мисленнево-мовленневої діяльності людини. Проте створення оптимальної системи програмного забезпечення перекладу, якість роботи якої відповідала б потребам користувачів і перекладачів, залишається і донині нерозв'язаним завданням.

**Зв'язок роботи з науковими програмами, планами, темами.** Дисертацію виконано відповідно до тематичного плану науково-дослідних робіт Національного педагогічного університету імені М.П. Драгоманова за напрямом “Дослідження проблем гуманітарних наук”. Дисертаційна робота є складовою частиною наукової теми кафедри загального мовознавства і германістики Інституту іноземної філології Національного педагогічного університету імені М.П. Драгоманова “Зіставно-типологічне вивчення мов у синхронії і діахронії” (тему дисертації затверджено на засіданні вченої ради НПУ імені М.П. Драгоманова, протокол № 6 від 25 січня 2012 р., перезатверджено на засіданні спеціалізованої вченої ради К 26.053.15 – протокол № 9.2 від 15 грудня 2014 року).

**Метою** дисертаційної роботи є лінгвістична характеристика комп'ютерного моделювання природної мови у системах програмного забезпечення перекладу.

Досягнення поставленої мети передбачає реалізацію таких **завдань**:

- сформулювати теоретичні засади вивчення комп'ютерного моделювання природної мови у сучасній лінгвістиці;
- розробити методику дослідження комп'ютерного моделювання природної мови у системах програмного забезпечення перекладу;
- схарактеризувати основні системи програмного забезпечення перекладу;
- виявити типові помилки машинного перекладу як наслідки некоректного моделювання англійської, української та російської мов у системах програмного забезпечення перекладу, встановити причини їх виникнення та можливі шляхи усунення;
- зіставити якість роботи різних систем програмного забезпечення перекладу для обрання найефективнішої програми для роботи з англо-українською та англо-російською мовними парами.

**Об'єктом дослідження** є моделі природної мови у системах програмного забезпечення перекладу.

**Предмет дослідження** становлять результати моделювання англійської, української та російської мов у системах програмного забезпечення перекладу Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex.

**Фактичним матеріалом дисертації** є вилучені з мережі інтернет 44 оригінальні тексти англійською мовою та їхні змодельовані машинні переклади українською і російською мовами, які належать до таких тематичних галузей, як авіація, автомобільна, банківська справа, Біблія, бізнес, біологія, будівництво, військова, гірнична, деревообробна, екологія, економіка, електроніка, енергетика, загальна, залізниця, зв'язок, інструкція, інтернет, інформатика, кіно, комерція, косметика, космонавтика, кулінарія, логістика, математика, машинобудування, медицина, металургійна, морська, музика, нафтогаз, подорож, поліграфія, право, секс, сільське господарство, спорт, текстиль, техніка, фізика, харчова, хімія, з яких було відібрано 11 перших речень із їхніми заголовками. Усього проаналізовано 1804 пари речень, серед яких 484 пари – в англо-українському перекладі для оцінки якості системи Pragma 6.3 Business та по 660 пар – в англо-українському та англо-російському перекладах для зіставлення ефективності роботи Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex.

**Методи дослідження.** Серед методів і прийомів дослідження – загальнонаукові (*описовий метод* – для класифікації та інтерпретації структурних і функціональних властивостей досліджуваних одиниць; *дедуктивно-індуктивний метод* – для аналізу і синтезу наукових теорій і концепцій; *суцільної вибірки*, за допомогою якого було зібрано фактичний матеріал дослідження; *математичних підрахунків* – для встановлення відсоткового співвідношення помилок та визначення найбільш проблемних мовних рівнів для систем програмного забезпечення перекладу) і спеціальні лінгвістичні (*структурний метод*, зокрема його методика аналізу за безпосередніми складниками, застосована для дослідження структури речень; методика *дистрибутивного аналізу* та методика *аналізу словникових дефініцій* – для виявлення оточення слів та їх визначень у словниках і подальшого розмежування багатозначних слів і лексичних омонімів; *зіставний метод* спрямовано на встановлення спільних і відмінних рис між текстами оригіналу і машинного перекладу з метою виявлення помилок на різних мовних рівнях (лексичному, морфологічному, синтаксичному).

**Наукова новизна** одержаних результатів полягає у тому, що в роботі *уперше представлено* результати комп'ютерного моделювання англійської, української та російської мов у системі машинного перекладу на основі правил (Pragma 6.3 Business), системах статистичного машинного перекладу (Google Translate, Babylon, Free Translation, Yandex) та гібридній системі машинного перекладу (Bing Translator); *удосконалено* методику оцінки якості систем програмного забезпечення перекладу на основі розроблених критеріїв виявлення лексичних, граматичних (морфологічних, синтаксичних) та лексико-граматичних помилок, які порушують норми української та російської мов і впливають на еквівалентність та адекватність машинного перекладу; *уведено* поняття лексико-граматичних помилок, які полягають у вживанні словосполучень у невласливому для контексту та тематики значенні, а також відхиленні від нормативно-узуального порядку слів, порушенні логіко-синтаксичного зв'язку між словами через неправильний вибір варіантів перекладу та граматичної трансформації перестановки; *розкрито* синкретичну природу виявлених лексичних, синтаксичних і лексико-граматичних помилок, яка полягає в порушенні однією ненормативною одиницею різних мовних норм; *укладено* тестові

збірки речень англійською мовою, які дають змогу перевіряти ефективність систем програмного забезпечення перекладу при роботі з текстами різного тематичного спрямування; на основі зіставного лінгвістичного аналізу машинних перекладів, виконаних Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex, *визнано* перевагу Yandex над іншими системами для англо-української та англо-російської мовних пар у кількості помилок (571 помилка та 585 помилок відповідно); у результаті оцінки якості систем програмного забезпечення перекладу *визначено* недоліки їхнього лінгвістичного та програмного забезпечення, які полягають у неповноті автоматичних словників та моделей мови, а також недосконалої роботи алгоритмів систем. *Установлено* домінування граматичних помилок над лексичними у машинних перекладах систем статистичного машинного перекладу (Google Translate – 59,3%, Free Translation – 64,0%, Yandex – 66,7%) та гібридній системі машинного перекладу Bing Translator (58,7%), а також лексичних помилок над граматичними у системі машинного перекладу на основі правил Pragma 6.3 Business (53,5%) та системі статистичного машинного перекладу Babylon (56,8%).

