

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

На правах рукопису

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

Дисертація на здобуття наукового ступеня кандидата філологічних наук

Науковий керівник:

Денисова Світлана Павлівна,

доктор філологічних наук, професор

КИЇВ – 2015

## ЗМІСТ

ВСТУП.....	4
РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ ВИВЧЕННЯ КОМП'ЮТЕРНОГО МОДЕЛЮВАННЯ ПРИРОДНОЇ МОВИ .....	13
1.1 Комп'ютерна лінгвістика з позицій прикладного мовознавства .....	13
1.2 Основи автоматичної обробки природної мови .....	16
1.3 Моделювання у комп'ютерній лінгвістиці .....	26
1.4 Становлення та подальший розвиток теорії штучного інтелекту.....	33
1.5 Провідні поняття теорії штучного інтелекту в різних наукових інтерпретаціях .....	37
1.6 Системи програмного забезпечення перекладу в аспекті теорії штучного інтелекту .....	41
Висновки до розділу 1.....	49
РОЗДІЛ 2. МЕТОДИКА ДОСЛІДЖЕННЯ КОМП'ЮТЕРНОГО МОДЕЛЮВАННЯ ПРИРОДНОЇ МОВИ В СИСТЕМАХ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ПЕРЕКЛАДУ.....	52
2.1 Антиномії комп'ютерної лінгвістики .....	52
2.2 Методи та способи дослідження моделювання природної мови у системах програмного забезпечення перекладу .....	55
2.3 Специфіка методики дослідження комп'ютерного моделювання природної мови у системах програмного забезпечення перекладу .....	62
Висновки до розділу 2.....	65
РОЗДІЛ 3. СИСТЕМИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ПЕРЕКЛАДУ ЯК ДОСВІД МОДЕЛЮВАННЯ ПРИРОДНОЇ МОВИ .....	67
3.1 Система Pragma 6.3 Business .....	67

3.2	Системи статистичного машинного перекладу Google Translate, Babylon, Free Translation, Yandex .....	68
3.3	Гібридна система машинного перекладу Bing Translator .....	70
3.4	Критерії оцінки якості систем програмного забезпечення перекладу .....	70
	Висновки до розділу 3 .....	75
	РОЗДІЛ 4. МОДЕЛЮВАННЯ ПРИРОДНОЇ МОВИ В РІЗНОМАНІТНИХ СИСТЕМАХ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ПЕРЕКЛАДУ .....	77
4.1	Оцінка якості системи Pragma 6.3 Business .....	77
4.2	Оцінка якості системи Google Translate .....	99
4.3	Оцінка якості системи Bing Translator.....	115
4.4	Оцінка якості системи Babylon .....	132
4.5	Оцінка якості системи Free Translation .....	151
4.6	Оцінка якості системи Yandex .....	171
4.7	Порівняння ефективності роботи систем Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation і Yandex в англо-українському напрямку перекладу .....	187
4.8	Порівняння ефективності роботи Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation і Yandex в англо-російському напрямку перекладу .....	189
	Висновки до розділу 4 .....	191
	ВИСНОВКИ .....	194
	СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ .....	198
	СПИСОК ЛЕКСИКОГРАФІЧНИХ ТА ДОВІДКОВИХ ДЖЕРЕЛ .....	219
	СПИСОК ДЖЕРЕЛ ІЛЮСТРАТИВНОГО МАТЕРІАЛУ .....	220
	ДОДАТКИ .....	225

## ВСТУП

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 методика 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. О. Моторін), *математики* (У. Вівер, Д. Ю. Панов), *філософії* (I. Bar-Hillel) 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. М. Баранов, Є. А. Карпіловська, М. П. Кочерган), а також їх моделювання у системах комп'ютерної обробки мови (Ю. М. Марчук, Л. Л. Нелюбін, З. В. Партико, Р. Г. 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. І. Коваль, М. Г. Мамедова, Л. 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. П. Бурдаєв, Г. А. Гайна, О. В. Гладченко, В. Г. Доля, О. В. Зубов, Є. А. Карпіловська, Ю. В. Рождественський, О. М. 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. В. Бірюков, Ю. М. 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. С. Бархударов, Ю. В. Ванников, Б. С. Виноградов, В. Н. Комісаров, О. Д. 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 мисленнєво-мовленнєвої 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 Національного педагогічного університету 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 р.).

**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 змодельовані тексти англійською мовою, що належать до таких тематичних галузей, як: авіація, автомобільна, банківська справа, Біблія, бізнес, біологія, будівництво, військова, 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 [230–273]. Для оцінки якості системи Pragma 6.3 Business було проаналізовано 484 пари речень в англо-українському напрямку перекладу. Матеріалом дослідження якості систем Google Translate, Bing Translator, Babylon, Free Translation та Yandex слугували 110 речень (10 перших речень та заголовки), відібраних з текстів 10 тематик, у яких було виявлено найбільшу кількість помилок [236; 238; 240; 242; 246; 251; 257; 259; 265; 269]. 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 системи 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 програмного забезпечення перекладу (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 не властивому для контексту та тематики значенні, а також відхиленні від нормативно-узусуального 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 недосконалості 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%).

На захист виносяться такі **положення**:

1. Сучасні системи програмного забезпечення перекладу, які працюють з англо-українською та англо-російською мовними парами, поділяються на системи, що базуються на правилах (Pragma 6.3 Business), системи статистичного машинного перекладу (Google Translate, Babylon та Free Translation) та гібридні системи машинного перекладу (Bing Translator).

2. Запропонована у роботі методика дослідження машинних перекладів відповідно до детального набору критеріїв виявлення лексичних, морфологічних, синтаксичних, лексико-граматичних та орфографічних помилок дозволила виявити неповноту лінгвістичного та програмного забезпечення програм, що впливає на адекватність та еквівалентність машинних перекладів українською та російською мовою.

3. У машинних перекладах системи Pragma 6.3 Business домінують лексичні помилки у перекладі слів та словосполучень. Причиною їхнього виникнення є неповнота лінгвістичного забезпечення системи. Покращити ефективність роботи програми можна шляхом поповнення автоматичного словника та внесення змін до алгоритму роботи.

4. У текстах, перекладених системами статистичного машинного перекладу (Google Translate, Babylon, Free Translation, Yandex), переважають граматичні

помилки. Причиною їхнього виникнення є невикористання програмами граматичних правил у процесі роботи. Для підвищення якості роботи програм необхідним є їхнє подальше навчання на основі двомовних корпусів текстів.

5. Найефективнішою програмою в англо-українському та англо-російському напрямках перекладу є Yandex, у машинних перекладах якої виявлено найменшу кількість помилок.

**Теоретичне значення** дисертації пов'язане із розширенням основних положень загального мовознавства, комп'ютерної лінгвістики та теорії штучного інтелекту, а також із визначенням теоретико-методологічних засад дослідження комп'ютерного моделювання природної мови у системах програмного забезпечення. Введення поняття лексико-граматичної помилки збагатило методіку оцінки якості систем програмного забезпечення перекладу.

**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”, “Грамматична 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 XXI 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.

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.

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.

Методика дослідження комп'ютерного моделювання англійської, української та російської мов у системах програмного забезпечення перекладу охоплює чотири етапи. Їх кількість і послідовність визначаються поставленою метою та завданнями дослідження. На кожному етапі було використано методи дослідження, специфіка застосування яких описана в цьому розділі.

У третьому розділі **“Системи програмного забезпечення перекладу як досвід моделювання природної мови”** подано коротку характеристику систем програмного забезпечення перекладу Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex, схарактеризовано критерії оцінки якості моделювання англійської, української та російської мов у системах програмного забезпечення перекладу.

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.

# РОЗДІЛ 1

## ТЕОРЕТИЧНІ ЗАСАДИ ВИВЧЕННЯ

### КОМП'ЮТЕРНОГО МОДЕЛЮВАННЯ ПРИРОДНОЇ МОВИ

#### 1.1 Комп'ютерна лінгвістика з позицій прикладного мовознавства

Сучасне мовознавство характеризується тенденцією до експансіонізму. Поглиблення фундаментальних теоретичних досліджень супроводжується розширенням сфери практичного застосування їхніх результатів у природничих науках і техніці. Упродовж останніх років інтерес до мови надзвичайно зріс. Лінгвістичними питаннями займаються не тільки мовознавці, але також філософи і математики, соціологи і кібернетики, психіатри і програмісти. Результатом цього стало виникнення таких стикових дисциплін, як комп'ютерна лінгвістика, кібернетична лінгвістика, математична лінгвістика тощо.

Комп'ютерна лінгвістика почала формуватися у 50-х роках ХХ ст. на основі комп'ютерної науки, теорії інформації і структурно-математичної лінгвістики та виокремилася в окрему галузь науки у 60-х роках ХХ ст. Основним стимулом до її виникнення була поява комп'ютерів, які стали універсальним засобом моделювання мови [224, с. 595]. Теоретичними орієнтирами комп'ютерної лінгвістики були спроби створення систем штучного інтелекту, систем машинного перекладу, інформаційно-пошукових систем і мов, які ґрунтувалися на розробках трансформаційного й дистрибутивного методів З. Харрися, генеративної концепції Н. Хомського, а також на мисленнєвих експериментах з метою перевірки наявності інтелекту в комп'ютера (А. Тьюринг) [95, с. 649]. Подальші дослідження визначили головне завдання комп'ютерної лінгвістики – досягнення схожості між інтелектуальною діяльністю людини та комп'ютерними моделями шляхом розробки формального апарату інтелекту як цілеспрямованої діючої символічної системи.

Дискусійним питанням залишається розмежування прикладної й комп'ютерної лінгвістики через майже одночасне їхнє становлення та спільне

спрямування на розв'язання прикладних завдань. Проте вони перебувають у різних епістемологічних площинах [224, с. 595]. Так, прикладна лінгвістика зорієнтована на результат і практичну діяльність людини, а комп'ютерна лінгвістика вказує на інструментарій досягнення наукового результату.

Прикладна лінгвістика опрацьовує методи розв'язання практичних завдань для задоволення суспільних потреб, пов'язаних із використанням мови [49, с. 261; 60, с. 18]. Предметом дослідження цього напрямку мовознавства є “застосування методів інших наукових дисциплін для вивчення будови мовної системи та закономірностей її реалізації, а також здобутків лінгвістики у розв'язанні завдань інших галузей науки та суспільної практики” [46, с. 18]. У прикладній лінгвістиці розробляються технології, призначені для використання мови як засобу масової комунікації, фіксації, збереження та передавання інформації, лінгвістичного опрацювання текстів [46, с. 28]. Вона спрямована на розв'язання практичних завдань різних галузей науки й техніки, повсякденного життя людини, суспільства на основі теоретичного доробку досліджень мови й мовлення [95, с. 640]. Сучасні проблемні напрямки прикладної лінгвістики охоплюють машинний переклад, лінгвістичне забезпечення автоматичних систем управління, автоматичний аналіз й синтез тексту, створення інформаційних мов, автоматичне анотування та індексування документів, лінгвістичне забезпечення роботи інформаційних систем, автоматичну переробку текстової інформації [60, с. 18] тощо.

Існують різні концепції комп'ютерної лінгвістики як галузі мовознавства. Спершу вона ототожнювалася з прикладною лінгвістикою [224, с. 595]. На сьогодні комп'ютерна лінгвістика розглядається як важлива частина прикладної лінгвістики [25, с. 10; 87, с. 29; 224, с. 595] та як самостійна лінгвістична дисципліна [47, с. 9], яка розв'язує теоретичні і прикладні завдання мовознавства за допомогою комп'ютера.

Комп'ютерна лінгвістика визначається як широка галузь використання комп'ютерних інструментів для моделювання функціонування мови в різноманітних умовах, ситуаціях, проблемних сферах, а також застосування

комп'ютерних моделей мови у лінгвістиці та суміжних дисциплінах [5, с. 13]. Вона розв'язує теоретичні й прикладні завдання мовознавства за допомогою комп'ютера.

Об'єктом дослідження комп'ютерної лінгвістики є мова в усіх трьох формах свого існування: 1) мовна система; 2) мовлення; 3) мовна діяльність [47, с. 30]. Предметом вивчення є будова, зміст та особливості функціонування мовних одиниць як результатів мовленнєвої та мовної діяльності людини, які можуть бути підґрунтям для їхнього моделювання і використання у процесі комп'ютерної обробки інформації [47, с. 8]. Метою комп'ютерної лінгвістики є забезпечення автоматизації інтелектуальних функцій і когнітивної діяльності людини, створення систем автоматизованого аналізу, обробки і синтезу мовлення [95, с. 648; 224, с. 595]. Існують три класи проблем, які досліджує комп'ютерна лінгвістика: 1) фундаментальні проблеми, пов'язані з репрезентацією і моделюванням мови і знань; 2) лінгвістичні проблеми штучного інтелекту; 3) лінгвістичне забезпечення прикладних систем, до яких належать системи машинного перекладу [224, с. 595].

Сучасна комп'ютерна лінгвістика розв'язує широке коло завдань з удосконалення комп'ютерних технологій моделювання мовної й комунікативної компетенції людини. У цьому аспекті вона перетинається з теорією штучного інтелекту [23, с. 15], однією з найважливіших і найскладніших проблем якої є розуміння та автоматична обробка природної мови, оптимізація спілкування людини і машини, що вписується в моделювання розумових здібностей людини. Завданням комп'ютерної лінгвістики є зовнішня формалізація системи мови, яка полягає у формалізації моделей мовних одиниць для їхнього представлення у доступній для комп'ютера формі – у вигляді інформаційної моделі [47, с. 15].

Комп'ютерна лінгвістика має міждисциплінарний характер. За об'єктом і предметом дослідження вона належатиме до лінгвістики, адже вивчає мову у різних формах. За інструментом вивчення та процедурами комп'ютерного опису й аналізу мовних одиниць – до інформатики і теорії штучного інтелекту [27, с. 106]. Комп'ютерна лінгвістика є частиною кібернетики, одним із напрямів дослідження якої є створення систем для опрацювання природних мов (комп'ютерних лінгвістичних систем). Цей напрям у кібернетиці повністю збігається з

комп'ютерною лінгвістикою [87, с. 31]. Адже створення комп'ютерних лінгвістичних систем – це такий самий процес, як і створення будь-яких інших систем.

Комп'ютерна лінгвістика є стиковою дисципліною, яка утворилася на основі взаємодії мовознавства з теорією штучного інтелекту, кібернетикою, теорією інформації. Завдання комп'ютерної лінгвістики полягає у дослідженні способів моделювання мови й мовлення як результатів когнітивної діяльності людини з метою створення комп'ютерних лінгвістичних систем, призначених для машинної обробки інформації у системах машинного перекладу і реферування, природно-мовних інтерфейсах, діалогових систем, систем автоматичного розпізнання мови і мовлення.

## **1.2 Основи автоматичної обробки природної мови**

Автоматична обробка природної мови є одним із головних прикладних напрямів комп'ютерної лінгвістики. І. Б. Штерн визначає автоматичну обробку мови як “загал теоретичних і прикладних напрямків лінгвістики, пов'язаних із комп'ютерним опрацюванням текстів” [126, с. 15]; до неї, з-поміж іншого, належить створення моделей та систем машинного перекладу, розуміння природної мови, дослідження та моделювання механізмів синтезу речень.

Дослідження комп'ютерної обробки природної мови пройшло чотири етапи [144; 163, с. 352]: 1) кінець 1940-х – кінець 1960-х років; 2) кінець 1960-х – кінець 1970-х років; 3) кінець 1970-х – кінець 1980-х років; 4) кінець 1980-х років – до сьогодні.

Перший етап зосереджувався на створенні систем машинного перекладу [163, с. 352]. Дослідження систем програмного забезпечення перекладу розпочалися в кінці 40-х років ХХ ст. та здійснювалися математиками й інженерами (У. Вівер, Н. Вінер). Цей період називають романтичною ерою досліджень [47, с. 150].

Поява ідеї машинного перекладу та створення перших систем машинного перекладу послуговували стимулом для розвитку лінгвістичної теорії перекладу [23,



с. 24; 55] та лінгвістичної думки загалом [77]. Є. А. Карпіловська зазначає: “Романтична ера в історії машинного перекладу пройшла під знаком розуміння мови як одного з різновидів математичного числення, однозначного й несуперечливого виведення одних одиниць з інших, створення жорстких і детальних моделей перетворення структур однієї мовної системи в структури іншої мовної системи” [47, с. 151]. Теоретичною основою був підхід до мови як кодової системи. Перші дослідники в галузі машинного перекладу (У. Вівер та Н. Вінер) розвивали теорію математизації мовного матеріалу – формалізації мови на основі математичних прийомів, аналогічних прийомам дешифрування [78, с. 22; 166, с. 707]. Поштовхом для створення систем машинного перекладу був меморандум американського фахівця з дешифрування У. Вівера (стаття під назвою “Переклад”) [211], хоча ще у 1936 р. У. Вівер й Е. Бут висунули ідею механічного перекладу, а у 1947 р. розробили першу програму послівного перекладу [224]. У меморандумі У. Вівер виклав концепцію про можливість перекладу з однієї мови іншою за допомогою комп’ютера [211]. Він розглядав проблему перекладу як аспект криптографії. Для цього машині необхідно було тільки розшифрувати код, щоб витягнути закладену в тексті інформацію. У меморандумі було наведено приклад дешифрування та вказано на інваріантність мовних характеристик на різних рівнях: “Можна було здогадатися, що при розшифруванні використовувалася частотність літер і комбінацій літер, які майже не залежать від того, якою мовою написаний зашифрований текст. Одразу виникає припущення, що на етапі, коли людина створювала і розвивала мови, виникли деякі інваріантні властивості, які знову ж неточно, проте до певного статистично корисного рівня притаманні всім мовам” [211]. Було розроблено детальний код для послівного перекладу (Е. Бут і Д. Бріттен) та запропоновано правила розбиття слів на основи та закінчення (Р. Річенс) [67, с. 10]. Перша спроба машинного перекладу базувалася на принципі прямого послівного перекладу політичного тексту з російської мови англійською на основі рядів еквівалентів двох мов (Джорджтаунський експеримент під керівництвом Л. Дорстера) [23, с. 174; 79, с. 352; 80, с. 10]. Згодом ідеї У. Вівера стали підґрунтям для підходу до машинного перекладу на основі концепції мови-

посередника (*interlingua*), відповідно до якої стадія передачі інформації поділялася на два етапи: спочатку речення перекладається на мову-посередник (створену на базі спрощеної англійської мови), згодом результат перекладу репрезентується засобами мови перекладу.

Значним імпульсом для розвитку машинного перекладу стало створення і широке застосування перших комп'ютерів, які переробляли інформацію за певними правилами чи алгоритмами [78, с. 20]. Саме тому за кілька років після опублікування меморандуму розпочалися серйозні дослідження у багатьох університетах Сполучених Штатів. З цього часу у США створюються колективи розробників машинного перекладу у Массачусетському технологічному інституті, Каліфорнійському університеті, Національному бюро стандартів у Лос-Анджелесі, у Техаському університеті. У СРСР із 1955 р. також розпочинаються роботи у сфері машинного перекладу в Інституті прикладної математики АН СРСР під керівництвом О. С. Кулагіної й І. О. Мельчука, а перші експерименти були проведені у 1956 р. колективом науковців під керівництвом Д. Ю. Панова й І. К. Бельської (з англійської на російську), О. А. Ляпунова, О. С. Кулагіної (із французької на російську), Т. М. Молошної (із російської на англійську) [95, с. 687].

Радянські лінгвісти, зокрема І. К. Бельська, були першими, хто запропонував ефективнішу концепцію “лінгвістичного” машинного перекладу. Стало зрозуміло, що мова значно відрізняється від коду неоднозначністю своїх одиниць та зміною їх значень у контексті [77, с. 81–82]. І. К. Бельська провела детальне дослідження тих лінгвістичних закономірностей тексту, які застосовує кваліфікований перекладач у своїй діяльності. На основі аналізу матеріалу вона створила робочий алгоритм автоматичного лінгвістичного аналізу англійського науково-технічного тексту [78, с. 22–23]. Дослідниця доводила можливість повної формалізації лінгвістичного аналізу будь-якого тексту, що було значним зрушенням порівняно з досвідом Л. Достерта.

У 50-ті роки ХХ ст. дослідження зосереджувалися на ідеї створення повністю автоматичного високоякісного перекладу, який би здійснювався шляхом бінарного трансферу або через універсальну мову-посередник [162]. У цей період переважала

розробка мови-посередника з метою зменшення лінгвістичних складнощів, притаманних кожній природній мові. Створювалася система запису, придатна для однозначного й достатньо зручного представлення змісту тексту в процесі машинного перекладу.

У ролі мови-посередника лінгвісти вбачали англійську мову, есперанто, санскрит, штучну мову [75, с. 173; 161]. М. Д. Андреев запропонував створити мову-посередник на основі граматики й словників різних мов, включивши в словник граматичні правила і лексичні одиниці, універсальні для більшості мов світу, та виразити його машинними символами [78, с. 27]. Проте такий підхід не став ефективним, адже, як стверджує Ю. В. Рождественський, “жодна штучна мова не може достовірно передати широкий спектр слів, виразів та значень, що постійно виникають в процесі мовлення” [93, с. 91]. Відомий ізраїльський філософ-логік І. Бар-Хіллель ще на початку 1950-х років звернув увагу на відмінність між штучною та природною мовами та висловив сумніви щодо створення повністю автоматичного високоякісного машинного перекладу [147, с. 309–311]. Це породило значні розчарування серед дослідників та ідеї про неможливість високоякісного повністю автоматичного машинного перекладу, що міг би конкурувати з “традиційним” перекладом.

Радянські вчені теж ставили за мету втілити ідею повністю автоматичного високоякісного перекладу, для якого слід було розробити логічну універсальну метамову. Їхні концепції базувалися на припущеннях Л. Єльмслева і Н. Хомського про те, що природна мова є численням, як штучні мови математики і логіки. На думку М. Д. Андреева, метамова повинна враховувати комбінаторні та ймовірнісні обмеження для кожного лексичного, морфологічного та синтаксичного явища [119, с. 108].

Перше покоління систем машинного перекладу базувалося на алгоритмах послідовного перекладу слів та фраз. Можливості таких систем залежали від обсягу наявних словників, що містилися у пам’яті комп’ютера. Переклад тексту здійснювався окремими реченнями, проте без врахування смислових зв’язків між ними. Такі системи отримали назву “системи прямого перекладу” [51, с. 98; 67,

с.10]. На межі 50–60-х років ХХ ст. американські дослідницькі групи не тільки організувати покази “грубого” машинного перекладу, але й виконали масовий послівно-зворотний переклад англійською мовою російських документів технічної, медичної, політичної та інших тематик. Проте до кінця 50-х років ХХ ст. поглибився розрив між теоретичною та практичною спрямованістю на повністю автоматичний високоякісний переклад та на реальні функціонуючі системи, здатні виконувати обмежене коло завдань. Ідея повної формалізації мови виявилася нездійсненною. Головною проблемою на цьому шляху залишалося розпізнання семантики [32]. Потрібно було також вирішити питання про формалізацію та розширення словників, розробку принципів синтаксичного аналізу та дослідження відмінностей між синтаксичними структурами мов. Як зазначає Н. Ю. Зайцева, “стало абсолютно очевидно, що створення метамови або логічного опису мови, що супроводжується формалізацією людських знань та психічних процесів, потребує декількох десятків років роботи сотень лінгвістів, психологів, математиків і програмістів” [36, с. 454].

У 60-х роках ХХ ст. дослідники в різних країнах розпочали пошук альтернативної стратегії та розробку нової інженерно-лінгвістичної технології. Такою стратегією став інформаційно-статистичний та багаторівневий підхід до завдання автоматичної обробки тексту. Одна з центральних ідей розвитку цієї методології полягала в імітації мовленнєвої поведінки людини – поетапному скороченні невизначеності у процесі роботи лінгвістичного апарата, спричиненої багатозначністю лексичних одиниць, морфологічних форм та синтаксичних структур тексту, а також браком мовних та енциклопедичних знань [36, с. 456]. Неоднозначність мала зніматися за допомогою вузької спрямованості системи на конкретну предметну область та ретельного відбору термінів і термінологічних словосполучень в автоматичні словники [192, с. 240–241]. У лінгвістичних дослідженнях назріла необхідність спеціального опису мовних рівнів, формального дослідження синтаксису й семантики, орієнтованих на алгоритмічну обробку та створення спеціальних словників. У цей час почали застосовуватись математичні методи (математична логіка, формальні граматики, теорія графів, теорія інформації,

статистика), що сприяло створенню математичних моделей одиниць мови, мовлення й текстів [87, с. 147]. Проводились статистичні й теоретично-інформаційні дослідження мови й мовлення, було створено ряд статистичних моделей розподілу лінгвістичних одиниць у текстах для потреб машинних словників [93, с. 91]. М. Д. Андрєєв розробив статистико-комбінаторний метод вивчення текстової інформації, що ґрунтувався на врахуванні статистичних та комбінаторних властивостей морфологічних, лексичних та синтаксичних одиниць. Важливе значення мало лінгвостатистичне вивчення спеціальних підмов, що здійснювалося на основі методики, розробленої в групі “Статистика мови” під керівництвом Р. Г. Піотровського [78, с. 17].

Радянські вчені та деякі американські колективи спрямували свої дослідження на розробку систем, що могли видати переклад, який, будучи не зовсім граматично й стилістично коректним, відображав би основний зміст оригіналу. Серед їхніх проектів слід виокремити Джорджтаунську систему (репрезентовану в грудні 1961 року в Євроатомі), систему Systran П. Тома (1964 рік), англо-російську систему для перекладу патентної документації (з 1964 року) [36, с. 455].

У цей час І. О. Мельчук, О. К. Жолковський та Ю. Д. Апресян розробили багаторівневу динамічну модель мови “Смисл $\leftrightarrow$ Текст” для детального опису перетворення смислів у мовні вираження (тексти), семантичного представлення для речення як цілого, а також тлумачення слова як сентенціальної форми [46]. Проте у СРСР наукова стратегія машинного перекладу базувалася на моделі “текст – текст”, що відображалось у створенні систем машинного перекладу двох типів: термінологічного та змішаного (з послівно-зворотним перекладом за допомогою машини).

У середині 1960-х років дослідження були практично призупинені через визнання неекономічності систем машинного перекладу. У 1966 році спеціально створений комітет Національної академії наук США ALPAC зробив офіційний висновок про нерентабельність машинного перекладу в порівнянні зі звичайним перекладом [171]. Підходи попередніх років не були ефективними, оскільки базувалися на принципових помилках. З одного боку, переоцінювався ступінь

близькості операцій комп'ютера до елементарних функцій людського інтелекту. З іншого – недооцінювалася фундаментальна складність природної мови як інструмента інтелектуальної діяльності людини й перекладу як одного з найскладніших видів цієї діяльності [119, с. 105]. Невдалі спроби створення високоякісного машинного перекладу та мови-посередника були спричинені неврахуванням нечіткої асоціативної природи мови [192, с. 237–239]. Комп'ютерний переклад тексту здійснювався під управлінням словника за допомогою однозначних процедур, що цілком підпорядковувались переходу від вхідної мови до вихідної та не використовували експліцитно виділені лінгвістичні правила.

Другий етап досліджень автоматичної обробки природної мови зосереджувався на моделюванні знань про світ та можливості їхньої інтеграції у процес маніпулювання формами представлення значень [163, с. 354]. Незважаючи на формальну орієнтацію та існування трансформаційного парсингу, домінуюча лінгвістична теорія кінця 60-х років ХХ ст., трансформаційна граматики, вважалася нерелевантною для комп'ютерного аналізу. Дослідники цього періоду вбачали семантичну обробку мови у розробці технологій шляхом залучення відмінкової граматики Ч. Філлмора, теорії концептуальної залежності Р. Шенка, семантичних мереж Квіліана, преференційної семантики Й. Вілкса. Найбільш ефективним способом вирішення тогочасної проблеми стала процедурна семантика Вудса, яка перенесла граматичну структуру на мови програмування [144]. Такі способи представлення знань пов'язували комп'ютерну обробку природної мови з теорією штучного інтелекту. Розробки цього періоду представлені діалоговими системами ELIZA Дж. Вейценбаума, BASEBALL Дж. Карбонела, системою смислового аналізу тексту LUNAR У. Вудса, системою мовленнєвого керування діями комп'ютера SHRDLU Т. Винограда, проектом SUR (Speech Understanding Research) ARPA, INTELLECT (Harris, 1984) [163].

Проте ці проекти не виправдали очікувань розробників, які намагалися знайти вирішення двох проблем – побудови користувацьких інтерфейсів із залученням спеціальних знань та забезпечення потреб користувачів у діалозі [163]. Перша

проблема привела до розробки модульних архітектур, загальних формалізмів та засобів забезпечення спеціального лексикону, семантики та моделі баз даних на довершення до стандартного синтаксису, а також моделі підмови для обробки текстів. Друга проблема стимулювала дослідження постановки цілей та планів користувача, необхідних для динамічної та масштабної взаємодії з експертними системами для консультування та спрямування команд.

На третьому етапі досліджень домінував граматично-логічний підхід, у рамках якого почала активно досліджуватися теорія машинної граматики та її взаємозв'язок з логікою представлення значень і знань [144; 163]. Стало зрозуміло, що неможливо створити передбачувані та масштабні системи обробки природної мови навіть для обмеженого застосування. Граматично-логічний підхід став реакцією на помилки побудови практичних систем, його виникнення стимулювало протягом 1970-х років розвиток граматичних теорій, а також спрямування науки на використання логіки для представлення знань та формування умовиводів у теорії штучного інтелекту. Дослідження зосереджувалися на декларативному підході та уніфікації як фундаментальному процесі, які співвідносилися із логічним програмуванням. Аналіз природної мови проводився у процесі складної синтаксично залежної інтерпретації в логічних формах. Такий підхід привів до розробки потужних процесорів речень, як SRI Core Language Engine, Alvey Natural Language Tools, та аналізу значення у рамках теорії репрезентації дискурсу, яка запропонувала засіб опрацювання більш широкого дискурсу в граматично-логічному контексті [163].

На сучасному етапі автоматична обробка природної мови використовується для пошуку та виділення інформації для розпізнавання, аналізу та тегування текстів, у довідкових системах (Google, Bing), діалогових системах та системах машинного перекладу [144]. Починаючи з 1980-х років і до сьогодні дослідження зосереджуються більше на лексиці, ніж на граматиці, та відзначаються домінуванням статистичного підходу до обробки даних, який реалізується у процесі аналізу (семантичної класифікації) даних та ймовірнісного парсингу.

На початку 90-х років ХХ ст. виник та почав активно розвиватися новий емпіричний підхід [201, с. 513], який базувався на обробці паралельних корпусів текстів. У цей час з'явився новий теоретико-методологічний напрям у машинному перекладі – переклад “на основі прикладів” (example-based machine translation) та його варіанти – переклад “за аналогією” (analogy-based machine translation) і переклад “за зразками” (case-based machine translation) [187; 201; 202; 212]. Їхньою спільною рисою є використання корпусу або бази даних оригіналів і перекладених зразків та здатність систем до самонавчання і формування синтаксичних правил. При цьому вхідний текст порівнюється з вмістом бази даних та виділяються відповідні приклади, які поєднуються за аналогією для створення правильного перекладу.

Машинний переклад на основі прикладів вперше був запропонований ще в 1981 році [185], проте почав розроблятися тільки з 1990 року. Основна ідея цього напрямку полягає у використанні за основу прикладів існуючих готових перекладів. У цьому контексті машинний переклад схожий до перекладацького інструмента, відомого як “пам'ять перекладів”. Проте, на відміну від останнього, автоматичний процес триває шляхом підбору відповідних фрагментів перекладу та їх поєднання для формування вихідного тексту. Таким чином, процес перекладу складається з трьох етапів: підбору (matching), вирівнювання (alignment) та поєднання (recombination). Приклади можуть зберігатися у формі повністю анотованих деревовидних структур (з експліцитними зв'язками між компонентами в двох мовах), рядів символів, частково впорядкованих множин тощо [201, с. 513–516]. Основним фактором, що впливає на якість роботи системи, є відповідність прикладів. Оскільки використовується реальний корпус наявних перекладів, існує можливість суперечливих прикладів або таких, що частково збігаються, тому їх потрібно редагувати вручну.

Статистичний підхід до машинного перекладу (statistical MT) базується на ідеї моделювання процесу перекладу за допомогою статистичних ймовірностей [157; 167; 169; 180; 201; 207; 215; 216]. Процедура роботи системи охоплює кілька етапів: вирівнювання фраз, груп слів та окремих слів з корпусу паралельних текстів;



обчислення ймовірностей використання кожного слова в реченнях вхідної та вихідної мов; генерування перекладу. Ймовірнісний розрахунок залежить від двох величин: моделі перекладу (translation model) та моделі вихідної мови (target language model). Ускладнення роботи системи може бути спричинене розмаїттям варіантів перекладу, що вирішується шляхом підрахунку ймовірностей для коротких послідовностей слів – n-грам [201, с. 517]. Характерною для цього підходу є повна відсутність лінгвістичних знань, що обмежує можливості вдосконалення системи. Тому група дослідників компанії ІВМ розпочала дослідження з введення в систему лінгвістичних знань на основі емпіричного підходу до статистичного представлення інформації [138; 168], проте результати виявилися невдалими, і експерименти припинилися в кінці 90-х років. Сучасні системи статистичного машинного перекладу містять моделі перекладу на основі слів (word-based translation models) [139; 198], які згодом були вдосконалені за рахунок введення моделей перекладу на основі фраз (phrase-based translation models) [142; 180; 216], синтаксичних конструкцій (syntax-based translation models) [215] та ієрархічних фраз (hierarchical phrase-based machine translation) [141]. На сьогоднішній день працюють системи статистичного машинного перекладу, деякі з них доступні в режимі онлайн [27; 150]. До них належать Babylon, Bing Translator, Google Translate, Prompt, Reverso, Systranet, Targoman, Wordlingo тощо.

За останнє десятиліття поглибилися дослідження у сфері обробки усного мовлення [131; 149; 181; 182]. Хоча в минулому вона вважалася надто складним завданням, вдосконалені технології аналізу мовлення в поєднанні з інноваційним дизайном дозволяються створити ряд систем машинного перекладу, багато з яких залишаються експериментальними розробками.

Становлення автоматичної обробки природної мови пройшло чотири етапи. Перший з них був присвячений дослідженню у сфері машинного перекладу. На другому етапі домінувало вивчення можливості застосування доробку штучного інтелекту до обробки мови. Третій характеризувався граматично-логічним підходом до обробки природної мови. Дослідження четвертого етапу зосереджуються на лексичних і корпусних даних.

На сучасному етапі спостерігається циклічність досліджень комп'ютерної обробки природної мови. Увага вчених спрямована на статистичний підхід до обробки мови та на виділення лінгвістичної інформації з корпусів текстів.

### 1.3 Моделювання у комп'ютерній лінгвістиці

Детальні дослідження мови й мовлення, спрямовані на створення алгоритмів їхньої комп'ютерної обробки, показали, що деякі лінгвістичні мікрооб'єкти (семантичні множники, смислові відношення) важко або неможливо простежити. Цих труднощів не знають класичні лінгвістичні дослідження, завдання яких полягають у вивченні й описі функціонування мови на рівні прямого спостереження, тобто на макрорівні відтворення і сприйняття мови людиною. Для вивчення об'єктів і зв'язків, прихованих від прямого спостереження, застосовують метод моделювання [89; 95; 224], відповідно до якого досліджується не сам об'єкт, а його формалізований і спрощений аналог – модель.

Термін “модель” має в науці різні тлумачення та визначається як: 1) система, яка має мисленнєве представлення, матеріальну реалізацію і здатність заміщати об'єкт дослідження, щоб його вивчення надавало нову інформацію про нього [224, с. 466]; 2) зразок будови мовних одиниць та процесів, у яких вони залучені [47, с. 20]; 3) абстрактний, формалізований, ідеалізований та спрощений знаковий образ об'єкта, за допомогою якого досліджуються його ознаки, структура і способи функціонування [224, с. 467]; 4) штучно створений лінгвістом пристрій (уявний чи матеріальний), який імітує функціонування об'єкта чи перебіг процесу [87, с. 180]. Таким чином, системи комп'ютерного перекладу є моделями перекладацької діяльності людини. Їхнє створення повинне підкорятися загальним закономірностям моделювання у сучасній структурній і прикладній лінгвістиці.

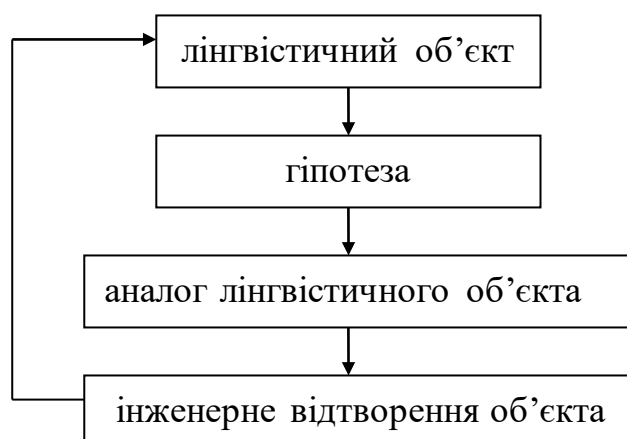
У лінгвістиці розрізняють пояснювальні і прикладні моделі [25, с. 14]. Пояснювальні моделі пов'язані з побудовою і перевіркою гіпотез про структуру мовної реальності. Головним критерієм оцінки їхньої ефективності є адекватність у відношенні до цієї реальності. Прикладні семантичні моделі, до яких належить

переклад, повинні характеризуватися прикладною адекватністю до розв'язання поставленого завдання та мати особливі специфічні риси, а саме: використовувати обмежені дані про мовні рівні; зосереджуватися більше на аналізі, ніж на синтезі; мати високий рівень формалізації та конкретики всіх елементів, пов'язаних з обмеженою підмовою; відображати лише практично необхідні семантичні явища на різноманітних рівнях глибини семантичного представлення, достатніх для розуміння тексту у конкретній ситуації; поєднувати лінгвістичну та екстралінгвістичну інформацію; бути функціональними та експериментально обґрунтованими [1, с. 94; 25, с. 14–15; 69, с. 5].