**Практичне значення** одержаних результатів полягає в можливості їхнього застосування у нормативних курсах із теоретичного мовознавства (розділи “Семантика”, “Морфологія”, “Синтаксис”), зі вступу до мовознавства (розділи “Лексика і фразеологія”, “Граматики”), загального мовознавства (розділи “Мова і мислення”, “Структура мови”, “Грамотична система мови”, “Лексико-семантична структура мови”), прикладної лінгвістики (“Комп’ютерне опрацювання природної мови”), теорії та практики перекладу (розділи “Письмовий переклад”, “Теорія та практика перекладу з української мови на англійську”, “Сучасні інформаційні технології у перекладі”), сучасної української літературної мови (розділи “Лексика і фразеологія”, “Граматики”). Положення та результати роботи можуть бути використані розробниками систем програмного забезпечення перекладу для підвищення загальної якості роботи програм, а також користувачами при проведенні первинної оцінки якості та виборі найбільш ефективної системи в англо-українському та англо-російському напрямках перекладу.

**Апробація результатів дослідження.** Основні положення дисертаційної роботи висвітлено у доповідях на *чотирьох* міжнародних наукових конференціях: “Мови і світ: дослідження та викладання” (Кіровоград, 2012, 2013), “Україна і світ: діалог мов та культур” (Київ, 2013), “Іноземна філологія у ХХІ столітті” (Запоріжжя, 2014); на *одній* всеукраїнській науковій конференції “Сучасний стан і перспективи лінгвістичних досліджень та проблеми перекладу” (Житомир, 2012); на *трьох* щорічних звітно-наукових конференціях викладачів, аспірантів, докторантів НПУ ім. М.П. Драгоманова “Єдність навчання і наукових досліджень – головний принцип університету” (Київ, 2012–2014).

**Публікації.** Проблематику, теоретичні й практичні результати дисертаційної роботи викладено у *семи* публікаціях: у *чотирьох* статтях, опублікованих у фахових наукових виданнях України, у *двох* статтях – у виданнях інших держав, а також тезах доповідей *однієї* міжнародної наукової конференції.

**Обсяг і структура роботи.** Дисертація складається зі вступу, чотирьох розділів, висновків, списку використаної літератури (217 найменувань, з яких 87 – іноземними мовами), списку лексикографічних та довідкових джерел (12 позицій),

списку джерел ілюстративного матеріалу (44 найменування) і додатків (9). Повний обсяг дисертації – 240 сторінок, основний зміст викладено на 197 сторінках.

## ОСНОВНИЙ ЗМІСТ

У **вступі** обґрунтовано актуальність теми дослідження, сформульовано мету, визначено завдання, об'єкт, предмет, описано матеріал і методи дослідження, наукову новизну одержаних результатів, окреслено практичне значення праці, подано відомості про апробацію основних положень дисертації, а також структуру та обсяг роботи.

У **першому розділі “Теоретичні засади вивчення комп’ютерного моделювання природної мови у сучасній лінгвістиці”** представлено ступінь розробки проблеми та викладено теоретичні положення щодо ключових понять дисертації, зокрема: схарактеризовано комп’ютерну лінгвістику з позицій прикладного мовознавства, описано основи автоматичної обробки природної мови, визначено особливості процесу моделювання у комп’ютерній лінгвістиці, проаналізовано системи програмного забезпечення перекладу в аспекті теорії штучного інтелекту.

Комп’ютерна лінгвістика як напрям прикладної лінгвістики на сучасному етапі її розвитку (А. Баранов, Є. Карпіловська, З. Партико, Р. Піотровський, І. Штерн) досліджує мову, знання та психічні процеси з метою моделювання мовної і комунікативної компетенції людини у системах автоматичної обробки мови, до яких належать системи машинного перекладу, системи автоматичного індексування і рубрикації, системи автоматичного анотування і реферування, інформаційно-пошукові системи, системи генерування текстів тощо.

Спільним для комп’ютерної лінгвістики та теорії штучного інтелекту є кінцева мета дослідження, що полягає у розробці комп’ютерних моделей окремих інтелектуальних функцій людини, до яких належить володіння мовою. Для вивчення об’єктів і зв’язків, прихованих від прямого спостереження, застосовують метод моделювання (О. Большакова, Є. Карпіловська, Е. Клишинський, Д. Ланде, Р. Піотровський, О. Селіванова), відповідно до якого досліджується не сам об’єкт, а його формалізований і спрощений аналог – модель. При цьому об’єктом моделювання виступає не мова, а лінгвістичні властивості текстів, опис яких організується як модель мови.

У комп’ютерній лінгвістиці розроблено відтворювальні інженерно-лінгвістичні моделі – штучно створені формальні системи, побудова та поведінка яких імітує мікроструктуру чи функціонування лінгвістичних об’єктів та дозволяє їхнє часткове відтворення (Р. Піотровський). Сучасні комп’ютерні моделі мови у системах автоматичної обробки інформації охоплюють структурну, редуковану статистичну та структурно-статистичну моделі (О. Большакова, Д. Ланде), у яких представлені різноманітні підходи до формального опису морфології, синтаксису та семантики у процесі аналізу та синтезу текстів (В. Касевич, Дж. Люгер, І. Bolshakov, Т. Pedersen, С. Samuelsson).

Комп’ютерне моделювання перекладу пов’язане з тим, що природна мова замінюється машинною мовою – штучною знаковою системою, яка функціонує у комп’ютері й моделює у процесі аналізу і синтезу тексту когнітивні процеси, розуміння, інтерпретацію і типи мислення людини у системах програмного

забезпечення перекладу. Системи програмного забезпечення перекладу – це комп'ютерні програми, які використовуються для перекладу тексту з однієї природної мови іншою. У роботі детально досліджено системи машинного перекладу – тип програмного забезпечення, який здійснює обробку та переклад тексту з однієї природної мови іншою з максимальним збереженням змісту та структури оригіналу.

Вивчення мисленнєвих процесів людини є складним завданням, адже потребує дослідження когніції перекладача як прояву розумових, інтелектуальних здібностей, необхідних для обробки і трансформації інформації з одного виду в інший (Ю. Зайченко, О. Кубрякова). Функціонування систем машинного перекладу відображає спробу формалізації мисленнєвих операцій перекладача.

Людський і машинний переклад складаються з кількох послідовних процедур: розуміння, обробки інформації та генерування готового продукту мовою перекладу (В. Гусєв, Л. Іомдін). Кожний тип систем програмного забезпечення перекладу характеризується певною семантичною силою, тобто має “здатність до смислоототожнення, що визначається багатством лексичної і логіко-семантичної системи, і смислорозрізнення, що визначається складністю граматичної системи процесора” (Е. Корольов) і залежить від повноти словників, ступеня розробки процедур аналізу та синтезу.

У процесі роботи системи моделюють механізми розуміння та створення текстів людиною. Залежно від застосування в роботі системи різних засобів аналізу, розуміння тексту може бути морфологічним, синтаксичним, семантичним, гіперсинтаксичним і комбінованим (М. Блехман). У системах машинного перекладу моделюються властиві людині типи мислення (О. Шаміс), серед яких перцептивне, когнітивне і практичне (поведінкове).

Система програмного забезпечення перекладу розглядається як інженерно-лінгвістична модель – *комплекс, який охоплює лінгвістичну модель мови та відповідне інженерне забезпечення (алгоритми і програми обробки на ЕОМ)* (Л. Нелюбін). Система програмного забезпечення перекладу містить предметний і динамічний компоненти (Ю. Шемакін), які відповідають декларативним і процедурним знанням людини (В. Бурдаєв). Моделювання перекладу розвивається не тільки шляхом уточнення чи розширення знань про мову або текст, а й через ускладнення комплексу процедурних складових системи. Це проявляється у розробці та вдосконаленні сучасних систем машинного перекладу на основі правил, статистичних і гібридних систем машинного перекладу (Р. Ісламов, Ю. Марчук, А. Міщенко, J. Hutchins, S. Nirenburg, S. Tripathi, A. Trujillo). У кожній програмі інтегрована власна інженерно-лінгвістична модель, прихована від прямого спостереження пересічного користувача. Саме тому необхідною є оцінка якості виконаних машинних перекладів для того, щоб визначити переваги й недоліки системи.

У другому розділі **“Методика дослідження комп'ютерного моделювання природної мови у системах програмного забезпечення перекладу”** розроблено методику дослідження комп'ютерного моделювання природної мови, а також описано етапи процедури застосування відповідних методів і прийомів аналізу.



У дисертації розроблено методику оцінки якості систем програмного забезпечення перекладу, що ґрунтується на підходах до оцінки якості машинних перекладів А. Бірюкова, М. Блехмана, Ю. Марчука, В. Новікова, П. Хроменкова, а також типології помилок у машинних перекладах (А. Бірюков, В. Новіков, П. Хроменков) та мовних помилок (Т. Бондаренко, В. Мельничайко), які порушують норми української мови. Специфіка внутрішньої організації та функціонування систем машинного перекладу закриті для доступу користувача. Саме тому оцінка якості передбачає випробування лінгвістичної потужності систем шляхом перевірки текстів перекладів як результатів інтерпретаційної діяльності програми на наявність помилок, виявлення причин їх виникнення та можливостей усунення. Це дозволить перевірити повноту моделі мови, закладеної у систему, та ефективність роботи програмного забезпечення. Для дослідження було обрано системи Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex, які застосовують різні підходи до моделювання природних мов в англо-українському й англо-російському напрямках перекладу і не були ще предметом зіставного вивчення у зазначених мовних парах.

Методика дослідження складається з чотирьох етапів. Перший етап передбачав підбір фактичного матеріалу дослідження за допомогою методу суцільної вибірки.

На другому етапі проаналізовано машинні переклади 44 текстів системи Pragma 6.3 Business. У результаті цього із матеріалу дослідження системи Pragma 6.3 Business було відібрано десять найбільш проблемних для перекладу тематик, які слугували матеріалом для подальшого вивчення систем Google Translate, Bing Translator, Babylon, Free Translation та Yandex. На цьому етапі було використано *структурний метод* із залученням елементів *методики аналізу за безпосередніми складниками* на синтаксичному рівні, *методики дистрибутивного аналізу* та *методики аналізу словникових дефініцій* на лексичному рівні, відповідно до яких на основі оточення слів та їх словникових визначень було розмежовано значення багатозначних слів і лексичних омонімів. *Зіставний метод* спрямовувався на виявлення спільних і відмінних рис між текстами оригіналу й машинного перекладу з метою виявлення помилок на різних мовних рівнях: лексичному, морфологічному, синтаксичному.

За допомогою розроблених критеріїв оцінки якості та *методу математичних підрахунків* було обчислено відсоткове співвідношення помилок і визначено найбільш проблемні мовні рівні для систем програмного забезпечення перекладу.

Третій етап пов'язаний з оцінкою якості роботи систем статистичного машинного перекладу Google Translate, Babylon, Free Translation, Yandex та гібридної системи машинного перекладу Bing Translator в англо-українському напрямку на матеріалі десяти текстів, у машинних перекладах яких, виконаних Pragma 6.3 Business, було виявлено найбільшу кількість помилок.

Четвертий етап був спрямований на зіставлення ефективності роботи програм Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex в англо-російському напрямку перекладу з метою визначення найкращої з них.

На третьому та четвертому етапах застосовано *метод зіставлення* машинних перекладів, відсоткових співвідношень та кількісних даних про помилки з метою

визначення найбільш ефективної системи машинного перекладу серед систем Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex.

У третьому розділі **“Системи програмного забезпечення перекладу як досвід моделювання природної мови”** подано характеристику систем програмного забезпечення перекладу Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex.

Pragma – це система машинного перекладу на основі правил, яка є багатомовною програмою і виконує переклади для 8 мов. Система має 56 напрямів перекладу і підтримує 44 тематики. Найновішою версією програми є Pragma 6.3 Business. Робота програми ґрунтується на численних інтегрованих лінгвістичних правилах вхідної та вихідної мов, а також двомовних словниках для кожної мовної пари. Pragma 6.3 Business є системою трансферного типу, де текст мовою оригіналу спочатку проходить морфологічний та синтаксичний аналіз, потім етап трансферу, який завершується процесом синтезу кінцевого варіанта перекладу через процедуру генерування (А. Бірюков, А. Міщенко). У процесі морфологічного аналізу відбувається лематизація лексичних одиниць, їхній пошук у словнику, аналіз морфем, визначення граматичних характеристик слів. Під час синтаксичного аналізу система розпізнає типи синтаксичних структур, реляційні зв'язки між окремими елементами структури. На етапі трансферу створюється проміжне представлення проаналізованого тексту, з якого на основі правил вихідної мови будуються синтаксичні структури синтезованих речень (Е. Корольов). Для ефективної роботи системи машинного перекладу Pragma 6.3 Business необхідні знання граматики, синтаксису, семантики та лексики мовних пар (Р. Ісламов), які містяться в одномовних модулях для аналізу та генерування речень і модулях трансферу. При цьому важливу роль відіграє ступінь розбіжності між структурами мов, що впливає на якість перекладу.

До програм статистичного машинного перекладу належать Google Translate, Babylon, Free Translation та Yandex, де використовується власне програмне забезпечення (М. Корonen). Позитивним у роботі програм є їхня здатність до самонавчання та самовдосконалення. Кожна система складається з трьох основних частин: моделі перекладу, моделі мови і декодера. Знання про мову перекладу містяться в моделі мови – списку вживань слів і словосполучень та ймовірностей їх використання. Після вибору кращого альтернативного перекладу програми враховують і запам'ятовують вибір користувача.