Для того, щоб модель була аналогом об'єкта, вона повинна бути спрощеним аналогом оригіналу, вільною від суперечностей, універсальною за характером, володіти пояснювальною силою та мати евристичні властивості [90, с. 25–26]. У зв'язку з цим, на відміну від ідеальних моделей, у комп'ютерній лінгвістиці застосовуються відтворювальні інженерно-лінгвістичні моделі. За визначенням Р. Г. Піотровського, моделі цього типу є штучно створеними формальними системами, побудова та поведінка яких імітує мікроструктуру чи функціонування лінгвістичного об'єкта та дозволяє бодай частково відтворити його [89, с. 8]. На відміну від ідеальних лінгвістичних моделей, відтворювальна модель будується за чотиричленною схемою [89, с. 9] (див. рис. 1.1), яка охоплює сильний зворотний зв'язок, функцію якого виконує відповідність між інженерним відтворенням і лінгвістичним об'єктом. Ступінь цієї відповідності є ефективним засобом перевірки самостійності лінгвістичної гіпотези і цінності всієї побудованої на її основі відтворювальної інженерно-лінгвістичної моделі.

За визначенням Л. Л. Нелюбіна, інженерно-лінгвістична модель – це комплекс, який охоплює лінгвістичну модель мови й відповідне інженерне забезпечення – алгоритми і програми обробки на комп'ютері [80, с. 17]. До найбільш типових відтворювальних інженерно-лінгвістичних моделей належать алгоритми автоматичної обробки текстів, які використовують у системах машинного анотування, реферування, перекладу, аналізу і синтезу усного мовлення і в людино-машинних діалогових системах штучного інтелекту. Відтворювальне

інженерно-лінгвістичне моделювання є найбільш розвинутою і завершеною формою “мікромоделювання” мови і механізмів текстотворення.



*Рис. 1.1* Схема відтворювальної інженерно-лінгвістичної моделі [89]

Моделювання перекладу вписується у моделювання інтелекту людини та її мисленнєвої діяльності [47, с. 153]. При цьому моделюється вся діяльність перекладача – від розпізнавання вхідного тексту до синтезу вихідного тексту. Для цього модель повинна складатися з послідовності етапів аналізу і синтезу, кожен з яких оперує певною інформацією та передає результат її обробки для наступного етапу. Алгоритми перекладу є лінгвістичною операцією перетворення різнорівневих мовних даних, які змінюються від етапу до етапу [71, с. 56]. Комп'ютерне моделювання перекладу пов'язане з тим, що природна мова замінюється машинною мовою – штучною знаковою системою, яка функціонує в комп'ютері і моделює дії людини у процесі виконання інтелектуального завдання. За словами Р. Г. Піотровського, “особливості побудови і функціонування пам'яті електронно-обчислювальної машини визначають спосіб побудови штучного машинного знака або, іншими словами, машинне моделювання лінгвістичного знака. Якщо в мозку людини знак записується у вигляді симультанної означуваної єдності і його прагматики, то в його машинній моделі складові знака виявляються роз'єднаними і записуються у вигляді ланцюжка наступних один за одним машинних кодів” [89, с. 26].

У штучних інформаційних системах використовуються штучні знаки, які поєднують ім'я і десигнат [71, с. 46]. У ролі імені виступає матеріальний носій інформації у вигляді набору цифр, літер або інших символів, роль десигната виконує лексико-граматична інформація, введена у комп'ютер. Проте специфіка природної мови перешкоджає повній її формалізації. Річ у тому, що переважна більшість лінгвістичних множин мають іншу природу, ніж ті множини, якими оперує теорія і практика програмування. Це проявляється у трьох особливостях лінгвістичних множин: толерантній організації, потенційній нескінченності і нечіткості (розмитості) їх меж. Саме тому об'єктом моделювання виступає не мова, а тексти як результати переважно письмового мовлення, які мають лінійну, одномірну структуру та передають відповідний смисл [1, с. 91].

Розробка комп'ютерних програм для автоматичної обробки текстів природною мовою передбачає опис лінгвістичних властивостей текстів у вигляді моделей мови з подальшим їх використанням у процесі роботи систем. Розрізняють структурну, редуковану та структурно-статистичну моделі мови [1, с. 94–95]. У рамках першої з них перетворення тексту проходить ряд етапів (графематичного, морфологічного, синтаксичного, семантичного і прагматичного аналізу й синтезу), у результаті яких у процесі аналізу кожне речення перетворюється у внутрішнє представлення його смислу і навпаки у процесі синтезу. Прикладом редукованої моделі є статистична модель обробки інформації, яка відображає частотність символів та їх сполучень (біграм, триграм). Така модель використовується для виявлення друкарських помилок, визначення мовної належності текстів, частиномовної належності слів, розв'язання лексичної неоднозначності [191; 197]. У структурно-статистичних моделях при обробці окремих рівнів мови враховуються статистичні дані про слова чи синтаксичні конструкції [1, с. 95].

У системах комп'ютерної обробки мови модульного типу на етапах аналізу та синтезу тексту використовуються відповідні моделі морфології, синтаксису та семантики [1, с. 95–96; 156]. Морфологічні моделі аналізу передбачають обробку словоформ та розрізняються за результатом роботи, методом аналізу, можливістю обробки словоформи лексеми, не внесеної до словника [29]. У процесі роботи

аналізується лема чи основа за набором морфологічних характеристик відповідної словоформи з опорою на словник словоформ чи словник основ або за допомогою безсловникового методу. У процесі морфологічного синтезу початковими даними є лексема і конкретні морфологічні характеристики відповідної словоформи. Результати морфологічного аналізу та синтезу неоднозначні.

Моделювання синтаксису характеризується різноманіттям методів та способів представлення синтаксичної структури речення. У рамках генеративного підходу синтаксичний аналіз виконується із застосуванням контекстно-вільної або мережевої граматики. У першому випадку у результаті послідовного лінійного членування описується фразова структура речення у формі ієрархічного дерева складових. Мережеві граматики передбачають аналіз речення на основі поняття кінцевого автомата, зокрема розширеної мережі переходів [1, с. 95–96; 48, с. 30].

Для представлення синтаксичної структури речення також використовуються дерева залежностей. При цьому порядок слів та синтаксичні зв'язки між ними аналізуються окремо [1, с. 96]. Проте спільною для обох випадків проблемою залишається представлення однорідних членів речення.

Моделі семантики мови найменш розроблені у комп'ютерній лінгвістиці. У процесі семантичного аналізу речень використовуються відмінкові граматики і семантичні відмінки [1, с. 96]. У цьому контексті семантика речення описується через зв'язок дієслова, як головного слова, з його семантичними актантами [137, с. 40]. Для моделювання семантики тексту використовуються формули підрахунку предикатів та семантичні мережі [66].

Моделювання перекладу розвивається не тільки шляхом уточнення чи розширення знань про мову чи текст, а й через ускладнення комплексу процедурних складових системи. Г. Сомерс розглядає системи машинного перекладу в рамках двох підходів – традиційного лінгвістичного та нового емпіричного [201]. Традиційний підхід до моделювання перекладу представлено у моделі трикутника Б. Вокуа (рис. 1.2) [74, с. 324; 75, с. 173; 207, с. 389]. Сюди належить розробка систем прямого машинного перекладу, систем машинного

перекладу на основі правил (машинного перекладу через інтерлінгву та за допомогою трансферу).

Новий емпіричний підхід виник у 90-х роках ХХ ст. та сформував системи статистичного машинного перекладу та гібридні системи. При цьому еволюціонувала не окрема галузь досліджень, а загальна схема роботи систем.

У перших системах використовувався “прямий” підхід до моделювання перекладу. Переклад розглядався як нетворчий рутинний процес, суть якого полягала у пошуку відповідників у словнику. Перетворення тексту оригіналу в текст перекладу розглядалося як операції міжмовного переходу [119, с. 105]. У кінці 50-х років ХХ ст. було започатковано підхід до роботи систем машинного перекладу на основі правил (rule-based approach), який охоплював дві стратегії перекладу – через інтерлінгву та за допомогою трансферу [119; 154, с. 336–337; 201].

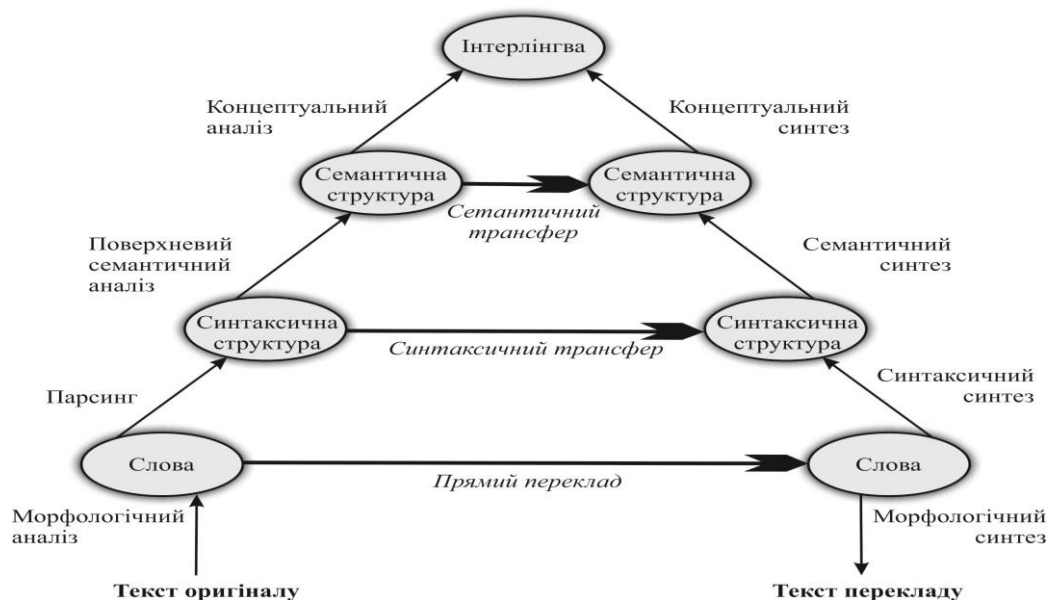


Рис. 1.2 Трикутник Б. Вокуа

Відповідно до першої стратегії процес перекладу складався з двох етапів: концептуального аналізу, під час якого відбувався перехід від тексту оригіналу до його семантичного представлення, та синтезу, у результаті якого на основі семантичного представлення формувався текст перекладу [119, с. 108].

Невдовзі стала очевидною неоптимальність такої стратегії, оскільки неможливо було створити універсальну мову-посередник. Окрім того, було зрозуміло, що людина працює іншим чином, тому в процесі машинного перекладу виникали смислові зсуви та помилки, спричинені засобами вираження та структурою тексту оригіналу. Дослідники почали розглядати існування та можливість введення у системи проміжного рівня представлення текстів, на якому відбувався власне переклад – заміна одиниць вхідного тексту їх перекладними еквівалентами. Таким чином, модель перекладу стала тричленною, складалася з аналізу, міжмовних операцій та синтезу із залученням граматики безпосередніх складників, граматики залежностей, уніфікаційних і лексико-функціональних грамастик [119, с. 110]. Останній вид доповнень до загальної тричленної схеми роботи системи машинного перекладу в її сучасних інтерпретаціях полягає у залученні “знань про світ” на доповнення до існуючих лінгвістичних даних та створенні систем на основі онтологічних знань (knowledge-based machine translation systems) [205; 213], що наближає їх до систем штучного інтелекту.

Сьогодні можна спостерігати перехід до моделей машинного перекладу, у яких пріоритет має не повнота охоплення всіх принципово можливих варіантів обробки, а можливість побудови необхідного варіанта без перебору або з мінімальним перебором хибних” [119, с. 114]. Розробляються системи, які містять однозначні стратегії із “упакованим” представленням, стратегії аналізу за “тактикою найкращих здогадок” (best guess policy) із застосуванням статистики та різних евристик для врахування ймовірнісних характеристик структур.

Моделювання мови загалом та перекладу зокрема, таким чином, повинно базуватися на уявленнях про мовний знак, інформаційні процеси, способи роботи інтелекту людини, які були б орієнтовані на достатньо формалізоване представлення знань і процесів обробки інформації. Очевидно, що в цьому випадку відтворенню може піддаватися не весь процес перекладу, а достатньо представницька його частина. Або ж, навпаки, відтворенням може бути весь процес, проте для достатньо специфічних умов або початкових даних. У цілому природна мова може бути оцінена як складний об’єкт для моделювання.



## 1.4 Становлення та подальший розвиток теорії штучного інтелекту

Сучасний етап науково-технічної революції ставить перед людьми широке коло завдань, які досить складно вирішити та формалізувати. Їх алгоритмічне розв'язання або не існує, або неможливе за сучасних умов. Проте вчені можуть здійснити їх аналіз за допомогою методів теорії штучного інтелекту, які моделюють розумову поведінку людини та сприяють дослідженню глибинних, прихованих від прямого спостереження об'єктів і зв'язків.

Термін “штучний інтелект” запропоновано у 1956 році на семінарі в Стенфордському університеті, присвяченому розробці логічних завдань [23, с. 5]. Проте ідея створення машини, схожої до людського розуму, з'явилася задовго до цього. Вперше її висловив Р. Луллій у XIV ст. Він намагався створити машину для вирішення різних завдань на основі всезагальної класифікації понять. У XVIII ст. Г. Лейбніц і Р. Декарт незалежно один від одного розвинули цю ідею, запропонувавши універсальні мови класифікації всіх наук [23; 62].

Штучний інтелект став повноправною науковою дисципліною з появою перших комп'ютерів у 40-х роках XX ст. та сформувався на основі доробків програмування, нейрофізіології, математичної логіки, психології, лінгвістики тощо [14, с. 5]. Теоретичним підґрунтям виникнення теорії штучного інтелекту став принцип зворотного зв'язку (Н. Вінер) і теорія діяльності людського мозку (У. Маккалоч, У. Піттс), які спонукали до формування парадигми про схожість мозку людини і функцій комп'ютера [23, с. 5; 190, с. 15]. Ці ідеї лягли в основу кібернетичного або нейромодельного підходу до моделювання нейронних мереж від найпростіших істот до людського мозку [14; 23].

У 1950 році у статті “Обчислювальні машини і розум” А. Тюрінг запропонував спеціальний тест для перевірки розумності машини, відомий під назвою “Тест Тюрінга”. У цей період було встановлено, що комп'ютери можуть керувати однаково успішно символами і числовими даними. Це сприяло розробці програм, призначених для розв'язання головоломок, математичних задач, доведення теорем, та ігрових програмних засобів (програма для гри в шашки

А. Самюеля) [31, с. 10–12; 62, с. 11]. У 1954 році під керівництвом академіка О. А. Ляпунова почав працювати семінар “Автомати й мислення”, у якому брали участь відомі дослідники у галузі фізіології, лінгвістики, психології та математики [14, с. 5–6].

Улітку 1956 року Ф. Ньюелл, Г. А. Саймон і Дж. К. Шоу продемонстрували свою розробку – комп’ютерну програму “Логік-теоретик”, здатну доводити теореми із символічної логіки. У 1957 році вони створили програму “GPS” (універсальний розв’язник задач) для пошуку рішень у просторі станів. Хоча їхні розробки не набули широкого застосування, вони привели до виникнення нової парадигми теорії штучного інтелекту – створення комп’ютерних програм, які могли моделювати логічні міркування людини [14, с. 6].

У 1960-ті роки відбувся стрибок у розвитку теорії штучного інтелекту у результаті прогресу у комп’ютерних технологіях та збільшення кількості розробок у цій галузі. Найбільш успішними виявилися дослідження зображення знань, побудов іграшкових світів, за допомогою яких тестувалися ідеї зі створення комп’ютерного зору, робототехніки та обробки природної мови. У цей період тривали розробки нейронних мереж [62, с. 12]. Так, було створено першу нейронну систему розпізнавання образів – перцептрон Розенблатта. Його робота базувалася на спрощеній моделі мозку людини у вигляді множини штучних нейронів, з’єднаних відповідно до існуючої на той час схеми взаємодії клітин людини. Проте ці дослідження зазнали критики у роботах М. Мінські та С. Пейперта, що значно підірвало інтерес до них. Тому новою панівною парадигмою стало створення комп’ютерних програм, які моделювали інтелект людини [14, с. 5].

У 60–70-х роках ХХ ст. виникла нова парадигма штучного інтелекту – евристичне програмування, яке полягало у розробці стратегій дій на основі задалегідь заданих правил (евристик) для розв’язання інтелектуальних завдань різних типів [14; 31, с. 16; 72, с. 11]. Значну увагу дослідники приділяли розробці евристичних моделей мислення, які втілювалися у розробці різноманітних програм індуктивного навчання (“Арки” П. Уінстона), самопрограмування бази знань з патологій рослин (“AQ11” Р. Михальського), навчання на власному досвіді у світі

елементарних чисел (“АМ” та “Евриско” Д. Б. Лената) [14, с. 6–7]. Розроблені моделі мислення стали основою для їхнього застосування у системах машинного перекладу мов, системах реферування текстів та системах відтворення текстів і музики. Зокрема, у системах машинного перекладу застосовувались логічні моделі співвіднесення словоформ мов оригіналу та перекладу. Відбувся поділ знань системи на дві частини – декларативну та процедурну [31, с. 16]. До першої частини належав словник, морфологія та синтаксис, до другої – програмне забезпечення комп’ютерних систем. Внесення змін до однієї частини не впливало на роботу іншої.

Суттєвий прорив у розвитку практичних застосувань систем штучного інтелекту відбувся, коли на зміну пошукам універсального алгоритму вирішення завдань прийшла ідея моделювати конкретні знання фахівців-експертів (Е. Фейгенбаум), яка знайшла своє вираження у розробці експертних систем – комп’ютерних програм, здатних приймати рішення у конкретній ситуації на рівні кваліфікованого фахівця. У результаті цього виникла нова парадигма – спеціалізація знань [14, с. 6]. Перші експертні системи Д. Лендерберга (“Dendral”) і Е. Шортліффа (“Mycin”) відрізнялися від традиційних систем опрацювання даних символьним способом подання інформації, виводу на знаннях і евристичним пошуком рішень завдань.

1970-ті роки ознаменувалися появою і подальшим удосконаленням людино-машинних систем, які поєднували інтелект людини й можливості комп’ютерної техніки для досягнення поставленої мети [31, с. 17]. За допомогою методу резолюцій Дж. Робінсон та його колеги створили формально-логічні моделі мислення. На підставі цього А. Колмрауер розробив мову логічного програмування “Prolog” (1973 рік), яка працювала з правилами та фактами [14 с. 6; 23, с. 7–8]. У 1979-х роках Д. Ленет із Стенфордського університету створив програми, які дали змогу відкрити нові теорії у галузі математики, зокрема у теорії чисел. Вперше було винайдено та застосовано нечітку логіку [62, с. 12]. У США були створені інтегральні роботи, які працювали у лабораторних умовах.

У 1970-х роках у рамках теорії штучного інтелекту проводились дослідження зі створення систем машинного перекладу на основі знань [214, с. 2]. Стало зрозуміло, що звичайні граматичні правила не були достатніми для отримання якісного результату. З цією метою для представлення природної мови у програмних засобах намагались використовувати логічний та лінгвістичний методи вираження змісту речення, які б відображали процес роботи людини. Проте не було створено жодної промислової системи, яка б містила бази знань. Незважаючи на це, дослідження у сфері машинного перекладу сприяли розвитку теорії штучного інтелекту [92, с. 70–71]. Так, аналіз особливостей мови стимулював відокремлення процедурних та декларативних знань, створення семантичних моделей для інтерпретації об'єктів та явищ за допомогою мовних засобів. Пошуки мови-посередика показали необхідність розробки мови для внутрішнього представлення знань.

У 1980-х роках почали відроджуватися дослідження нейронних мереж, що дістали застосування в процесі вирішення різноманітних проблем, зокрема в розпізнаванні мови та у з'ясуванні можливості самонавчання машини [62, с. 12]. Було оголошено глобальну програму розвитку нових технологій “Esprit” [31, с. 18], яка охоплювала проблематику створення комп'ютерних систем штучного інтелекту та інтелектуальних комп'ютерів.

У 1990-х роках дослідження зосереджувалися на розробці слабкого штучного інтелекту (широкого діапазону технологій штучного інтелекту), на відміну від сильного штучного інтелекту (створення програмного забезпечення, завдяки якому комп'ютери зможуть думати так само як і люди) [62, с. 13]. Елементи штучного інтелекту були інтегровані в низку додатків, серед яких системи автоматичного планування, системи видобутку даних з баз даних, системи персоналізації тощо.

В Україні теорія штучного інтелекту розпочалася із досліджень М. М. Амосова і В. М. Глушкова [14, с. 4]. М. М. Амосов вивчав моделювання мислення та психіки людини. В. М. Глушков розробив та реалізував на практиці автоматизовані системи управління. Сучасний розвиток теорії штучного інтелекту підтримується низкою науково-дослідних інститутів, серед яких Державний

університет інформатики й штучного інтелекту та Інститут проблем штучного інтелекту.

На сучасному етапі в теорії штучного інтелекту застосовується нейромодельний підхід до дослідження інтелекту як ідеального типу та розробки нейрокомп'ютерів, оскільки інтелектуальні системи необхідно створювати як системи подання знань та автоматизованих міркувань [14, с. 75; 44, с. 45]. Напрямки фундаментальних досліджень також охоплюють розробку інтелектуальних систем, природно-мовних інтерфейсів, систем машинного перекладу, розуміння та генерації природної мови, обробки та розпізнавання візуальної інформації, ігри й машинну творчість тощо [37; 47, с. 112].

Таким чином, теорія штучного інтелекту пройшла довгу історію розвитку. Кожен період можна охарактеризувати конкретним панівним підходом. Сформований на початковому етапі розвитку кібернетичний чи нейромодельний підхід змінили дослідження, спрямовані на створення комп'ютерних програм, які моделюють інтелект людини. У 60–70-х роках ХХ ст. широко використовувалося евристичне програмування. На сучасному етапі застосовується нейромодельний підхід до дослідження інтелекту людини.

### **1.5 Провідні поняття теорії штучного інтелекту в різних наукових інтерпретаціях**

Сучасна теорія штучного інтелекту досліджує широке коло завдань з моделювання розумової поведінки людини з метою створення комп'ютерної програмно-апаратної моделі інтелекту людини, здатної відтворювати різні прояви людини мислити і розв'язувати певні логічні завдання. Для цього застосовується цілий комплекс понять, які склалися в науках гуманітарного циклу.

Слово “*інтелект*” має латинське походження (від лат. *Intellectus* “поняття, розум, пізнання, спостереження”) [47, с. 107]. У тлумачному “Словнику української мови” терміну “*інтелект*” має таку дефініцію: “Розум, здатність людини думати, мислити; // Рівень розумового розвитку” [Словник української мови, с. 35–36].

Є. А. Карпіловська зазначає, що лінгвістичні словники тлумачать *інтелект* як субстанцію або здатність мозку керувати діями, рішеннями та поведінкою людини в певних ситуаціях [47, с. 107]. Інтелект також визначають як базову категорію когнітивної науки [224, с. 202], алгоритм дії свідомої біологічної системи [122, с. 47], здатність правильно реагувати на нову ситуацію [23, с. 12] та прогнозувати стан зовнішнього середовища [31], обсяг знань в певній області разом та здатність виробляти нові факти і знання [124, с. 18].

Було чимало спроб дати формальне визначення поняттю *інтелект*. *Інтелект* – це: а) алгоритм дії біологічної системи, яка знаходиться у свідомому стані, здатна самостійно спілкуватися з навколишнім світом, набувати досвіду, здійснювати постановку і розв’язання завдань відповідно до обраної мети [122, с. 12]; б) здатність людини прогнозувати стан зовнішнього середовища, досягати успіху під час пошуку однієї з багатьох можливих цілей у різноманітті зовнішнього світу, вміння перетворювати виниклу думку у визначену реакцію, що веде до поставленої мети тощо [31, с. 19]; в) здатність до пізнання, навчання, логічного мислення, систематизації інформації через її аналіз та критичну оцінку, а також до отримання нових знань [224, с. 202]

У 2000 році редакцією журналу “Штучний інтелект” було ініційовано дискусію з визначення поняття “*природний інтелект*” у телеконференціях системи Internet. Було висловлено різноманітні погляди і запропоновано понад 100 визначень цього поняття [31, с. 19]. Спільним для них є те, що *інтелект* розглядається як здатність людини до розумної й раціональної поведінки у проблемних ситуаціях, з метою отримання нових знань. На думку В. К. Фінна, якщо розглядати інтелект як сукупність здібностей, то можна вказати такі принципи: виділення істотного в знанні, здатність до міркувань, рефлексія, висунення мети і вибір засобів її досягнення, пізнавальна активність, адаптація до ситуації, формування узагальнень і навчання на прикладах, синтез пізнавальних процедур [116].

На сьогодні існує чимало визначень терміна “*штучний інтелект*”. *Штучний інтелект* – це: а) автоматизація розумової поведінки [62, с. 7]; б) алгоритм роботи

високоорганізованої матерії, яка характеризується індивідуальністю і вмінням отримувати, зберігати та опрацьовувати вхідні дані, синтезувати знання та приймати рішення [122, с. 12]; в) розробка інтелектуальних комп'ютерних систем, які володіють можливостями, традиційно пов'язаними з людським розумом, – розуміння мови, здатність до навчання, міркування, вирішення проблем тощо [31, с. 23]; г) властивість інтелектуальної системи здійснювати творчі процеси, притаманні людині [223, с. 48]; д) властивість штучних (інтелектуальних) систем виконувати функції, які імітують інтелектуальну діяльність людини і традиційно вважаються людською прерогативою [23]; е) комп'ютерна програмно-апаратна модель інтелекту людини, що здатна брати на себе функції мислення, вибирати й приймати оптимальні рішення на основі аналізу впливів навколишнього середовища і попереднього досвіду, навчатися на основі особливої організації пам'яті, робити висновки на основі суджень тощо [31, с. 23].

Розрізняють дві гіпотези штучного інтелекту, які називають гіпотезами сильного і слабого штучного інтелекту [124, с. 21]. Відповідно до першої з них, інтелектуальна комп'ютерна система повинна не лише моделювати розум людини, а й бути його повноцінним замінником. У контексті другої гіпотези програма розглядається як засіб задоволення окремих практичних потреб людини, який у процесі роботи не застосовує увесь спектр інтелектуальних здібностей.

Одним з головних напрямів штучного інтелекту є створення інтелектуальних систем, заснованих на збереженні знань. Як зазначає В. П. Бурдаєв, *знання* – це сприйнята живою істотою інформація із зовнішнього світу, яка має суб'єктивний характер. Вони є результатом розумової діяльності людини, спрямованої на узагальнення її досвіду, отриманого в результаті практичної діяльності [14, с. 16]. Знання формують систематизовану інформаційну основу інтелектуальних систем, яку можна поповнювати та використовувати для створення нових знань [23, с. 32]. Розрізняють декларативні знання (факти), процедурні знання (правила) та метазнання [23, с. 12]. Д. О. Поспелов визначає шість чинників, існування яких вказує на те, що йдеться про *знання*, а саме: внутрішня інтерпретованість,

структурованість, зв'язність, шкалювання, семантична метрика та наявність активності [92, с. 70].

У теорії баз даних під *знаннями* розуміють сукупність понять, організовану в певну структуру за допомогою відповідних відношень [23, с. 12]. При такому підході розрізняють поняття “знання”, “дані” та “інформація”. Знання пов'язані з даними, ґрунтуються на них, але є результатом діяльності мислення людини, узагальнюючи її досвід, отриманий під час виконання будь-якої практичної діяльності. Вони отримуються емпіричним шляхом.

Знання закладаються в інтелектуальні комп'ютерні системи у вигляді програми чи бази даних. Дані тлумачать як інформацію, яка використовується для опису об'єктів, процесів, явищ та їхніх властивостей [14, с. 17]. Д. П. Шишков визначає *дані* як абстрактне представлення інформації за допомогою літер певного алфавіту [124, с. 13]. Вони є формальним об'єктом, продуктом писемності. У комп'ютерній системі дані – це математичні символи, рядки нулів та одиниць (двійкова система), над якими виконуються тільки формальні операції заміщення.

*Обробка даних* – це опосередкована обробка інформації [124, с. 13], яка відображається в алгоритмах, що складаються з набору операцій, та у програмах. Дані самі собою не несуть інформації і можуть передавати її лише при наявності механізму інтерпретації даних або механізму декодування. *Інформація* – це не просто нові знання, а зміна знань, розширення уявлень спостерігача відносно мети та шляхів її досягнення [124, с. 12–13]. Це слід конкретної або абстрактної взаємодії між матеріальними структурами, явищами, процесами, моделями (включаючи поняття, ідеї) у природі і суспільстві. Синонімом інформації є знання про об'єкт, явище або процес у реальному або абстрактному світі [124, с. 8].

Штучний інтелект є предметом дослідження прикладної комп'ютерної науки, яка займається вивченням і автоматизацією розумової поведінки з метою створення комп'ютерної програмно-апаратної моделі інтелекту, що може моделювати різні прояви здатності людини мислити і розв'язувати певні логічні завдання. Розробка комп'ютерних інтелектуальних систем здійснюється на основі нових досягнень у



дослідженнях механізмів пізнання, пов'язаних з процесом опрацювання інформації в мозку людини.

У системах штучного інтелекту функції спілкування людини з самою системою виконує лінгвістичне забезпечення або лінгвістичний інтерфейс, оскільки вважається, що мовна компетенція належить до сфери інтелекту. Основним завданням сучасних дослідників є створення такого штучного інтелекту, який би міг спілкуватися з користувачами природною мовою та опрацьовувати тексти з такою ж ефективністю, як людина. Проте поки що наука не знає всіх механізмів мислення людини, методів її творчості, механізмів мовного спілкування. Тому створення штучного інтелекту в наш час перебуває на стадії дослідження й створення невеликих експериментальних систем, а також удосконалення комп'ютерних лінгвістичних систем.

## **1.6 Системи програмного забезпечення перекладу в аспекті теорії штучного інтелекту**

Проблематика систем програмного забезпечення перекладу охоплює не тільки аспекти програмування, визначення оптимальних конфігурацій системи, а й лінгвістичні проблеми: вирішення питань семантичної теорії та неоднозначності мовних одиниць, автоматизоване представлення контексту, подолання толерантної організації, потенційної нескінченності та нечіткості меж лінгвістичних множин природної мови для потреб машинного перекладу.

Незважаючи на застосування різних підходів до розробки систем перекладу [7; 8; 39; 68; 69; 70; 71; 89], на сьогодні день не існує єдиного вирішення проблеми створення універсальної автоматичної програми перекладу.

Термін *“системи програмного забезпечення перекладу”* не має єдиного визначення. На позначення програмного забезпечення, яке виконує переклад, у літературі застосовуються синонімічні терміни *програмне забезпечення перекладацької діяльності* [42], *системи машинного перекладу*, *комп'ютерні перекладацькі інструменти* [161], *програми автоматичного перекладу текстів*

[67]. Так, українська дослідниця Т. І. Коваль подає визначення окремих видів програмного забезпечення перекладацької діяльності – системного і прикладного [42]. *Системне програмне забезпечення* – це набір програм, які виконують базові функції для організації процесу пошуку та обробки даних, обслуговування компонентів комп'ютера і надають користувачеві зручні способи діалогу. *Прикладне програмне забезпечення* – це комплекс прикладних програм, за допомогою яких перекладач виконує управлінські, творчі, перекладацькі, навчальні та інші завдання. До прикладного програмного забезпечення належать текстові процесори, графічні редактори, системне управління базами даних, електронні таблиці, веб-редактори, системи автоматичного перекладу, електронні словники та інші [42, с. 27].

При визначенні систем програмного забезпечення перекладу важливе значення має одиниця перекладу. Таким чином, система програмного забезпечення називається “*системою машинного перекладу*” (machine translation system), якщо вона приймає на вході повні речення однією мовою й одночасно генерує відповідні повні речення іншою мовою (не обов'язково виконуючи переклад високої якості) [161]. Система програмного забезпечення називається “*перекладацьким інструментом*” (computer-based translation support tool), якщо вона приймає на вході одне слово або фразу одночасно і синтезує перекладні еквіваленти, які користувач може вкласти в речення і тексти [161].

*Програми автоматичного перекладу текстів* – це спеціальне програмне забезпечення, що використовується для комп'ютерного перекладу – автоматичного перекладу тексту з однієї природної мови іншою [67]. При цьому рівень якості перекладу залежить від програмного забезпечення та структури тексту.

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

Термін “*машинний переклад*” вживається для позначення комп’ютерної програми, яка виконує переклад з однієї природної мови на іншу з або без допомоги людини [161; 162]. Системи машинного перекладу розглядають як один з трьох основних типів засобів комп’ютеризації перекладу (computerized translation tools), до яких також належать системи людино-машинного перекладу і термінологічні банки даних [82]. Розрізняють *системи машинного перекладу за участю людини* (human-aided machine translation [161], human-assisted machine translation [82]) та *системи людського перекладу за участю машини* (machine-aided human translation [161], machine-assisted human translation [82]). Такі системи узагальнено називають системами людино-машинного перекладу [82]. Дж. Хатчінс зазначає, що системи людського перекладу за допомогою машини охоплюють комп’ютерні засоби перекладу, які допомагають перекладачам у процесі роботи, надаючи доступ до онлайн словників, дистанційних термінологічних банків даних, пересилання та прийому текстів, баз попередньо перекладених текстів (пам’яті перекладів) та інтегрованих ресурсів, які прийнято називати “автоматизоване робоче місце перекладача” (translator workstations or translator workbenches) [160].

Е. І. Корольов визначає системи машинного перекладу як один з важливих типів автоматизованих систем обробки текстової інформації [59]. Системи машинного перекладу і перекладу за допомогою машини розглядаються Л. Л. Нелюбіним з позицій використання технічних засобів, рівня взаємодії людини і машини, а також ступеня застосування комп’ютера у процесі роботи. Залежно від залучення технічних засобів до процесу перекладу системи машинного перекладу можуть бути механізованими, автоматизованими і автоматичними [78]. Механізована система передбачає наявність технічних засобів, що розширюють можливості обробки інформації. Автоматизована система перекладу працює на основі чітко внормованих дій із застосуванням комп’ютера та інших інструментів, що забезпечують мінімальне використання можливостей людини і машини. Автоматична система виконує переклад без участі людини. У цьому контексті існуючі методики машинного перекладу – лексичного, граматичного, семантичного – є етапами його автоматизації, що дозволяє здійснити лише

автоматизовану і/або напівавтоматичну обробку тексту в умовах двомовної ситуації. Система є повністю автоматичною лише за умови аудіовізуального вводу і виводу адекватно конфронтованої інформації з першоджерела без участі людини [78, с. 16]. Тому термін “система автоматичного машинного перекладу” у своєму повному і прямому значенні характеризує найвищий і досі недосяжний рівень.

З позицій рівня взаємодії людини і машини, системи машинного перекладу є автоматизованими, оскільки комп’ютер забезпечує мінімальне використання можливостей людини (перед- і постредагування). Системи перекладу за допомогою машини, у яких комп’ютер виступає в ролі помічника, консультанта перекладача, тобто є тим технічним засобом, який лише розширює можливості обробки інформації людиною, є механізованими системами.

Залежно від ступеня використання комп’ютера у процесі перекладу, машинний переклад виконується практично повністю без участі людини, оскільки перед- і постредагування не переривають його, тому процес машинного перекладу є автоматичним. Під час перекладу за допомогою машини комп’ютер використовується як допоміжний засіб, оскільки різноманітні правки є його невід’ємними частинами. Таким чином, переклад за допомогою машини є автоматизованим процесом [78, с. 16–18].

Системи програмного забезпечення перекладу досліджуються з позицій теорії штучного інтелекту. Ю. В. Рождественський розглядає системи автоматичного перекладу як середню ланку між базовими і складними системами штучного інтелекту [93, с. 87], що містять систему тлумачення тексту для моделювання змісту тексту та словник для довідок редактора. Таким чином, системи автоматичного перекладу з’єднуються з моделюючими та експертними системами.

Системи машинного перекладу призначені для виконання перекладу без втручання людини, хоча при цьому не виключається попередня підготовка тексту (розмітка текстів, зняття різних видів омонімії) та постредагування (що є звичним при перекладі людиною). Проте процес перекладу – з моменту введення тексту оригіналу і до виведення тексту перекладу – повністю здійснюється системою машинного перекладу без участі людини [82, с. 360].

Когнітивно-інтерпретаційний аспект систем програмного забезпечення перекладу полягає в моделюванні та автоматизації інтелектуальних функцій і когнітивної діяльності людини, адже машинний переклад є одним з напрямків теорії штучного інтелекту [14; 21; 23; 31; 39; 46; 47; 93; 98], в основі якої лежить переконання про можливість моделювання на комп'ютері когніції людини та механізмів обробки інформації нею з метою досягнення схожості між моделлю та людськими процедурами.

Основними завданнями в цьому контексті є аналіз і синтез тексту природною мовою, які моделюють різні рівні особистості, її когнітивні процеси, розуміння, інтерпретацію і типи мислення.

Машинний переклад вважається вищим видом моделювання перекладу, який може пояснити й експлікувати приховані механізми цього виду мовної діяльності людини [69, с. 190]. Проте вивчення мисленнєвих операцій, що визначають розуміння і вибір мовних засобів у процесі перекладу, є складним завданням. Його вирішення пов'язане зі значними труднощами, оскільки передбачає втручання в мисленнєву діяльність перекладача, дослідження його когніції як прояву розумових, інтелектуальних здібностей. У процесі роботи людина використовує певну стратегію перекладу, яка забезпечує ефективний розподіл її когнітивних можливостей [28, с. 27–28]. Оскільки когнітивні процеси перетворення тексту з однієї мови на іншу не піддаються прямому спостереженню, функціонування систем перекладу певним чином відображає спробу формалізації мисленнєвих операцій у свідомості перекладача.