Гібридні системи машинного перекладу поєднують принципи роботи систем машинного перекладу на основі правил і систем статистичного машинного перекладу, до яких належить Bing Translator.

Навчання як статистичних систем машинного перекладу, так і гібридних систем машинного перекладу базується на двомовних корпусах текстів і не потребує глибокого й складного контрастивного лінгвістичного аналізу.

Ефективність роботи систем Google Translate, Bing Translator, Babylon, Yandex, Free Translation залежить від обсягу доступних їм корпусів паралельних текстів. Проте готові переклади, виконані людиною чи машиною, часом є далекими від ідеалу. Це впливає на якість перекладу Google Translate, Babylon, Bing Translator,

Free Translation та Yandex. Тому в машинних перекладах можуть виникати лексичні та граматичні помилки. Дослідження машинних перекладів програм Google Translate, Bing Translator, Babylon, Yandex (О. Тихомирова, В. Dhakar, К. Kais), виконаних у різних мовних напрямках, показали наявність різних типів помилок.

При вивченні лінгвістичної потужності систем машинного перекладу пропонуємо виділяти лексичні, граматичні, лексико-граматичні та орфографічні помилки, які порушують *норми літературної мови* – сукупність загальноприйнятих, усталених правил, якими керуються мовці в усному й писемному мовленні (М. Зубков, О. Пономарів) та впливають на еквівалентність і адекватність машинного перекладу по відношенню до оригіналу.

У нашому дослідженні під *еквівалентністю машинного перекладу* розуміємо семантичну та структурну відповідність одиниць основних рівнів мови (морфологічного, лексико-семантичного і синтаксичного) у текстах оригіналу та перекладу. *Адекватність машинного перекладу* тлумачимо як забезпечення прагматичних засад перекладу відповідно до поставленої мети на максимально можливому рівні еквівалентності, при цьому зберігаючи форму оригіналу і якнайменше порушуючи норми мови перекладу.

Лексичні помилки розглядались на рівні слів і словосполучень. У групі граматичних помилок виділено морфологічні та синтаксичні помилки. До групи синтаксичних помилок належать порушення в побудові словосполучень, простих і складних речень. Окремі групи формують лексико-граматичні й орфографічні помилки.

До *помилки у перекладі слів* відносимо: 1) неперекладені слова – слова, залишені в оригінальному варіанті, оскільки не було знайдено їх відповідників у мові перекладу; 2) вибір нерелевантних перекладних відповідників – вживання слів у невласивому для контексту та тематики значенні через неадекватне вирішення лексичної полісемії, лексичної та лексико-граматичної омонімії, уживання русизмів, неадекватний переклад власних назв та аббревіатур, які слід було залишити в оригінальному варіанті.

До *помилки у перекладі словосполучень* належать: 1) неадекватний переклад словосполучень як семантичних єдностей – уживання словосполучень у невласивому для контексту та тематики значенні через неправильний вибір відповідника в мові перекладу; 2) неадекватний переклад фразеологічних сполучень – переформатування значенневого наповнення стійких словосполук, що полягає у заміні їхніх компонентів у результаті неадекватного вибору відповідників у мові перекладу; 3) дослівний переклад іншомовних висловів, які мають свої відповідники в українській та російській мовах; 4) неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті.

До оцінки якості систем програмного забезпечення перекладу вводимо поняття *лексико-граматичних помилок* – вживання словосполучень у невласивому для контексту та тематики значенні, а також відхилення від нормативно-узуального порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір варіантів перекладу та граматичної трансформації.

*Морфологічні помилки* пов'язані з неправильним визначенням омонімічних граматичних форм англійської мови та словозмінних категорій української та російської мов (числа, відмінка, ступеня порівняння, часу, стану та способу). До них

належать: 1) помилки при знятті граматичної омонімії – неправильне визначення частиномовної належності слів; 2) ненормативне утворення пасивних конструкцій дієслів; 3) порушення категорії числа іменників; 4) ненормативне утворення відмінкових форм іменників; 5) порушення категорії часу дієслів; 6) порушення категорії способу дієслів; 7) ненормативне утворення ступенів порівняння прикметників; 8) вживання ненормативних форм дієприкметників; 9) вживання ненормативних форм дієприслівників.

*Синтаксичні помилки* полягають у порушенні побудови словосполучень та речень. Помилки у побудові *словосполучень* включають: 1) відхилення від нормативно-узвального порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір граматичної трансформації; 2) порушення керування – неправильний вибір відмінкової форми залежних слів після керівних слів; 3) відсутність необхідного прийменника; 4) неправильний переклад прийменника через неадекватний вибір відповідника в українській та російській мовах; 5) порушення узгодження – порушення синтаксичного зв'язку між залежним та стрижневим словом у роді, числі та відмінку; 6) вживання у тексті перекладу зайвого прийменника, якого немає в оригіналі або який не перекладається в мові перекладу.

До помилок у побудові *простих речень* належать: 1) порушення синтаксичних зв'язків між членами речення; 2) порушення координації між підметом і присудком, коли форма присудка вимагає підмета у формі називного відмінка, а форма присудка залежить від підмета у формах числа, роду та особи; 3) граматична неоднотипність однорідних членів речення через необґрунтоване застосування в мові перекладу різних частин мови, неправильний вибір відмінкових форм іменників; 4) неправильний порядок слів; 5) пропущені члени речення; 6) штучне ускладнення речення – необґрунтоване застосування сполучника, що трансформує просте речення у складне; 7) вживання зайвих членів речення; 8) невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення

До помилок у побудові *складних речень* належать: 1) порушення синтаксичних зв'язків між членами речення; 2) порушення координації між підметом і присудком; 3) граматична неоднотипність однорідних членів речення; 4) штучне ускладнення речення – необґрунтоване застосування сполучника, який відсутній у тексті оригіналу; 5) невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення; 6) вживання сполучників, які суперечать смисловим зв'язкам між частинами речення; 7) пропущені члени речення; 8) вживання зайвих членів речення, відсутніх в оригіналі.

*Орфографічні помилки* – це неправильне написання слів.

Запропоновані критерії оцінки якості лінгвістичної потужності систем Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex дозволяють більш комплексно проаналізувати тексти машинних перекладів та визначити типові помилки. Це допоможе обрати систему, яка найбільш ефективно виконує переклад у межах англо-українського та англо-російського мовних напрямків. Приклади помилок наведено в розділі 4.

У четвертому розділі “**Моделювання природної мови в різноманітних системах програмного забезпечення перекладу**” описано результати лінгвістичної оцінки якості роботи систем програмного забезпечення перекладу Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex; охарактеризовано природу лексичних, граматичних, лексико-граматичних та орфографічних помилок; здійснено порівняння якості роботи Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex в англо-українському та англо-російському напрямках перекладу.

Лінгвістичний аналіз текстів машинного перекладу програми Pragma 6.3 Business показав, що у текстах машинних перекладів переважають лексичні помилки – 53,5% від загального обсягу помилок.

Помилки у перекладі слів включають: 1) вибір нерелевантних перекладних відповідників, що не відповідають контексту та тематичному спрямуванню тексту: а) неадекватний вибір варіантних відповідників багатозначних слів: *area* (Авіація) – *область (площа)*; б) помилки при вирішенні лексико-граматичної омонімії: *measures* (Загальна) – *зважає (заходи)*; в) неадекватний переклад власних назв та аббревіатур, які слід було залишити в оригінальному варіанті *SAP* (Техніка) – *СОК*; г) помилки при вирішенні лексичної омонімії: *pile* (Текстиль) – *куна (ворс)*; д) неадекватний переклад слів через наявність помилки в оригіналі: *casting* (Будівництво) – *відливання (casing)*; е) вживання русизмів: *itching* (Медицина) – *зудить*; 2) неперекладені слова, які охоплюють: а) терміни: *extraction* (Екологія); б) власні назви: *Metinvest* (Гірничка); в) скорочення: *Mtonnes* (Гірничка, Металургійна); г) аббревіатури: *LCD* (Електроніка).

До помилок у перекладі словосполучень належать: 1) дослівний переклад іншомовних висловів, які в українській мові мають свої відповідники: *finely dispersed state* (Харчова) – *дрібно розігнав державу*; 2) неадекватний переклад словосполучень як семантичних єдностей: *particulate filter* (Фізика) – *сипкий фільтр*; 3) неправильний переклад фразеологічних сполучень: *take care about* (Подорож) – *брати турботу про*; 4) неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті: *CFM International CFM56-7B* (Авіація) – *ПІДТВЕРДЖУЮТЬ Міжнародний CFM56-7B*.

До морфологічних помилок, виявлених у машинних перекладах програми Pragma 6.3 Business належать: 1) помилки при знятті граматичної омонімії: *A “kit” of ready-made parts created in space* (Космонавтика) – *“Комплект” готових частин створив в космічному*; 2) ненормативне утворення пасивної конструкції дієслова: *An armed attack against one or more of them in Europe or North America shall be considered an attack* (Морська) – *озброєно нападають проти один або більше з них в Європі або Північна Америка має вважатися нападом*; 3) порушення категорії числа іменника: *further installation of finished products* (Будівництво) – *подальша установка готовий виріб*; 4) ненормативне утворення відмінкової форми іменника: *Equipment allows to produce very large diameter logs* (Деревообробка). – *Устаткування дозволяє робити дуже великі колоди діаметру*; 5) порушення категорії часу дієслова: *to restore and maintain the security of the North Atlantic area* (Право) – *щоб відновив і підтримує безпека північний атлантичний область*; 6) порушення категорії способу дієслова: *Do not expose this equipment to rain* (Інструкція) – *не піддають це устаткування дощу*; 7) ненормативне утворення ступеня порівняння прикметника:

*always the best* (Кулінарія) – завжди **кращий**; 8) використання ненормативної форми дієприкметника: *air policing duties* (Морська) – повітряні **охороняючі** обов'язки; 9) вживання ненормативної форми дієприслівника: *starting* (Сільське господарство) – **починаючись**.

Помилки у побудові словосполучень включають: 1) відхилення від нормативно-узвального порядку слів, порушення логіко-синтаксичного зв'язку між словами: *analog format* (Техніка) – формат аналога (аналоговий формат); 2) порушення керування: *have an enormous effect on business efficiency* (Загальна) – мати величезний ефект на діловій **ефективності**; 3) відсутність необхідного прийменника: *The previous loan programme has contributed to reduced emissions of carbon dioxide by some 22,000 tonnes* (Банківська справа). – Попередня програма кредиту сприяла зменшенню емісії вуглекислоти **приблизно 22,000 метричних тонн**; 4) неправильний переклад прийменника: *UN Framework convention on climate change* (Екологія) – угода Структури ООН на зміні клімату; 5) порушення узгодження між словами: а) у роді: *newly arrived company's aircraft is presented* (Авіація) – **новоприбулий авіація** компанії є був присутнім; б) у числі та відмінку: *26 billion* (Нафтогаз) – 26 **більйона**; в) у роді та відмінку: *you need highly skilled, committed personnel* (Загальна) – вам треба **висококваліфікованим, вчинив персонал**; г) у відмінку: *to develop its own fleet of aircraft, ground handling and maintenance base* (Авіація) – розвивати його власний флот авіації, **грунтової, управляючи і ремонтна основа**; б) вживання зайвого прийменника: *44% of Ukrainians* (Секс) – **44% з українців**.

До помилок у побудові простих речень належать: 1) порушення синтаксичних зв'язків між членами речення: *The new Panamera S and Panamera 4S are fitted with a 4.8 liter naturally aspirated engine with direct fuel injection, the VarioCam Plus system for regulating gas distribution phases and a variable intake collector* (АВТО). – Нові Panamera S і Panamera 4S придатні з **4.8 літром природно вимовив з придухом двигун з прямою паливною ін'єкцією, система VarioCam Плюс, бо регулюючи фази поширення газу і змінна впускають колектор**; 2) граматична неоднотипність однорідних членів речення: *Groundwork is trenching, trench excavation and pitting for isolated supports, cellars and strip foundations; trench making for underground pipelines, transportation and loosening of grounds, pads planning, making earthfills and ground ramming, pile sinking and hammering, making of pile foundation* (Будівництво). – **Фундамент копає траншею, траншейна земляна робота і риючи яму для ізольованої підтримки, ляхів і strip принципів; траншейне створення для заглиблених трубопроводів, перевезення і послабляючи земель, планування втулок, роблячи earthfills і обгрунтовують трамбівку, pile пониження і стук, роблячи pile основи**; 3) порушення координації між підметом і присудком: *Bank Lviv will for its part contribute with EUR 2 million for the loan facility* (Банківська справа). – **Банківський Львів буде для його частини вносять вклад з мільйоном EUR 2 для засобу кредиту**; 4) неправильний порядок слів: *Ready deck has a "cup" or "half cup" for the connection logs together, compensatory incision for the reduction of longitudinal cracks, "bed" for laying logs on each other* (Деревообробна). – **Готовий поміст має приймач "приймача" або "половини для connection колод разом, компенсаційне розрізання для зниження подовжніх тріщин, "ліжко" для поклав колоди один на одного**.