Існують різні способи членування процесу перекладу, який виконує людина і машина. О. В. Зубов виділяє в процесі перекладу тексту людиною 3 основні етапи: 1) осягнення (розуміння) тексту вхідною мовою; 2) інтерпретація тексту вхідною мовою; 3) перевираження тексту вихідною мовою і створення тексту мовою перекладу [39, с. 251–252]. Суть осягнення змісту вхідного тексту полягає в розумінні того, про що йдеться в тексті оригіналу, і може бути дослівним, стилістичним, а також сприйняттям ідейного задуму автора [39]. Процес перевираження тексту з однієї мови на іншу носить виключно творчий характер.

Базуючись на більш широкому матеріалі, Л. Л. Іомдін зазначає, що людський і машинний переклад складаються з двох послідовних завдань: розуміння тексту (осягнення його змісту) і генерації тексту (переведення змісту в слова, тобто створення нового тексту) [40, с. 56].

Переклад також розглядається як послідовність формально-логічних операцій, які доступні обчислювальній машині відповідної потужності. Правда, строго логічні, алгоритмізовані операції можуть супроводжуватися суто людськими інтуїтивно-евристичними діями, що спонукаються когнітивними асоціаціями до здогадок і непередбачуваними осяяннями, які ґрунтуються на тезаурусі когнітивного досвіду [118, с. 23].

У процесі перекладу людині необхідні знання відповідної лексики та граматики, предметного змісту тексту і правил перетворення. Ця інформація використовується у процесі морфологічного аналізу кожного слова речення та синтаксичного аналізу кожного речення вхідного тексту з їх подальшим перетворенням і синтезом мовою перекладу. Людина виконує ці дії, спираючись на знання мови і досвід. На думку О. В. Зубова, “комп’ютер, який здійснює переклад текстів, теж повинен вміти виконувати ті ж самі дії” [39, с. 252–253].

На сучасному етапі когнітивний підхід до моделювання перекладу актуальний для опису як початкової фази процесу перекладу – розуміння, так і для інтерпретації його наступних фаз – зміни коду та видачі готового продукту. Будь-яке розуміння базується перш за все на контекстуальній інтерпретації мисленнєвих моделей людини. При цьому, основне для перекладу поняття інтерпретації трактується як когнітивні дії, об’єктом яких є продукти мовленнєвої діяльності [28, с. 26; 50, с. 29; 76, с. 11]. Інтерпретація виступає як ментальна діяльність, предметом якої є текст, об’єктом – зміст тексту, метою – отримання змісту тексту, результатом – розуміння тексту на різних рівнях [33]. Вона розглядається як когнітивний процес і одночасно результат встановлення смислу мовленнєвих і/або немовленнєвих повідомлень [30, с. 43; 50, с. 29]. Оскільки когнітивні процеси перетворення тексту з однієї мови на іншу не піддаються прямому спостереженню,

функціонування систем перекладу певним чином відображає спробу формалізації мисленнєвих операцій у свідомості перекладача.

Кожний тип систем програмного забезпечення перекладу характеризується певною семантичною силою, тобто має “здатність до *сміслоототоження*, що визначається багатством лексичної і логіко-семантичної системи процесора, і *сміслорозрізнення*, що визначається складністю граматичної системи процесора” [59, с. 9], яка залежить від повноти словників, ступеня розробки процедур аналізу та синтезу. Дослідник зазначає, що засоби сміслорозрізнення більш розвинені, ніж засоби сміслоототоження, оскільки завданням автоматизованих систем перекладу є передача смислу тексту, вираженого мовою оригіналу, засобами іншої мови з максимальним збереженням лексико-граматичної структури речення. Для цього інколи потрібні зміни при перекладі, зокрема у випадках, коли лексико-граматична структура мови оригіналу відсутня або небажана в мові перекладу.

Під час аналізу вхідної мови системи машинного перекладу використовують засоби моделювання механізмів розуміння людини, а в разі синтезу вихідної мови – засоби моделювання механізмів створення тексту. У системах автоматизованого перекладу текстів моделюються властиві людині типи мислення, серед яких перцептивне, когнітивне і практичне (поведінкове) [120]. Перцептивне мислення реалізується в процесі сприйняття тексту системою і його подальшого перетворення в машинне подання на різних рівнях глибини. Когнітивне мислення актуалізується у вигляді обмеженої моделі предметної області та способі взаємодії з нею як з певною базою знань про світ. У цьому випадку застосовують емпіричний підхід, який полягає у використанні готових мовних творів носіїв мови, а також пар мовних творів авторів оригіналів текстів та їх перекладачів для складання лексико-семантичної та граматичної бази знань системи. Суть моделювання поведінкового мислення полягає в способах передачі окремих лексико-граматичних конструкцій з однієї мови на іншу. Таким чином, моделю практичного мислення можуть виступати тезауруси відповідностей слів і словосполучень у двох мовах, можливі обмеження на їх використання, а також правила перенесення в системах відповідного типу [4, с. 86].

Сутність етапів аналізу і синтезу визначається як розуміння перекладачем значення (суми або системи значень) вхідного тексту і вираження цього ж значення (тієї ж суми або системи значень) засобами мови перекладу [6, с. 233]. Р. Шенк виокремлює шість рівнів розуміння мови, серед яких: синтаксичний, семантичний, контекстний, діалогічний, енциклопедичний, культурний [224, с. 9]. Комп'ютерне розуміння слід розглядати як розуміння, обмежене певною метою (“зрозуміти, щоб виконати синтаксичний аналіз речення”), оскільки повне розуміння речення, фрази, тексту поза досить широким контекстом неможливе навіть для людини [39]. Для позначення ефективності системи М. С. Блехман вводить поняття “рівень розуміння тексту”, що позначає таку “переробку тексту, яка забезпечує певну міру користі для користувача, що прагне задовольнити свою інформаційну потребу або інформаційну потребу іншої людини – кінцевого користувача” [11, с. 4]. При цьому він виділяє два нижчих рівня розуміння: нульовий і перший. Нульовий “рівень розуміння” означає, що у вихідному записі відображається все, що є в явному вигляді в оригіналі. При першому “рівні розуміння” вихідний запис містить деяку імпліцитну інформацію, яка в явному вигляді не доступна в оригінальному тексті. Залежно від застосування у роботі системи характеристик різних рівнів мови, розуміння тексту може бути морфологічним, синтаксичним, семантичним і гіперсинтаксичним (коли система “розуміє” відношення між реченнями тексту) [11]. Одна система може комбінувати вищезазначені різновиди розуміння. Рівень розуміння тим вищий, чим більший відсоток наявної в тексті інформації користувач може одержати від інформаційної системи. Максимальним може бути такий рівень розуміння системи, який відповідав би рівню розуміння його людиною-фахівцем у цій галузі.

Система програмного забезпечення перекладу містить предметний і динамічний компоненти [96, с. 69; 123, с. 69], які відповідають декларативним та процедурним знанням людини [14]. Предметний компонент включає лексику, граматику і семантику, а також перекладні відповідники (перекладний компонент) та формує структуру, на яку накладається мережа перекладного компонента. Динамічний компонент відповідає за конструювання вихідного тексту. Основною



одиницею перекладу є речення, проте за необхідності її межі можуть звужуватися або розширюватися.

Кожна програма застосовує свою власну стратегію перекладу, приховану від прямого спостереження пересічного користувача. Саме тому необхідна оцінка якості виконаних машинних перекладів для того, щоб визначити переваги та недоліки системи, а також обрати ту програму, яка найкраще працює у зазначеній мовній парі.

## **Висновки до розділу 1**

Комп'ютерна лінгвістика – напрям прикладної лінгвістики, який досліджує мову, знання та психічні процеси з метою досягнення схожості між інтелектуальною діяльністю людини та комп'ютерними системами, які моделюють мовну й комунікативну компетенції людини.

Автоматична обробка природної мови є одним із основним прикладних напрямків комп'ютерної лінгвістики, який зосереджується на створенні систем машинного перекладу, систем автоматичного індексування і рубрикації, систем автоматичного анотування і реферування, інформаційно-пошукових систем, систем генерування текстів [65, с. 21] тощо.

Спільним для комп'ютерної лінгвістики та теорії штучного інтелекту є метод комп'ютерного моделювання, а також кінцева ціль, яка полягає у розробці комп'ютерних моделей окремих інтелектуальних функцій людини, до яких належить володіння мовою. Метод моделювання застосовується для створення формалізованих і спрощених аналогів лінгвістичних мікрооб'єктів та мисленнєвих процесів людини, які не піддаються безпосередньому спостереженню. У цьому контексті досліджуються мовні рівні з метою створення алгоритмів їхньої комп'ютерної обробки. Перешкодою на шляху до повної формалізації лінгвістичних одиниць є різноманітність природних мов, кожна з яких характеризується специфічною лексикою, морфологією, синтаксисом та способами вираження значень.

Сучасні комп'ютерні моделі мови у системах автоматичної обробки інформації охоплюють структурну, редуковану статистичну та структурно-статистичну моделі, у яких представлені різноманітні підходи до формального опису морфології, синтаксису та семантики у процесі аналізу та синтезу текстів.

Досягнення у дослідженнях машинного перекладу сприяли розвитку теорії штучного інтелекту, зокрема показали необхідність відділення декларативних знань від процедурних, створення семантичних моделей представлення знань, а також стимулювали поділ процесів аналізу і синтезу інформації на ряд послідовних етапів.

Моделювання перекладу пов'язане з тим, що у комп'ютері природна мова замінюється машинною мовою – штучною знаковою системою, яка функціонує у відтворювальних інженерно-лінгвістичних моделях – штучно створених формальних системах, побудова та поведінка яких імітує мікроструктуру чи функціонування лінгвістичних об'єктів та дозволяє їхнє часткове відтворення.

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

Функціонування систем машинного перекладу відображає спробу формалізації мисленнєвих операцій перекладача. У процесі роботи програми моделюють когніцію, механізми розуміння та створення текстів людиною, її здатність до смислоототожнення та смислорозрізнення, а також властиві їй типи мислення та знання.

Система програмного забезпечення перекладу – це інженерно-лінгвістична модель – комплекс, який охоплює лінгвістичну модель мови та відповідне інженерне забезпечення. Вона містить предметний і динамічний компоненти, які відповідають декларативним і процедурним знанням людини.

Моделювання перекладу розвивається шляхом уточнення чи розширення знань про мову або текст, а також через ускладнення комплексу процедурних складових частин системи. Це проявляється у розробці та вдосконаленні сучасних

систем машинного перекладу на основі правил, статистичних і гібридних систем машинного перекладу. 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.

Основні положення розділу відображено у відповідних публікаціях автора [102; 103; 104; 105].

## РОЗДІЛ 2

### МЕТОДИКА ДОСЛІДЖЕННЯ

### КОМП'ЮТЕРНОГО МОДЕЛЮВАННЯ ПРИРОДНОЇ МОВИ

### В СИСТЕМАХ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ПЕРЕКЛАДУ

#### 2.1 Антиномії комп'ютерної лінгвістики

Інтерес до перекладу та комп'ютерної лінгвістики виник у 50-х роках ХХ ст., коли поява структурно-математичного мовознавства і винахід комп'ютера створили передумови для здійснення різних видів автоматичної обробки тексту, перш за все машинного перекладу. Саме до цього “романтичного” періоду відносяться експерименти з перекладу на комп'ютері заздалегідь підібраних уривків іноземного тексту [47, с. 151]. У цей же період висувалися ідеї створення універсальних семантичних кодів або мов-посередників [78, с. 27; 162], за допомогою яких можна було б отримувати високоякісні переклади з однієї мови будь-якою іншою.

Незабаром, однак, з'ясувалося, що побудова універсального семантичного коду наштовхується на багатозначність і нечіткість лінгвістичного матеріалу. Алгоритми, зорієнтовані на заздалегідь підібрані тексти, виявилися абсолютно непридатними для вирішення початкового практичного завдання, яке полягало у побудові програм, здатних здійснювати хоч і грубу, зате масову обробку потоку науково-технічної та ділової документації [89, с. 45]. Стало зрозуміло, що на шляху лінгвістики, математики і обчислювальної техніки існує бар'єр, який перешкоджає формалізації мови у комп'ютері.

Кожен лінгвістичний об'єкт розглядають або як елемент мовної чи мовленнєвої сукупності (множини), або як безліч інших елементів. Проте переважна більшість лінгвістичних множин має іншу природу, ніж ті множини класичної математики, якими оперує теорія і практика програмування. Специфіка природної мови, що породжує ефект відторгнення, проявляється в трьох особливостях лінгвістичних множин: толерантній організації, потенційній нескінченності і нечіткості (розмитості) їх меж.

На основі згаданих вище розбіжностей Р. Г. Піотровський визначає п'ять парадоксальних антиномій, які виникають при конфронтації природної і штучних мов, у тому числі і мови програмування [88, с. 159–161; 89, с. 47].

Суть першої антиномії – толерантності природної мови – еквівалентності комп'ютерної лінгвістики полягає в невідповідності між лінгвістичними толерантними множинами, позбавленими властивості транзитивності, та еквівалентними множинами, які складають основу мови комп'ютера [88, с. 159; 89, с. 47].

Друга антиномія – антиномія потенційної нескінченності лінгвістичних множин – полягає в тому, що нескінченність і поліморфність лінгвістичних множин поки що не вдається відтворити у програмі, яка використовує, як правило, кінцеві множини [88, с. 161; 89, с. 48].

Третя антиномія базується на безперервності діакронії системи мови [88, с. 161]. При формалізації мови і побудові лінгвістичних автоматів розробники орієнтуються на синхронний опис мови, який збігається в часі з початком розробки програми, до моменту реалізації якої цей опис стає застарілим. Особливо відчутною є дія цього парадоксу при складанні галузевих автоматичних словників, на виготовлення і програмування яких зазвичай витрачають кілька років. Як наслідок, закладений у програму автоматичний словник стає застарілим до моменту свого завершення [18, с. 14; 89, с. 48].

Четверта антиномія – антиномія зворотного зв'язку – виникає на основі традиційного лінгвістичного парадоксу мови та ідіолекту [18, с. 15; 89, с. 49]. Кожен носій мови або група носіїв, володіючи індивідуальним тезаурусом і власною інтерпретацією системи мови, виробляє своє індивідуальне становлення цілей і готовність до ухвалення та розуміння повідомлення. Наявність парадоксу мови та ідіолекту обумовлює багатоцільову спрямованість інформаційного процесу. Для досягнення ефективності спілкування необхідним є існування зворотного зв'язку між відправником і отримувачем повідомлення. У комунікативній системі людина – машина – людина (Л1–М–Л2) створення цього зворотного зв'язку

натрапляє на серйозні труднощі, оскільки передача даних розпадається на два інформаційні процеси:  $L1 > M$ ,  $M > L2$  [89, с. 49].

Оскільки лінгвістичний автомат позбавлений індивідуального мовного і життєвого досвіду та оперує стандартизованим колективним описом мови і зовнішнього світу, то перший інформаційний процес має одноцільовий характер: людина-відправник може завжди однозначно прогнозувати інформаційну готовність і цілі автомата-абонента.

Ситуація змінюється при передачі людині вироблених автоматом результатів. Цей другий інформаційний процес завжди імпліцитно багатоцільовий, оскільки вироблена інформація лінгвістичного автомата адресована абонентам, які володіють різними ідіолектами, професійно-життєвими тезаурусами, цілями і рівнями готовності. Сучасні лінгвістичні автомати майже повністю позбавлені здатності прогнозувати і контролювати багатоцільову спрямованість другого інформаційного процесу і тим самим враховувати інтереси отримувача та адаптуватися до них. Тому ні між людиною-абонентом і лінгвістичним автоматом, ні між абонентом ( $L2$ ) і відправником повідомлення ( $L1$ ) не встановлюється зворотний зв'язок. А це призводить до втрати частини інформації, що міститься в повідомленні, обробленому на комп'ютері, підсилюючи тим самим ефект відторгнення.

Найгострішою антиномією є розбіжності між нечіткістю лінгвістичних об'єктів і чіткістю елементів, які використовують у штучних мовах.

Розвиток комп'ютерної лінгвістики показує, що усунення описаних парадоксів потрібно починати з вирішення антиномії мова – ідіолект. При такому підході, за Л. Заде, всі нечіткі лінгвістичні множини перетворюються за допомогою математичної операції інтенсифікації (згущування) в традиційні чіткі множини [35, с. 39]. До них входять тільки ті лінгвістичні елементи, які мають високий коефіцієнт належності. Проте орієнтування на усереднений опис мови у відтворювальній інженерно-лінгвістичній моделі дає значну втрату тієї інформації, яка пов'язана з новизною повідомлення. Річ у тому, що ця новизна для конкретного абонента

зазвичай закладена в тих лінгвістичних елементах, які, маючи незначну належність, знаходяться на периферії розмитих мовних множин.

Виходячи з описаних антиномій природної і штучної мов, основні зусилля сучасної комп'ютерної лінгвістики повинні бути спрямовані на створення таких систем переробки тексту, які моделювали б інтелектуальні здібності людини розпізнавати новизну (рему) повідомлення. Для розв'язання поставлених завдань створюються відтворювальні інженерно-лінгвістичні моделі [89], які повинні володіти здатністю цілеспрямованого пошуку інформації, мати блок адаптації системи до інтересів живого абонента та можливість самонавчання.

Фундаментальна проблематика комп'ютерної лінгвістики полягає в тому, щоб знайти шляхи подолання або принаймні ослаблення описаних вище відмінностей між побудовою природної мови і організацією штучної мови машини.

## **2.2 Методи та способи дослідження моделювання природної мови у системах програмного забезпечення перекладу**

Сучасна комп'ютерна лінгвістика розв'язує широке коло завдань з удосконалення комп'ютерних технологій моделювання мовної й комунікативної компетенції людини. Одним із її головних прикладних напрямків є автоматична обробка природної мови [224, с. 7]. І. Штерн залучає до цього напрямку створення моделей та систем автоматичного перекладу і лінгвістичних процесорів [126, с. 253]. Автоматичний аналіз мовних структур передбачає аналіз текстів, що є складовою операцією машинного перекладу з однієї мови на іншу. Один з аспектів проблеми вдосконалення систем машинного перекладу полягає у вивченні специфіки процесу перекладу як різновиду комунікації – процесу обміну інформацією та знаннями. У цій царині залишається без відповіді ще ціла низка важливих питань, зокрема досягнення адекватності та еквівалентності перекладу.

Ці питання вже багато років активно вивчаються в лінгвістиці такими мовознавцями, як Л. С. Бархударов, Б. С. Виноградов, В. Н. Комісаров,

О. Д. Швейцер [6; 16; 52; 53; 54; 55; 64; 121] і багато інших, однак досі залишаються малодослідженими в контексті комп'ютерної лінгвістики.

Традиційні терміни “адекватність перекладу” та “еквівалентність перекладу” є багатозначними і суперечливими. У теорії перекладу та лінгвістиці існують різні підходи щодо визначення цих понять та їх класифікації.

*Адекватність перекладу* – це: а) близькість оцінок змісту текстів їхніми адресатами або відповідність поставленій перед перекладачем меті [95, с. 675]; б) забезпечення прагматичного завдання на максимально можливому рівні еквівалентності із дотриманням жанрово-стилістичних вимог і відповідаючи конвенційній нормі перекладу [55, с. 51]; в) відповідність перекладу як процесу певним комунікативним умовам [121, с. 95]; г) відповідність вибору мовних знаків в мові перекладу тому виміру початкового тексту, який вибирається як основний орієнтир процесу перекладу. Це таке співвідношення початкового і кінцевого текстів, при якому послідовно враховується мета перекладу [121, с. 92]; д) відтворення змісту і форми оригіналу засобами іншої мови, що є метою художнього перекладу; переклад, який викликає в іншомовного одержувача реакцію, яка відповідає комунікативній спрямованості відправника; такий переклад, у якому збережені всі наміри автора щодо певного впливу на читача, з дотриманням усіх авторських ресурсів образності, колориту, ритму [83, с. 272].

У роботах А. Б. Федорова та Р. Якобсона була закладена концепція семантико-стилістичної адекватності перекладу, у рамках якої адекватність перекладу зводилася до категорій семантичної (сислової) повноти і точності та доповнювалась стилістичною еквівалентністю, що включає, зокрема, принцип підпорядкування тексту перекладу функціонально-стилістичним нормам мови перекладу [15, с. 37]. Згодом розробка та введення в експлуатацію систем машинного перекладу сприяли формуванню концепції функціонально-прагматичної адекватності перекладу [6]. Відповідно до неї потрібна не повна і точна передача всього смислового змісту та стилістичних особливостей оригіналу, узгоджених з функціонально-стилістичними нормами мови перекладу, але лише правильна передача основної комунікативної функції оригіналу. Машинному перекладу



піддаються не всі тексти, а тільки тексти документів, наукової і технічної літератури та тексти масової інформації та інформатики.

Досягнення адекватності перекладу неможливе без урахування еквівалентності перекладу. Вчені пропонують різні визначення еквівалентності. Наведемо деякі з них. *Еквівалентність перекладу* – це: а) смислова спільність прирівнюваних один до одного одиниць мови й мовлення [55, с. 51]; б) збереження відносної рівності змістової, смислової, семантичної, стилістичної і функціонально-комунікативної інформації, що міститься в оригіналі та перекладі [16, с. 18]; в) відповідність тексту перекладу оригіналу [121, с. 95]; г) збалансоване співвідношення двох найбільш важливих характеристик текстів оригіналу й перекладу: повноти й точності змісту, що передається [95, с. 144]. Еквівалентним є переклад, що здійснюється на рівні, необхідному й достатньому для передачі незмінного плану змісту при дотриманні норм мови [6, с. 186].

До основних концепцій еквівалентності відносять концепцію формальної та динамічної еквівалентності Ю. Я. Найди, концепцію рівнів еквівалентності О. Д. Швейцера, концепцію комунікативної еквівалентності Г. Єгера, види еквівалентності В. Коллера [63, с. 9; 121, с. 78–80], концепцію рівнів еквівалентності В. Н. Комісарова [52, с. 59; 121, с. 81], концепцію “скопос” К. Райса та Х. Вермеєра [53, с. 81–82], “описову” концепцію Г. Турі [53, с. 125–129], концепцію функціональної та функціонально-змістової еквівалентності Л. К. Латишева [63, с. 34–35].

Будь-який адекватний переклад повинен бути еквівалентним змісту оригіналу, однак не кожний еквівалентний переклад можна визнати адекватним, а тільки той, що крім норми еквівалентності відповідає й іншим нормативним вимогам.

Адекватність та еквівалентність перекладу – поняття відносні [16, с. 18]. Їх рівень та специфіка залежать від способу перекладу і типу тексту, що перекладається. Типи текстів визначають підходи і вимоги до перекладу, впливають на вибір прийомів перекладу і визначення ступенів еквівалентності та адекватності перекладу оригіналу. Ступінь наближення до оригіналу залежить від багатьох

факторів – від майстерності перекладача, від особливостей зіставлення мов і культур, характеру перекладених текстів.

В. Н. Комісаров зазначає, що оцінка якості перекладу може вироблятися з більшим чи меншим ступенем деталізації. Для загальної характеристики результатів перекладацького процесу використовуються терміни “*адекватний переклад*”, “*еквівалентний переклад*”, “*точний переклад*”, “*буквальний переклад*” і “*вільний переклад*” [55, с. 233].

Проте такі поняття не зовсім коректно використовувати при оцінці якості машинного перекладу, що набуває все більшого поширення в контексті посилення процесів глобалізації та інформатизації суспільства. Експерименти з оцінки якості машинного перекладу, спрямовані на визначення користі від застосування машинного перекладу та ефективності задоволення специфічних вимог користувача, відбуваються у межах макрооцінки, де на перший план виходить не безпосередня якість перекладу, а ступінь задоволення потреб користувачів, та мікрооцінки, на текстовому рівні.

Машинний переклад, як і традиційний, складається безпосередньо з процесу перекладу, протягом якого здійснюється переклад повідомлення з однієї мови іншою, та результату цієї роботи – тексту перекладу. Говорячи про оцінку процесу машинного перекладу, до уваги беруться всі складові частини системи, яка виконує переклад. Тут до оцінки входять такі чинники, як зручність інтерфейсу системи, кількість словників, напрямки перекладу, вартість системи, швидкість перекладу та інші. У дослідженні зосереджуємось на оцінці кінцевого результату перекладу – тексту, який видає машина.

Кінцевий продукт, тобто перекладений текст, може розглядатися з двох боків: як переклад, виконаний для подальшого використання, та як інформативний переклад для внутрішнього користування. Ю. М. Марчук розглядає градацію перекладів за О. Каде, а саме: “чорновий переклад”, “робочий переклад” та “переклад для публікування”. У вказаній градації машинний переклад розглядається як чорновий [71, с. 171]. Вже на цьому рівні його можна використовувати як сигнальну інформацію. За таких обставин текст не редагують,

сприймають як відносно еквівалентний та частково адекватний переклад і використовують як кінцевий продукт. Останній чинник суттєво відрізняє машинний переклад від традиційного. Перекладів, які людина-перекладач виконувала б лише для себе, не існує. Використання систем машинного перекладу для інформативного ознайомлення зі змістом документу, є на сьогоднішній момент чи не найбільшою сферою застосування подібних систем. Тому здається переконливим, що оцінка машинного перекладу має суттєво відрізнитись від традиційного перекладу.

Розглядаючи наявні способи оцінки якості машинного перекладу, Ю. М. Марчук пропонує при оцінці якості перекладу враховувати еквівалентність перекладу за рівнями, адже в оцінці якості машинного перекладу ігнорується структура досягнення еквівалентності за рівнями “переклад слів” – “переклад словосполучень” – “переклад речень” – “переклад тексту”. Машинний переклад можна досить точно оцінювати на кожному з цих рівнів, навіть включаючи рівень морфем [71, с. 170].

Дослідження якості машинного перекладу охоплює різноманітні методи та способи її оцінки, зокрема суб’єктивну оцінку людини та анкетування інформантів (А. В. Бірюков, Ю. М. Марчук, D. Arnold, M. Federico, C. Girardi, M. L. Glenn, J. M. Lopez, D. Vilar) [9; 68; 132; 148; 152; 153; 176; 210], зворотний переклад (Р. Koehn) [168; 169], порівняння тексту перекладу з текстом оригіналу за набором критеріїв, визначення адекватності та еквівалентності перекладу та оригіналу (Л. С. Бархударов, Ю. В. Ванников, Б. С. Виноградов, В. Н. Комісаров, О. Д. Швейцер) [6; 15; 16; 52; 53; 54; 55; 121], застосування автоматичних метрик оцінки якості машинних перекладів (В. Яковина, G. Doddington, N. Madhani, I. Melamed, J. Olive, M. Snover, L. Specia, C. Tillmann) [129; 146; 183; 188; 199; 200; 203; 204] та систем кількісної оцінки якості перекладу (А. Gerasimov, G. Saldanha, A. Zerfass) [151; 178; 196; 217].

Основні критерії, які виділяють іноземні дослідники для оцінки машинного перекладу, це: читабельність, зрозумілість перекладу; тип тексту; сприйнятливість для користувача; формальна структура, тема-рематична і образна структура, функціональний стиль та прагматична функція; структурно-лінгвістичний критерій,

критерії комунікативного ефекту та джерела помилки; орієнтація на перевірку певного мовного феномена за допомогою штучно створених тестів [6; 9; 140; 153; 164; 165; 194; 196]. Що стосується останнього, то мається на увазі створення текстів, які б містили у собі речення для перевірки правильності перекладу певних синтаксичних конструкцій.

Об'єднати всі ці критерії при оцінці конкретного перекладу людиною дуже важко, оскільки необхідно щоразу виконувати великий обсяг праці для порівняння оригіналу та перекладу. Перераховані вище способи оцінки теж недостатньо ефективні та мають ряд недоліків, а саме: висока суб'єктивність оцінки, низька кореляція з оцінкою людини-експерта, тривалість у часі та складність виставлення інтегральної оцінки. Постає питання розробки такої системи оцінки якості, яка б була універсальною та застосовувалась для різних типів тексту.

На сьогодні все більшої популярності набувають системи кількісної оцінки якості перекладу (QA Distiller, Déjà Vu, SDLX QA Check, Star Transit, Trados QA Checker, Wordfast, ErrorSpy, XBench) [133; 136; 151; 152; 155; 179; 184; 196; 217]. Програми не редагують перекладений текст, а тільки дозволяють визначити кількість помилок у зразку тексту відповідно до заданих параметрів та “відносної” ваги помилок. Вони показують високу ефективність при перевірці термінології, синтаксису, пунктуації, орфографії та оформлення. Тобто тих аспектів мови, які піддаються відносній формалізації. Якість перекладу оцінюється за таким принципом: чим більшою буде частка серйозних помилок, маркованих позначкою “5”, тим нижчим буде рівень еквівалентності перекладу.

Можливість автоматичної оцінки якості систем програмного забезпечення перекладу є важливим аспектом дослідження їхньої якості. Спільним у роботі автоматичних метрик є те, що вони використовують переклад, виконаний людиною, як еталон. Основним принципом оцінки є схожість автоматичного та людського перекладу. Для того, щоб проаналізувати рівень схожості, метрики визначають відсоток часткового збігу слів чи словосполучень між машинним та еталонним перекладом. Більшість метрик можуть працювати з одним або навіть кількома еталонними перекладами, адже речення може мати кілька правильних перекладів.

Для визначення часткових збігів метрики використовують леми замість форм слів та враховують перефрази або імплікації на рівні слів та речень.

Для автоматичної оцінки якості виконаного комп'ютером перекладу застосовують два типи метрик: а) метрики, що визначають обсяг редагування (Edit Rate Metrics); б) метрики, орієнтовані на точність (Precision-Oriented Metrics) [129; 203]. До першої групи належать метрики TER (Translation Edit Rate) [199; 200], WER (Word Error Rate) [188], PER (Position-Independent Word Error Rate) [204] і HTER (Human-Mediated Translation Error Rate) [199].

Метрики, орієнтовані на точність, зосереджуються на лексичній точності – частці лексичних одиниць (зазвичай n-gram різного розміру) в автоматичному перекладі, які трапляються в еталонних перекладах. До них належать метрики BLEU (Bilingual Evaluation Understudy) [189], METEOR (Metric for Evaluation of Translation with Explicit Ordering) [135; 143], GTM (General Text Matcher) [209], CDER (Cover Disjoint Error Rate) [172], STM (Syntax Tree Based Metric) [173–175] тощо.

Усі згадані метрики потребують набору еталонних перекладів, що унеможливило повністю автоматичну оцінку машинних перекладів, які з'являються в результаті роботи систем програмного забезпечення перекладу. На жаль, серед них не існує єдиної універсальної і досконалої метрики, тому вони потребують подальшого доопрацювання. Системи автоматичної оцінки якості перекладу виконують 50% загального обсягу роботи. Вони не враховують граматику, стиль, регістр та прагматичний ефект (зрозумілість, логічність, зв'язність). Також вони не розкривають лінгвістичну природу помилок та не пояснюють причини їх виникнення. Визначення цих характеристик і надання перекладу статусу адекватного залишається прерогативою людини.

Для прагматичних цілей слід проводити лінгвістичний аналіз помилок на рівні тексту, проте одного лише підрахунку їх не достатньо. Більш ефективною буде лінгвістична категоризація помилок, пошук причин їхнього виникнення та визначення способів усунення. Такий спосіб оцінки якості машинних перекладів надасть інформацію про обсяг необхідного постредагування [159, с. 164]. Для

перевірки ефективності роботи систем машинного перекладу та категоризації мовних помилок застосовують тестові збірки [9; 195]. Вони дозволяють дослідити, які лінгвістичні феномени успішно опрацьовуються системою, а які призводять до помилок. Тестові збірки складаються з граматично правильних речень мови оригіналу та містять набір мовних феноменів для перевірки лексичних та граматичних можливостей системи.

А. В. Бірюков [9] провів оцінку якості машинного перекладу систем “Плай 4.0” та “Прагма 4.4” для німецько-українського та російсько-українського мовних напрямків на основі порівняльного аналізу перекладів речень із тестової збірки обох систем. Для цього він розробив детальну класифікацію помилок з визначенням “ваги” кожної окремої помилки та її впливу на загальну якість перекладу. Додатково було визначено ймовірний етап роботи системи, що спричинив конкретну помилку в процесі аналізу, трансферу або синтезу.

Аналіз практичних оцінювань машинного перекладу ще раз доводить, що на сьогодні поки що відсутні єдині критерії оцінки, неоднакові й підходи при побудові самої концепції її проведення. Тому пропонуємо лінгвістичну оцінку якості машинного перекладу за мовними рівнями на матеріалі живих текстів, яку можна застосовувати для дослідження будь-якої системи програмного забезпечення перекладу.

### **2.3 Специфіка методики дослідження комп’ютерного моделювання природної мови у системах програмного забезпечення перекладу**

Аналіз систем програмного забезпечення перекладу полягає в перевірці їхньої лінгвістичної потужності, тобто можливості обробки певних мовних явищ шляхом “сліпого” перекладу реальних текстів у рамках англо-українського та англо-російського напрямків перекладу. При цьому аналізуються система Pragma 6.3 Business, яка є системою машинного перекладу на основі правил, а також он-лайн системи статистичного машинного перекладу (Google Translate, Babylon, Free Translation, Yandex) та гібридного машинного перекладу (Bing Translator).

У дисертаційному дослідженні запропоновано методику оцінки якості систем програмного забезпечення перекладу, що ґрунтується на підходах до оцінки якості машинних перекладів А. В. Бірюкова, М. С. Блехмана, Ю. М. Марчука, В. О. Новікова, П. М. Хроменкова [9; 10; 11; 68; 81; 117], а також типології помилок у машинних перекладах А. В. Бірюкова, В. О. Новікова, П. М. Хроменкова [9; 81; 117] та мовних помилок Т. Г. Бондаренко, М. А. Куніловської, В. Я. Мельничайка, О. П. Сосніної [13; 61; 73; 97], які порушують норми літературної мови. Оцінка якості проводиться відповідно до стратегії мікрооцінки [9], яка передбачає випробовування лінгвістичної потужності системи шляхом перевірки тексту перекладу на наявність помилок, виявлення причин їх виникнення та можливостей усунення.

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.

Запропонована методика аналізу систем програмного забезпечення перекладу дозволить визначити найбільшу кількість помилок та рівень, на якому системи роблять найсуттєвіші помилки. Врешті-решт все це допоможе обрати найкращу систему, яка найкращіше виконує переклад у межах англо-українського мовного напрямку. Це дозволить також віднайти архітектурні недоліки систем як певний вид



зворотного зв'язку для розробників з метою виправлення помилок та подальшого покращення якості машинного перекладу.

## **Висновки до розділу 2**

Моделювання природної мови у системах автоматичної обробки інформації ускладнюється характеристиками мови, які формують п'ять антиномій комп'ютерної лінгвістики: антиномію толерантності природної мови, антиномію потенційної нескінченності лінгвістичних множин, антиномію синхронії – діахронії, антиномію зворотного зв'язку та антиномію нечіткості лінгвістичних об'єктів. Для їхнього подолання повинні створюватися такі відтворювальні інженерно-лінгвістичні моделі, які б могли цілеспрямовано шукати необхідну інформацію, мали блок адаптації системи до інтересів користувача та володіли здатністю до самонавчання.

На сучасному етапі розвитку мовознавчої науки простежується тенденція до застосування різноманітних методів та способів оцінки якості машинних перекладів як результатів комп'ютерного моделювання природної мови у системах програмного забезпечення перекладу, до яких належить суб'єктивна оцінка людини й анкетування інформантів, зворотний переклад, зіставлення тексту перекладу з текстом оригіналу, застосування автоматичних метрик оцінки якості машинних перекладів та систем кількісної оцінки якості перекладу. Проте не існує єдиної методики оцінки якості систем програмного забезпечення перекладу, яка б пояснювала лінгвістичну природу помилок у машинних перекладах та можливі причини їхнього виникнення.

У дисертаційному дослідженні запропоновано методику оцінки якості систем програмного забезпечення перекладу, відповідно до якої виявляються помилки на різних мовних рівнях.

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

На другому етапі аналізуються машинні переклади 44 текстів системи Pragma 6.3 Business, із яких було відібрано десять найбільш проблемних для перекладу тематик. Вони слугуватимуть матеріалом для подальшого вивчення систем Google Translate, Bing Translator, Babylon, Free Translation та Yandex. На цьому етапі застосовуються спеціальні лінгвістичні методи – структурний метод (зокрема його методики аналізу за безпосередніми складниками, дистрибутивного аналізу та аналізу словникових дефініцій) та зіставний метод.

Третій та четвертий етапи пов'язані з оцінкою якості роботи систем Google Translate, Bing Translator, Free Translation, Babylon та Yandex в англо-українському та англо-російському напрямках перекладу. На третьому та четвертому етапах застосовується метод зіставлення машинних перекладів та відсоткових співвідношень помилок з метою визначення найбільш ефективної системи машинного перекладу серед систем Pragma 6.3 Business, Google Translate, Bing Translator, Babylon та Free Translation.

Розроблена методика дозволить провести оцінку якості машинних перекладів текстів, виконаних системами програмного забезпечення перекладу, як результатів інтерпретаційної діяльності програми, перевірити повноту закладених у систему декларативних та ефективність процедурних знань.

Основні положення розділу відображено у відповідній публікації автора [99].

## РОЗДІЛ 3

### СИСТЕМИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ПЕРЕКЛАДУ ЯК ДОСВІД МОДЕЛЮВАННЯ ПРИРОДНОЇ МОВИ

#### 3.1 Система Pragma 6.3 Business

Система Pragma 6.x належить до українських програмних засобів для перекладу, які працюють на основі правил. Це пояснюється її модульною структурою. Програма складається з базового модуля (обов'язкового) та мовних модулів для 8 мов. Завдяки цьому зміни всередині модуля, наприклад поповнення словника, не повинні впливати на загальний процес роботи системи. Система має 56 напрямів перекладу і підтримує близько 50 тематик [58]. Pragma 6.3 Business є найновішою версією програми, яка підтримує три мови – англійську, російську та українську – з можливістю перекладу в шести напрямках: англо-російський, російсько-англійський, англо-український, українсько-англійський, російсько-український, українсько-російський [193]. Pragma 6.3 Business є політематичною. Словник, включений до її складу, має тематичну розмітку та охоплює 44 тематики.

Робота програми ґрунтується на численних інтегрованих лінгвістичних правилах вхідної та вихідної мов, а також двомовних словниках для кожної мовної пари. Модульна структура Pragma 6.3 Business дозволяє припустити, що вона є системою трансферного типу [68; 159; 186]. У ній текст мовою оригіналу спочатку проходить морфологічний та синтаксичний аналіз, потім етап трансферу, який завершується процесом синтезу кінцевого варіанта перекладу через процедуру генерування [75; 208]. У процесі морфологічного аналізу відбувається лематизація лексичних одиниць, їхній пошук у словнику, аналіз морфем, розпізнавання граматичних характеристик слів тощо. Під час синтаксичного аналізу система визначає типи синтаксичних структур, реляційні зв'язки між окремими елементами структури [75]. На етапі трансферу створюється проміжне представлення проаналізованого тексту, з якого на основі правил вихідної мови будуються синтаксичні структури синтезованих речень [59, с. 13]. Таким чином, для

ефективної роботи систем машинного перекладу на основі правил потрібні “знання” граматики, синтаксису, семантики та лексики мовних пар [41], які містяться в одномовних модулях для аналізу та генерування речень і модулях трансферу [208]. При цьому важливу роль відіграє ступінь розходження між структурами мов, що впливає на якість перекладу. Правила та словники розробляються лінгвістами та програмістами, що потребує тривалої кропіткої праці. Для покращення роботи системи користувач може поповнювати уже наявні словники та створювати свої власні. Проте такі доповнення можуть забирати багато часу та інколи погіршувати якість перекладу через неоднозначність термінів. Система також може робити граматичні помилки, оскільки розроблені правила не мають універсального характеру, адже складно врахувати всі винятки та випадки їхнього застосування.

### **3.2 Системи статистичного машинного перекладу Google Translate, Babylon, Free Translation, Yandex**

Основою роботи статистичних систем машинного перекладу є автоматична екстракція схожих сегментів мовних пар з двомовних повнотекстових корпусів, доступних у мережі Інтернет. Робота системи базується на статистичних методах і теорії ймовірності. Перекладач визначає величину ймовірності кожного перекладу і на основі цього, відповідно до моделі мови, виконує вибір з численних можливих варіантів. До програм статистичного машинного перекладу, які працюють он-лайн в англо-українському та англо-російському напрямках перекладу, належать Google Translate, Babylon, Free Translation та Yandex.

Google Translate – безкоштовна он-лайн служба компанії Google для автоматичного перекладу слів, речень, текстів та веб-сторінок. Перекладач підтримує 90 мов. У процесі роботи Google Translate пропонує альтернативні варіанти перекладу. Вибір користувачем одного з можливих еквівалентів сприяє навчанню програми. При перекладі окремих слів чи фраз Google Translate пропонує словник із визначенням слів, прикладами їхнього використання, антонімами та

синонімами. Існує можливість збереження перекладів слів та фраз у розмовнику, доступна опція покращення перекладу у вікні з вихідним текстом та його фонетичного читання. Перекладач підтримує голосове введення оригіналу. Користувач може також запропонувати обраний або власний варіант перекладу, що допоможе покращити якість роботи системи у майбутньому [158]. Ефективність останньої залежить від обсягу доступних їй корпусів текстів. Проте готові переклади, виконані людиною чи машиною, інколи бувають далекі від ідеалу. Це впливає на якість перекладу Google Translate, яка не має лінгвістичних правил і тому може робити граматичні помилки.

Babylon – он-лайн-перекладач, який працює на основі моделі SaaS (Software-as-a-Service) компанії Language Weaver [170]. Програма виконує багатомовний автоматичний переклад слів, фраз, текстів тощо та підтримує 75 мов. Для пошуку термінів використовується база даних, яка містить понад 1600 словників та глосаріїв з різноманітних сфер. Перекладач також пропонує пошук визначення, синонімів та антонімів термінів [134]. Програма надає послуги голосового виводу інформації, завдяки чому користувач може прослухати правильну вимову слів і текстів.

Free Translation – сайт для перекладу текстів та веб-сторінок у режимі он-лайн, розроблений компанією Transparent Language Inc. [206]. Програма підтримує 43 мови та використовує у роботі платформу перекладу природної мови TranscendRT для перекладу текстів, документів (у форматах .doc(x), .ppt(x), .odt, .pdf, .txt обсягом до 5MB) та веб-сайтів. Система містить опцію для миттєвого перекладу та прослуховування вихідного тексту.

Yandex – веб-сервіс компанії Яндекс, призначений для перекладу текстів та веб-сторінок на основі технології фразового статистичного машинного перекладу [130]. Програма підтримує 44 мови та має опції автоматичного визначення мови, синхронного перекладу, підказок при наборі тексту та перевірки правопису [85]. Окрім повнотекстового перекладу, сервіс забезпечує послугу машинного словника, який містить систематизовані варіанти перекладу, приклади використання слів та словосполучень, транскрипцію для англійських слів та інформацію про частиномовну належність слів. Як і інші системи статистичного машинного

перекладу, Yandex складається з трьох основних частин: моделі перекладу, моделі мови і декодера [130]. Для кожної мовної пари у моделі перекладу зберігається таблиця зі словами і фразами однією мовою та їхніми можливими варіантами перекладу іншою мовою із зазначенням ймовірності виникнення останніх у текстах. Знання системи про мову перекладу містяться у моделі мови – списку всіх вживань слів і словосполучень та ймовірностей їх використання. Процес перекладу відбувається у декодері, який знаходить всі можливі варіанти перекладу та обирає серед них найбільш частотний, який складається з підібраних слів та фраз, скомбінованих на основі моделі мови та відповідно до моделі перекладу.

### **3.3 Система гібридного машинного перекладу Bing Translator**

Bing Translator – це перекладач, що працює на платформі Microsoft Translator [84]. Bing підтримує 45 мов та може застосовуватись для перекладу текстів, веб-сторінок тощо. Проте обсяг тексту не має перевищувати 5000 знаків. Bing Translator поєднує алгоритми статистичного машинного перекладу та лінгвістичні правила [177]. Дослідницька група формувала лінгвістичну складову системи шляхом отримання даних з двомовних корпусів текстів. Згодом був залучений підхід на основі правил, що сформувало гібридний підхід до машинного перекладу у межах системи Bing Translator. У програмі існує опція синтезу мовлення.

### **3.4 Критерії оцінки якості систем програмного забезпечення перекладу**

Дослідження машинних перекладів програм Pragma, Google Translate, Bing Translator, Babylon, Yandex [9; 114; 145; 164], виконаних у різних мовних напрямках, показали наявність у машинних перекладах таких помилок, як неперекладені слова, неадекватний вибір перекладних відповідників багатозначних слів, порушення порядку слів у реченнях, пропущені члени речення, неадекватний переклад та порушення у побудові складних синтаксичних конструкцій.

При вивченні лінгвістичної потужності систем машинного перекладу було виділено лексичні, граматичні, лексико-граматичні та орфографічні помилки, які порушують *норми літературної мови* – сукупність загальноприйнятих, усталених правил, якими керуються мовці в усному та писемному мовленні [38, с. 56; 91, с. 19–20; 115, с. 120] та які впливають на еквівалентність і адекватність машинного перекладу у відношенні 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 збереженні форми оригіналу та якнайменшому порушенні норм мови перекладу.

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

Лексичні помилки порушують норми слововживання у властивих їм значеннях за змістом [38, с. 56] та включають помилки у перекладі слів та словосполучень як семантичних єдностей [86, с. 225].

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 не властивому для контексту та тематики значенні через неадекватне вирішення лексичної полісемії, лексичної та лексико-граматичної омонімії, уживання русизмів, неадекватний переклад власних назв та абревіатур, які слід було залишити в оригінальному варіанті.

Неперекладені слова – це слова, які програма залишила в оригінальному варіанті, оскільки не знайшла їм відповідників у мові перекладу. Вибір

нерелевантних перекладних відповідників – вживання слів у не властивому для контексту та тематики значенні через неправильне вирішення полісемії, лексичної та лексико-граматичної омонімії, вживання русизмів, неадекватний переклад власних назв та аббревіатур, які слід було залишити в оригінальному варіанті.

Полісемія характеризується наявністю кількох значень в одному слові [91, с. 119; 108, с. 48; 125, с. 217]. Багатозначні слова є багатоеквівалентними. Вони мають декілька перекладних відповідників, які залежать від кількості їхніх лексико-семантичних варіантів [45, с. 279]. Такі відповідники називаються варіантними та передають у мові перекладу якесь одне значення слова [70, с. 152]. Варіанти перекладу можуть бути внесені до автоматичного словника чи моделі мови систем машинного перекладу. У такому випадку перед програмою постає завдання обрати правильний варіант перекладу, який має співвідноситись із контекстом речення і тематичним спрямуванням усього тексту.

Лексичні омоніми – це слова, які мають однакові граматичні характеристики, проте різні лексичні значення [12, с. 26; 91, с. 154; 106, с. 123; 107, с. 50; 108, с. 234; 128, с. 185–186]. Лексико-граматичні омоніми – це слова чи їхні окремі граматичні форми, які збігаються формально, проте мають різні лексичні та граматико-категоріальні значення [3, с. 169; 12, с. 29; 94, с. 474; 106, с. 234]. До відступу від лексичних норм належить вживання незасвоєних та невмотивованих русизмів [91, с. 27], які мають в українській мові відповідні лексичні позначення.

*Помилки у перекладі словосполучень* як семантичних єдностей полягають у неадекватному виборі відповідників складених термінів та фразеологічних сполучень у мові перекладу, дослівному перекладі іншомовних висловів, які порушують норми лексико-семантичної сполучуваності слів у межах словосполучень. До них належать: 1) неадекватний переклад словосполучень як семантичних єдностей – уживання словосполучень у не властивому для контексту та тематики значенні через неправильний вибір відповідника в мові перекладу; 2) неадекватний переклад фразеологічних сполучень – переформатування значеннєвого 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.

Фразеологічні сполучення – це семантично подільні фразеологічні звороти, цілісне значення яких випливає із семантики окремих слів з вільним і фразеологічно зв'язаним значенням [12, с. 54; 91, с. 122; 109, с. 339; 112, с. 96; 128, с. 231]. У фразеологічних сполученнях синтаксичні зв'язки відтворюються у структурі готових, закріплених у вживанні фразеологічних одиниць [2, с. 186; 107; 108, с. 139; 125, с. 254].

До оцінки якості систем програмного забезпечення перекладу вводимо поняття *лексико-граматичних помилок* – вживання словосполучень у не властивому для контексту та тематики значенні, а також відхилення від нормативно-узусуального 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) [110, с. 13]. 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; б) порушення категорії способу дієслів; 7) ненормативне утворення ступенів порівняння прикметників; 8) вживання ненормативних форм дієприкметників; 9) вживання ненормативних форм дієприслівників. Граматичні омоніми – це слова, які належать до різних частин мови, проте мають змістову спільність [12, с. 29; 45, с. 17; 106, с. 234].

*Синтаксичні помилки* полягають у порушенні норм побудови словосполучень, простих і складних речень та відсутності синтаксичних зв'язків між компонентами синтаксичних одиниць.

Розрізняють три типи синтаксичних зв'язків: предикативний, підрядний та сурядний [34, с. 18; 127, с. 19]. Предикативний зв'язок формує структурну основу простого двоскладного речення та визначає взаємозалежність головних членів речення. Способом предикативного зв'язку є координація, її особливість – взаємозв'язок підмета і присудка, який виражається формами словами у відмінку, роді, числі. Визначальною категорією підмета є називний відмінок, присудка – актуалізована особа [17, с. 65; 127, с. 20–21].

Підрядний зв'язок – тип синтаксичного зв'язку, який вказує на граматичну залежність одного компонента від іншого у словосполученні чи реченні [127, с. 21]. Формами підрядного зв'язку є узгодження, керування і прилягання [111, с. 66; 112, с. 282–283; 127, с. 21]. Узгодження – форма підрядного зв'язку між компонентами словосполучення, при якому форми залежного слова уподібнюються до граматичних форм головного (стрижневого) слова [112, с. 282; 127, с. 34]. Іншою формою підрядного зв'язку є керування, при якому залежне слово вживається у тій відмінковій або прийменниково-відмінковій формі, якого вимагає головне слово [91, с. 283; 127, с. 35].

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 *помилоч* 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 синтаксичних помилок у побудові складних реченнях 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, що суперечить 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.

### **Висновки до розділу 3**

Програми, які здійснюють автоматичний переклад в англо-українському та англо-російському напрямках, включають систему Pragma 6.3 Business, яка працює на основі правил, системи статистичного машинного перекладу Google Translate, Babylon, Free Translation, Yandex та систему гібридного машинного перекладу Bing

Translator. Дослідження машинних перекладів програм Google Translate, Bing Translator, Babylon, Yandex, виконаних у різних мовних напрямках, показали наявність різних типів помилок

Для оцінки якості роботи систем Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex запропоновано критерії виявлення лексичних, морфологічних, синтаксичних, лексико-граматичних та орфографічних помилок, які порушують норми літературної мови, а також негативно впливають на еквівалентність та адекватність машинних перекладів.

Еквівалентність машинного перекладу – це семантична та структурна 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 збереженні форми оригіналу та якнайменшому порушенні норм мови перекладу.

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

Лексико-граматичні помилки – це вживання словосполучень у не властивому для контексту та тематики значенні, а також відхилення від нормативно-узусуального порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір варіантів перекладу та граматичної трансформації.

## РОЗДІЛ 4

### МОДЕЛЮВАННЯ ПРИРОДНОЇ МОВИ В РІЗНОМАНІТНИХ СИСТЕМАХ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ПЕРЕКЛАДУ

#### 4.1 Оцінка якості системи Pragma 6.3 Business

Оцінка якості системи Pragma 6.3 Business передбачає випробовування лінгвістичної потужності системи шляхом перевірки тексту перекладу на наявність помилок, виявлення причин їх виникнення та можливостей усунення. Матеріалом дослідження слугували укладені нами тестові збірки речень на основі текстів англійською мовою з мережі Інтернет, яка найбільш динамічно реагує на потреби інформаційного суспільства та характеризується жанровим різноманіттям [20; 56; 57]. З кожного тексту, який відповідає 44 тематикам, зазначеним виробником програми Pragma 6.3 Business [230–273], було відібрано по 11 перших речень включно із заголовками. Таким чином, усього проаналізовано 484 пари речень в англо-українському напрямку перекладу.

Результати аналізу помилок у текстах машинних перекладів представлені в таблиці 4.1, більш детальна таблиця подана в Додатку А.

*Таблиця 4.1*

#### Кількісне співвідношення помилок системи Pragma 6.3 Business в англо-українському напрямку перекладу

Типи помилок	Кількість помилок	
1. Лексичні помилки	1759	53,5%
1.1. Помилки у перекладі слів	1218	37,0%
1.2. Помилки у перекладі словосполучень	541	16,5%
2. Граматичні помилки	1423	43,2%
2.1. Морфологічні помилки	465	14,1%

## Продовження таблиці 4.1

Типи помилок	Кількість помилок	
2.2 Синтаксичні помилки	958	29,1%
2.2.1 Помилки у побудові словосполучень	358	10,9%
2.2.2 Помилки у побудові речень	600	18,2%
а) Помилки у побудові простих речень	320	9,7%
б) Помилки у побудові складних речень	280	8,5%
3. Лексико-граматичні помилки	100	3,1%
4. Орфографічні помилки	5	0,2%
Усього	3287	100%

Як видно з таблиці 4.1, у текстах машинних перекладів переважають лексичні помилки – 1759 помилок, що становить 53,5% від загального обсягу помилок. Граматичні помилки є менш численними та становлять 43,2% від загального обсягу помилок. Програма зробила найменше лексико-граматичних помилок – 3,1% та орфографічних помилок – 0,2%.

У таблиці 4.2 показано **лексичні помилки**, які зробила програма Pragma 6.3 Business у перекладі слів та словосполучень.

Таблиця 4.2

**Лексичні помилки системи Pragma 6.3 Business**

Типи помилок	Кількість помилок	
1. Помилки у перекладі слів	1218	69,2%
1.1 Вибір нерелевантних перекладних відповідників	924	52,5%
1.1.1 Неправильне вирішення лексичної полісемії	733	41,7%

## Продовження таблиці 4.2

Типи помилок	Кількість помилок	
1.1.2 Неправильне вирішення лексико-граматичної омонімії	173	9,8%
1.1.3 Неадекватний переклад власних назв та аббревіатур, які слід було залишити в оригінальному варіанті	8	0,5%
1.1.4 Неправильне вирішення лексичної омонімії	5	0,3%
1.1.5 Вибір нерелевантних перекладних відповідників через наявність помилки в оригіналі	3	0,1%
1.1.6 Вживання русизмів	2	0,1%
1.2 Неперекладені слова	294	16,7%
2. Помилки у перекладі словосполучень	541	30,8%
2.1. Дослівний переклад іншомовних висловів	247	14,1%
2.2. Неадекватний переклад словосполучень як семантичних єдностей	190	10,8%
2.3. Неадекватний переклад фразеологічних сполучень	95	5,4%
2.4. Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті	9	0,5%
Усього	1759	100%

Серед лексичних помилок було виявлено найбільше помилок у перекладі **слів** (69,2%), а саме:

1. Вибір нерелевантних перекладних відповідників – вживання слів у не властивому для контексту та тематики значенні через неправильне вирішення

лексичної полісемії, лексичної та лексико-граматичної омонімії, уживання русизмів.

Значення багатозначного слова в кожному конкретному випадку його уживання можна встановити тільки з урахуванням контексту. Проте така операція недоступна програмі перекладу, адже в процесі роботи вона застосовує тільки наявні у словнику відповідники.

Так, термін *area* [230] Pragma 6.3 Business переклала як *область*. У словнику *area* має кілька значень: 1) площа, простір; ділянка; 2) район, область; зона, край; площа; 3) сфера, галузь; 4) розмах, масштаб [218, с. 72]. Проте правильним варіантом при перекладі речення *Revised wing has a 25% larger area, the spread is increased by 4.88 m, which increased fuel capacity by 30%* [230] є *площа*, а не *область*.

У цьому ж реченні термін *spread* [230] був перекладений як *поширення*. У словнику відображено кілька варіантів його значення: 1) поширення; ріст, зростання; збільшення; 2) розмах; відстань; ширина; протяжність; 3) накритий стіл; 4) бенкет, щедре частування; 5) розкіш напоказ; 6) паста, пастоподібні продукти; масло; джем; паштет; 7) покривало; скатерка; простирadlo; 8) газетний (журнальний) матеріал, поданий на розвороті; 9) розворот (книги, газети); 10) різниця, розрив (між цінами, курсами тощо); 11) розсіювання; 12) діапазон відхилень [219, с. 417].

Програма обрала найчастотніший варіант перекладу, проте у цьому контексті йдеться про *розмах* крил, а не їх *поширення*.

У результаті неправильного визначення частиномовної належності слів Pragma 6.3 Business обрала їм неадекватні відповідники в українській мові. Так, у реченні *What measures does your company take in order to develop its employees?* [244] програма переклала іменник *measures* прикметником *зважене*. Проте правильним варіантом є іменник *заходи*, що відповідає контексту речення.

У реченні *Governed by the modern global business models, Corum Group offers its clients the efficient and safe integrated solutions* [257] Pragma переклала



прикметник *efficient* формою іменника *чинник*, проте правильним є варіант *ефективні*.

Програма не розрізняє лексичні омоніми, яких у досліджуваних текстах було виявлено 5 одиниць, та не аналізує контекст у процесі перекладу, тому обирає найчастотніший еквівалент, який не відповідає тематичному спрямуванню тексту. Система переклала слово *pile* [269] як *купа*, проте правильний варіант – *ворс*. Ці слова, за даними словника “Collins English Dictionary”, є гетерогенними омонімами, які виникли внаслідок збігу етимологічно різних слів. Так, слово *pile* зі значенням *купа, велика кількість чогось, громаддя будинків, ядерний реактор, вістря, похоронне багаття, багатство* походить від давньофранцузького та латинського слова *pila*, що означає *кам'яний стовп*. А *pile* – *ворс, вовна, волосся* – від англо-нормандського *pyle*, від латинського *pilus* – *волосся* [229].

Термін *stills* [250] програма переклала як *дистилятори*, проте правильний варіант – *кадри*. Ці слова, за даними словника “Collins English Dictionary”, є гетерогенними омонімами, які виникли внаслідок збігу етимологічно різних слів. Так, слово *still* зі значенням *дистилятор* було запозичене в 16 столітті та походить від давньофранцузького слова *stiller* (*канати*), від латинського *stillare*, від *stilla* – *крапля*. Слово *still* зі значенням *кадр* походить від давньоанглійського слова *stille*; пов'язаного з давньосаксонським, давньогерманським *stilli*, голландським *stolen* *застигати*, санскритського *sthanus* зі значенням *нерухомий* [229].

У тексті машинного перекладу також наявне вживання слів у не властивому для контексту та тематики значенні через помилку в оригіналі. Так, термін *casting* [236] отримав переклад *відливання*. Проте в тексті йдеться про *опалубку*, словесним відповідником якої в англійській мові є слово *casing* [221, с. 86].

Наявність у перекладі русизмів пояснюється використанням російської мови як проміжної ланки при перекладі з англійської мови на українську. Наприклад: *itching* [258] – *зудить*; *scale* [261] – *гамма*. Правильними варіантами перекладу є *свербіж* та *гама* відповідно.

2. Неперекладені слова займають 16,7% від лексичних помилок. Програма залишила без перекладу 294 слова, серед яких: а) терміни: *extraction* [240], *earthfills*

[236]; б) власні назви: *Metinvest* [238], *Howarth* [240] *Azovstal* [259], *Deloitte* [246; 251]; в) скорочення: *Mtonnes* [238; 259], *mmtry* [238], *mm* [259]; г) аббревіатури: *GOK* [238], *DSTU*, *GOST* [259], *LCD*, *DC*, *VDC*, *VAC* [242].

3. Оскільки *Pragma 6.3 Business* не розрізняє власні та загальні назви, вона неадекватно перекладає слова, які слід було залишити в оригінальному варіанті. Наприклад: *SAP* [270] – *СОК*, *SIM card* [270] – *карта СИМА*.

**Помилки у перекладі словосполучень** становлять 30,8% від обсягу лексичних помилок. До них належать:

1. Дослівний переклад іншомовних висловів, які в українській мові мають свої відповідники, через відсутність готового варіанта у словнику системи: *burning sensations* [258] – *палить чутливість (печіння)*; *finely dispersed state* [272] – *дрібно розігнав державу (дрібнодисперсний стан)* [222, с. 425].

Причиною дослівного перекладу словосполучення може також бути наявність помилки в тексті оригіналу: *ready made Italian pasta* [254] – *готові зроблені Італійські макаронні вироби*. Слово *ready-made* пишеться через дефіс, тоді як в оригіналі воно написано окремо. Виправлене словосполучення *ready-made Italian pasta* система перекладає правильно *готова італійська паста*.

2. Неадекватний переклад словосполучень як семантичних єдностей – уживання словосполучень у не властивому для контексту та тематики значенні через неправильний вибір відповідника в мові перекладу. Наприклад: *particulate filter* [271] – *сипкий фільтр (сажовий фільтр)* [221, с. 466]; *essential amino acids* [272] – *істотні амінокислоти (незамінні амінокислоти)* [221, с. 16].

3. Неадекватний переклад фразеологічних сполучень через їхню відсутність у словнику програми.

Програма дослівно переклала сталий зворот *take care about* [263] – *брати турботу про*. Правильним є переклад – *піклуватися про* [218, с. 176; 220, с. 914].

Неадекватний переклад отримало словосполучення *enter into force* у реченні *on the date when the Treaty entered into force* [260] – *на даті, коли Договір увійшов всередину змушують*. Правильний варіант – *набути чинності* [220, с. 305].

Причиною неправильного перекладу фразеологічних сполучень може бути наявність помилки в тексті оригіналу. Наприклад: *Modern woodworking equipment allows to produce very large diameter logs (even more than 30 cm), which is not inferior logs, prepared manually* [239]. – Сучасне деревообробне устаткування дозволяє робити дуже великі колоди (навіть більш ніж 30 см) діаметру, який не є підпорядкованими колодами, готовими вручну.

В оригіналі пропущений прийменник *to*, що впливає на зміст речення. Правильний варіант оригіналу – *is not inferior to logs*, тоді переклад Pragma матиме вигляд – *не підлеглий колодам*, що є неадекватним перекладом. Правильний варіант перекладу – *що не поступаються за якістю колодам*.

4. Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті: *CFM International CFM56-7B* [230] – ПІДТВЕРДЖУЮТЬ Міжнародний CFM56-7B; *Boeing 737 Classic* [230] – Класичний Боїнг 737.

**Граматичні помилки** в текстах машинного перекладу становлять 43,2% від загального обсягу помилок. Програма зробила їх внаслідок неправильної роботи модулів морфологічного та синтаксичного аналізу. Граматичні помилки системи Pragma 6.3 Business представлені в таблиці 4.3.

Таблиця 4.3

### Граматичні помилки системи Pragma 6.3 Business

Типи помилок	Кількість помилок	
	Кількість	Відсоток
1 Синтаксичні помилки	957	67,3%
1.1 Помилки у побудові речень	600	42,2%
1.1.1 Помилки у побудові простих речень	320	22,5%
1.1.2 Помилки у побудові складних речень	280	19,7%
1.2 Помилки у побудові словосполучень	357	25,1%
2 Морфологічні помилки	466	32,7%
Усього	1423	100%

Найчисленнішими серед граматичних помилок є синтаксичні помилки (67,3%), зокрема у побудові простих та складних речень (42,2%).

Найбільше помилок програма зробила у побудові простих речень (Таблиця 4.4), які виникли через неправильно виконаний програмою синтаксичний аналіз речень.

Таблиця 4.4

#### Помилки у побудові простих речень системи Pragma 6.3 Business

Типи помилок	Кількість помилок	
1 Порушення синтаксичних зв'язків між членами речення	161	50,3%
2 Граматична неоднотипність однорідних членів речення	86	26,9%
3 Порушення координації між підметом і присудком	47	14,7%
4 Неправильний порядок слів	16	5,0%
5 Штучне ускладнення речення	4	1,3%
6 Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	2	0,6%
7 Вживання зайвих членів речення	2	0,6%
8 Пропущені члени речення	2	0,6%
Усього	320	100%

1. Найпоширенішою помилкою є порушення синтаксичних зв'язків між членами речення (50,3%):

*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* [231]. – *Нові Panamera S і Panamera 4S* *придатні з 4.8 літром природно вимовив з придином двигун з прямою паливною ін'єкцією, система VarioCam Плюс, бо регулюючі фази поширення газу і змінна впускають колектор.*

Незважаючи на те, що правильно визначені підмет та присудок, порушені синтаксичні зв'язки керування між присудком *придатні* та додатками *двигун, система, колектор*. Причиною цього є неправильний вибір відмінкових форм іменників *двигун, система, колектор*.

Також на адекватність перекладу вплинуло неправильне визначення частиномовної належності прийменника *for*, іменника *regulating*, а також дослівний переклад словосполучень *naturally aspirated engine, a variable intake collector*. У наведеному прикладі наявне штучне ускладнення простого речення, яке проявляється у вживанні сполучника *бо* замість прийменника *для*.

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* [236]. – *Фундамент копає траншею, траншейна земляна робота і риючи яму для ізольованої підтримки, льохів і strip принципів; траншейне створення для заглиблених трубопроводів, перевезення і послабляючи земель, планування втулок, роблячи earthfills і обґрунтовують трамбівку, pile пониження і стук, роблячи pile основи.*

У реченні порушені синтаксичні зв'язки між однорідними присудками *копає, траншейна земляна робота, риючи, траншейне створення, перевезення, послабляючи, планування, роблячи, обґрунтовують, пониження, стук і роблячи* через те, що програма неправильно виконала синтаксичний аналіз речення та визначила частиномовну належність слів *trenching, pitting, loosening, making, ground*, які, окрім цього, отримали неадекватний переклад.

3. Порушення координації між підметом і присудком:

*Bank Lviv will for its part contribute with EUR 2 million for the loan facility* [232]. – *Банківський Львів буде для його частини вносять вклад з мільйоном EUR 2 для засобу кредиту.*

У цьому реченні відсутня координація між підметом *Львів* і присудком *буде вносять вклад* у числі та особі.

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* [239]. – *Готовий поміст має приймач” “приймача” або “половини для connection колод разом, компенсаційне розрізання для зниження подовжніх тріщин, “ліжко” для поклав колоди один на одного.*

Програма неправильно виконала синтаксичний розбір речення та здійснила необґрунтовану граматичну трансформацію перестановки, що вплинуло на адекватність перекладу.

5. Штучне ускладнення речення:

*The unique properties of Graftobian HD Make-up cosmetics allow modern make-up artists working in beauty salons, at shows, on television, in fashion and in movie-making industry to easily observe the principles of harmonious beauty and naturalness* [252]. – *Унікальне майно Graftobian косметик Косметики HD дозволяє сучасних гримерів, працюючих в салонах краси, в показах, на телебаченні, за модою і у виготовленні кінофільму промисловості, щоб легко спостерігати компоненти гармонійної краси і природності.*

Програма трансформувала просте речення у складне, необґрунтовано застосувавши сполучник *щоб*.

6. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення:

*What measures does your company take in order to develop its employees?* [244]. – *Що зважене ваша компанія приймає наказ розвивати його працівників?*

Оскільки програма не аналізує гіперсинтаксичні зв'язки в межах речення, це призвело до неправильного визначення роду іменника **компанія** та вживання займенника **його** в другій частині речення.

7. Використання зайвих членів речення:

*On March 27, 2013 as part of the program of the route network development UTair-Ukraine completed the first flight Kiev – Sharm-el-Sheikh using the newly arrived Boeing 737-800 [130]. – 27 березня, 2013, оскільки частина програми UTair-Ukraine розвитку мережі маршруту завершила перший Київ польоту – Sharm-el-sheikh надземної залізної дороги, користуючись новоприбулим Боїнгом 737-800.*

Програма застосувала зайві слова **надземної залізної дороги**, яких немає в тексті оригіналу.

8. Пропущені члени речення:

*Piracy in the Gulf of Aden, off the Horn of Africa and in the Indian Ocean is undermining international humanitarian efforts in Africa and the safety of one of the busiest and most important maritime routes in the world – the gateway in and out of the Suez Canal [260]. – Піратство в аденській Затоці, від Horn Африки і в Індійському Океані підриває міжнародні гуманітарні зусилля в Африці і безпеку з найзайнятішого і найбільш важливі морські маршрути у світі - вхід в і поза суецьким Каналом.*

У процесі синтезу речення програма пропустила повнозначний член речення – означення **one**, яке відсутнє у машинному перекладі.

Помилки у побудові **складних речень** Pragma 6.3 Business подано в таблиці 4.5.

До помилок у побудові складних речень належать:

1. Порухення 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% [238]. – 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%*

У реченні порушений синтаксичний зв'язок керування між присудком *оцінює* та обставиною *горизонтальний*, оскільки програма неправильно визначила частиномовну належність іменника *level* (*рівень*) та переклала його формою прикметника *горизонтальний*. У другій частині речення наявні два присудки *виділяють* та *оцінюється*, іменник *газ* виступає у ролі додатка стосовно першого з них і підмета стосовно другого. Причиною цих помилок є не тільки неправильно виконаний синтаксичний аналіз та синтез речення, а й неправильне визначення частиномовної належності іменників *level*, *gas*, прикметника *natural*.

Таблиця 4.5

#### Помилки у побудові складних речень системи Pragma 6.3 Business

Типи помилок	Кількість помилок	
1. Порушення синтаксичних зв'язків між членами речення	117	41,8%
2. Граматична неоднотипність однорідних членів речення	49	17,5%
3. Порушення координації між підметом і присудком	47	16,8%
4. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	17	6,0%
5. Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення	16	5,7%
6. Штучне ускладнення речення	12	4,3%
7. Неправильний порядок слів	12	4,3%
8. Вживання зайвих членів речення	8	2,9%



## Продовження таблиці 4.5

Типи помилок	Кількість помилок	
9. Пропущені члени речення	2	0,7%
Усього	280	100%

## 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 [242]. – Спасибі до захисту проти перевантаження, перегріваючи і IP65, заливка герметиком класу, вони можуть бути широко використані в саморушній промисловості, де джерело живлення – батарея або інший подібний пристрій.*

У наведеному прикладі відсутнє узгодження між однорідними обставинами **захисту** і **заливка** у відмінку, а також однорідними означеннями **перевантаження** і **перегрівуючи** до обставини **захисту** в частиномовній приналежності. Причиною його може бути неадекватно проведений морфологічний та синтаксичний аналіз речення.

## 3. Порушення координації між підметом і присудком:

*Domestic demand, which has demonstrated strong growth over the past few years, was another key factor [241]. – Домашній **вимагають**, який продемонстрував стійке зростання за минулі декілька років, **був** іншим ключовим **чинником**.*

У наведеному прикладі порушена координація між підметом **вимагають** і присудком **був чинником** внаслідок неадекватного визначення частиномовної належності іменника **demand** (noun).

## 4. 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 [249]. – На завершальній стадії, складна*

**інтерпретація** усієї доступної геологічної і геофізичної інформації здійснюється, **який** дозволяє ефективно визначальні основні характеристики об'єктів під вивченням.

Оскільки програма не аналізує контекст та не враховує гіперсинтаксичні зв'язки між простими реченнями у складному, у перекладі займенник **який** не узгоджується з іменником **інтерпретація** в іншій частині речення.

5. 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 [251].* –

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.

Програма обрала неправильний відповідник для багатозначного сполучника **as**, переклавши його **оскільки**. Проте правильним є вживання сполучника **як**.

6. Неправильний порядок слів через необґрунтоване застосування граматичної трансформації перестановки:

*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 [253].* –

Вони витягнули кожен зразок з рішенням мурашиної кислоти і **пробігли рідинний chromatograph їх**, інструмент, який відділяє суміш сумішей.

7. Штучне ускладнення 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 [251].* –

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.

Програма штучно ускладнила речення, застосувавши сполучник **bo**. Причиною цьому є неправильне визначення частиномовної належності слова **for**, яке слід було перекласти як прийменник **для**.

8. Вживання зайвих членів речення, яких немає в тексті оригіналу:

*This year's predictions cover a range of topics, including that 400 million new phones with forward-facing cameras will be sold and that people are not yet ready to make video a regular part of their lives [246].* – Пророцтва цього року покривають дальність тем, у тому числі, що 400 мільйонів нових телефонів з переслати-облицювальними камерами продаватиметься і що люди не **проміжні системи** ще готові зробити відео регулярною частиною їх життів.

*To prevent electric shock, do not remove the cover (or back); there are no user serviceable parts inside [247].* – Щоб запобігти удару від електричного розряду, не переміщують кришку (чи назад); не **амплітудні модуляції** ніякі призначені для користувача корисні частини усередині.

У машинному перекладі дієслово **are** мало залишитись без перекладу, проте воно отримало еквіваленти **проміжні системи** і **амплітудні модуляції**, що призвело до появи зайвих членів речення.

9. Пропущені члени речення, які впливають на адекватність та еквівалентність перекладу:

*Right triangles conform to the Pythagorean Theorem, wherein the sum of the squares of the two legs is equal to the square of the hypotenuse, i.e.,  $a^2 + b^2 = c^2$ , where **a and b** are the legs and **c** is the hypotenuse [256].* – Правильні трикутники узгоджуються з Піфагорійською Теоремою, в чому сума квадратів двох ніг однакова до квадрата гіпотенузи, тобто,  $a^2 + b^2 = c^2$ , де **i b** – ноги і **c** – гіпотенуза.

Внаслідок неадекватного синтаксичного аналізу речення Pragma 6.3 Business пропустила підмет **a**, що призвело до порушення координації між підметом **b** та присудком **ноги** у числі.

**Помилки у побудові словосполучень** займають 25,1% від кількості граматичних помилок та представлені у таблиці 4.6.

1. **Порушення керування:**

*Have an enormous effect on business efficiency* [244] – *мати величезний ефект на діловій ефективності.*

Відповідно до норм української мови прийменник **на** може вимагати знахідного та місцевого відмінків. У цьому випадку програма помилково вибрала місцевий відмінок. Причиною виникнення помилки є відсутність заданого у словнику відмінкового керування слова.

Причиною неправильного вибору відмінкової форми може бути помилка в тексті оригіналу: *Sometimes consulates can doubt and accommodation – hotel reservation* [263]. – *Іноді консульства можуть сумніватися і акомодация – готельне бронювання.* Правильний варіант: *sometimes consulates can doubt an accommodation*, тоді машинний переклад матиме вигляд – *іноді консульства можуть сумніватися в акомодациї.*

Таблиця 4.6

**Помилки у побудові словосполучень системи Pragma 6.3 Business**

Типи помилок	Кількість помилок	
	Кількість	Відсоток
1 Порушення керування	119	33,2%
2 Відхилення від нормативно-узвального порядку слів	79	22,1%
3 Неправильний вибір прийменника	74	20,7%
4 Порушення узгодження між словами	47	13,1%
5 Відсутність необхідного прийменника	25	7,0%
6 Вживання зайвого прийменника	14	3,9%
Усього	358	100%

2. Відхилення від нормативно-узвального порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір граматичної трансформації: *analog format* [270] – *формат аналога (аналоговий формат)*; *lighting cables* [236] – *кабелі освітлення (освітлювальні кабелі)*.