До помилок у побудові складних речень належать: 1) порушення синтаксичних зв'язків між членами речення: *In general, Robert Howarth estimates methane emissions during the shale gas extraction at the level 3.6%–7.9% of total extraction volume, while during extraction of conventional natural gas the loss rate is estimated at the level of 1.7%–6%* (Гірнична). – Взагалі, Роберт Howarth оцінює емісії метану впродовж відбору сланцевого газу в горизонтальний **3.6%–7.9% повного extraction об'єму, поки впродовж витягання обумовленого бекара виділяють газ втраченої норми оцінюється на рівні 1.7%–6%**; 2) граматична неоднотипність однорідних членів речення: *Thanks to the protection against overload, overheating and IP65 sealing class, they can be widely used in automotive industry where the power source is a battery or other similar device* (Електроніка). – Спасибі до захисту проти **перевантаження, перегрівання і IP65, заливка герметиком класу**, вони можуть бути широко використані в саморушній промисловості, де джерело живлення – батарея або інший подібний пристрій; 3) невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення: *At the final stage, the complex interpretation of all available geological and geophysical information is carried out, which allows effective determining basic characteristics of the objects under study* (Інформатика). – На завершальній стадії, складна **інтерпретація** усієї доступної геологічної і геофізичної інформації здійснюється, **який** дозволяє ефективні визначальні основні характеристики об'єктів під вивченням; 4) вживання сполучника, що суперечить смисловим зв'язкам між частинами речення: *Being the dominant operating system provider in the emerging non-PC market would be a tremendous prize for a company but no single player will become the de facto standard in 2011, as has happened in other computing markets in the past* (Комерція). – Існування домінуючого операційного системного постачальника на некомп'ютерному ринку появи було б величезним призом, бо компанія але ніякий єдиний гравець стане де-факто стандартом в 2011, **оскільки** стався на інших обчислювальних ринках в минулому; 5) неправильний порядок слів через необґрунтоване застосування граматичної трансформації: *They extracted each sample with a solution of formic acid and ran them through a liquid chromatograph, an instrument that separates a mixture of compounds* (Космонавтика). – Вони витягнули кожен зразок з рішенням мурашиної кислоти і **пробігли рідинний chromatograph їх**, інструмент, який відділяє суміші сумішей.

У машинних перекладах програми Pragma 6.3 Business було знайдено лексико-граматичні помилки: *timber industry* (Хімія) – промисловість лісоматеріалу (лісопереробна промисловість) та орфографічні помилки: *geophysical* (Бізнес) – геофізичний.

Кількісне співвідношення помилок систем Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation і Yandex в англо-українському напрямку перекладу подано в таблиці 1.

Таблиця 1

**Кількісне співвідношення помилок систем  
Pragma 6.3 Business, Google Translate, Bing Translator, Babylon,  
Free Translation і Yandex в англо-українському напрямку перекладу**

Типи помилок				Babylon		Yandex
--------------	--	--	--	---------	--	--------

	Pragma 6.3 Business		Google Translate		Bing Translator				Free Translation			
1 Лексичні	603	53,4%	266	39,5%	371	39,8%	762	56,7%	327	35,6%	189	33,1%
1.1 Помилки у перекладі слів	398	35,3%	158	23,4%	217	23,2%	517	38,5%	182	19,8%	107	18,7%
1.2 Помилки у перекладі словосполучень	205	18,1%	108	16,1%	154	16,6%	245	18,2%	145	15,8%	82	14,4%
2 Граматичні	484	42,9%	400	59,3%	548	58,7%	563	41,9%	587	64,0%	382	66,9%
2.1 Морфологічні	176	15,6%	63	9,3%	91	9,8%	181	13,5%	110	12,0%	46	8,1%
2.2 Синтаксичні	308	27,3%	337	50,0%	457	48,9%	382	28,4%	477	52,0%	336	58,8%
2.2.1 Помилки у побудові словосполучень	118	10,5%	171	25,4%	248	26,5%	174	12,9%	220	24,0%	171	29,9%
2.2.2 Помилки у побудові речень:	190	16,8%	166	24,6%	209	22,4%	208	15,5%	257	28,0%	165	28,9%
а) помилки у побудові простих речень	91	8,0%	82	12,2%	104	11,1%	98	7,3%	124	13,5%	82	14,4%
б) помилки у побудові складних речень	99	8,8%	84	12,4%	105	11,3%	110	8,2%	133	14,5%	83	14,5%
3 Лексико-граматичні	42	3,7%	8	1,2%	14	1,5%	16	1,2%	2	0,2%	0	0,0%
4 Орфографічні	0	0,0%	0	0,0%	0	0,0%	2	0,2%	2	0,2%	0	0,0%
Усього:	1129	100%	674	100%	933	100%	1343	100%	918	100%	571	100%

Найбільшу кількість помилок знайдено у текстах, перекладених програмою Babylon – 1343 помилки, серед яких переважають лексичні помилки – 762. Порівняно з іншими програмами в машинних перекладах програми Free Translation превалюють граматичні помилки – 587, серед яких – синтаксичні помилки в побудові словосполучень, простих і складних речень.

У текстах машинних перекладів Pragma 6.3 Business було виявлено найбільшу кількість лексико-граматичних помилок (42 помилки).

У машинних перекладах програм Babylon та Free Translation було виявлено по дві орфографічні помилки внаслідок відсутності блоків перевірки орфографії у системах. Найменшу кількість помилок було виявлено у текстах, перекладених програмою Yandex – 571 помилка. Саме тому вона є найбільш ефективною для виконання перекладу з англійської мови українською.

Кількісне співвідношення помилок, виявлених у машинних перекладах систем Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex, в англо-російському напрямку перекладу подано у таблиці 2.

Таблиця 2

**Кількісне співвідношення помилок систем Pragma 6.3 Business, Google Translate,**



**Bing Translator, Babylon, Free Translation і Yandex  
в англо-російському напрямку перекладу**

Типи помилок	Pragma 6.3 Business		Google Translate		Bing Translator		Babylon		Free Translation		Yandex	
1 Лексичні	601	53,7%	280	41,0%	326	40,8%	467	44,6%	400	37,9%	198	33,8%
1.1 Помилки у перекладі слів	396	35,4%	169	25,0%	192	24,0%	273	26,1%	230	21,8%	120	20,5%
1.2 Помилки у перекладі словосполучень	205	18,3%	111	16,0%	134	16,8%	194	18,5%	170	16,1%	78	13,3%
2 Граматичні	476	42,5%	392	57,5%	468	58,6%	574	54,8%	651	61,8%	385	65,8%
2.1 Морфологічні	172	15,4%	57	8,3%	60	7,5%	106	10,1%	125	11,9%	41	7,0%
2.2 Синтаксичні	304	27,1%	335	49,2%	408	51,1%	468	44,7%	526	49,9%	344	58,8%
2.2.1 Помилки у побудові словосполучень	115	10,3%	177	26,0%	200	25,0%	211	20,2%	249	23,6%	177	30,3%
2.2.2 Помилки у побудові речень:												
а) помилки у побудові простих речень	91	8,1%	79	11,6%	106	13,3%	112	10,7%	140	13,3%	81	13,8%
б) помилки у побудові складних речень	98	8,7%	79	11,6%	102	12,8%	145	13,8%	137	13,0%	86	14,7%
3 Лексико-граматичні	42	3,8%	8	1,2%	5	0,6%	7	0,6%	3	0,3%	1	0,2%
4 Орфографічні	0	0,0%	2	0,3%	0	0,0%	0	0,0%	0	0,0%	1	0,2%
Усього:	1119	100%	682	100%	799	100%	1047	100%	1054	100%	585	100%

Найбільшу кількість помилок виявлено в машинних перекладах програми Pragma 6.3 Business (1119 помилок). Граматичні помилки є найчисленнішими у перекладах програми Free Translation – 651 помилка, серед яких переважають синтаксичні помилки (49,9% від загальної кількості помилок програми).