3. Неправильний вибір прийменника:

*UN Framework convention on climate change* [240] – *угода Структури ООН на зміні клімату*. У словнику прийменник *on* має кілька значень: *на, по, в, біля, коло, після, під час, у, з приводу, про, за, при, з, від, під, над, до* [219, с. 47]. При перекладі програма обрала найчастотніший варіант перекладу, не врахувавши контекст. Правильний варіант – *рамкова конвенція зі зміни клімату*.

*Finally, the carbon dioxide is emitted when the natural gas from shale is combusted by its final consumers* [238]. – *Нарешті, вуглекислота випущена, коли природний газ від сланцю запалюється його завершальними споживачами*. Прийменник *from* є багатозначним. Його найчастотнішим відповідником в українській мові є *від* [218, с. 454]. Проте в конкретному випадку правильним варіантом є *зі*.

4. Порушення узгодження між словами у словосполученні:

а) у роді: *newly arrived company's aircraft is presented* [230] – *новоприбулий авіація компанії є був присутнім*;

б) у числі та відмінку: *26 billion* [262] – *26 більйона*; *18 billion cu m* [262] – *18 більйона cu м*; *at 6,000 rpm* [231] – *в 6,000 обороті в хвилину*;

в) у роді та відмінку: *you need highly skilled, committed personnel* [244] – *вам треба висококваліфікованим, вчинив персонал*;

г) у відмінку: *to develop its own fleet of aircraft, ground handling and maintenance base* [230] – *розвивати його власний флот авіації, ґрунтової, управляючи і ремонтна основа*.

5. Відсутність необхідного прийменника, оскільки у словнику не вказане прийменникове керування:

*The previous loan programme has contributed to reduced emissions of carbon dioxide by some 22,000 tonnes and decreased energy consumption by approximately*

*16,000 megawatt hours* [232]. – Попередня програма кредиту сприяла зменшенню емісій вуглекислоти **приблизно 22,000 метричних тонн і зменшив енергоспоживання приблизно 16,000 годинами мегавата.**

В обох словосполучення пропущений прийменник *на*, який є відповідником англійського прийменника *by*.

6. Вживання зайвого прийменника.

Програма дослівно перекладає словосполучення, без урахування особливостей керування в англійській та українській мовах. У результаті цього в тексті перекладу вживаються зайві прийменники, які порушують норми української мови: *to substitute Russian gas with domestic shale gas* [262] – **замінити Російський газ з комунально-побутовим сланцевим газом**; *44% of Ukrainians* [266] – **44% з українців.**

**Морфологічні помилки**, виявлені у машинних перекладах програми Pragma 6.3 Business, представлені в таблиці 4.7. До морфологічних помилок, зроблених програмою Pragma 6.3 Business, належать:

1. Помилки при знятті граматичної омонімії.

*A “kit” of ready-made parts created in space and delivered to Earth by meteorite* [253]. – **“Комплект” готових частин створив в космічному і доставленому, щоб Заземлити метеоритом.**

Програма неправильно визначила частиномовну належність дієприкметника (*created*) та двох іменників (*space, Earth*) і натомість переклала їх формами дієслів **створив, заземлити** та прикметника **космічному**. Правильними варіантами є вживання дієприкметника **створений** та іменників **космос і Земля**.

Причиною неправильного визначення частиномовної належності слів може бути наявність помилки в тексті оригіналу. Наприклад:

*Ensures optimal quality voice transmission* [270] – **гарантує оптимальну якісну переговорну передачу**. Правильний варіант оригіналу – *ensures optimal quality of voice transmission*, тоді машинний переклад матиме вигляд – **гарантує оптимальну якість голосової передачі**.

Таблиця 4.7

## Морфологічні помилки системи Pragma 6.3 Business

Типи помилок	Кількість помилок	
1 Помилки при знятті граматичної омонімії	392	84,3%
2 Ненормативне утворення пасивної конструкції дієслова	15	3,2%
3 Порухення категорії числа іменника	15	3,2%
4 Ненормативне утворення ступеня порівняння прикметника	11	2,4%
5 Порухення категорії часу дієслова	9	1,9%
6 Ненормативне утворення відмінкової форми іменника	8	1,7%
7 Вживання ненормативної форми дієприкметника	8	1,7%
8 Порухення категорії способу дієслова	4	0,9%
9 Вживання ненормативної форми дієприслівника	3	0,7%
Усього	465	100%

2. Порухення категорії числа іменника:

*Further installation of finished products* [236] – подальша установка **готовий** **виріб**.

*Installation of lifts* [236] – **установки** **підйомників**.

В оригіналі іменники були вжиті в множині, проте Pragma 6.3 Business використала однину через неадекватно виконаний морфологічний аналіз слів.

3. Ненормативне утворення пасивної конструкції дієслова.

*An armed attack against one or more of them in Europe or North America shall be considered an attack* [260]. – *Озброєно нападають проти один або більше з них в Європі або Північна Америка має вважати нападом*. Програма переклала пасивну конструкцію активною, що вплинуло на адекватність перекладу. Правильний варіант – *буде вважатися*.

*The protein structure of milk is presented by casein* [272] – *protein структура молока є був присутнім казеїном*. Програма не тільки неправильно утворила пасивну конструкцію дієслова, а й неадекватно переклала дієслово *present*. Правильний варіант – *представлений*.

#### 4. Ненормативне утворення ступеня порівняння прикметника:

*Always the best* [254] – *завжди кращий*. В оригіналі вживається прикметник у найвищому ступені порівняння, проте програма використала вищий ступінь, що є помилкою. Правильний варіант – *найкращий*.

*Diseases are becoming more and more widespread* [Медицина] – *хвороби стають ще і ще широкі поширеними*. Програма неправильно утворила вищий ступінь порівняння прикметника *поширений*, застосувавши прислівник *ще*, проте правильним є використання прислівника *більш*: *більш поширені*.

#### 5. Порушення категорії часу дієслова:

*To restore and maintain the security of the North Atlantic area* [265] – *щоб відновив і підтримує безпека північний атлантичний область*.

В оригіналі дієслово *to restore* вживається в теперішньому часі, проте програма переклала його формою минулого часу.

*And when those came who were hired about the eleventh hour* [233] – *і коли ті прийшли, хто наймається про одинадцяту годину*.

В оригіналі дієслово *to hire* вживається у формі пасивного стану минулого часу, проте програма переклала його формою теперішнього часу.

#### 6. Ненормативне утворення відмінкової форми іменника:

*Volatilization of ammonia nitrogen* [273] – *звітрявання амонійного азоту від карбонату амонія*.



Іменник **амоній** належить до м'якої групи II відміни [26, с. 75]. Залежно від значення у родовому відмінку одники іменники чоловічого роду можуть мати закінчення *-а*, *-я* або *-у*, *-ю*. У м'якій групі закінчення *-у*, *-ю* набувають іменники, які означають речовину, масу або матеріал [227, с. 84]. Тому правильний варіант – **амонію**.

*Equipment allows to produce very large diameter logs (even more than 30 cm)* [239] – *устаткування дозволяє робити дуже великі колоди (навіть більш ніж 30 см) діаметру*

Іменник **діаметр** належить до твердої групи II відміни [26, с. 75]. Іменники чоловічого роду II відміни твердої групи у родовому відмінку можуть мати закінчення *-а* або *-у*. Відповідно до “Українського правопису” “терміни іншомовного походження, які означають елементи будови чогось, конкретні предмети, геометричні фігури та їх частини” мають закінчення *-а* [227, с. 83], тому правильний варіант – **діаметра**.

7. Використання ненормативної форми дієприкметника:

*Coking coal* [238] – **коксуєче вугілля**.

*The United States took over air policing duties From Belgium* [260] – *США прийняли повітряні охороняючі обов'язки з Бельгії*.

В українській мові відсутні дієприкметники **коксуєче**, **охороняючі**, тому при перекладі застосовуються інші частини мови: **коксівне вугілля**; *США перебрали від Бельгії обов'язки з охорони повітряного простору*.

8. Порухення категорії способу дієслова:

*As this equipment gets hot during use, operate it in a well-ventilated place* [247]. – *Оскільки це устаткування добирається гаряче впродовж використання, діють з цим в well-ventilated місці*.

*Do not expose this equipment to rain* [247] – **не піддають це устаткування дощу**.

У тексті інструкції вживається наказовий спосіб дієслів, проте програма при перекладі використовує особову форму дієслова дійсного способу теперішнього часу, що впливає на еквівалентність та адекватність перекладу.

### 9. Вживання ненормативної форми дієприслівника:

*We are responsible for the quality of our meat, starting from the stage of crop cultivation* [267]. – *Ми відповідаємо за якість нашого м'яса, починаючись від стадії урожайного культивування.*

*Call the labourers, and give them their wages, beginning with the last to the first* [233]. – *Закликають чорноробів, і дають їм їх заробітні плати, розпочинаючись з витримки до першого.*

В українській мові відсутні дієприслівники *починаючись, розпочинаючись*. Правильними є варіанти – *починаючи, почавши* відповідно.

Лексико-граматичні та орфографічні помилки є найменш чисельними помилками та становлять 3,1% та 0,2% відповідно.

Програма зробила **лексико-граматичні помилки**, які полягають у вживанні словосполучень у не властивому для контексту та тематики значенні, а також відхиленні від нормативно-узвального порядку слів, порушенні логіко-синтаксичного зв'язку між словами через неправильний вибір варіантів перекладу та граматичної трансформації: *field survey* [249] – *огляд поля (польові роботи)*; *timber industry* [273] – *промисловість лісоматеріалу (лісонеробна промисловість)*.

Pragma 6.3 Business зробила три **орфографічні помилки**: *geophysical* [234] – *геофізичний*, *longitudinal* [239] – *подовжні*, *compared to* [234] – *була порівнянна з*.

Найбільшу кількість помилок було знайдено у машинних перекладах десятих текстів таких тематичних спрямувань, як: будівництво, металургійна, право, електроніка, машинобудування, зв'язок, комерція, текстиль, гірнична, екологія [236; 238; 240; 242; 246; 251; 257; 259; 265; 269] (див. Додаток Б). Вони слугуватимуть основою для подальшої оцінки якості статистичних систем машинного перекладу Google Translate, Bing Translator, Babylon, Free Translation та Yandex.

За результатами лінгвістичної оцінки помилок програми Pragma 6.3 Business можна зробити висновок, що найчисленнішими у машинних перекладах є лексичні помилки – 53,5% від загальної кількості помилок, серед яких переважають помилки у перекладі слів (37,0%) – неадекватний вибір перекладних відповідників внаслідок помилок при вирішенні лексичної полісемії та лексико-граматичної омонімії.

Причиною цього є недостатня повнота предметного компонента системи – автоматичного словника.

Серед граматичних помилок переважають синтаксичні помилки (29,1% від загальної кількості помилок), зокрема у побудові простих речень – порушення синтаксичних зв'язків між членами речення внаслідок неадекватно виконаного синтаксичного аналізу. Лексико-граматичні та орфографічні помилки займають найменшу частку від загальної кількості помилок та становлять 3,1% та 0,2% відповідно.

Для покращення роботи системи необхідним є поповнення словника користувача та вдосконалення роботи алгоритму, що є завданням розробників.

Найбільш проблемні для перекладу тексти слугуватимуть основою для подальшої оцінки якості статистичних систем машинного перекладу Google Translate, Babylon, Free Translation, Yandex та гібридної системи машинного перекладу Bing Translator.

## **4.2 Оцінка якості системи Google Translate**

Оцінка якості програми Google Translate перевірялася на матеріалі 10 текстів, що належать до тематик, у яких Pragma 6.3 Business зробила найбільше помилок [236; 238; 240; 242; 246; 251; 257; 259; 265; 269]. Усього було проаналізовано 110 пар речень.

Результати лінгвістичного аналізу текстів машинного перекладу подано в таблиці 4.8, більш детальна таблиця – див. Додаток В.

Лінгвістичний аналіз машинних перекладів показав, що програма зробила найбільше граматичних помилок (59,3%), серед яких переважають синтаксичні (50,0%). Морфологічні помилки менш чисельні – 9,3% від загальної кількості помилок. Лексичні помилки охоплюють 39,5% від усього обсягу помилок. Найменшу частку становлять лексико-граматичні помилки – 1,2%.

Таблиця 4.8

## Кількісне співвідношення помилок Google Translate

Типи помилок	Кількість помилок	
1. Лексичні помилки	266	39,5%
1.1 Помилки у перекладі слів	158	23,4%
1.2 Помилки у перекладі словосполучень	108	16,1%
2. Граматичні помилки	400	59,3%
2.1. Морфологічні помилки	63	9,3%
2.2. Синтаксичні помилки	337	50,0%
2.2.1 Помилки у побудові словосполучень	171	25,4%
2.2.2 Помилки у побудові речень:	166	24,6%
а) помилки у побудові простих речень	82	12,2%
б) помилки у побудові складних речень	84	12,4%
3. Лексико-граматичні помилки	8	1,2%
Усього	674	100%

Серед граматичних помилок переважають помилки у побудові словосполучень (25,4% від загального обсягу помилок), які подані в Таблиці 4.9.

**Помилки у побудові словосполучень** включають:

1. Порушення керування:

*Demand for data* [246] – *попит на даних*. Відповідно до норм української мови прийменник *на* може вимагати знахідного або місцевого відмінка. Програма переклала словосполучення, застосувавши форму місцевого відмінка, проте відповідно правильною є форма знахідного відмінка – *попит на дані*.

*Through Wi-fi hotspots* [246] – *через Wi-Fi точок доступу*. Система застосувала форму родового відмінка, проте прийменник *через* вимагає знахідного відмінка – *через Wi-Fi точки доступу*.

Оскільки програма не використовує в процесі роботи граматичні правила та не містить словників із вказаним прийменниковим керуванням, вона робить помилки у виборі відмінкової форми іменників, що входять до складу словосполучень.

Таблиця 4.9

### Помилки у побудові словосполучень Google Translate

Типи помилок	Кількість помилок	
1 Порушення керування	102	59,6%
2 Порушення узгодження між словами	31	18,1%
3 Неправильний вибір прийменника	18	10,5%
4 Відхилення від нормативно-узуального порядку слів	12	7,1%
5 Вживання зайвого прийменника	5	2,9%
6 Відсутність необхідного прийменника	3	1,8%
Усього	171	100%

2. Порушення узгодження – порушення синтаксичного зв'язку між залежним та стрижневим словом:

а) у роді: *a high performance steelmaking company with fully integrated production process* [259] – *висока продуктивність сталеливарна компанія з повністю інтегрованої виробничого процесу;*

б) у роді та відмінку: *the two wireless technologies* [246] – *два бездротових технологій;*

в) у роді та числі: *free in-store Wi-Fi access* [246] – *безкоштовних в магазині Wi-Fi доступ;*

г) у числі та відмінку: *comparing to the tufted carpets* [369] – *порівняно з чубатого килимів.*

3. Неправильний вибір прийменника.

Неправильний аналіз контексту речення призводить до вибору нерелевантного відповідника прийменника у словосполученні.

*Global predictions for the telecommunications sector* [246] – глобальні прогнози на телекомунікаційному секторі.

*Public's appetite for smartphones* [246] – апетит громадськості для смартфонів (правильний варіант – апетит громадськості **на** смартфони).

Прийменник *for* у словнику має такі варіанти перекладу: 1) для; 2) протягом; 3) до; в; у; на; 4) задля, заради за, на, для; 5) від, через, з-за, за, з; 6) замість, за; 7) на; 8) на честь; 9) (для того) щоб; 10) до [218, с. 438]. У першому випадку правильним є вживання прийменника **для**, а в другому – **на**.

4. Відхилення від нормативно-узвального порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір граматичної трансформації: *large diameter pipes manufacturing* [259] – труб великого діаметру виробництва.

Програма необґрунтовано застосувала граматичну трансформацію перестановки, що вплинуло на адекватність перекладу словосполучення. Правильний варіант – *виробництво труб великого діаметра*.

5. Вживання зайвого прийменника.

*The top five developers of operating systems have billions in cash* [251]. – **В першу** п'ятірку розробники операційних систем мають мільярди готівкою

У цьому випадку програма використала прийменник **в**, проте в тексті оригіналу його немає. Правильний варіант – *перша п'ятірка*.

6. Відсутність необхідного прийменника, який програма в процесі роботи пропустила:

*Year 2011 is likely to be very exciting to Ukraine's telecommunication sector* [246]. – 2011 рік, ймовірно, буде дуже **цікаво** телекомунікаційного сектора.

Незважаючи на те, що в оригіналі є прийменник *to* (для), програма не переклала його українською мовою, що спричинило помилку.

*Enter Chinese and South African markets* [257] – **входити** Китаю і Південної Африки ринки

У цьому випадку прийменник вживається відповідно до норм української мови. Правильний варіант – *входити на ринки Китаю та Південної Африки*. Оскільки система не використовує граматичні правила в процесі роботи, вона робить помилки в керуванні.

До помилок у побудові речень належать помилки у перекладі та побудові простих та складних речень, серед яких переважають помилки у побудові складних речень (12,4 %), подані в таблиці 4.10.

Таблиця 4.10

### Помилки у побудові складних речень системи Google Translate

Типи помилок	Кількість помилок	
	Кількість	Відсоток
1 Порушення синтаксичних зв'язків між членами речення	30	35,7%
2 Порушення координації між підметом і присудком	24	28,6%
3 Граматична неоднотипність однорідних членів речення	18	21,4%
4 Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	5	5,9%
5 Пропущені члени речення	3	3,6%
6 Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення	2	2,4%
7 Штучне ускладнення речення	1	1,2%
8 Неправильний порядок слів	1	1,2%
Усього	84	100%

До помилок у побудові складних речень належать:

1. Порухення синтаксичних зв'язків між членами речення:

*As Ukraine's technology sector is moving more towards the service model Saulius expects a double digit growth of IT services sector mainly driven by international demand and supported by certain tax privileges contained in the new tax code which became effective 1 January 2011 [251].* – Оскільки **технологія сектор України рухається більше до сервісної моделі Саулюс очікує зростання двозначними цифрами ІТ-послуг, сектор** головним чином за рахунок міжнародного попиту та підтримки певних податкових пільг, що містяться в новому податковому кодексі, який набув чинності 1 січня 2011 року.

У цьому реченні порушені синтаксичні зв'язки між членами речення через неправильне визначення частиномовної належності прикметника **technology**, неправильний вибір відмінкової форми іменника **сектор** та необгрунтовану пунктуацію між словами **ІТ-послуг** та **сектор**. Таким чином, у реченні порушене узгодження між означенням **технологія** та підметом **сектор**, також порушене керування між присудком **очікує** та додатком **сектор**.

2. Порухення координації між підметом і присудком:

*This year's predictions cover a range of topics, including that 400 million new phones with forward-facing cameras will be sold and that people are not yet ready to make video a regular part of their lives [246].* – **Прогнози** цього року **охоплює** коло питань, у тому числі **400 мільйонів нових телефонів** по напрямку руху камер **буде продаватися** і що люди ще не готові, щоб зробити відео невід'ємною частиною їхнього життя.

У реченні порушена координація між підметами **прогнози, 400 мільйонів телефонів** та присудками **охоплює, буде продаватися** відповідно у числі.

3. Граматична неоднотипність однорідних членів речення:

*It is the producer of high quality steel plate 6-200 mm thick and as wide as 1500-3200 mm that is broadly used in shipbuilding, power and special machine building, bridge and offshore deep sea platforms construction, manufacture of large diameter*



*pipes for gas and oil heavy duty pipelines* [259]. – Це є виробником високоякісного сталевого листа товщиною 6-200 мм і шириною, 1500-3200 мм, що широко використовується в суднобудуванні, влади і спеціального машинобудування, моста і на шельфі будівництва глибоких морських платформ, виробництво труб великого діаметру на газ і нафту важких трубопроводів.

У реченні порушені синтаксичні зв'язки між однорідними означеннями *товщиною* і *шириною*, а також однорідними обставинами *суднобудуванні*, *машинобудування*, *моста*, *шельфі*, *виробництво* через неправильний вибір відмінкових форм іменників. Неадекватне вирішення граматичної омонімії прикметника *power* та його переклад формою іменника *влади* призвели до порушення граматичної однотипності однорідних означень *влади* і *спеціального*.

4. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення:

*After discharging, the charger automatically turns into cells charging with the current twice as high as the one with which they were discharged* [242]. – Після вивантаження, зарядний пристрій автоматично включається в клітини зарядки зі струмом в два рази вища за *ту*, з якою вони були випусані.

У процесі роботи програма неправильно визначила рід іменника *струм*, тому в іншій частині речення вживаються займенники *ту*, *якою*, які характерні для іменника жіночого роду.

5. Пропущені члени речення:

*Tufted carpets are made in the process of tufting – it is such a technique of manufacturing carpets in which the raw material is inserted, by means of the needle, in some sort of the textile base (primary backing) thus making the carpet pile* [269]. – **Ворсові килими в процесі тафтинга** – це такий спосіб виготовлення килимів, в якій сировинний матеріал, вставлених за допомогою голки, в свого роду текстильної основи (первинної основи) в результаті чого килим купу.

У машинному перекладі відсутній присудок *виготовляються*.

6. Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення:

*Such measures shall be terminated **when** the Security Council has taken the measures necessary to restore and maintain international peace and security [265]. – Такі заходи будуть припинені, як Рада Безпеки вживе заходів, необхідних для відновлення і підтримання міжнародного миру та безпеки.*

Сполучник може виражати різні зв'язки між частинами складного речення, тому сполучник **when** має кілька значень у словнику [219, с. 674]. Для вибору правильного варіанта перекладу програмі слід здійснити смисловий аналіз речення, проте ця операція не закладена в алгоритмі її роботи. У наведеному прикладі правильним варіантом є вживання сполучника **коли**.

#### 7. Штучне ускладнення речення:

*The Parties undertake, as set forth in the Charter of the United Nations, **to settle any international dispute** in which they may be involved by peaceful means in such a manner that international peace and security and justice are not endangered, and to refrain in their international relations from the threat or use of force in any manner inconsistent with the purposes of the United Nations [265]. – Сторони зобов'язуються, як це передбачено в Статуті Організації Об'єднаних Націй, **щоб вирішувати всі міжнародні спори**, в яких вони можуть бути залучені мирними засобами таким чином, що міжнародний мир і безпеку і справедливість не піддавати загрозі, і утримуватися в їх міжнародних відносинах від загрози силою або її застосування в будь-якому чином, несумісним з цілями Організації Об'єднаних Націй.*

Програма необґрунтовано використала сполучник **щоб**, що призвело до штучного ускладнення речення.

#### 8. Неправильний порядок слів:

*Yet 2011 will mark the tipping point as the growth of applications for non-PC items outstrips **traditional software sales and consumers** embrace a wider variety of devices [251]. – Тим не менш, 2011 ознаменує переломний момент в якості росту заявок на непродукції ПК випереджає **традиційні продажі і споживачі програмного забезпечення** охопити більш широке коло пристроїв.*

Програма здійснила необґрунтовану зміну порядку слів *програмного забезпечення*, що вплинуло на зміст та адекватність перекладу речення. Правильний варіант – *продаж традиційного програмного забезпечення*.

Помилки у побудові простих речень, зроблені Google Translate, подані в таблиці 4.11.

Таблиця 4.11

### Помилки у побудові простих речень системи Google Translate

Типи помилок	Кількість помилок	
1. Порушення синтаксичних зв'язків між членами речення	39	47,6%
2. Граматична неоднотипність однорідних членів речення	31	37,8%
3. Порушення координації між підметом і присудком	7	8,5%
4. Пропущені члени речення	5	6,1%
Усього	82	100%

У побудові **простих речень** було виявлено такі помилки, як:

1. Порушення синтаксичних зв'язків між членами речення:

*Charger enables cells discharging with 100mA, 200mA, 350mA currents* [242]. – *Зарядний пристрій дозволяє клітини розвантажувальні з 100 мА, 200 мА, 350 мА струмів.*

У результаті вибору нерелевантних перекладних відповідників та неправильного визначення частиномовної належності іменника *discharging* (*розрядка*), який програма переклала формою дієприкметника *розвантажувальні*, та неправильного вибору відмінкової форми іменника *струмів*, у реченні спостерігаємо порушення узгодження між додатком *клітини* і означенням *струмів*.

2. Граматична неоднотипність однорідних членів речення:

*Corum's regular clients include over 30 major mining holdings with more than 400 mines and open pits, enrichment plants, steel and heavy engineering companies* [257]. – *Постійними клієнтами Corum включає більше 30 великих гірничодобувних холдингів з більш ніж 400 шахт і кар'єрів, збагачувальних фабриках, стали і важких інжинірингових компаній.*

У реченні порушені синтаксичні зв'язки між однорідними додатками *шахт, кар'єрів, збагачувальних фабриках і компаній* у зв'язку з вживанням іменників у різних відмінкових формах. Також наявна граматична неоднотипність однорідних означень *стали і важких інжинірингових* у результаті їхньої належності до різних частин мови внаслідок неправильного морфологічного аналізу прикметника *steel* та перекладу його формою дієслова *стали*. У наведеному прикладі також порушена координація між підметом *клієнтами* і присудком *включає* у числі, особі та відмінку через неправильно виконаний синтаксичний аналіз речення.

У реченні також порушені синтаксичні зв'язки керування між присудком *включає* та додатками *шахт і кар'єрів, збагачувальних фабриках і компаній*, які проявляються у вживанні іменників у неправильних відмінкових формах.

3. Порушення координації між підметом і присудком (див. попереднє речення).

4. Пропущені члени речення:

*The quality of Metinvest's steel products is monitored at every stage of an integrated production process, from coal and ore mining to finished steel production* [259]. – *Якість сталевих продукції Метінвесту відстежується на кожному етапі комплексного виробничого процесу, від вугілля і гірничо до готового виробництва сталі.*

У наведеному прикладі пропущений додаток *mining* – *видобуток*, що вплинуло на зміст машинного перекладу.

**Морфологічні помилки** пов'язані з неправильним визначенням омонімічних граматичних форм англійської мови, порушенням словозмінних граматичних категорій української мови (числа, відмінка, часу). Причина виникнення

морфологічних помилок у результаті перекладу Google Translate полягає в тому, що система не використовує граматичні правила для перекладу. Морфологічні помилки, які виникли в результаті роботи програми Google Translate, подані в таблиці 4.12.

Таблиця 4.12

### Морфологічні помилки системи Google Translate

Типи помилок	Кількість помилок	
1 Помилки при знятті граматичної омонімії	52	82,5%
2 Ненормативне утворення відмінкової форми іменника	5	7,9%
3 Порухення категорії числа іменника	3	4,8%
4 Порухення категорії часу дієслова	3	4,8%
Усього	63	100%

1. Серед морфологічних помилок найбільш чисельними є помилки при знятті граматичної омонімії, які полягають у неправильному визначенні частиномовної належності слів. Наприклад: *a white felt* [269] – білий *відчував*. Програма не розпізнала граматичні характеристики іменника *felt* та переклала його формою дієслова в минулому часі, оскільки друга форма неправильного дієслова *to feel* (*відчувати*) та іменник *felt* (*фільтр*) є омоформами – словами, які мають спільну графічну форму, проте розрізняються семантикою і частиномовною належністю. У контексті є інформація, необхідна для врахування омонімії. Так, перед *felt* стоїть прикметник *white* з неозначеним артиклем *a*. Проте програма не має блоку морфологічного аналізу, а тільки обирає найбільш частотний еквівалент без урахування граматичних характеристик слова.

*The charger repeats **discharging** and **charging*** [242] – зарядний пристрій повторює *розвантажувальні і зарядка*.

У цьому випадку іменник *discharging* (розрядка) перекладений як прикметник.

2. У тексті машинного перекладу було виявлено випадки ненормативного утворення відмінкових форм іменників, наприклад:

*Large diameter pipes manufacturing* [259] – труб великого діаметру виробництва.

*Large diameter pipes* [259] – труби великого діаметру.

*Manufacture of large diameter pipes* [259] – виробництво труб великого діаметру.

Відповідно до норм української мови іменник *діаметр* належить до твердої групи II відміни, є терміном іншомовного походження та позначає частину геометричної фігури. Тому у родовому відмінку він набуває закінчення *-а* [227, с. 83]. Правильний варіант – *діаметра*.

3. Порушення категорії числа іменників спричинене неправильною роботою алгоритму системи. Наприклад:

*More than half of all computers aren't computers anymore* [251] – більше половини всіх комп'ютерів не комп'ютер більше

*25 percent of North American big box and anchor tenant retailers* [246] – 25 відсотків північноамериканського великий ящик і якірним орендарем роздрібною торгівлі

У наведених прикладах в оригіналі іменники вживаються у множині, проте у машинному перекладі вони мають форму однини.

4. Порушення категорії часу дієслова:

*The worldwide developments forecasted in 2011 will definitely have impact on or will correlate with the developments in Ukraine* [246]. – В усьому світі події прогнозовані в 2011 році, безумовно, впливають на або буде корелювати з подіями в Україні.

У наведеному прикладі в оригіналі дієслово вживається у майбутньому часі, на що вказує використання допоміжного дієслова *will*. Проте програма переклала їх формами дієслова в теперішньому часі. Правильний варіант – *матимуть вплив*

**Лексичні помилки** займають 39,5% від усіх помилок, виявлених у машинних перекладах системи Google Translate. Детально вони представлені в таблиці 4.13.

Таблиця 4.13

### Лексичні помилки системи Google Translate

Типи помилок	Кількість помилок	
	Кількість	Відсоток
1. Помилки у перекладі слів	158	59,4%
1.1. Вибір нерелевантних перекладних відповідників	136	51,1%
1.1.1 Неправильне вирішення лексичної полісемії	100	37,6%
1.1.2 Неправильне вирішення лексико-граматичної омонімії	27	10,2%
1.1.3 Вживання русизмів	7	2,6%
1.1.4 Неправильне вирішення лексичної омонімії	2	0,7%
1.2. Неперекладені слова	22	8,3%
2. Помилки у перекладі словосполучень	108	40,6%
2.1. Дослівний переклад іншомовних висловів	59	22,2%
2.2. Неадекватний переклад словосполучень як семантичних єдностей	43	16,2%
2.3. Неадекватний переклад фразеологічних сполучень	6	2,2%
Усього	266	100%

Помилки у перекладі слів є найбільш чисельними серед лексичних помилок (59,4%). Серед них переважає вибір нерелевантних перекладних відповідників

через неправильне вирішення лексичної полісемії, лексико-граматичної та лексичної омонімії, а також вживання русизмів.

Значення багатозначного слова в кожному конкретному випадку його вживання можна встановити тільки з урахуванням контексту. Проте така операція недоступна програмі перекладу. У процесі роботи вона вираховує величину ймовірності вживання слова та обирає найчастотніший варіант перекладу, який не відповідає контексту.

Так, термін *mill* [259] система переклала як *млин*, проте правильним варіантом є *прокатний стан* [218, с. 727; 222, с. 62]. У словнику це слово має кілька значень, а саме: 1) млин, дробарка; 2) фабрика; 3) (металургійний) завод; 4) прядильна фабрика; 5) лісопильний завод; 6) прокатний стан; 7) прес (олійниці тощо); 8) фреза; 9) фрезерний верстат; 10) змагання з боксу; 11) кулачний бій; 12) в'язниця; гауптвахта; 13) двигун літака; 14) тисячна частка долара [218, с. 727]. Таким чином, система вибрала найбільш частотний варіант перекладу.

Слово *holder* [236] система переклала як *держатель*, проте правильним варіантом перекладу є *опора*. У словнику *holder* має кілька значень: 1) орендар; 2) власник, держатель (цінних паперів); пред'явник; 3) володар призу (почесного звання); 4) ручка, держак; 5) тех. штатив; опора; кронштейн; патрон; обойма; 6) касета; адаптер; 7) газгольдер; 8) ікло; 9) хватальний орган; 10) трюмний [218, с. 543]. Програма знову обрала одне з перших значень.

Програма неадекватно переклала 27 слів через неправильне вирішення лексико-граматичної омонімії. Так, іменник *well-being* [265] отримав переклад *добре бути*, проте правильним варіантом є іменник *добробут*.

У реченні *It currently mines iron ore from its one open-pit quartzite field through a process of drilling and blasting, and by the removal of overburden to external dumps* [238] програма неправильно визначила частиномовну належність дієслова *mines* та переклала його іменником *шахти*. Правильний варіант – *видобуває*.

У текстах машинних перекладів було виявлено 7 русизмів, що пояснюється використанням російської мови як проміжної ланки при перекладі з англійської



мови на українську. Або ж система не знайшла українських відповідників англійським словам. Наприклад:

*acid-resistant brickwork* [236] – *кислотостійкі цегляні кладки*;

*tarnishing* [236] – *потускнення*;

*vessels* [265] – *суди*;

*non-destructive* [259] – *неразрушаючий*;

*LCD display* [242] – *ЖК-дисплей*;

*needlepunched carpets* [269] – *іглопробівні килими*;

*needling* [269] – *прошивання*.

Програма не розрізняє лексичні омоніми та не аналізує контекст у процесі перекладу, тому обирає найчастотніший еквівалент, який не відповідає тематичному спрямуванню тексту. Система переклала слово *pile* як *куна* [236; 269], проте правильний варіант – *наля* для будівельної тематики та *ворс* для текстилю [219, с. 121; 222, с. 163]. Слова *pile* (*куна*) та *pile* (*наля*) є гетерогенними омонімами, які виникли внаслідок збігу етимологічно різних слів [229].

Google Translate залишила без перекладу 22 слова, серед яких: а) терміни: *earthfills*, *transportwork* [236], *Power Equipment Plant* [257]; б) власні назви: *Deloitte* [246; 251]; в) скорочення: *Mtonnes* [259]; г) аббревіатури: *DC*, *V* [242], *PC* [251].

Помилки у перекладі словосполучень становлять 40,6% від усієї кількості лексичних помилок. Найчисленнішими серед них є випадки дослівного перекладу іношомовних висловів, які мають власні відповідники в українській мові (22,4%).

Словосполучення *heavy engineering companies* [257] отримало дослівний переклад – *важкі інжинірингові компанії*, проте правильним є варіант – *машинобудівні підприємства* [221, с. 412].

Словосполучення *panel carriers* [236] система переклала як *носії панелі*, а мало б бути – *панелевози* [221, с. 171].

Вживання словосполучень у не властивому для контексту та тематики значенні через неадекватний вибір еквівалента в українській мові займає 16,2% від загальної кількості лексичних помилок.

Словосполучення *mining machines* [257] було перекладене як *гірські машини*. У словнику *mining machine* перекладається як *врубова машина, гірнична машина* [218, с. 729]. *Гірські машини* – це калька з російської “горные машины”.

*Armed force* [265] отримало переклад *озброєний напад*, проте правильним варіантом є *збройний напад* [218, с. 74]. У словнику *armed* має кілька варіантів перекладу: *озброєний; збройний; укріплений; підсилений* [218, с. 74]. Таким чином, можемо припустити, що система не знайшла українського відповідника усього словосполучення та переклала його дослівно або ж у доступному корпусі текстів вказано неправильний варіант перекладу.

Неправильний переклад фразеологічних сполучень є найменш чисельною помилкою серед лексичних помилок і становить 2,2%.

Машинний переклад звороту *all measures taken as a result thereof* [265] – *про всі прийнятих в результаті нього заходи* – не є нормативним, оскільки *take measures* перекладається українською як *вжити заходів* [220, с. 918]. Тому правильний варіант – *про всі заходи, вжиті у зв'язку з ним*.

**Лексико-граматичні помилки** становлять 1,2% від загального обсягу помилок. Причиною їхнього виникнення є неадекватний вибір відповідника словосполучення у мові перекладу та необґрунтоване застосування граматичної трансформації перестановки:

*Reinforcing cage installation* [236] – *зміцнення установка клітина*, правильний варіант – *установка каркасу з арматури* [221, с. 151].

Отже, лінгвістичний аналіз помилок, знайдених у текстах машинних перекладів Google Translate, показав переважання граматичних помилок над лексичними – 59,3% та 39,5% відповідно. Причиною цього є невикористання програмою граматичних правил у процесі роботи. Серед граматичних помилок домінують синтаксичні помилки (50,0%), зокрема помилки у побудові словосполучень (25,4%). Лексичні помилки становлять 39,5% від загальної кількості помилок. Це переважно вибір нерелевантних перекладних відповідників (20,1%) внаслідок неадекватного вибору еквівалентів багатозначних слів.

Причиною виникнення цих помилок є недостатня повнота паралельних корпусів текстів, на основі яких працює програма.

Покращити якість роботи Google Translate можна за допомогою вдосконалення алгоритмів функціонування програми та її подальшого навчання на основі паралельних корпусів текстів.

### 4.3 Оцінка якості системи Bing Translator

Оцінка якості програми Bing Translator проводилася на матеріалі 10 текстів різного тематичного спрямування [236; 238; 240; 242; 246; 251; 257; 259; 265; 269]. Усього було проаналізовано 110 пар речень.

Кількісне співвідношення зроблених програмою помилок подане в таблиці 4.14, більш детально помилки системи репрезентовані в Додатку Д.

У машинних перекладах системи Bing Translator виявлено більшу кількість граматичних помилок (58,7%), ніж лексичних (39,8%). Домінують помилки у побудові словосполучень – 26,5% від загальної кількості помилок.

Таблиця 4.14

#### Кількісне співвідношення помилок системи Bing Translator

Типи помилок	Кількість помилок	
1. Лексичні помилки	371	39,8%
1.1. Помилки у перекладі слів	217	23,3%
1.2. Помилки у перекладі словосполучень	154	16,5%
2. Граматичні помилки	548	58,7%
2.1. Морфологічні помилки	91	9,8%
2.2. Синтаксичні помилки	457	48,9%
2.2.1 Помилки у побудові словосполучень	248	26,5%
2.2.2 Помилки у побудові речень:	209	22,4%

## Продовження таблиці 4.14

Типи помилок	Кількість помилок	
а) помилки у побудові простих речень	104	11,1%
б) помилки у побудові складних речень	105	11,3%
3. Лексико-граматичні помилки	14	1,5%
Усього	933	100%

Помилки у побудові словосполучень подано в таблиці 4.15.

Таблиця 4.15

## Помилки у побудові словосполучень системи Bing Translator

Типи помилок	Кількість помилок	
1. Порушення керування	145	58,5%
2. Порушення узгодження між словами	63	25,4%
3. Неправильний вибір прийменника	18	7,3%
4. Відхилення від нормативно-узвального порядку слів	11	4,4%
5. Відсутність необхідного прийменника	8	3,2%
6. Вживання зайвого прийменника	3	1,2%
Усього	248	100%

До помилок у побудові словосполучень належать:

1. Порушення керування:

*Ukraine's technology sector is moving more towards the service model* [251] – технологія в Україні сектора рухається більше до сервісну модель.

У цьому словосполученні програма використала знахідний відмінок, проте, відповідно до норм української мови прийменник **до** вимагає родового відмінка.

*What Is the Difference between Tufted and Needle-punched Carpets?* [269] – У чому полягає **різниця між** східні і Needle-punched килими?

У цьому випадку після прийменника **між** Bing Translator використав іменник **килими** у називному відмінку, проте правильним є вживання іменника в орудному відмінку.

2. Порушення узгодження між словами – порушення синтаксичного зв'язку між залежним та стрижневим словом:

а) у роді: *make the final structure* [269] – *робити остаточний структуру*;

б) у числі: *additional needling* [269] – *додаткові пункція*;

в) у відмінку: *supported by certain tax privileges* [251] – *підтримується певні податкових пільг*;

г) у числі та відмінку: *Saulius expects a double digit growth of IT services sector* [251]. – *Саулюс очікує двозначними зростання його послуги сектор*;

д) у роді та відмінку: *25 percent of North American big box* [246] – *25 відсотків від північноамериканського велику коробку*.

3. Неправильний вибір прийменника:

Вибір правильного еквівалента прийменника залежить від контексту, проте Bing Translator використовує найчастотніший варіант перекладу. Наприклад:

*in other computing markets* [251] – *в інших обчислень ринки*

У словнику у просторовому значенні прийменник **in** має кілька варіантів перекладу: у, в, на [218, с. 576]. У наведеному вище прикладі слід було вжити прийменник **на**.

Прийменник **for** є багатозначним, у словнику подано такі варіанти його перекладу: 1) для; 2) протягом; 3) до; в; у; на; 4) задля, заради за, на, для; 5) від, через, з-за, за, з; 6) замість, за; 7) на; 8) на честь; 9) (для того) щоб; 10) до [218, с. 438].

При перекладі словосполучення *a headache for application developers* [251] програма використала перше значення – *головний біль для розробників*. Проте слід було застосувати прийменник **у**.

4. Відхилення від нормативно-узвального порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір граматичної трансформації, що вплинуло на адекватність перекладу словосполучення:

*Azovstal's quality management system* [259] – *Азовсталь, система управління якістю (система управління якістю Азовсталь)*.

*Potassium salts* [257] – *солі калію (калійні солі)*.

5. Відсутність необхідного прийменника:

*After their installation in the living area* [269] – *після їх встановлення зоною вітальні (після їхнього встановлення у зоні вітальні)*.

*From the development of technologies to production and servicing of equipment* [257] – *з розвитку технологій виробництва і обслуговування обладнання (від розробки технологій до виробництва та обслуговування обладнання)*.

6. Вживання зайвого прийменника:

*To provide a customer with a seamless experience* [246] – *забезпечити клієнту з безшовної досвід*.

*On the request of UN Secretary-General* [265] – *за запитом з Генеральним секретарем ООН*

У вищенаведених прикладах причиною вживання зайвого прийменника з є дослівний переклад стійких конструкцій програмою. Окрім цього, у першому прикладі порушене узгодження між словами у роді та відмінку.

Серед помилок у побудові речень переважають помилки у побудові складних речень (11,3%), які подані в таблиці 4.16.

Програма зробила такі помилки у побудові **складних речень**:

1. Порушення синтаксичних зв'язків між членами речення:

*Cabinet and wood works mainly consist of transportation and further installation of finished products; there can be both bars and laths used in construction erection and windows, doors and plank and parquet floors* [236]. – *Скабінет і деревини робіт в основному складаються з транспорту і далі установка готової продукції; може існувати барів і рейки використовується в будівництві ерекції і вікон, дверей та дощата і паркетною підлогою*.

У машинному перекладі порушена координація між підметом *деревини робіт* і присудком *складаються* у відмінку. Також порушений підрядний зв'язок між присудком *складаються* і додатком *установка* через неправильний вибір відмінкової форми іменника, що спричинило граматичну неоднотипність однорідних додатків *транспорту* і *установка*. У другій частині речення порушена координація між однорідними підметами *барів і рейки, вікон, дверей, підлогою* та присудком *може існувати* у числі та відмінку, також спостерігаємо граматичну неоднотипність однорідних підметів через необґрунтоване застосування іменників *барів, дверей* у родовому відмінку та іменника *підлогою* в орудному відмінку. У реченні порушено зв'язок між однорідними означеннями *дощата, паркетною* через відсутність узгодження означення *дощата* з підметом *підлогою* у відмінку.

2. Порушення координації між підметом і присудком (див. попереднє речення).

3. Граматична неоднотипність однорідних членів речення (див. попереднє речення).

Таблиця 4.16

### Помилки у побудові складних речень системи Bing Translator

Типи помилок	Кількість помилок	
1. Порушення синтаксичних зв'язків між членами речення	42	40,0%
2. Порушення координації між підметом і присудком	29	27,6%
3. Граматична неоднотипність однорідних членів речення	20	19,1%
4. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	5	4,8%

## Продовження таблиці 4.16

Типи помилок	Кількість помилок	
5. Пропущені члени речення	4	3,8%
6. Неправильний порядок слів	3	2,9%
7. Вживання зайвих членів речення	1	0,9%
8. Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення	1	0,9%
Усього	105	100%

4. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення:

*The Parties undertake, as set forth in the Charter of the United Nations, to settle any **international dispute in which** they may be involved by peaceful means in such a manner that international peace and security and justice are not endangered, and to refrain in their international relations from the threat or use of force in any manner inconsistent with the purposes of the United Nations [265]. – Сторони зобов'язуються, як набір викладені в Статуті організації О'єднаних Націй, вирішувати всі **міжнародні спори, в якій** вони можуть бути залучені, мирними засобами і таким чином, що міжнародного миру та безпеки і юстиції є не ставити під загрозу а також утримуватись у своїх міжнародних відносинах від погроз або застосування сили у будь-який спосіб, несумісний з цілями організації Об'єднаних Націй.*

У реченні порушені синтаксичні зв'язки між членами речення через неправильний морфологічний аналіз та переклад фрази *as set forth* – як набір викладені, через невідповідність займенника *якій* іменнику *спори* в іншій частині речення, а також через порушення синтаксичних зв'язків керування між присудком *не ставити під загрозу* і додатками *миру, безпеки, юстиції*.

5. Пропущені члени речення:

*The issues which Ukraine's telecommunication sector will be dealing with are **different and unique from global trends** but the worldwide developments forecasted in*



*2011 will definitely have impact on or will correlate with the developments in Ukraine* [246]. – *Питання, які сектори телекомунікацій України буде мати справу з унікальних з світовим тенденціям, але по всьому світу розробок, прогнозують у 2011 безумовно позначиться на або буде співвідносяться з події в Україні.*

У реченні машинного перекладу пропущений повнозначний член речення *different*, який наявний у тексті оригіналу. Окрім цього, порушена координація між підметом *сектори телекомунікацій* і присудком *буде мати справу* у відмінку, а також підметом *питання* та присудком *унікальних*, підметом *розробок* та однорідними присудками *позначиться* і *буде співвідносяться*, що призвело до їхньої граматичної неоднотипності. Також у результаті неправильного синтаксичного аналізу речення порушені синтаксичні зв'язки між другорядними членами речення: неправильно використано прийменник *з* та відмінкову форму додатка *тенденціям* після нього, а також недоцільно вжито відмінкову форму додатка *події*.

6. Неправильний порядок слів:

*In case of cells unused for a long time or while formatting new ones, the charger repeats discharging and charging until cells reach their maximum capacity* [242]. – *У разі не використовується протягом тривалого часу, або під час форматування нові клітини зарядний пристрій повторює розрядження і зарядки поки клітини досягти максимальної ємності.*

Через необгрунтоване застосування граматичної трансформації перестановки програма змінила порядок слів, що вплинуло на адекватність перекладу. У реченні також спостерігаємо порушення координації між підметом *клітини* і присудком *досягти* та синтаксичних зв'язків між другорядними членами речення.

7. Зайвий член речення:

*The Parties undertake, as set forth in the Charter of the United Nations, to settle any international dispute in which they may be involved by peaceful means in such a manner that international peace and security and justice are not endangered, and to refrain in their international relations from the threat or use of force in any manner inconsistent with the purposes of the United Nations* [265]. – *Сторони зобов'язуються,*

як набір викладені в Статуті організації Об'єднаних Націй, вирішувати всі міжнародні спори, в якій вони можуть бути залучені, мирними засобами і таким чином, що міжнародного миру та безпеки і юстиції є не ставити під загрозу а також утримуватись у своїх міжнародних відносинах від погроз або застосування сили у будь-який спосіб, несумісний з цілями організації Об'єднаних Націй.

У результаті того, що програма дослівно переклала пасивну конструкцію *international peace and security and justice are not endangered*, у реченні з'явився зайвий член речення – *є*, який ускладнює присудок *не ставити під загрозу*.

8. Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення:

*The Parties undertake, as set forth in the Charter of the United Nations, to settle any international dispute in which they may be involved by peaceful means in such a manner **that** international peace and security and justice are not endangered, and to refrain in their international relations from the threat or use of force in any manner inconsistent with the purposes of the United Nations [265].* – Сторони зобов'язуються, як набір викладені в Статуті організації Об'єднаних Націй, вирішувати всі міжнародні спори, в якій вони можуть бути залучені, мирними засобами і таким чином, **що** міжнародного миру та безпеки і юстиції є не ставити під загрозу а також утримуватись у своїх міжнародних відносинах від погроз або застосування сили у будь-який спосіб, несумісний з цілями організації Об'єднаних Націй.

Сполучник *that* має кілька варіантів перекладу: *що*; *щоб*; *так*, *щоб* [219, с. 508]. У процесі роботи програма обрала перший з них – *що*, проте правильно було б використати *щоб*.

Помилки, які система зробила у побудові простих речень, подані в таблиці 4.17.

У побудові **простих речень** система зробила такі помилки:

1. Порушення синтаксичних зв'язків між членами речення:

*Corum Group (formerly known as Mining Machines) is an expert in mining business [257].* – **Група Corum (раніше відомий як гірські машини) є експертом у гірничо-шахтної справи.**

У реченні порушений зв'язок між присудком *є експертом* та додатком *гірничо-шахтної справи* через неправильний вибір відмінкової форми іменника і недоцільне використання прийменника *у*. Також порушене узгодження між підметом *Група Corum* та вставленою конструкцією *раніше відомий* у роді.

Таблиця 4.17

### Помилки у побудові простих речень системи Bing Translator

Типи помилок	Кількість помилок	
1. Порушення синтаксичних зв'язків між членами речення	52	50,0%
2. Граматична неоднотипність однорідних членів речення	32	30,7%
3. Порушення координації між підметом і присудком	14	13,5%
4. Пропущені члени речення	4	3,8%
5. Зайві члени речення	1	1,0%
6. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	1	1,0%
Усього	104	100%

2. Граматична неоднотипність однорідних членів речення:

*DC (300mA) LED power supply is characterized with the input **voltage** from 7-24 VDC or 7-21 VAC and **self-adjusting output voltage** in the range from 3 to 21 VDC [242]. – Постійного струму (300mA) світлодіодні блок живлення характеризується вхідної **напруги** з 7-24 в постійного струму або 7-21 VAC і **самоналаштуванням вихідна напруга** в діапазоні від 3 до 21 в постійного струму.*

У реченні порушено синтаксичний зв'язок між присудком *характеризується* та додатками *напруги, напруга* через неправильним вибір відмінкових форм іменників, що спричинило граматичну неоднотипність однорідних додатків у відмінку. Також програма неправильно визначила частиномовну належність прикметника *self-adjusting* та переклала його формою іменника *самоналаштуванням*, чим порушила узгодження між означенням *самоналаштуванням* та додатком *напруг* і спричинила граматичну неоднотипність однорідних означень *самоналаштуванням* і *вихідна*.

Окрім цього, слід відзначити наявність іншої помилки: порушення узгодження між означенням *світлодіодні* та підметом *блок живлення* у числі.

3. Порушення координації між підметом і присудком:

*In 2012, the proportion should continue to rise in North America and start to spread around the world* [246]. – *У 2012 році частка повинні продовжувати рости в Північній Америці і почати поширювати по всьому світу.*

У реченні порушена координація між підметом *частка* і однорідними присудками *повинні продовжувати рости* і *почати поширювати* у числі.

4. Пропущений член речення:

*Possibilities for designing the needlepunched carpets are less comparing to the tufted carpets* [269]. – *Для проектування needlepunched килими це менше в порівнянні з килими східні.*

У процесі роботи програма пропустила підмет *possibilities*, що вплинуло на якість перекладу.

5. Зайвий член речення:

*2011 is likely to be very exciting to Ukraine's telecommunication sector* [246]. – *2011 рік є може бути дуже цікаво секторі телекомунікацій України*

У цьому реченні зайва дієслівна зв'язка *є*, яка з'явилася в перекладі у результаті дослівного перекладу фрази *is likely to be*. Окрім цього, у реченні порушена координація між підметом *2011* та присудком *може бути цікаво* через те, що програма неправильно визначила частиномовну належність прикметника

*exciting*, а також пропустила прийменник *to* та обрала неправильну відмінкову форму додатка *секторі*.

6. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення:

*Any such armed attack and all measures taken as a result thereof shall immediately be reported to the Security Council [265]. – Кожний такий збройний напад і про всі заходи, прийняті в результаті їх буде негайно повідомлено Раду Безпеки ООН.*

Програма не проаналізувала контекст речення та неправильно здійснила його синтаксичний аналіз, що призвело до використання займенника *їх*, який не узгоджується з іменником *напад* у першій частині речення. Правильним варіантом перекладу є займенник *ним*.

**Морфологічні помилки** становлять 9,8% від загальної кількості помилок програми. Вони представлені в таблиці 4.18.

Таблиця 4.18

### Морфологічні помилки системи Bing Translator

Типи помилок	Кількість помилок	
1. Помилки при знятті граматичної омонімії	75	82,4%
2. Порушення категорії числа іменника	10	11,0%
3. Порушення категорії часу дієслова	3	3,3%
4. Ненормативне утворення відмінкової форми іменника	2	2,2%
5. Ненормативне утворення ступенів порівняння прикметника	1	1,1%
Усього	91	100%

До морфологічних помилок програми Bing Translator належать:

1. Помилки при знятті граматичної омонімії (82,4%). Зазвичай у контексті є інформація, необхідна для врахування омонімії. Проте програма не має блоку морфологічного аналізу, а тільки обирає найбільш частотний еквівалент до слова без урахування його граматичних характеристик та контексту речення:

*2011 is likely to be very exciting to Ukraine's telecommunication sector [246] – 2011 рік є може бути дуже цікаво секторі телекомунікацій України.*

Програма неправильно визначила частиномовну належність прикметника *exciting* та переклала його прислівником *цікаво*.

*The worldwide developments forecasted in 2011 [246] – по всьому світу розробок, прогнозують у 2011.*

Дієприслівник *forecasted* був перекладений формою дієслова *прогнозують*, проте правильний варіант – *досягнення, прогнозовані на 2011 рік*.

2. Порушення категорії числа іменників:

*Plank and parquet floors [236] – дощата і паркетною підлогою.*

*Apron feeders [257] – пластинчастий конвеєр.*

У наведених прикладах іменники *floors* і *feeders* в оригіналі вживаються у множині, проте в перекладі їхні відповідники вжиті в однині – *підлогою, конвеєр*.

*A white felt providing additional sound and thermal insulation [269] – там, як правило, білого кольору відчував, забезпечуючи додатковий звук і утеплювачів.*

В оригіналі іменник *insulation* має форму однини, проте в перекладі його відповідник, окрім неправильно обраного еквівалента в українській мові, отримав форму множини – *утеплювачів*. Правильний варіант – *теплоізоляція* [221, с. 611].

3. Порушення категорії часу дієслова:

*Demand for tablets for personal use in Ukraine also will increase [251] – попит на таблетки для особистого користування в Україні також збільшується.*

В оригіналі *will increase* має форму майбутнього часу, на що вказує застосування допоміжного дієслова *will*. Проте у машинному перекладі використано дієслово *збільшується* у теперішньому часі. Правильний варіант – *збільшиться*.

*The worldwide developments forecasted in 2011 will definitely have impact on or will correlate with the developments in Ukraine* [246]. – По всьому світу розробок, прогнозують у 2011 безумовно позначиться на або буде співвідносяться з події в Україні.

У наведеному прикладі в оригіналі дієслово вживається в майбутньому часі. Помилка полягає в тому, що при утворенні складеної форми майбутнього часу допоміжне дієслово вживається у майбутньому часі, проте у неправильній особовій формі, а замість інфінітива основного дієслова програма використала особову форму дієслова в теперішньому часі. Правильний варіант – **будуть співвідноситися**.

4. Ненормативне утворення відмінкової форми іменника, наприклад:

*Large diameter pipes manufacturing* [259] – великого діаметру труби виробництва.

*Manufacture of large diameter pipes* [259] – виробництво труб великого діаметру.

Відповідно до норм української мови іменник **діаметр** належить до твердої групи II відміни, тому при відмінюванні у родовому відмінку він повинен мати закінчення **-а** [227, с. 83].

5. Ненормативне утворення ступеня порівняння прикметника:

*A wider variety of devices* [251] – **широкий** спектр пристроїв.

В оригіналі прикметник **wider** має вищий ступінь порівняння, проте це не відображено в перекладі. Правильний варіант – **ширший** спектр пристроїв.

**Лексичні помилки** займають 39,8% від загальної кількості помилок та детально показані в таблиці 4.19.

До них належать:

1. Неправильне вирішення лексичної полісемії.

Термін **deposit** [238] система переклала як **застава**. У словнику слово **deposit** має кілька значень, а саме: 1) депозит, вклад у банку; 2) внесок (грошовий); 3) завдаток, застава; 4) вкладання, внесення; здавання на зберігання, депонування; 5) склад, сховище; 6) відкладення; відстій; осад; 7) наліт, нагар; накип; 8) поклад,

родовище [218, с. 282]. У контексті гірничої тематики правильним варіантом перекладу є *родовище* [228, с. 124].

Слово *party* [265] система переклала як *партія*. У словнику *party* має кілька значень: 1) партія; 2) загін, команда; група, партія; 3) компанія; 4) прийом гостей; вечір, вечірка; пікнік; прогулянка в компанії; 5) особи, які супроводжують; 6) учасник; 7) особа; людина; 8) сторона; 9) доступна дівчина; 10) вечірка з поцілунками; обіймання [219, с. 88]. Програма обрала перше значення, не врахувавши тематичне спрямування тексту та контекст речення. Проте правильним варіантом перекладу для юридичної сфери є *сторона*.

Інколи неврахування контексту при перекладі багатозначних слів приводить до тавтології. Наприклад: *are not ready or willing to abandon the market* [251] – *не готові або не готові відмовитися від ринку*. У словнику слова *ready* та *willing* справді мають значення *готовий* [219, с. 684]. Проте слово *willing* означає *що бажає*, яке слід було використати в перекладі. Правильний варіант – *і не готові та не бажають залишити ринок*.

Таблиця 4.19

### Лексичні помилки системи Bing Translator

Типи помилок	Кількість помилок	
1. Помилки у перекладі слів	217	58,5%
1.1 Вибір нерелевантних перекладних відповідників	174	46,9%
1.1.1 Неправильне вирішення лексичної полісемії	124	33,4%
1.1.2 Неправильне вирішення лексико-граматичної омонімії	42	11,3%
1.1.3 Неправильне вирішення лексичної омонімії	5	1,3%



## Продовження таблиці 4.19

Типи помилок	Кількість помилок	
1.1.4 Неадекватний переклад власних назв та аббревіатур, які слід було залишити в оригінальному варіанті	2	0,6%
1.1.5 Вживання русизмів	1	0,3%
1.2. Неперекладені слова	43	11,6%
2. Помилки у перекладі словосполучень	154	41,5%
2.1. Дослівний переклад іншомовних висловів	86	23,1%
2.2. Неадекватний переклад словосполучень як семантичних єдностей	50	13,5%
2.3. Неадекватний переклад фразеологічних сполучень	10	2,7%
2.4. Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті	8	2,2%
Усього	371	100%

## 2. Неправильне вирішення лексичної омонімії.

Система переклала термін *plants* [259] як *рослини*, проте правильний варіант – *заводи* [219, с. 129; 222, с. 173; 228, с. 336]. У словнику *plant* має кілька варіантів перекладу: 1) рослина; саджанець; 2) новачок; юнак; нащадок; 3) урожай; 4) ріст; 5) поза; позиція; 6) похорон; 7) завод, фабрика; підприємство; 8) установка, устаткування, комплект машин; 9) капітал, багаж; 10) агрегат, механізм; 11) рибна сажалка; 12) сищик, детектив; 13) поліцейська засідка; 14) крадене; склад краденого; 15) шахрайство; ошуканство. Значення *рослина* є первинним, а значення *завод, фабрика, підприємство* слово набуло у ХХ столітті [229].

3. Помилки при знятті лексико-граматичної омонімії, що призвели до неправильного перекладу слів. Програма не змогла правильно визначити

частиномовну належність деяких слів, у результаті чого підбрала неадекватний відповідник, наприклад:

*current* [242] – *поточний*, правильний варіант – *струм* [218, с. 256];

*implement* [236] – *реалізувати*, правильний варіант – *інструмент* [218, с. 573];

*ramming* [236] – *набивний*, правильний варіант – *ущільнення* [219, с. 205].

4. Вживання русизмів:

*predictions* – *предсказания* [251]; *in some sort of* – *в какой-то* [269].

Причиною цього може також бути те, що система не знайшла українських відповідників англійським словам.

**Неперекладені слова** займають 11,6% від загальної кількості лексичних помилок. Система залишила без перекладу 43 слова, серед яких:

а) терміни: *earthfills*, *cabinet* [236]; *sinter* [259]; *needlepunched*, *tufting*, *web* [269];

б) власні назви: *Bakas* [246; 251];

в) скорочення: *Mtonnes*, *mm* [259]; *V*, *mA* [242]; *non-PC* [251];

г) аббревіатура: *VAC* [242].

**Неадекватний переклад аббревіатур**, які слід було залишити в оригінальному варіанті: *BC-700* [251] – *Н.е.-700*, *TME* [251] – *ТЗН*.

**Помилки у перекладі словосполучень** становлять 41,5% від усієї кількості лексичних помилок. До них належать:

1. Дослівний переклад іншомовних висловів, які в українській мові мають свої відповідники.

Словосполучення *open pits* [257] було перекладене – *відкрити ями*, що є його дослівним перекладом. Проте правильним є варіант – *кар'єри* [222, с. 170].

Зворот *for the first time* [251] отримав переклад *в перший раз*, що є калькою з російської мови. Правильний варіант – *вперше*.

2. Неадекватний переклад словосполучень як семантичних єдностей.

Програма переклала словосполучення *external dumps* [238] як *зовнішні звалища*. У словнику термін *dump* має кілька значень, серед яких: 1) звалище;

2) перекидач; 3) склад; 4) відвал (породи) [228, с. 146]. У контексті речення йдеться не про *звалища*, а про *відвали породи*, тому правильний варіант – *зовнішні відвали*.

Словосполучення *loss rate* [240] отримало переклад *коефіцієнт втрат*, проте відповідно до тематичного спрямування тексту правильним є варіант – *рівень витоків* [228, с. 269].

3. Неадекватний переклад фразеологічних сполучень.

Програма переклала фразеологічне сполучення *as set forth in* [265] дослівно – *як набір викладені в*. Зворот *set forth* має кілька значень: 1) вирушати; 2) виставляти (напоказ); 3) висловлювати, визначати, формулювати; 4) видавати, друкувати; 5) прикрашати; 5) розгортати війська [219, с. 332]. Відповідно до тематичного спрямування тексту та контексту речення правильним є варіант *як це визначено у*.

*Become effective* [251] – *ставати ефективним*. Програма зробила дослівний переклад цього звороту, проте він має інший відповідник в українській мові – *набирати чинності, сили (про постанову, закон, угоду)* [220, с. 82].

4. Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті:

*Corum Quality* [257] – *Corum якості*;

*Corum Source* [257] – *Corum джерело* – *Corum Source*;

*Corum Group* [257] – *Corum Група*.

**Лексико-граматичні помилки** становлять 1,5% від загальної кількості помилок. Програма не тільки обрала неправильні відповідники в українській мові, а й необґрунтовано застосувала граматичну трансформацію перестановки. Наприклад: *a heavy section mill* [257] – *млин важких розділ*. Правильний варіант – *цех крупного сорту* [228, с. 287].

Програма не зробила жодної орфографічної помилки.

Лінгвістичний аналіз помилок, виявлених у текстах машинних перекладів системи Bing Translator показав переважання граматичних помилок над лексичними – 58,7% та 39,8% відповідно.

Серед граматичних помилок домінують помилки у побудові словосполучень – 26,5% від загальної кількості помилок. Серед помилок у побудові речень переважають помилки у побудові складних речень (11,3%). Лексико-граматичні помилки є найменш чисельними та становлять 1,5%. У текстах машинних перекладів не було виявлено орфографічних помилок.

Причиною домінування граматичних помилок є недосконалість роботи алгоритмів програми. Покращити якість роботи Bing Translator можна за допомогою вдосконалення її програмного компонента та подальшого навчання на основі паралельних корпусів текстів.

#### 4.4 Оцінка якості системи Babylon

Система Babylon оцінювалася на матеріалі 10 текстів, що належать до таких тематик, як: будівництво, металургійна, право, електроніка, машинобудування, зв'язок, комерція, текстиль, гірнична, екологія [236; 238; 240; 242; 246; 251; 257; 259; 265; 269]. Усього було проаналізовано 110 пар речень.

Результати лінгвістичної оцінки подані в таблиці 4.20 та більш детально представлені в додатку Е.

Таблиця 4.20

#### Кількісне співвідношення помилок системи Babylon

Типи помилок	Кількість помилок	
	Кількість	Відсоток
1. Лексичні помилки	762	56,8%
1.1. Помилки у перекладі слів	517	38,6%
1.2. Помилки у перекладі словосполучень	245	18,2%
2. Граматичні помилки	563	41,9%
2.1. Морфологічні помилки	181	13,5%
2.2. Синтаксичні помилки:	382	28,4%

## Продовження таблиці 4.20

Типи помилок	Кількість помилок	
2.2.1 Помилки у побудові словосполучень	174	12,9%
2.2.2 Помилки у побудові речень:	208	15,5%
а) помилки у побудові простих речень	98	7,3%
б) помилки у побудові складних речень	110	8,2%
3. Лексико-граматичні помилки	16	1,2%
4. Орфографічні помилки	2	0,1%
Усього	1343	100%

Програма зробила більше лексичних помилок, ніж граматичних, – 56,8% і 41,9% відповідно. Причиною цього є те, що вона працює на основі недостатньо повного паралельного корпусу текстів. Лексико-граматичні та орфографічні помилки є найменш чисельними та становлять 1,2% та 0,1% відповідно.

Лексичні помилки представлені в таблиці 4.21.

Таблиця 4.21

## Лексичні помилки системи Babylon

Типи помилок	Кількість помилок	
1. Помилки у перекладі слів	517	67,8%
1.1 Вибір нерелевантних перекладних відповідників	286	37,5%
1.1.1 Неправильне вирішення лексичної полісемії	190	24,9%
1.1.2 Неправильне вирішення лексико-граматичної омонімії	89	11,7%
1.1.3 Неправильне вирішення лексичної омонімії	4	0,5%

## Продовження таблиці 4.21

Типи помилок	Кількість помилок	
1.1.4 Вживання русизмів	2	0,3%
1.1.5 Неадекватний переклад власних назв та аббревіатур, які слід було залишити в оригінальному варіанті	1	0,1%
1.2 Неперекладені слова	231	30,3%
2. Помилки у перекладі словосполучень	245	32,2%
2.1 Дослівний переклад іншомовних висловів	108	14,2%
2.2 Неадекватний переклад словосполучень як семантичних єдностей	93	12,2%
2.3. Неадекватний переклад фразеологічних сполучень	41	5,4%
2.4. Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті	3	0,4%
Усього	762	100%

Серед лексичних помилок переважають помилки у перекладі слів – 67,8%, серед яких:

1. Вибір нерелевантних перекладних відповідників.

Програма зробила найбільше помилок при вирішенні лексичної полісемії (24,9%). Причиною цього є те, що в процесі роботи вона не аналізує контекст речення, а тільки обирає найчастотніший відповідник в українській мові.

Так, термін *current* [242] система переклала як *течія*, проте правильним варіантом є *струм*. У словнику *current* має кілька значень, а саме: 1) течія; потік; 2) струмінь, потік; 3) припливна (неприпливна) течія; 4) хід (подій тощо); 5) (електричний) струм [218, с. 256].

Термін *painting* [236] був перекладений як *картина*, проте правильний варіант – *фарбування*. У словнику воно має кілька варіантів перекладу: 1) картина;

2) живопис; 3) зображення; розпис; 4) змалювання; опис; 5) розмальовування, розфарбовування; малювання; фарбування; забарвлення; 6) малярська справа [219, с. 76].

У результаті неправильного вибору еквівалента в українській мові та через неврахування контексту системою у перекладених текстах трапляються випадки тавтології: *Demand for data passing through wi-fi hotspots is set to outstrip the amount of traffic carried over mobile networks by between 25 per cent and 50 per cent next year* [246]. – *Вимоги для даних винесення через wi-fi hotspots береться за outstrip сума перевезення перевезеного над пересувними мережами між 25 відсотком та 50 відсотком наступного року*

У словнику слово **traffic** має кілька значень: 1) рух; транспорт; сполучення; 2) перевезення; вантажообіг; 3) торгівля; 4) кількість перевезених пасажирів (вантажів) за певний період; фрахт, вантажі; трафік, потік (інформаційного) обміну, робоче навантаження (лінії зв'язку); 5) кількість телеграм (телефонних переговорів) за певний період; 6) торговельні угоди [218, с. 545].

Фразове дієслово **carry over** перекладається як: 1) перевозити; 2) переносити, передавати [218, с. 178].

Саме тому правильними варіантами перекладу є **трафік** та **передавати** відповідно.

Програма не змогла правильно визначити частиномовну належність деяких слів, у результаті чого вони отримали неадекватний переклад. Так, у реченні *It is still a narrow range that pales in comparison to the magnitude of different prices* [251] програма переклала дієслово **pales** формою іменника **товаришиі**: *досі вузький діапазон що товаришиі у порівнянні до величини різних цін*. Проте правильним є варіант: **тьмяніє** у порівнянні з розмаїттям цін.

У словосполученні **charging/discharging time** [242] іменники **charging** і **discharging** були перекладені формами дієприслівника **звинувачувавши** і дієслова **звільнюють**. Проте правильним є вживання іменників **зарядка** [228, с. 79] і **розрядка** [228, с. 132] відповідно.

Програма не розрізняє лексичні омоніми та не аналізує контекст у процесі перекладу, тому обирає найчастотніший еквівалент, який не відповідає тематичному спрямуванню тексту.

Так, термін *channel* програма переклала як *ламаниш* [242], проте правильним варіантом є *канал*. Комп'ютер не розпізнав лексичні омоніми – географічну назву *Channel*, яка перекладається як *Ла-Маниш* [218, с.189], та багатозначне слово *channel*, одним зі значень якого є *канал* [218, с. 189].

Програма неадекватно переклала аббревіатуру *IT* у словосполученнях *IT services sector* [251] та *IT departments* [251]. При перекладі вона використала займенник *це*. Проте правильними є варіанти *сектор IT послуг* та *IT-департаменти*.

Наявність в перекладі русизмів пояснюється використанням російської мови як проміжної ланки при перекладі з англійської мови українською. Іншою причиною виникнення помилки може бути те, що система не знайшла українських відповідників англійським словам. Наприклад:

*AB web which makes the carpet keep starting dimensional stability and shape* [269]. – *AB навутиння котрий робить килим держать вихідну стабільність розмірів та форму*. Відповідниками до слова *keep* в українській мові є дієслова *тримати, зберігати* [219, с. 635].

*By the removal of overburden to external dumps* [238] – *усуненням переобтяжують до зовнішніх свалки*. Правильним варіантом перекладу у цьому випадку є термін *відвали* [228, с. 146].

## 2. Неперекладені слова.

Програма залишила без перекладу 231 слово, що становить 30,3% від загальної кількості помилок, серед яких:

а) терміни: *engineering, tech, specialized, feeders, conveyors, loaders, spreaders, stackers, potassium, directorate, potassium* [257]; *cellular broadband* [246], *fragmentation* [251];

б) власні назви: *Deloitte* [246; 251], *Ukrtelecom* [246], *Svet Shakhtyora, Donetsgormach* [257];



в) географічні назви: *Gorlovka, Druzhkovka, Donetsk, Krivoy Rog, Belarus, India, Macedonia, Bosnia, Herzegovina, Montenegro* [257];

г) скорочення: *Mtonnes, mm* [259]; *V, mA* [242]; *PC* [251];

д) аббревіатури: *VDC, VAC* [242]; *DSTU, ISO, GOST* [259].

**Помилки у перекладі словосполучень** складають 32,2% від усієї кількості лексичних помилок.

1. Найчисленнішими є помилки, які виникають у результаті дослівного перекладу словосполучень (14,2%).

Так, словосполучення *magnetic and floatation metal content upgrading facility* [238] було перекладене як *магнітного та floatation метал вміст покращуючи устаткування*, проте правильним є варіант *комплекс магніто-флотаційного доведення концентрату* [221, с. 439].

Словосполучення *gas extraction* [240] отримало переклад *газова екстракція*, проте правильний варіант – *видобуток газу* [228, с. 163].

2. Неадекватний переклад словосполучень як семантичних єдностей становить 12,2% від усієї кількості лексичних помилок. Причиною виникнення цієї помилки є неправильний вибір еквівалента для термінологічної одиниці в українській мові.

Так, термін *tax code* [251] був перекладений як *податковий код*, проте правильний варіант – *податковий кодекс* [218, с. 212].

Програма переклала *economic policies* [265] словосполученням *економічні поліси*. У словнику термін *policy* має кілька значень, серед яких: 1) політика; 2) лінія поведінки, курс; 3) політичність, розсудливість; 4) хитрість, спритність; 5) страховий поліс; 6) парк (при садибі, маєтку); 7) азартна гра [219, с. 140]. Проте у словосполученні зі словом *economic* він має значення *політика*, тому правильним є варіант *економічна політика*.

Babylon дослівно переклала 41 **фразеологічне сполучення**, що призвело до виникнення помилок на лексичному рівні. Наприклад:

*In exercise of the right* [265] – *у вправі права*, правильний варіант – *здійснюючи своє законне право* [218, с. 379].

*In concert with the other Parties* – у концерті з іншими Сторонами, правильний варіант – спільно з іншими Сторонами [218, с. 222].

3. Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті: *Corum Quality* [257] – *Corum Якість*; *Corum Source* [257] – *Corum Джерело*; *Corum Group* [257] – *Corum Групують*.

**Граматичні помилки** програми Babylon становлять 41,9% від загальної кількості помилок. Серед них найчисленнішими є синтаксичні помилки – 28,4% з домінуванням помилок у побудові складних речень (8,2%), які подані в таблиці 4.22.

Таблиця 4.22

### Помилки у побудові складних речень системи Babylon

Типи помилок	Кількість помилок	
1. Порушення синтаксичних зв'язків між членами речення	45	40,9%
2. Порушення координації між підметом і присудком	27	24,6%
3. Граматична неоднотипність однорідних членів речення	20	18,2%
4. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	9	8,2%
5. Пропущені члени речення	3	2,7%
6. Неправильний порядок слів	2	1,8%
7. Вживання зайвих членів речення	2	1,8%
8. Штучне ускладнення речення	1	0,9%
9. Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення	1	0,9%
Усього	110	100%

Система Babylon зробила такі помилки у побудові **складних речень**:

1. Порухення синтаксичних зв'язків між членами речення:

*Shale gas extraction is accompanied by the greenhouse gases' emissions; their inventory and abatement are important to minimize the impact of shale gas extraction on climate* [240]. – *Shale газова екстракція супроводжується greenhouse газами емісія; їхній інвентар та анулювання є важливе мінімізувати вплив shale газу екстракцію на кліматі.*

У реченні порушено синтаксичних зв'язок між присудком *супроводжується* та додатком *газами емісія*. Іменник *емісія* слід було вжити в орудному відмінку, а залежне від нього *газами* – у родовому. Ця помилка виникла також через неадекватний дослівний переклад словосполучення *gases' emissions* – *газами емісія*, правильний варіант – *викиди газів*. Було не лише обрано неправильні відповідники в українській мові, а й порушено морфологічні норми – неправильно визначено число іменника *emissions* та не розпізнано присвійний відмінок англійського іменника *gases'*.

Також порушена координація між підметами *інвентар, анулювання* і присудком *є важливе* у числі. Окрім цього, пропущений сполучник *щоб* у частині *їхній інвентар та анулювання є важливе мінімізувати*.

2. Порухення координації між підметом і присудком (див. попереднє речення).

3. Граматична неоднотипність однорідних членів речення:

*In case of cells unused for a long time or while formatting new ones, the charger repeats **discharging** and **charging** until cells reach their maximum capacity* [242]. – *У випадку клітин unused довгий час або той час, як форматуючи нові, charger повторюється **звільнюють** та **звинувачуючі** до клітин досягають їхньої максимальної потужності.*

Внаслідок того, що програма Babylon неправильно визначила частиномовну належність слів *discharging* і *charging* та переклала їх дієсловом *звільнюють* і дієприслівником *звинувачуючи* відповідно, виникла граматична неоднотипність однорідних додатків.