Лексико-граматичні помилки домінують у машинних перекладах програми Pragma 6.3 Business (42 помилки). У текстах, перекладених Google Translate та Yandex, виявлено дві та одну орфографічні помилки відповідно. Встановлено, що найбільш ефективною для виконання перекладів в англо-українському та англо-російському напрямках є програма Yandex.

Загальні результати дослідження дають підстави зробити такі **висновки**:

Теоретична концепція дисертації ґрунтується на таких положеннях:

- 1) у системах програмного забезпечення перекладу моделюються процеси розуміння, інтерпретації та створення текстів людиною, її здатність до смислоототожнення та смислорозрізнення, перцептивне, когнітивне й поведінкове мислення людини, декларативні та процедурні знання (О. Бабіна, М. Блехман, В. Бурдаєв, Е. Корольов, О. Шаміс, Ю. Шемакін);
- 2) системи програмного забезпечення перекладу охоплюють

лінгвістичне забезпечення та алгоритми обробки мов (Ю. Марчук, Л. Нелюбін, Р. Піотровський), створені з метою автоматизації інтелектуальних функцій і когнітивної діяльності людини, механізмів обробки інформації.

Методологічною основою роботи стали праці, які ґрунтуються на підходах до оцінки якості машинних перекладів (А. Бірюков, М. Блехман, Ю. Марчук, П. Хроменков) та типології мовних помилок (Т. Бондаренко, М. Зубков). Запропонована у роботі методика дослідження дала змогу виявити неповноту лінгвістичного та програмного забезпечення систем, що впливає на адекватність та еквівалентність машинних перекладів українською та російською мовами. Для цього було використано структурний метод, зокрема його методику аналізу за безпосередніми складниками, методику дистрибутивного аналізу та методику аналізу словникових дефініцій, а також зіставний метод.

Англо-українська й англо-російська мовні пари змодельовані у шести системах програмного забезпечення перекладу, серед яких Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation, Yandex. У програмі Pragma 6.3 Business застосовано підхід до машинного перекладу на основі правил. Bing Translator – система гібридного машинного перекладу, яка у процесі роботи поєднує правила та статистичні методи перекладу. Google Translate, Babylon, Free Translation та Yandex є системами статистичного машинного перекладу, які працюють на основі ймовірнісного визначення перекладних відповідників із корпусів паралельних текстів.

Результати роботи систем програмного забезпечення перекладу в англо-українському та англо-російському напрямках дозволяють зробити висновок про те, що типовими для всіх програм є лексичні, граматичні та лексико-граматичні помилки. Проте їхня кількість залежить від тематичного спрямування текстів, напрямку перекладу, лінгвістичного та програмного забезпечення системи. Орфографічні помилки становлять менше 1% та зустрічаються в українських перекладах програм Babylon (0,1%) і Free Translation (0,2%), а також у російських перекладах програм Google Translate (0,3%) і Yandex (0,2%).

В англо-українському й англо-російському напрямках у машинних перекладах системи Pragma 6.3 Business домінують лексичні помилки (53,4% та 53,7% відповідно) – неадекватний переклад слів і словосполучень. Причиною їхнього виникнення є нездатність програми аналізувати контекст і тематичне спрямування тексту та неповнота предметного компонента програми, зокрема автоматичного словника, що призводить до появи неперекладених слів, неадекватно перекладених слів через невирішення полісемії, лексичної та лексико-граматичної омонімії, а також дослівного перекладу фразеологічних сполучень.

Покращити ефективність роботи програм можна шляхом поповнення автоматичного словника та внесення змін до алгоритмів роботи.

У текстах, перекладених системами статистичного машинного перекладу, переважають граматичні помилки: Google Translate – 59,3% в англо-українському та 57,5% в англо-російському напрямках перекладу; Free Translation – 64,0% та 61,8% в англо-українському та англо-російському перекладах відповідно; Yandex – 66,7% та 65,5% в англо-українському та англо-російському перекладах відповідно. Причиною їхнього виникнення є невикористання програмами граматичних правил у процесі

роботи.

Виняток становить програма Babylon, у машинних перекладах якої домінують лексичні помилки (56,8%). Причиною їхнього виникнення може бути недостатня повнота паралельного корпусу текстів, на основі якого працює програма.

Найменшу кількість помилок виявлено в машинних перекладах, виконаних програмою Yandex. Саме тому вона визнана найбільш ефективною системою для роботи з англо-українською та англо-російською мовними парами.

Серед лексичних, синтаксичних і лексико-граматичних помилок зустрічаються синкретичні помилки, природа яких полягає в порушенні однією ненормативною одиницею різних мовних норм.

Перспективним напрямом подальших досліджень оцінки якості систем програмного забезпечення перекладу може бути, на наш погляд, зіставлення якості окремих версій однієї системи з метою визначення покращення або погіршення якості перекладу в новітніх версіях комп'ютерних перекладачів, а також застосування нових мовних пар до їхньої оцінки якості.