Окрім цього, у реченні порушені синтаксичні зв'язки між членами речення через наявність неперекладених слів, неправильний вибір прийменника **до** (що суперечить смисловим зв'язкам між членами речення) та порушення координації між підметом **клітин** і присудком **досягають** у відмінку.

4. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення:

*On the back there is usually a white felt providing additional sound and thermal insulation as well as better comfort, or so called AB web which makes the carpet keep starting dimensional stability and shape for a long time after their installation in the living area [269].* – *На спині є звичайно білий відчувала забезпечуючи додатковий звук та термічна ізоляція так же, як і краще комфорт, або так закликане AB павутиння котрий робить килим держать вихідну стабільність розмірів та форму довгий час після їхньої установки у живій області.*

Програма використала займенник **котрий** щодо іменника **павутиння**, що не відповідає середньому роду останнього.

5. Пропущені члени речення:

*The Parties agree that an armed attack **against one or more of them** in Europe or North America shall be considered an attack against them [265].* – *Сторони погоджуються що збройний напад **проти або більше з них** у Європі або північній Америці повинні бути розглянуті атака проти них всі.*

*For the purpose of Article 5, an armed attack **on one or more of the Parties** is deemed to include an armed attack [265].* – *Для мети Статті 5, збройної напад **на або більше Сторін** вважається щоб включити збройний напад.*

У поданих реченнях, окрім порушення синтаксичних зв'язків між їхніми членами, також відсутні додатки **one**.

6. Неправильний порядок слів:

*However **this diversity leads to fragmentation** and complexity at the hardware level which causes a headache for application developers, media companies and IT departments [251].* – *Але **ці свинець різноманітності до fragmentation** та*

*складності у апаратурі рівень котрий викликає headache для розробників застосування, засобів масової інформації компаній та ЦЕ департаменти.*

Система Babylon неправильно виконала синтаксичний аналіз речення та застосувала трансформацію перестановки в результаті неправильного морфологічного аналізу, що призвело до порушення порядку слів. Так, дієслово *leads* було перекладене іменником *свинець*. Внаслідок цього у реченні відсутній присудок. Також порушений синтаксичний зв'язок між додатком *складності* та обставиною *апаратурі рівень* у відмінку через неправильний вибір прийменника. Відсутній синтаксичний зв'язок між присудком *викликає* та додатком *ЦЕ департаменти*, який проявляється у неправильному виборі відмінкової форми іменника *департаменти*. Замість називного відмінка програма мала застосувати родовий. Ця неточність спричинила граматичну неоднотипність однорідних додатків *розробників застосування, засобів масової інформації компаній та ЦЕ департаменти*.

7. Вживання зайвих членів речення:

*The issues which Ukraine's telecommunication sector will be dealing with are different and unique from global trends but the worldwide developments forecasted in 2011 will definitely have impact on or will correlate with the developments in Ukraine* [246]. – *Питання котрий сектор телекомунікації України буде мати справу з різний та унікальний з глобальних тенденцій але всесвітніх розвитків forecasted у 2011 буде напевно мати вплив на або буде буде співвідноситись з розвитками в Україні.*

У результаті неправильно виконаного синтаксичного аналізу речення програма двічі використала дієслово *буде*.

8. Штучне ускладнення речення:

*The Parties undertake, as set forth in the Charter of the United Nations, to settle any international dispute in which they may be involved by peaceful means in such a manner that international peace and security and justice are not endangered, and to refrain in their international relations from the threat or use of force in any manner inconsistent with the purposes of the United Nations* [265]. – *Сторони заходяться, як*

висунуте у Хартії Організації Об'єднаних націй, **щоб** улагодити будь-яку міжнародну суперечку у котрій вони можуть бути включені миролюбними засобами у такому способі що міжнародний мир та безпека та справедливість не загрожуються, та щоб утриматися у їхніх міжнародних відношеннях з загрози або користування силою у будь-якому способі не відповідаючому цілям Організації Об'єднаних націй.

У процесі роботи програма використала сполучник **щоб**, який відсутній у реченні оригіналу. Це призвело до штучного ускладнення речення.

9. Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення:

*Yet 2011 will mark the tipping point as the growth of applications for non-PC items outstrips traditional software sales and consumers embrace a wider variety of devices* [251]. – *Все ще 2011 поставить оцінку tipping пункт як ріст застосувань для не-pc пунктів outstrips традиційний збут програмного забезпечення та споживачи охоплюють більш широкий різновид приладів.*

Сполучник **as** є багатозначним та має такі варіанти перекладу: 1) тоді, коли, у той час як; 2) як, у ролі; 3) тому що, оскільки, через те що, бо; 4) хоч, як би не [218, с. 78]. Програма обрала одне зі значень – **як**. Однак, оскільки Babylon не аналізує контекст речення, вона не змогла обрати правильний варіант перекладу – **оскільки**. Окрім цієї помилки, у реченні порушені синтаксичні зв'язки між його членами.

Помилки, зроблені у побудові простих речень, становлять 7,3% від загальної кількості помилок. Вони представлені в таблиці 4.23.

Таблиця 4.23

### Помилки у побудові простих речень системи Babylon

Типи помилок	Кількість помилок	
1. Порушення синтаксичних зв'язків між членами речення	55	56,1%

## Продовження таблиці 4.23

Типи помилок	Кількість помилок	
2. Граматична неоднотипність однорідних членів речення	31	31,6%
3. Порушення координації між підметом і присудком	11	11,2%
4. Неправильний порядок слів	1	1,1%
Усього	98	100%

У простих реченнях система зробила такі помилки, як:

1. Порушення синтаксичних зв'язків між членами речення:

*Manufacturing and installing of reinforced concrete constructions and concrete frames is using of concrete mix (transportation and laying in a form with subsequent compression of the mix (casing)); keeping the conditions for better concrete curing; making solid sections [236]. – Виробництво та встановлююче конструкцій залізобетону та конкретних каркасів використовує бетону змішуються (перевезення та покладаюче у формі з наступним стиском змішуються (склад акторів)); keeping умови для кращого бетону лікуючого; укладення твердих секцій.*

Неправильно визначена частиномовна належність ряду слів призвела до порушення синтаксичних зв'язків між членами речення. Так, іменники *installing*, *laying* і *curing* були перекладені формами дієприкметників *встановлююче*, *покладаюче* і *лікуючого* відповідно. Це призвело до граматичної неоднотипності однорідних підметів *виробництво*, *встановлюючи*, *перевезення*, *покладаюче*. Іменник *mix* був двічі перекладений формою дієслова *змішуються*. У результаті цього порушився підрядний зв'язок між присудком *використовує* та додатком *змішуються*, між підметом *покладаюче* і додатком *змішуються*, між підметом *keeping* і означенням *лікуючого*. Неправильне визначення частиномовної належності слів та наявність неперекладеного слова *keeping* спричинили

граматичну неоднотипність однорідних присудків *використовує, keeping* і *укладення*. На порушення синтаксичних зв'язків у реченні також вплинув неправильний переклад прикметника *concrete* у словосполученні *concrete mix* формою іменника *бетон*, а також дослівний переклад словосполучення *concrete curing* – *бетону лікуючого*.

2. Граматична неоднотипність однорідних членів речення (див. попереднє речення).

3. Порушення координації між підметом і присудком:

*Wi-Fi complements cellular broadband for “data on the move”* [246]. – *Wi-Fi додатки cellular broadband* для “інформації про зміну місця”.

У цьому реченні порушена координація між підметом і присудком, оскільки програма неправильно визначила частиномовну належність дієслова *complements* (*доповнювати*) та переклала його іменником *додатки*.

4. Неправильний порядок слів:

*The company focuses on hi-tech integrated solutions, production and maintenance of equipment used in mining, processing and transportation of mineral resources, as well as construction of mines* [257]. – *Фокуси компанії на вітаю-tech інтегрованих рішень, виробництво та підтримка устаткування використаного у гірництві, обробці та перевезенні надр, так же, як і будівництва шахт*.

Програма неправильно визначила частиномовну належність дієслова *focuses* і переклала його формою іменника *фокуси*. Також була необґрунтовано використана граматична трансформація перестановки, що вплинуло на порядок слів у реченні і спричинило відсутність присудка.

**Помилки у побудові словосполучень** складають 12,9% від загальної кількості помилок. Вони представлені в таблиці 4.24.

Помилки у побудові словосполучень включають:

1. Порушення керування:

*The charger allows charging of cells made in the NiMH and NiCd technology with the AAA and AA battery size* [242]. – *Charger дозволяє звинувачують клітин зроблених у Nimh та Nicd технології з AAA та AA величина батареї*.



У цьому реченні іменник *клітин* вжито у родовому відмінку, проте правильним є знахідний відмінок. Після прийменника з іменник *величина* вжитий у називному відмінку, проте слід було використати орудний відмінок.

Таблиця 4.24

### Помилки у побудові словосполучень системи Babylon

Типи помилок	Кількість помилок	
1. Порушення керування	97	55,7%
2. Порушення узгодження між словами	32	18,4%
3. Неправильний вибір прийменника	27	15,5%
4. Відсутність необхідного прийменника	10	5,7%
5. Відхилення від нормативно-узвального порядку слів	6	3,5%
6. Вживання зайвого прийменника	2	1,2%
Усього	174	100%

2. Порушення узгодження – порушення синтаксичного зв'язку між залежним та стрижневим словом:

а) у числі: *trench making for underground pipelines* [236] – траншея спрямовуючих до *підземних трубопроводу*; *any such armed attack* [265] – будь-які такі збройні напад; *groundwork is trenching, trench excavation and pitting for isolated supports* [236] – *groundwork* рие, риють розкопку та *pitting* для *ізолюваних підтримки*;

б) у роді: *On the back there is usually a white felt providing additional sound and thermal insulation as well as better comfort* [269]. – На спині є звичайно білий відчувала забезпечуючи додатковий звук та термічна ізоляція так же, як і *краще комфорт*;

3. Неправильний вибір прийменника.

Прийменники є багатозначними словами, тому для вибору правильного варіанта перекладу слід аналізувати контекст. Оскільки програма не робить цього, виникають помилки у виборі прийменника. Наприклад:

*Both are installed **from** wall to wall [269] – обидва встановлюються з стіни до стіни.*

*DC (300mA) LED power supply is characterized with the input voltage **from** 7-24 VDC or 7-21 VAC and self-adjusting output voltage in the range **from** 3 to 21 VDC [242]. – DC (300mA) ВІВ енергетичне постачання характеризується з введенням voltage з 7-24 VDC або 7-21 VAC та самого-лагодячого випуску voltage у діапазоні з 3 до 21 VDC.*

У всіх випадках прийменник **from** перекладений як з. У словнику він має кілька значень: а) у просторовому значенні – від, з, із; б) у значенні відправної точки – від, з, на [218, с. 454].

Отже, у всіх випадках слід було вжити у перекладі прийменник **від**.

4. Відсутність необхідного прийменника. Наприклад:

*However the PC will not vanish with sales set to rise **by** 15 percent next year [251]. – Але PC не зникнуть з збутом братися за збільшення **15 процентом** наступного року.*

В оригіналі є прийменник **by**, проте в перекладі він відсутній. Правильний варіант – *зросте на 15%*.

*From coke and sinter production **to** steelmaking [259] – з coke та sinter виробництво steelmaking.*

У цьому випадку відсутній переклад прийменника **to** – **до**. Окрім цього, програма неправильно переклала прийменник **from** – **від**, обравши варіант з.

5. Відхилення від нормативно-узуального порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір граматичної трансформації:

*Installing of **reinforced concrete constructions** [236] – встановлююче **конструкцій залізобетону**, правильний варіант – встановлення **залізобетонних конструкцій**.*

*The growth of applications for non-PC items outstrips traditional software sales* [251] – *ріст застосувань для не-рс пунктів outstrips традиційний збут програмного забезпечення*, правильний варіант – *продаж традиційного програмного забезпечення*.

Програма необґрунтовано застосувала граматичну трансформацію перестановки, що вплинуло на адекватність перекладу словосполучень.

6. Вживання зайвого прийменника:

*New phones with forward-facing cameras* [246] – *нові телефони з переднім-зустрічаючими з камерами*.

У словосполученні є зайвим прийменник *з*, якого немає в оригіналі. Окрім цього, програма неправильно переклала сам термін *forward-facing cameras*, тому тут також наявна лексична помилка. Правильний переклад – *телефони з камерою на передній панелі*.

**Морфологічні помилки** пов'язані з неправильним визначенням омонімічних граматичних форм англійської мови, порушенням словозмінних граматичних категорій української мови (числа, відмінка, часу). Вони становлять 13,5% від загальної кількості помилок та представлені в таблиці 4.25.

До **морфологічних помилок** належать:

1. Помилки при знятті граматичної омонімії, які є найчисленнішими серед морфологічних помилок програми – 84,5%. Наприклад:

*The Divisions of Corum Group produce coal-cutting, heading, stationary and electrical equipment, mine transport* [257]. – *Відділення Corum Групують продукцію вугілля-розрізання, заголовки, стаціонарне та електричне устаткування, шахта транспортують*.

У реченні програма неправильно визначила частиномовну належність іменників *Group, transport*, прикметників *heading, mine*, дієслова *produce* та обрала їм неправильні еквіваленти в українській мові. Правильними варіантами є *виробляють, прохідницьке* та *шахтний транспорт* відповідно.

Таблиця 4.25

## Морфологічні помилки системи Babylon

Типи помилок	Кількість помилок	
1. Помилки при знятті граматичної омонімії	153	84,5%
2. Порушення категорії числа іменника	15	8,3%
3. Ненормативне утворення відмінкової форми іменника	5	2,8%
4. Ненормативне утворення ступенів порівняння прикметників	3	1,7%
5. Вживання ненормативної форми дієприслівника	3	1,7%
6. Вживання ненормативної форми дієприкметника	1	0,5%
7. Порушення категорії часу дієслова	1	0,5%
Усього	181	100%

2. Порушення категорії числа іменників:

*Charger enables cells discharging with 100mA, 200mA, 350mA currents* [242]. – *Charger дозволяє клітини звільнюючі з 100та, 200та, 350та мечія.*

*In 2012, the proportion should continue to rise in North America and start to spread around the world* [246]. – *У 2012, пропорції мусить продовжити брати початок у північну Америку та початок розповсюдитися навколо світу.*

У першому реченні в оригіналі іменник *currents* вживається у множині, проте у машинному перекладі його відповідник має форму однини *мечія*. У другому реченні іменник *proportion* вжито в однині, а у перекладі він має форму множини *пропорції*.

3. Ненормативне утворення відмінкової форми іменника, наприклад:

*On the territory of or on the Islands under the jurisdiction of any of the Parties in the North Atlantic area north of the **Tropic of Cancer** [265] – на території або на Островах під юрисдикцією будь-яких Сторін у північній атлантичній області північ **Тропіку Рака***

Іменник **Тропік** належить до твердої групи II відміни, тому у родовому відмінку однини він має мати закінчення **-а**.

*In exercise of the right of individual or collective self-defence recognised by **Article 51 of the Charter of the United Nations** [265] – у справі права особи або колективного самого-захисту визнаного **Статтею 51 Хартії Організації Об'єднаних націй***

Іменник **стаття** належить до м'якої групи I відміни, у якій іменники в орудному відмінку можуть мати закінчення **-єю, -єю**. У слові **стаття** правильним є закінчення **-єю**.

4. Ненормативне утворення ступенів порівняння прикметника:

*In order **more effectively** to achieve the objectives of this Treaty [265] – на порядку денному **більше ефективно** досягнути мети цієї Угоди*

Відповідно до норм української мови при утворенні вищого ступеня порівняння прикметників вживається прислівник **більш**, а не **більше**.

5. Вживання ненормативної форми дієприслівника:

***Charging /discharging current** [242] – **звинувачувавши**/звільнюють течію.*

***Charging /discharging time** [242] – **звинувачувавши**/звільнюють час.*

В обох випадках вжито дієприслівник **звинувачувавши**, відсутній в українській мові. Правильно – **звинувачуючи**. Проте і цей варіант є некоректним, оскільки у словосполученнях вживається не дієприслівник, а іменник **charging**, еквівалентом якого в українській мові є термін **зарядка**. Тому правильними є варіанти: *струм зарядки / розрядки і час зарядки / розрядки*.

6. Вживання ненормативної форми дієприкметника:

*Being the dominant operating system provider in the **emerging non-PC market** [251] – являючи собою домінуючу діючу систему провайдер у **з'являючому не-р**ринку.*

У машинному перекладі вжито дієприкметник *з'являючому*, який відсутній в українській мові. Програма неправильно визначила частиномовну належність прикметника *emerging* та здійснила його неадекватний переклад. Правильний варіант: *новий ринок*.

#### 7. Порухення категорії часу дієслова:

*The Parties, separately and jointly, by means of continuous and effective self-help and mutual aid, will maintain and develop their individual and collective capacity to resist armed attack* [265]. – *Сторони, окремо та об'єднано, за допомогою безперервного та ефективного самовдосконалення та взаємної допомоги, підтримають та розвивають їхню особу та колективну потужність опиратися збройні напад*.

У реченні дієслово *develop* вживається в майбутньому часі, на що вказує наявність допоміжного дієслова *will*. Проте при перекладі програма використала дієслово *розвивають* у формі теперішнього часу, що є помилкою. Правильний варіант: *будуть розвивати*.

**Лексико-граматичні помилки** становлять 1,2% від загальної кількості помилок. Їхнє виникнення спричинене неправильним вибором відповідників в українській мові та необґрунтованим застосуванням граматичної трансформації, що порушує нормативно-узв'язаний і логіко-синтаксичний порядок слів:

*Pile foundation* [236] – *фундація купу*.

Вживаючись поза контекстом слова *pile* і *foundation* справді мають значення *купа* і *фундація*, проте разом вони утворюють словосполучення *pile foundation*, яке перекладається як *пальовий фундамент* [228, с. 183].

*Field studies* [257] – *навчання сфери*.

Програма окремо та неадекватно переклала кожне слово, яке входить до складу словосполучення. Окрім цього, вона змінила порядок слів. Правильний варіант – *польові дослідження* [228, с. 170].

**Орфографічні помилки** становлять 0,2% від загальної кількості мовних помилок, виявлених у текстах машинних перекладів. До них належить неправильне написання таких слів:

*Data on the move* (Зв'язок) – інформація про зміну **місьця**.

Неправильним є написання слова **місьця** з м'яким знаком. Правильний варіант – **міця**.

*Yet 2011 will mark the tipping point as the growth of applications for non-PC items outstrips traditional software sales and consumers embrace a wider variety of devices* [251]. – *Все ще 2011 поставить оцінку tipping пункт як ріст застосувань для нер-с пунктів outstrips традиційний збут програмного забезпечення та споживачи охоплюють більш широкий різновид приладів.*

У цьому реченні є дві орфографічні помилки:

- 1) написання слова **застосувань** із подвоєнням літери **н**;
- 2) написання слова **споживачи** із закінченням **-и** замість **-і**.

*400 million new phones* (Зв'язок) – **400 мільону**

Слово **мільону** написано без літери **й** перед **о**.

Лінгвістичний аналіз помилок, знайдених у текстах машинних перекладів системи Babylon, показав переважання лексичних помилок над граматичними – 56,7% та 41,9% відповідно. Це переважно помилки у перекладі слів (38,5%), зокрема вибір нерелевантних перекладних відповідників (21,3%) внаслідок неправильного вирішення лексичної полісемії (14,1%). Причиною їх виникнення є недостатня повнота паралельних корпусів текстів, на основі яких працює програма. Лексико-граматичні та орфографічні помилки найменш чисельні і становлять 1,2% та 0,2% відповідно.

#### **4.5 Оцінка якості системи Free Translation**

Оцінка якості програми Free Translation проводилася на матеріалі 110 речень машинного перекладу. Усього було проаналізовано 110 пар речень. Результати дослідження подано в таблиці 4.26. Більш детально помилки системи представлені в додатку Ж.

Таблиця 4.26

**Кількісне співвідношення помилок системи Free Translation**

<b>Типи помилок</b>	<b>Кількість помилок</b>	
1. Лексичні помилки	327	35,6%
1.1. Помилки у перекладі	182	19,8%
1.2. Помилки у перекладі словосполучень	145	15,8%
2. Граматичні помилки	587	64,0%
2.1. Морфологічні помилки	110	12,0%
2.2. Синтаксичні помилки:	477	52,0%
2.2.1 Помилки в побудові словосполучень	220	24,0%
2.2.2 Помилки в побудові речень:	257	28,0%
а) помилки у побудові простих речень	124	13,5%
б) помилки у побудові складних речень	133	14,5%
3. Лексико-граматичні помилки	2	0,2%
4. Орфографічні помилки	2	0,2%
Усього	918	100%

Помилки у побудові складних речень програми представлені в таблиці 4.27.

Таблиця 4.27

**Помилки у побудові складних речень системи Free Translation**

<b>Типи помилок</b>	<b>Кількість помилок</b>	
1. Порушення координації між підметом і присудком	44	33,1%
2. Порушення синтаксичних зв'язків між членами речення	43	32,3%
3. Граматична неоднотипність однорідних членів речення	20	15,0%



## Продовження таблиці 4.27

Типи помилок	Кількість помилок	
4. Пропущені члени речення	8	6,0%
5. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	5	3,8%
6. Неправильний порядок слів	4	3,0%
7. Вживання зайвих членів речення	4	3,0%
8. Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення	4	3,0%
9. Штучне ускладнення речення	1	0,8%
Усього	133	100%

У побудові **складних речень** було виявлено такі помилки:

1. Порухення координації між підметом і присудком:

*Roofwork is attic paving, built-up roofing; when making attic coating, steel sheets are often used and in case of a built-up roof, roll materials are welded to the prepared base [236].* – *Roofwork є горища тротуарної плитки, вбудовані до покрівлі; при прийнятті горища покриття, сталевих листів часто використовується і у випадку забудована даху, рулонних матеріалів, зварних до підготовлених бази.*

У результаті неадекватного морфологічного аналізу та неправильного перекладу словосполучень *attic paving* і *built-up roofing* у реченні порушена координація між підметом *roofwork*, який програма не переклала, і однорідними присудками – *горища тротуарної плитки, вбудовані до покрівлі*. У машинному перекладі присудок *вбудовані до покрівлі* виконує функцію означення, що не відповідає тексту оригіналу. У другій частині речення порушена координація підмета *листів* і присудка *використовується* у числі та відмінку.

Також порушено синтаксичний зв'язок підмета *матеріалів* і присудка *зварних*. Оскільки програма неправильно виконала синтаксичний і морфологічний

аналіз, дієслово *are welded* було перекладене формою дієприкметника *зварних*, який у машинному перекладі виконує функцію означення.

У реченні наявні інші помилки. Так, програма неправильно визначила частиномовну належність прикметника *attic* та переклала його іменником *горища*, тому у частині *при прийнятті горища покриття* є два додатки і відсутнє означення.

У реченні також порушене узгодження між означенням *забудована* і додатком *даху* у роді та відмінку, а також між означенням *підготовлених* і додатком *бази* у числі та відмінку.

2. Порушення синтаксичних зв'язків між членами речення:

*The Parties will contribute toward the further development of peaceful and friendly international relations by strengthening their free institutions, by bringing about a better understanding of the principles upon which these institutions are founded, and by promoting conditions of stability and well-being [265].* – *Сторони будуть сприяти подальшому розвитку мирного та дружнього міжнародних відносин шляхом посилення їх безкоштовно установ, шляхом залучення про краще розуміння принципів, на яких засновано цих установ, а також сприяння розвитку умов стабільності і добробуту.*

У машинному перекладі речення порушене узгодження між означеннями *мирного* і *дружнього* та додатком *відносин* у числі. Це призвело до граматичної неоднотипності однорідних означень *мирного, дружнього* і *міжнародних*. Також відсутнє узгодження між означенням *безкоштовно* і додатком *установ* через переклад прикметника *free* формою прислівника *безкоштовно*. Окрім цього, порушено зв'язок керування між додатком *установ* і присудком *засновано* у відмінку.

3. Граматична неоднотипність однорідних членів речення:

*Corum Group successfully operates in Ukraine, Russia, Belarus, Kazakhstan, Poland, Vietnam, India, and enters Chinese and South African markets; the company supplies its products to Romania, Georgia, Armenia, Estonia, Macedonia, Bosnia and Herzegovina, Montenegro and other countries [257].* – *Corum група успішно працює в Україні, Росії, Білорусі, Казахстану, Польщі, в'єтнамі, Індії та входить до*

*Китаю та Південної Африки ринках; компанія постачає свою продукцію до Румунії, Грузії, Вірменії, Естонія, Македонія, Боснія і Герцеговина, Чорногорія і інших країн.*

У наведеному прикладі простежується граматична неоднотипність однорідних обставин *Україні, Росії, Білорусі, Казахстану, Польщі, В'єтнамі, Індії* через неправильний вибір відмінкової форми *Казахстану*. У той час, як однорідні обставини вживаються в місцевому відмінку, *Казахстан* має форму родового відмінка. В іншій частині речення теж порушена граматична однотипність однорідних обставин *Румунії, Грузії, Вірменії, Естонія, Македонія, Боснія і Герцеговина, Чорногорія і країн* через вживання *Естонія, Македонія, Боснія і Герцеговина, Чорногорія* в називному відмінку, коли інші іменники мають форму родового відмінка.

#### 4. Пропущені члени речення:

*It is the producer of high quality steel plate 6-200 mm thick and as wide as 1500-3200 mm that is broadly used in shipbuilding, power and special machine building, bridge and offshore deep sea platforms construction, manufacture of large diameter pipes for gas and oil heavy duty pipelines [259].* – Це є виробником високоякісного товстолистового прокату товщиною 6-200 мм і шириною 1500-3200 мм для суднобудування, енергетичного й спеціального машинобудування, мостобудування, глибоководних, виробництво труб великого діаметра для магістральних газо-тяжких трубопроводів.

У перекладі пропущені повнозначні члени речення оригіналу *deep sea platforms construction* і *and oil*.

5. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення:

*The basic difference between them is the manufacturing process itself in which they are made [269].* – Основна відмінність між ними є сам процес виробництва в яких вони робляться.

У реченні порушене узгодження між сполучним словом – займенником *яких*, що вживається у формі місцевого відмінка множини, та іменним присудком *процес*

**виробництва**, який має форму називного відмінка однини. Таким чином, займенник та іменник не узгоджуються у числі та відмінку.

6. Неправильний порядок слів:

*PC sales are set to hit 400 million in 2011, yet the combined sales of smartphones, tablets and non-PC netbooks will exceed that number for the first time [251].* – Продаж комп'ютер встановлено хіт 400 млн. у 2011 році, проте об'єднані продаж смартфонів, пігулок і не буде перевищувати кількість комп'ютера нетбуками вперше.

У результаті неправильного синтаксичного аналізу речення, виконаного програмою, та необгрунтованої граматичної трансформації перестановки у машинному перекладі речення спостерігаємо неправильний порядок слів. Так, присудок *не буде перевищувати* вживається перед підметом *комп'ютера нетбуками*, що порушує синтаксичні зв'язки в реченні. Також внаслідок перестановки речення стало заперечним, тоді як в оригіналі воно стверджувальне.

7. Зайві члени речення:

*The Technology, Media and Telecommunications (TMT) practice at Deloitte today announces its global predictions for the telecommunications sector, forecasting that video calling will become cheaper, more widely available and of a much improved quality in 2011, yet the long-awaited boom in demand will not happen [246].* – На технології, медіа та телекомунікації (англ. ) Інші бюлетені) практики компанії “Делойт” оголошує сьогодні глобальний прогнозів на телекомунікаційній сфері, прогнозування, відео телефонує, буде дешевшати, більш широко доступні і в 2011 році значно поліпшити якість, але довгоочікуваного бум у попит не станеться.

У машинному перекладі є слова (англ. ) *Інші бюлетені*), які відсутні в оригіналі. Програма використала їх замість аббревіатури *TMT*.

8. Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення:

*However this diversity leads to fragmentation and complexity at the hardware level which causes a headache for application developers, media companies and IT departments [251].* – Але це розмаїття призводить до фрагментації і складності на

обладнання рівня, **який** викликає головний біль для заяви девелоперів, медіа-компаній та відділів.

Сполучник **which** має кілька значень – *що, хто; який, котрий* [219, с. 675]. У машинному перекладі програма використала сполучник **який**, проте в цьому випадку він співвідноситься зі словом **обладнання**. В оригіналі **which** стосується додатка **complexity** – **складності**. Тому, відповідно до контексту, правильним є сполучник **що**.

#### 9. Штучне ускладнення речення:

*The researcher notes that greenhouse gases emissions during extraction of shale gas are greater than in case of conventional natural gas extraction, and are even close to emissions associated with coal extraction* [240]. – Дослідник зазначає, що сукупні викиди парникових газів під час видобування сланцевого газу, **що є більшою**, ніж у разі звичайного природного газу, і навіть є близьким до викидів пов'язані з видобутку вугілля.

В українському реченні програма використала сполучник **що**, який відсутній в англійському варіанті. Це призвело до штучного ускладнення першої частини речення, у результаті чого була порушена координація між підметом і присудком у числі.

Помилки у побудові **простих речень** становлять 14,5% від загальної кількості помилок. Вони подані в таблиці 4.28.

Таблиця 4.28

#### Помилки у побудові простих речень системи Free Translation

Типи помилок	Кількість помилок	
1. Порушення синтаксичних зв'язків між членами речення	51	41,1%
2. Граматична неоднотипність однорідних членів речення	31	25,0%

Продовження таблиці 4.28

Типи помилок	Кількість помилок	
3. Порушення координації між підметом і присудком	16	12,9%
4. Пропущені члени речення	11	8,9%
5. Неправильний порядок слів	9	7,3%
6. Зайві члени речення	5	4,0%
7. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	1	0,8%
Усього	124	100%

У машинних перекладах, виконаних програмою Free Translation, було виявлено такі помилки у побудові **простих речень**:

1. Порушення синтаксичних зв'язків між членами речення:

a) *It currently mines iron ore from its one open-pit quartzite field through a process of drilling and blasting, and by the removal of overburden to external dumps [238]. – Зараз вона шахт залізорудного відкритими з одним піт-кварцитного щебеню поле через процес буріння та потужним, і на усунення розкривних порід у зовнішні відвали.*

Неправильний синтаксичний аналіз речення та морфологічний аналіз його членів призвели до порушення синтаксичних зв'язків у машинному перекладі. Так, порушена координація між підметом і присудком через неправильне визначення частиномовної належності та неадекватний переклад дієслова *mines* – *шахти*, правильний варіант – *видобуває*. Неправильний морфологічний аналіз словосполучення *iron ore* – *залізорудний* призвів до використання його у функції означення, проте іменник *залізна руда* виконує функцію додатка. Дослівний переклад та необґрунтована граматична трансформація перестановки частини *from*

*its one open-pit quartzite field* призвели до порушення синтаксичного зв'язку між присудком *шахт* та обставиною *відкритими з одним піт-кварцитного щебеню поле*. Програма неправильно визначила частиномовну належність іменника *blasting* та переклала його прикметником *потужний*. Це призвело до порушення синтаксичного зв'язку між присудком *шахт* та обставиною *потужним*. Окрім цього, у реченні наявна граматична неоднотипність однорідних обставин *буріння* і *потужним*.

б) *The amount of energy for shale gas extraction, in turn, depends on specific geology conditions and depth of shale gas reserves* [238]. – *Кількість енергії для сланцевого газу, у свою чергу, залежить від конкретних умов геології і глибина сланцевого газу.*

У реченні порушений синтаксичний зв'язок між присудком *залежить* та одним з однорідних додатків *глибина сланцевого газу*. Причиною є неправильний вибір відмінкової форми додатка. Замість родового відмінка іменника *глибина* програма використала називний. Ця помилка спричинила також граматичну неоднотипність однорідних додатків *умов геології* та *глибина сланцевого газу*.

2. Граматична неоднотипність однорідних членів речення (див. приклади 1а, 1б).

3. Порушення координації між підметом і присудком (див. приклад 1а).

4. Пропущені члени речення:

*Metinvest mining division comprises three iron ore enrichment works, a flux and mining company and two coking coal companies* [238]. – *Гірничорудний дивізіон Групи Метінвест входять три ГЗК, гірничо-потоку і два коксівного вугілля.*

У процесі роботи програма пропустила два члени речення – *company, companies*. Окрім цього, у реченні порушена координація між підметом *дивізіон* та присудком *входять* у числі через неправильний синтаксичний аналіз речення та неадекватний переклад дієслова *comprise* – *включає*.

5. Неправильний порядок слів (див. приклад 1а).

6. Зайві члени речення:

*The Divisions of Corum Group produce coal-cutting, heading, stationary and electrical equipment, mine transport, machinery for open-pit mining, including bucket wheel excavators, apron feeders, conveyors, loaders, spreaders, stackers, etc. [257]. – Підрозділи Групи Corum видобувати вугілля-різання, заголовки, стаціонарні та електричного обладнання, шахтного транспорту, обладнання для відкритого піт-корисних копалин, у тому числі роторні екскаватори, перону живильники, транспортерів, навантажувачі, розповсюджувачів, штабелери і укладальники розділових прокладок, тощо.*

У машинному перекладі вживається зайвий додаток – **укладальники розділових прокладок**, якого немає в оригіналі. Окрім цього, порушена координація між підметом **підрозділи** і присудком **видобувати** в числі, особі та часі.

Також наявні граматично неоднотипні однорідні члени речення:

а) означення – **вугілля-різання, заголовки, стаціонарні та електричного**, які не узгоджуються в частиномовній належності та відмінку;

б) додатки – **обладнання, транспорту, обладнання, роторні екскаватори, перону живильники, транспортерів, навантажувачі, розповсюджувачів, штабелери**, через неправильний вибір відмінкової форми **транспорту, транспортерів і розповсюджувачів**.

7. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення:

*These include, but not limited to, the **power** (and **its** sources) at sites for the work of pumps, combusting (and type) of fuel for extraction and transportation of water, sand, waste rock etc [240]. – Вони включають, але не обмежуються, **силу** (і **його** джерела) на об'єктах роботи насоси, потрібно спалювати останню (і тип) палива для видобування та транспортування води, з піском, відходів рок-тощо.*

У машинному перекладі займенник **його** не співвідноситься з іменником **сила** у роді. Неправильне визначення частиномовної належності іменника **combusting** та його неадекватний переклад словосполученням **потрібно спалювати останню**, використання зайвого прийменника **з** перед словом **піском**, а також неправильний вибір відмінкової форми і переклад **waste rock – відходів рок** призвели до



граматичної неоднотипності однорідних додатків: *силу, потрібно спалювати останню (і тип) палива; води, з піском, відходів рок-тощо.*

Помилки у побудові словосполучень представлені в таблиці 4.29.

Таблиця 4.29

### Помилки у побудові словосполучень системи Free Translation

Типи помилок	Кількість помилок	
1. Порушення керування	109	49,5%
2. Порушення узгодження між словами	66	30,0%
3. Неправильний вибір прийменника	16	7,3%
4. Вживання зайвого прийменника	16	7,3%
5. Відсутність необхідного прийменника	7	3,2%
6. Відхилення від нормативно-узвального порядку слів	6	2,7%
Усього	220	100%

Помилки у побудові словосполучень становлять 24,0% від загальної кількості помилок. Вони включають:

#### 1. Порушення керування.

*The company focuses on hi-tech integrated solutions, production and maintenance of equipment used in mining, processing and transportation of mineral resources, as well as construction of mines [257].* – **Компанія концентрується на hi-tech інтегровані рішення, виробництва і обслуговування обладнання, що використовується при видобутку, переробки і транспортування корисних копалин, а також будівництво шахт.**

У машинному перекладі після прийменника **на** іменники **рішення, виробництва і обслуговування** вжито в називному чи знахідному відмінку, проте правильним є вживання місцевого відмінка.

Прийменник *при* вимагає після себе місцевого відмінка, проте іменники *переробки* і *транспортування* вживаються родовому.

*Video exceeds the needs of most businesses* [246] – відео *перевищує потребам* бізнесу.

У машинному перекладі іменник *потребам* вживається в давальному відмінку, проте правильним є його використання в знахідному відмінку: *відео перевищує потреби*.

2. Порушення узгодження між залежним та стрижневим словом:

а) у відмінку: *develop their individual and collective capacity* [265] – *розвивати їх індивідуальних і колективних спроможності*;

б) у числі: *Deloitte announces its global predictions* [251] – “Делойт” оголошує *про глобальну прогнози*;

в) у числі та відмінку: *self-adjusting output voltage* [242] – *саморегульованими вихідна напруга*;

г) у роді: *by means of continuous and effective self-help* [265] – *за допомогою безперервної та ефективного самоврядування*.

3. Неправильний вибір прийменника:

*Building of constructions (holders, vaults, walls, pillars) from single-piece units and stones* [236] – *будівництво споруд (власників, сейфи, стін, колон) від одного-шматок одиниць і каменів*.

Коли прийменник *from* вказує на матеріал, речовину, з яких щось виготовляють, він перекладається словами *з, із* [218, с. 454], а український відповідник *від* вказує на просторове значення, що є помилковим.

*The battery charger can work independently for each of the four channels* [242] – *зарядний пристрій може працювати самостійно по кожному з чотирьох каналів*

У машинному перекладі прийменник *for* не має українського відповідника *по*. Його первинним значенням є *для* [218, с. 438].

4. Вживання зайвого прийменника:

*In 2012, the proportion should continue to rise in North America* [246] – *у 2012 році в пропорції продовжують зростати у Північній Америці*.

*Corum Group offers its clients the efficient and safe integrated solutions* [257]. – *Corum Group пропонує своїм клієнтам в оперативного та безпечні комплексні рішення.*

У машинних перекладах зайвим є прийменник *в*, який відсутній у текстах оригіналу.