### **СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ**

1. Стахмич Ю. С. Системи програмного забезпечення перекладу в науковій рефлексії : історія і сучасність / Ю. С. Стахмич // Наукові записки. Сер.: “Філологічні науки (мовознавство)” : у 2 ч. : [зб. наук. пр.] / [відп. ред. О. Семенюк]. — Кіровоград : РВВ КДПУ ім. В. Винниченка, 2012. — Вип. 104 (1). — С. 323—327.
2. Стахмич Ю. С. Системи програмного забезпечення перекладу як лінгвістичний об'єкт / Ю. С. Стахмич // Наукові записки. Сер.: “Філологічні науки (мовознавство)” : [зб. наук. пр.] / [відп. ред. О. Семенюк]. — Кіровоград : РВВ КДПУ ім. В. Винниченка, 2013. — Вип. 116. — С. 353—357.
3. Стахмич Ю. С. Адекватність та еквівалентність перекладу в контексті комп'ютерної лінгвістики / Ю. С. Стахмич // Вісник Житомирського державного університету імені Івана Франка : [наук. журн.] / [відп. ред. Н. А. Сейко]. — № 66. — Житомир : Вид-во ЖДУ ім. І. Франка, 2013. — С. 235—238.
4. Стахмич Ю. С. Оцінка якості систем Pragma 6.3 Business і Google Translate / Ю. С. Стахмич // Нова філологія : [зб. наук. пр.] / Міністерство освіти і науки України, Запорізький національний університет / [відп. ред. М. В. Воробйова]. — Запоріжжя : ЗНУ, 2014. — № 65. — С. 246—252.
5. Стахмич Ю. С. Системы программного обеспечения перевода: когнитивно-интерпретационный аспект / Ю. С. Стахмич // Филология, искусствоведение и культурология в XXI веке : [материалы междун. заочной науч.-практ. конф., (Новосибирск, 18 марта 2013 г.)]. — Новосибирск : Изд. “СибАК”, 2013. — С. 55—63.
6. Стахмич Ю. С. Оценка качества систем статистического машинного перевода / Ю. С. Стахмич // European Applied Sciences : [wissenschaftliche Zeitschrift] / [Chefredakteur S. Hezberg]. — Stuttgart : ORT Publishing, 2014. — № 9 (September). — S. 91—93. — ISSN 2195-2183.
7. Стахмич Ю. С. Рівні розуміння систем машинного перекладу / Ю. С. Стахмич // Україна і світ : діалог мов та культур : [матеріали міжнар. наук.-практ. конф., (Київ, 03—05 квіт. 2013 р.)]. — К. : Вид. центр КНЛУ, 2013. — С. 363—365.

### **АНОТАЦІЯ**

Стахмич Ю. С. Комп'ютерне моделювання природної мови (на матеріалі систем програмного забезпечення перекладу). – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата філологічних наук за спеціальністю 10.02.15 – загальне мовознавство. – Національний педагогічний університет імені М.П. Драгоманова, Київ, 2015.

У дисертації запропоновано новий підхід до вивчення комп'ютерного моделювання англійської, української та російської мов у системах програмного забезпечення перекладу Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation, Yandex. Удосконалено методику оцінки якості систем програмного забезпечення перекладу на основі розроблених критеріїв виявлення лексичних, граматичних (морфологічних, синтаксичних) та лексико-граматичних помилок. Укладено тестові збірки речень англійською мовою, які дають змогу перевіряти ефективність систем програмного забезпечення перекладу при роботі з текстами різного тематичного спрямування в англо-українському та англо-російському перекладах. Визнано перевагу Yandex над іншими системами для англо-української та англо-російської мовних пар за кількістю помилок. Установлено, що для оптимізації роботи систем перекладу слід збагачувати лінгвістичне забезпечення систем, удосконалювати алгоритми їх роботи і продовжувати навчання на основі багатомовних корпусів текстів.

Ключові слова: комп'ютерне моделювання, природна мова, системи програмного забезпечення перекладу, машинний переклад, мовні помилки, оцінка якості систем програмного забезпечення перекладу.

### **АННОТАЦИЯ**

Стахмич Ю. С. Компьютерное моделирование естественного языка (на материале систем программного обеспечения перевода). – На правах рукописи.

Диссертация на соискание ученой степени кандидата филологических наук по специальности 10.02.15 – общее языкознание. – Национальный педагогический университет имени М.П. Драгоманова, Киев, 2015.

В диссертационной работе представлено исследование компьютерного моделирования естественного языка в системах программного обеспечения англо-украинского и англо-русского перевода.

Актуальность диссертационной работы обусловлена общими тенденциями теоретической и прикладной лингвистики к изучению структур представления знаний человека, операций мышления и сознания с целью их моделирования в созданных формальных системах, к которым относятся системы программного обеспечения перевода.

У работе впервые исследовано результаты компьютерного моделирования английского, украинского и русского языков в системе программного обеспечения перевода на основе правил (Pragma 6.3 Business), системах статистического машинного перевода (Google Translate, Babylon, Free Translation, Yandex) и

гибридной системе машинного перевода (Bing Translator). Усовершенствовано методику оценки качества систем программного обеспечения перевода на основе разработанных детальных критериев выявления ошибок на лексическом, морфологическом и синтаксическом уровнях языка. Введено понятие лексико-грамматической ошибки, которая заключается в употреблении словосочетания в нерелевантном для контекста и тематики смысле, а также уклонении от нормативно-визуального порядка слов, нарушении логико-синтаксической связи между словами в результате неправильного выбора вариантов перевода и грамматической трансформации перестановки. Разработанные тестовые подборки предложений на английском языке позволили проверить эффективность систем программного обеспечения перевода при работе с текстами разного тематического направления. Результаты работы систем программного обеспечения перевода в англо-украинском и англо-русском направлениях показали, что типическими для всех программ являются лексические, грамматические (морфологические и синтаксические), лексико-грамматические и орфографические ошибки. Их количество зависит от тематического направления текстов, направления перевода, лингвистического и программного обеспечения систем. Недостатки лингвистического и программного обеспечения систем машинного перевода заключаются в неполноте автоматических словарей, моделей языков, а также в несовершенстве работы алгоритмов систем. На основании сопоставительного лингвистического анализа машинных переводов, выполненных Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation и Yandex, признано преимущество Yandex над другими системами англо-украинского и англо-русской языковых пар в количестве ошибок.

Ключевые слова: компьютерное моделирование, естественный язык, системы программного обеспечения перевода, машинный перевод, языковые ошибки, оценка качества систем программного обеспечения перевода.

### **RESUME**

Stakhmych Yu. S. Computer Modeling of Natural Languages (a Case Study of Translation Software Systems). – Manuscript.

Thesis for a Candidate's Degree in 10.02.15 – General Linguistics. – National Pedagogical Dragomanov University, Kyiv, 2015.

The thesis suggests a new approach to the investigation of computer modeling of English, Ukrainian and Russian in translation software systems Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation, Yandex. The methodology of quality assessment of translation software systems is improved by the detailed criteria used to identify lexical, grammatical (morphological and syntactic) and lexico-grammatical mistakes. The introduced English text samples allow to assess the efficiency of translation software systems in translating texts on different topics from English into Ukrainian and from English into Russian. The prevalence of Yandex over other systems in the number of mistakes in English-Ukrainian and English-Russian machine translations is determined. The optimization of translation software systems is defined to require the enrichment of their linguistic components, improvement of algorithms and further study on the basis of multilingual text corpora.

Key words: computer modeling, natural language, translation software systems, machine translation, language mistakes, quality assessment of translation software systems

.