5. Відсутність необхідного прийменника:

*While during extraction of conventional natural gas* [240] – *у той час видобуток природного газу.* Правильний варіант: *у той час як при видобутку газу.*

*Corum Source for after-sales servicing of equipment* [257] – *Corum джерело сервісне обслуговування обладнання.* Правильний варіант: *Corum Source для сервісного обслуговування обладнання.*

6. Відхилення від нормативно-узуального порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір граматичної трансформації перестановки:

*BC-700 Microprocessor charger* [242]. – *Зарядний пристрій BC-700 мікропроцесора.* Правильний варіант: *Мікропроцесорний зарядний пристрій BC-700*

*The entire amount of flat rolled steel produced undergoes 100% non-destructive ultrasonic quality control* [259]. – *Весь обсяг плоский прокат піддається 100% сталі неруйнівного контролю якості ультразвукових.* Правильний варіант: *неруйнівний ультразвуковий контроль.*

Програма необґрунтовано застосувала граматичну трансформацію перестановки, що вплинуло на адекватність перекладу словосполучень.

Причина виникнення **морфологічних помилок** (таблиця 4.30) у машинних перекладах програми Free Translation полягає в тому, що система не використовує граматичні правила у процесі роботи.

До **морфологічних помилок** належать:

1) помилки при знятті граматичної омонімії – неправильне визначення частиномовної належності слів. Оскільки програма не містить блоку морфологічного аналізу та не враховує граматичні характеристики слова та

контекст, вона обирає найбільш частотний еквівалент, що призводить до виникнення помилок у машинних перекладах.

*PC sales are set to hit 400 million in 2011 [251] – продаж комп'ютер встановлено **xim** 400 млн. у 2011 році.* Дієслово *to hit* (*зростати*) – було перекладене формою іменника *xim*.

*The Technology, Media and Telecommunications (TMT) practice at Deloitte today announces its global predictions for the telecommunications sector, forecasting that video calling will become cheaper, more widely available and of a much improved quality in 2011 [246]. – На технології, медіа та телекомунікації (англ. ) Інші бюлетені) практики компанії “Делойт” оголошує сьогодні глобальний прогнозів на телекомунікаційній сфері, прогнозування, відео телефонує, буде дешевшати, більш широко доступні і в 2011 році значно поліпшити якість, але довгоочікуваного бум у попит не станеться.*

У машинному перекладі речення програма неправильно визначила частиномовну належність ряду слів:

- а) дієприслівник *forecasting* (*прогнозуючи*) переклала формою іменника *прогнозування*;
- б) іменник *calling* (*дзвінки*) переклала формою дієслова *телефонує*;
- в) прикметник *improved* (*кращий*), отримав форму дієслова *improved*.

Таблиця 4.30

### Морфологічні помилки системи Free Translation

Типи помилок	Кількість помилок	
1. Помилки при знятті граматичної омонімії	93	84,6%
2. Порухення категорії числа іменника	6	5,5%
3. Порухення категорії часу дієслова	5	4,5%
4. Ненормативне утворення пасивної конструкції дієслова	4	3,6%

## Продовження таблиці 4.30

Типи помилок	Кількість помилок	
5. Ненормативне утворення відмінкової форми іменника	2	1,8%
Усього	110	100%

## 2. Порушення категорії числа іменника:

*Only a nearly irresistible **force** could produce a standard operating system [251] – тільки майже нестримна **сили** могли б виробляти стандартного операційної системи.*

В оригіналі іменник *force* вживається в однині, проте в перекладі він має форму множини *сили*.

## 3. Порушення категорії часу дієслова:

*The worldwide developments forecasted in 2011 **will definitely have impact on or will correlate** with the developments in Ukraine [246]. – В усьому світі розробок прогнозованих у 2011 році обов'язково **мати вплив** на або **будуть корелюють** з розробками в Україні.*

В оригіналі дієслова *will have impact* і *will correlate* вживаються в майбутньому часі. Проте в машинному перекладі часові форми дієслів *мати вплив* та *будуть корелюють* утворені неправильно. Відповідно до норм української мови складена форма дієслів майбутнього часу недоконаного виду утворюється за допомогою особових форм допоміжного дієслова *бути* та інфінітива особового дієслова. У першому випадку відсутня особова форма допоміжного дієслова *бути*, у другому – інфінітив особового дієслова. Правильні варіанти перекладу: *будуть мати вплив, будуть корелювати*.

*Although the public's appetite for smartphones **shows** no sign of slowing down [246] – незважаючи на те, що суспільна апетит для смартфонів **буде показано** ніяких ознак уповільнюється.*

В оригіналі дієслово *shows* вживається в теперішньому часі, проте в машинному перекладі його відповідник використаний у майбутньому часі. Правильний варіант: **не показує**.

4. Ненормативне утворення пасивної конструкції дієслова:

*The technology, which was first demonstrated in 1964, does not yet appear ready for the mass market* [246]. – **Технологія, яка була вперше продемонструвала у 1964 році, поки не з'явиться готовий до масового ринку.**

*Any such armed attack and all measures taken as a result thereof shall immediately be reported to the Security Council* [265]. – **Будь-які з цих збройний напад і всі заходи, які вживаються як наслідок цього має бути негайно повідомляють Раду Безпеки.**

У наведених прикладах спостерігаємо ненормативне утворення пасивних конструкцій **була продемонструвала** та **має бути повідомляють**. Українською мовою їх слід було перекласти формами дієслів доконаного виду: **була продемонстрована, має бути повідомлено**.

5. Ненормативне утворення відмінкової форми іменника:

*On the territory of or on the Islands under the jurisdiction of any of the Parties in the North Atlantic area north of the Tropic of Cancer* [265]. – **На території або на островах у віданні будь-якої зі сторін у Північноатлантичному регіоні північної тропічна раку**

*The Treaty entered into force or the Mediterranean Sea or the North Atlantic area north of the Tropic of Cancer* [265]. – **Договір набув чинності або Середземного моря або в Північноатлантичному регіоні північної тропічна раку**

Іменник **Рак** належить до твердої групи II відміни, тому у родовому відмінку однини він має мати закінчення **-а**.

**Лексичні помилки** становлять 35,6% від загальної кількості помилок. Вони представлені в таблиці 4.31.

**Помилки у перекладі слів** займають 55,7% від усієї кількості лексичних помилок. До них належать:

1. Вибір нерелевантних перекладних відповідників – вживання слів у не властивому для контексту та тематики значенні через неправильне вирішення

лексичної полісемії, лексико-граматичної омонімії, а також неадекватний переклад слів, які слід було залишити в оригінальному варіанті.

Багатозначний термін *vault* [236] був перекладене як *сейф*. У словнику він має кілька значень: 1) склепіння; 2) небозвід; 3) підвал, льох, склеп (винний тощо); 4) сховище; (сталева) камера; сейф; 5) фільмосховище; 6) печера; 7) стрибок (через перешкоду); 8) опорний стрибок; стрибок з жердиною [219, с. 625]. Програма обрала одне з останніх значень, проте у будівельній тематиці правильним варіантом є *склепіння*.

Слово *charter* [265] отримало переклад *хартія*. У словнику *charter* має п'ять значень: 1) хартія; грамота; 2) право, привілей; пільга; 3) патент; 4) статут; 5) чартер [218, с. 191].

Програма обрала перше значення, проте у тестовому реченні правильним відповідником є *статут*.

Таблиця 4.31

### Лексичні помилки системи Free Translation

Типи помилок	Кількість помилок	
	Кількість	Відсоток
1 Помилки у перекладі слів	182	55,7%
1.1 Вибір нерелевантних перекладних відповідників	154	47,1%
1.1.1 Неправильне вирішення лексичної полісемії	117	35,8%
1.1.2 Неправильне вирішення лексико-граматичної омонімії	36	1,1%
1.1.3 Неадекватний переклад власних назв та аббревіатур, які слід було залишити в оригінальному варіанті	1	0,3%
1.2 Неперекладені слова	28	8,6%

## Продовження таблиці 4.31

Типи помилок	Кількість помилок	
2 Помилки у перекладі словосполучень	145	44,3%
2.1 Неадекватний переклад словосполучень як семантичних єдностей	78	23,9%
2.2 Дослівний переклад іншомовних висловів	44	13,4%
2.3 Неадекватний переклад фразеологічних сполучень	15	4,6%
2.4 Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті	8	2,4%
Усього	327	100%

У результаті неправильного визначення частиномовної належності слів *Free Translation* обрала їм неадекватні відповідники в українській мові. Так, у реченні *their inventory and abatement are important to minimize the impact of shale gas extraction on climate* [240] іменник *abatement* був перекладений дієслівним словосполученням *відповідає пріоритетам*. Проте правильним варіантом є іменник *зменшення* [228, с. 12], що відповідає контексту речення.

У словосполученні *to increase production of high quality merchant iron ore concentrate* [238] прикметник *merchant* (товарний) був перекладений формою іменника *комерсанта*: *збільшити випуск високоякісного концентрату з Комерсанта вмісту Fe*.

Програма неадекватно переклала одну аббревіатуру – *TMT* [246] – як *англ. Інші бюлетені* замість того, щоб залишити її в оригінальному варіанті.

2. Неперекладені слова. Система залишила без перекладу 28 слів, серед яких:

а) власні назви: *Bakas* [246; 251]; *Howarth* [240];

б) скорочення: *Mtonnes* [259];

в) аббревіатури: *CA, V, VAC* [242];



г) терміни: *earthfills, stonework, cabinet, roofwork, transportwork* [236], *forthwith* [265]; *hi-tech* [257]; *needlepunched, tufting, needling, band* [269].

**Помилки у перекладі словосполучень** становлять 44,3% від загального обсягу лексичних помилок. До них належать:

1. Неадекватний переклад словосполучень як семантичних єдностей.

Термін *concrete frames* [236] був перекладений як **конкретні кадри**, проте правильний варіант – **бетонні конструкції**. Обидва компоненти цього словосполучення є багатозначними словами. У словнику *concrete* має такі варіанти перекладу: 1) бетонний; 2) конкретний [218, с. 223]. *Frame* може означати: 1) рама, рамка; 2) каркас, кістяк, остов; 3) корпус, станина; 4) конструкція; споруда; будова; 5) будова, побудова; система; структура; 6) будова тіла, статура, фігура, конституція; 7) характер, склад; 8) парник; 9) ткацький верстат; 10) ферма, кроква; 11) шпангоут; елемент набору; 12) опока; 13) система координат; 14) (окремих) кадр [218, с. 448]. Програма обрала неправильні еквіваленти кожному зі слів, що привело до неадекватного перекладу усього словосполучення.

Словосполучення *machinery for open-pit mining* [257] було перекладене як **обладнання для відкритого піт-корисних копалин**, проте правильним є варіант – **обладнання для відкритих гірничих робіт** [228, с. 273].

Система переклала словосполучення *primary backing* як **первинний підкріплення** [269]. Проте цей варіант не відповідає контексту речення *Tufted carpets are made in the process of tufting – it is such a technique of manufacturing carpets in which the raw material is inserted, by means of the needle, in some sort of the textile base (primary backing) thus making the carpet pile* [269], яке було перекладене як *Прошивним килимки, зроблені в процесі tufting - це така техніка виготовлення серветок в яких сировина встановлено, за допомогою стержня, в якусь текстильної бази (первинний підкріплення) таким чином, здійснення ворс килима*. У словнику *backing* має кілька варіантів перекладу: 1) підтримка, допомога; схвалення; підкріплення; 2) прихильники; 3) зміцнення; закріплення; 4) опора, підпірка, підпора; 5) субсидування, фінансування; 6) покриття; 7) зворотний бік, зворот; виворіт; 8) підкладка; 9) спинка; 10) задник; 11) відступ; рух назад;

зворотний хід; задній хід; 12) об'їждження (коня); 13) сідання (на коня); 14) друкування на зворотному боці аркуша; 15) затримка, запізнення; сповільнення [218, с. 94].

Саме тому правильним варіантом перекладу є *первинна підкладка*.

2. Дослівний переклад іншомовних висловів, які в українській мові мають свої відповідники:

*Steelmaking company* [259] – виплавка компанії, правильний варіант – *металургійне підприємство* [222, с. 432].

*Apron feeders* [257] – перону живильники, правильний варіант – *пластинчасті живильники* [221, с. 451].

3. Неадекватний переклад фразеологічних сполучень.

Машинний переклад звороту *in cash* [251] – *в готівку* є неправильним, оскільки він перекладається українською як *готівкою, з грошми* [220, с. 525].

*Just to mention few* [246] система переклала як *просто кажучи вже про кілька*. Правильним є переклад – *згадати кілька*.

4. Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті: *Corum Quality* [257] – *Corum якості*; *Corum Source* [257] – *Corum джерело*; *Corum Group* [257] – *Corum Група*.

**Лексико-граматичні помилки** займають 0,2% від загальної кількості помилок. Наприклад:

*Manufacturing and steel erection* [236] – *виробництво сталі і зведення*, правильний варіант – *виготовлення і монтаж металоконструкцій* [228, с. 158].

У наведеному прикладі програма не тільки неправильно вибрала еквіваленти до англійських словосполучень, а й необґрунтовано використала граматичну трансформацію перестановки.

**Орфографічні помилки** становлять 0,2% від загального обсягу помилок та включають два випадки неправильного написання слів:

*Number of public hotspots* [246] – *кількість державних Точкии публічного доступу*.

*All kinds of field studies in mining geology together with the leading specialised research institutions [257] – всі види польових досліджень у гірничо-геологічного разом з провідними спеціалізованих науково-дослідних установ.*

Лінгвістичний аналіз помилок у текстах машинного перекладу Free Translation показав переважання граматичних помилок над лексичними – 64,0% та 35,6% відповідно. Причиною цього є невикористання програмою граматичних правил у процесі роботи. Серед граматичних помилок домінують синтаксичні помилки (52,0%), зокрема помилки у побудові речень (28,0%). Лексико-граматичні та орфографічні помилки є найменш чисельними та становлять по 0,2%. Покращити якість роботи Free Translation можна за допомогою вдосконалення алгоритмів функціонування системи та її подальшого навчання на основі паралельних корпусів текстів.

#### 4.6 Оцінка якості системи Yandex

Оцінка якості системи Yandex проводилась на матеріалі 110 речень, які належать до таких тематик: будівництво, металургійна, право, електроніка, машинобудування, зв'язок, комерція, текстиль, гірнична, екологія [236; 238; 240; 242; 246; 251; 257; 259; 265; 269]. Усього було проаналізовано 110 пар речень.

Результати лінгвістичного аналізу текстів машинного перекладу подано в таблиці 4.32, детальна таблиця представлена у додатку 3.

Таблиця 4.32

#### Кількісне співвідношення помилок системи Yandex

Типи помилок	Кількість помилок	
1. Лексичні помилки	189	33,1%
1.1 Помилки у перекладі слів	109	18,7%
1.2 Помилки у перекладі словосполучень	80	14,4%

Продовження таблиці 4.32

Типи помилок	Кількість помилок	
2. Граматичні помилки	381	66,9%
2.1 Морфологічні помилки	46	8,1%
2.2 Синтаксичні помилки	335	58,8%
2.2.1 Помилки у побудові словосполучень	170	29,9%
2.2.2 Помилки у побудові речень:	165	28,9%
а) помилки у побудові простих речень	82	14,4%
б) помилки у побудові складних речень	83	14,5%
3. Лексико-граматичні помилки	0	0,0%
4. Орфографічні помилки	0	0,0%
Усього	571	100%

Аналіз помилок машинних перекладів, виконаних системою Yandex, показав превалювання в них граматичних помилок – 66,9%, серед яких переважають синтаксичні помилки у побудові словосполучень та речень (58,8%). Морфологічні помилки менш чисельні і становлять всього 8,1%. Лексичні помилки охоплюють 33,1% від загального обсягу помилок. У машинних перекладах не було виявлено лексико-граматичних та орфографічних помилок.

Серед граматичних помилок переважають помилки у побудові словосполучень (29,9%), які подано у таблиці 4.33.

Таблиця 4.33

#### Помилки у побудові словосполучень системи Yandex

Типи помилок	Кількість помилок	
1 Порушення керування	102	60,0%
2 Порушення узгодження між словами	33	19,4%

3 Неправильний вибір прийменника	14	8,2%
----------------------------------	----	------

Продовження таблиці 4.33

Типи помилок	Кількість помилок	
4 Вживання зайвого прийменника	9	5,3%
5 Відсутність необхідного прийменника	7	4,1%
6 Відхилення від нормативно-узвального порядку слів	5	3,0%
Усього	170	100%

Помилки у побудові словосполучень включають:

1. Порушення керування (60,0%). Оскільки програма не використовує у процесі роботи граматичні правила та не містить словників з вказаним прийменниковим керуванням, вона робить помилки у виборі відмінкової форми іменників, що входять до складу словосполучень. Наприклад:

*Seek to eliminate conflict in their international economic policies* [265] – прагнути до усунення суперечностей у своїй міжнародній економічній політики.

Відповідно до норм української мови прийменник *у* вимагає вживання місцевого відмінка. Програма переклала словосполучення, застосувавши форму родового відмінка, що є помилкою.

*Together with the leading specialized research institutions* [257] – спільно з провідними спеціалізованими науково-дослідних установ.

У вказаному прикладі вжито форму родового відмінка іменника *установ*, проте після прийменника з іменники слід вживати в орудному відмінку.

2. Порушення узгодження між залежним та стрижневим словом:

а) у відмінку: *together with the leading specialized research institutions* [257] – спільно з провідними спеціалізованими науково-дослідних установ;

б) у числі та відмінку: *Northern GOK mines iron ore from two open pit quartzite fields, the Annovskyi and the Pervomayskyi deposits* [238]. – Північний ГЗК видобуток руди двох кар'єр кварцитів родовищ, Annovskyi та Первомайському депозити;

в) у числі: *400 million new phones with forward-facing cameras* [246] – 400 млн нових телефонів з фронтальна камери;

г) у роді: *Ingulets GOK has an annual iron ore concentrate production capacity of 14.8 Mtonnes* [238]. – Річна виробництво залізородного концентрату потужністю 14,8 млн. тонн.

### 3. Неправильний вибір прийменника:

*It currently mines iron ore from its one open-pit quartzite field through a process of drilling and blasting, and by the removal of overburden to external dumps* [238]. – В даний час видобуток руди на один кар'єр кварцитів родовища через процес буріння і вибухових робіт, і вивезенням розкривних порід під зовнішні відвали.

У наведеному прикладі в українському перекладі неправильно обрано еквіваленти до англійських прийменників *from* – на і *to* – під. Правильними варіантами відповідно до контексту речення є прийменник з [218, с. 453], який застосовується у просторовому значенні, та у [219, с. 531], який вживається на позначення напрямку.

### 4. Вживання зайвого прийменника:

*An armed attack against one or more of them in Europe or North America shall be considered an attack against them all* [265]. – збройний напад на одну або кількох із них у Європі або Північній Америці буде розглядатися як напад на них в **все**.

В оригіналі відсутній відповідник українського прийменника **в**, тому вживання зайвого прийменника є помилковим у перекладі.

5. Відсутність необхідного прийменника, бо програма у процесі роботи його прийменник. Наприклад:

*Groundwork is trenching, trench excavation and pitting for isolated supports* [236]. – Фундамент – це риття траншеї, риття траншеї та виразка ізольовані опори.

Незважаючи на те, що в оригіналі є прийменник *for*, програма не переклала його українською мовою, що спричинило помилку.

6. Відхилення від нормативно-узвального порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір граматичної трансформації:

*The growth of applications for non-PC items outstrips traditional software sales* [251]. – Зростання додатків для ПК, але і на елементи випереджає **традиційну з продажу програмного забезпечення.**

Програма необґрунтовано застосувала граматичну трансформацію перестановки, що вплинуло на адекватність перекладу словосполучення. Правильний варіант – *продаж традиційного програмного забезпечення.*

Серед помилок у побудові речень (28,9%) переважають помилки у побудові складних речень (14,5%), подані в таблиці 4.34.

Таблиця 4.34

#### Помилки у побудові складних речень системи Yandex

Типи помилок	Кількість помилок	
1 Порушення синтаксичних зв'язків між членами речення	36	43,4%
2 Порушення координації між підметом і присудком	19	22,9%
3 Граматична неоднотипність однорідних членів речення	16	19,3%
4 Пропущені члени речення	5	6,0%
5 Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	4	4,8%
6 Зайві члени речення	2	2,4%
7 Штучне ускладнення речення	1	1,2%
Усього	83	100%

До помилок у побудові **складних речень** належать:

1. Порухення синтаксичних зв'язків між членами речення:

*Cabinet and wood works mainly consist of transportation and further installation of finished products; there can be both bars and laths used in construction erection and windows, doors and plank and parquet floors* [236]. – **Кабінет деревини і працює, в основному, складаються з транспортування і подальшого монтажу готових виробів; там може бути як бари і планок, що використовуються в будівництві, монтажних і вікна, двері і дошки та паркетні підлоги.**

У цьому реченні порушені синтаксичні зв'язки між членами речення через неправильне визначення частиномовної належності, неадекватний переклад прикметників *cabinet* і *wood*, іменника *works* та необґрунтовану граматичну трансформацію перестановки, що призвело до неправильного порядку слів. Правильним варіантом є термін **столярні й деревообробні роботи**. Ця помилка також призвела до граматичної неоднотипності однорідних підметів **кабінет деревини і працює** та порушення координації між ними і присудком **складаються**.

У машинному перекладі також наявна граматична неоднотипність однорідних підметів **бари, планок, вікна, двері і дошки та паркетні підлоги**, що призвело до порушення координації між ними та присудком **може бути** у числі та відмінку.

Внаслідок неправильного визначення частиномовної належності іменника *erection* та неадекватного перекладу усього словосполучення *construction erection* як **будівництві, монтажних** виникло порушення синтаксичного зв'язку між присудком **використовуються** та обставинами **будівництві, монтажних**.

2. Порухення координації між підметом і присудком (див. попереднє речення).

3. Граматична неоднотипність однорідних членів речення (див. попереднє речення).

4. Пропущені члени речення.

У процесі роботи програма пропустила повнозначні слова, які є в тексті оригіналу, проте відсутні в тексті перекладу. Наприклад:



*The Parties undertake, as set forth in the Charter of the United Nations, to settle any international dispute **in which they may be involved** by peaceful means in such a manner that international peace and security and **justice are not endangered**, and to refrain in their international relations from the threat or use of force **in any manner inconsistent with the purposes of the United Nations** [265]. – Сторони беруть на себе зобов'язання, викладені в статуті Організації Об'єднаних Націй, щоб вирішувати всі міжнародні спори, **які можуть бути залучені**, мирними засобами таким чином, що міжнародний мир і безпеку і **справедливість**, а також утримуватись у міжнародних відносинах від загрози силою або її застосування або чином, несумісним з цілями Об'єднаних Націй.*

5. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення:

*It is the producer of high quality steel **plate** 6-200 mm thick and as wide as 1500-3200 mm **that** is broadly used in shipbuilding, power and special machine building, bridge and offshore deep sea platforms construction, manufacture of large diameter pipes for gas and oil heavy duty pipelines [259]. – Є виробником високоякісної листової **сталі** 6-200 мм товщиною і шириною 1500-3200 мм, **який** широко використовується в галузі суднобудування, енергетичного і спеціального машинобудування, мостів і морських глибоководних платформ, будівництво, виробництво труб великого діаметру для газової і нафтової надпотужних трубопроводів.*

У процесі роботи програма неправильно визначила рід іменника **сталь**, тому в іншій частині речення вживається займенник **який**, характерний для іменника чоловічого роду.

6. Вживання зайвих членів речення, відсутніх у тексті оригіналу.

*Finally, the carbon dioxide is emitted when the natural gas from shale is combusted by its final consumers [240]. – Нарешті, вуглекислий газ вивільняється, коли природний газ із сланців, **а потім** спалюють його кінцевим споживачам.*

У машинному перекладі є зайві члени речення – сполучники *a* і прислівник *потім*, які відсутні в тексті оригіналу. Ця помилка також призвела до штучного ускладнення речення.

7. Штучне ускладнення речення (див. попереднє речення).

Помилки у побудові простих речень, виявлені в машинних перекладах програми Yandex, подано в таблиці 4.35.

Таблиця 4.35

### Помилки у побудові простих речень системи Yandex

Типи помилок	Кількість помилок	
1. Порушення синтаксичних зв'язків між членами речення	38	46,3%
2. Граматична неоднотипність однорідних членів речення	20	24,4%
3. Порушення координації між підметом і присудком	11	13,4%
4. Зайві члени речення	6	7,3%
5. Пропущені члени речення	4	4,9%
6. Неправильний порядок слів	2	2,4%
7. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	1	1,3%
Усього	82	100%

У побудові простих речень програма Yandex зробила такі помилки:

1. Порушення синтаксичних зв'язків між членами речення:

*Wi-Fi complements cellular broadband for “data on the move”* [246]. – *Wi-Fi доповнює бездротового широкопasmового зв'язку для даних “на ходу”*.

У реченні порушений синтаксичний зв'язок між присудком *доповнює* та додатком *зв'язку* у результаті неправильного вибору відмінкової форми іменника – родового відмінка замість знахідного.

2. Граматична неоднотипність однорідних членів речення.

*The Divisions of Corum Group produce coal-cutting, heading, stationary and electrical equipment, mine transport, machinery for open-pit mining, including bucket wheel excavators, apron feeders, conveyors, loaders, spreaders, stackers, etc.* [257]. – **Підрозділів Corum Group видобувати вугілля-різка, Заголовок, стаціонарних та електротехнічного устаткування, рудникового транспорту, машин для відкритих гірничих робіт, в тому числі роторних екскаваторів, пластинчастих живильників, конвеєрів, перевантажувачів, отвалообразователів, перевантажувачів та ін.**

У машинному перекладі простежується граматична неоднотипність однорідних означень *вугілля-різка, Заголовок, стаціонарних та електротехнічного* у зв'язку з їхньою належністю до різних частин мови, що є результатом неправильного морфологічного та синтаксичного аналізу речення.

У реченні порушено синтаксичний зв'язок між присудком *видобувати* та додатками *устаткування, транспорту, машин, екскаваторів, живильників, конвеєрів, перевантажувачів, отвалообразователів, перевантажувачів* через неправильний вибір відмінкових форм іменників. Окрім цього, у реченні порушена координація між підметом *підрозділів* і присудком *видобувати* в особі, числі, відмінку і часі.

3. Порушення координації між підметом і присудком.

*Possibilities for designing the needlepunched carpets are less comparing to the tufted carpets* [269]. – **Можливості для проектування голкопробивні килими менше порівняно з тафтингові килими.**

У наведеному прикладі відсутня координація між підметом *можливості* і присудком *менше*, оскільки Програма неправильно визначила частиномовну належність слова *less* та використала прислівник *менше* замість прикметника *меншими*. Окрім цього, у реченні порушені синтаксичні зв'язки між підметом

*можливості* та додатком *проектування голкопробивні килими*, а також присудком *менше* та додатком *килими* через неправильний вибір відмінкової форми іменника *килими* в обох випадках.

4. Вживання зайвих членів речення.

*For each of the channels, the LCD display shows the information about: voltage, charging/discharging current, charging/discharging time, the quantity of accumulated/discharged electric charge [242].* – Для кожного каналу, РК-дисплей показує інформацію **про систему**: напруга зарядки/розрядки, струм зарядки/розрядки, часу, кількості накопиченого/скидаються електричний заряд.

У наведеному прикладі вживається зайвий додаток *систему*, який відсутній в оригіналі.

5. Пропущені члени речення.

*Ingulets GOK has an annual iron ore concentrate production capacity of 14.8 Mtonnes [238].* – Річна виробництво залізородного концентрату потужністю 14,8 млн. тонн.

У машинному перекладі пропущені головні члени речення, які є в оригіналі, а саме: підмет *Ingulets GOK* та присудок *has*. Окрім цього, у наведеному прикладі порушений порядок слів у результаті необґрунтованого застосування граматичної трансформації перестановки та неадекватного перекладу словосполучення *iron ore concentrate production capacity*. Також простежуємо порушення узгодження між означенням *річна* та додатком *виробництво* у роді.

6. Неправильний порядок слів:

*When constructing solid reinforced concrete constructions, installation of reinforcement and casing works are performed as well (casing and reinforcing cage installation) [236].* – Коли створення міцних залізобетонних конструкцій, установка арматури і опалубки виконуються роботи (опалубки та арматурного каркаса з установкою).

У машинному перекладі є неправильний порядок слів у частині *опалубки виконуються роботи*. Внаслідок неадекватного перекладу означень *reinforcement* і *casing* формами іменників *установка арматури* і *опалубки*, а також неправильного

порядку слів виникло порушення синтаксичних зв'язків між членами речення. Так, в оригіналі підметом є іменник *works*, тоді як в перекладі з'являються два підмети – *установка*, який не координується з присудком *виконуються* у числі, та *роботи*. Окрім цього, дієслово *constructing* отримало неадекватний переклад формою іменника *створення*.

7. Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення:

*Major greenhouse gases emissions in shale gas projects include methane emissions during its extraction and transportation (hereafter leakage emissions), and, indirectly, carbon dioxide emissions from the use of energy and power at shale gas extraction (hereafter indirect emissions) [240].* – Основних парникових газів в проекти видобутку сланцевого газу включають викиди *метану* в процесі *її* видобутку і транспортування (далі-викиди), і, побічно, викиди вуглекислого газу від використання енергії та потужності у видобутку сланцевого газу (далі-непрямі викиди).

У результаті неправильного визначення роду іменника *метану* в іншій частині речення було вжито займенник *її*, який вживається з іменниками жіночого роду.

**Морфологічні помилки** пов'язані з неправильним визначенням омонімічних граматичних форм англійської мови, порушенням словозмінних граматичних категорій української мови (числа, відмінка, часу). Причина виникнення морфологічних помилок у результаті перекладу програми Yandex полягає в тому, що система не використовує граматичні правила для роботи. Морфологічні помилки, які виникли у результаті роботи програми Yandex, подані в таблиці 4.36.

1. Серед морфологічних помилок найбільш чисельними є помилки при знятті граматичної омонімії – неправильне визначення частиномовної належності слів. У більшості випадків у контексті є інформація, необхідна для врахування омонімії. Проте програма не має блоку морфологічного аналізу, а тільки обирає найбільш частотний еквівалент до слова без урахування граматичних характеристик слова та контексту.

*Corum Group massively **upgrades** production plants, expands the range of test facilities for mining equipment, and invests in developing own engineering capacities [257]. – Corum Group масово **модернізації** виробництва, розширює асортимент випробувальних установок для гірничо-шахтного обладнання, та інвестує в розвиток власного інженерного потенціалу.*

У машинному перекладі дієслово **upgrades** було представлене іменником **модернізації**, проте правильною є форма дієслова **модернізує**.

*It currently **mines** iron ore from its one open-pit quartzite field through a process of drilling and blasting, and by the removal of overburden to external dumps [238]. – В даний час **видобуток** руди на один кар'єр кварцитів родовища через процес буріння і вибухових робіт, і вивезенням розкритих порід під зовнішні відвали.*

Програма Yandex переклала дієслово **mines** формою іменника **видобуток**, що призвело до порушення граматичних зв'язків між членами речення, зокрема через відсутність присудка **видобуває**. У машинному перекладі також пропущений підмет **it – вона**.

Таблиця 4.36

### Морфологічні помилки системи Yandex

Типи помилок	Кількість помилок	
1 Помилки при знятті граматичної омонімії	32	69,6%
2 Порушення категорії числа іменника	7	15,2%
3 Ненормативне утворення відмінкової форми іменника	3	6,5%
4 Порушення категорії часу дієслова	3	6,5%
5 Ненормативне утворення ступеня порівняння прикметника	1	2,2%
Усього	46	100%

2. Порухення категорії числа іменників спричинене неправильною роботою алгоритму системи. Наприклад:

*Underground Mining **Division** [257] – підземні гірничі **дивізії**.*

В оригіналі іменник *division* вживається в однині, проте в машинному перекладі він має форму множини.

*Corum Group successfully operates in Ukraine, Russia, Belarus, Kazakhstan, Poland, Vietnam, India, and enters Chinese and South African **markets** [257]. – Corum Group успішно працює в Україні, Росії, Білорусі, Казахстану, Польщі, В'єтнаму, Індії і входить в китайській і Південно-африканський **ринок**.*

У тексті оригіналу іменник *markets* має форму множини, проте в машинному перекладі він вживається у формі однини.

3. У текстах машинних перекладів зафіксовано випадки неправильного утворення відмінкових форм іменників:

*Large **diameter** pipes manufacturing [259] – труб великого **діаметру** виробництва.*

*Large **diameter** pipes [259] – труби великого **діаметру**.*

*Manufacture of large **diameter** pipes [259] – виробництво труб великого **діаметру**.*

Відповідно до норм української мови, іменник *діаметр* належить до твердої групи II відміни, яка при відмінюванні у родовому відмінку мають закінчення *-а*. Тому правильний варіант – *діаметра*.

4. Порухення категорії часу дієслова:

*The worldwide developments forecasted in 2011 **will** definitely **have impact** on or **will** correlate with the developments in Ukraine [246]. – В усьому світі події, прогнозовані в 2011 році, безумовно, **мають вплив** на чи буде корелювати з подіями в Україні.*

У наведеному прикладі в оригіналі дієслово вживається в майбутньому часі, на що вказує використання допоміжного дієслова *will*. Проте у машинному перекладі бачимо форму дієслова в теперішньому часі. Правильний варіант – *матимуть вплив*.

*Laid fibres are first intermingled and entangled by means of needles, then additional needling makes the final structure of the product [269]. – Поклав волокон спочатку змішалися і заплуталися з допомогою голки, потім додаткові проколювання робить будова кінцевого продукту.*

В оригіналі дієслова *are intermingled, and entangled* мають форму пасивного стану теперішнього часу, проте у машинному перекладі вони вживаються в активному стані минулого часу.

5. Ненормативне утворення ступеня порівняння прикметника:

*It utilises the most advanced technologies, including floatation beneficiation of iron ore, which is unique within the Ukrainian industry [238]. – Вона використовує **передові** технології флотаційного збагачення залізної руди, який є унікальним в українській промисловості.*

В оригіналі прикметник *the most advanced* вживається у формі найвищого ступеня, проте в машинному перекладі його український відповідник *передові* має форму звичайного ступеня порівняння, що є помилкою. Правильний варіант – *найбільш передові*.

**Лексичні помилки** займають 33,1% від усіх помилок, виявлених у машинних перекладах системи Yandex. Детально вони представлені в таблиці 4.37.

Таблиця 4.37

### Лексичні помилки програми Yandex

Типи помилок	Кількість помилок	
1. Помилки у перекладі слів	107	56,6%
1.1 Вибір нерелевантних перекладних відповідників	87	46,0%
1.1.1 Неправильне вирішення лексичної полісемії	70	37,0%
1.1.2 Неправильне вирішення лексико-граматичної омонімії	15	7,9%



## Продовження таблиці 4.37

Типи помилок	Кількість помилок	
1.1.3 Вживання русизмів	2	1,1%
1.2 Неперекладені слова	20	10,6%
2. Помилки у перекладі словосполучень	82	43,4%
2.1 Дослівний переклад іншомовних висловів	51	27,0%
2.2 Неадекватний переклад словосполучень як семантичних єдностей	26	13,8%
2.3 Неадекватний переклад фразеологічних сполучень	3	1,6%
2.4 Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті	2	1,1%
Усього	189	100%

Помилки у перекладі слів є найчисельнішими серед лексичних помилок (56,6%). Серед них домінують неадекватно перекладені слова через неправильне вирішення лексичної полісемії (37,0%).

Так, термін *casing* [236] був перекладений як *корпус*, проте правильним варіантом є *облицювання*. У словнику це слово має кілька значень, а саме: 1) обшивка, оббивка, оболонка; облицювання; 2) опалубка, покриття; 3) станина; кожух; корпус; рама; картер; капот; 4) оправа, футляр; покришка; 5) обсадні труби; 6) оклад (ікони) [218, с.179]. Таким чином, система вибрала найбільш частотний варіант перекладу.

Слово *developments* [246] було перекладене як *події*, проте правильним варіантом перекладу є *розробки*. У словнику *development* має п'ятнадцять значень [218, с. 288], серед яких *подія* є п'ятим за рахунком, а *розробка* – сьоме. Таким чином, програма обрала більш частотний, проте неадекватний варіант перекладу.

Програма неадекватно переклала 15 слів через неправильне вирішення граматичної омонімії. Так, дієприслівник *meaning* у реченні *The vast majority of calls will continue to be voice-based meaning video exceeds the needs of most businesses* [246]

був перекладений формою іменника *сенси*. Проте правильним варіантом є конструкція *що означає*.

У реченні *It currently mines iron ore from its one open-pit quartzite field through a process of drilling and blasting, and by the removal of overburden to external dumps* [238] програма неправильно визначила частиномовну належність дієслова *mines* та переклала його іменником *шахти*. Проте правильний варіант – *видобуває*.

У тексті перекладу було знайдено 2 русизми: *whenever* [265] – *всякий раз*, *spreader* [257] – *отвалообразователь*.

Їхня наявність у текстах машинних перекладів пояснюється використанням російської мови як проміжної ланки при перекладі з англійської мови українською. Система також могла не знайти українських відповідників до англійських слів у паралельних корпусах текстів.

У машинних перекладах системи Yandex було виявлено 20 неперекладених слів, серед яких: а) терміни: *earthfills* [236], *negative delta V* [242], *Technology Media and Telecommunications, non-PC* [251]; б) власні назви: *Deloitte* [246; 251], *Annovsky* [238], *Donetskgoramash* [257]; в) скорочення: *mmtpy* [238]; г) аббревіатури: *TMT* [251], *VDC*, *VAC* [242].

Помилки у перекладі словосполучень становлять 43,4% від усієї кількості лексичних помилок. Найчисленнішими серед них є випадки дослівного перекладу іншомовних висловів, які мають власні відповідники в українській мові. Наприклад: *coking coal companies* [238] – *коксівного вугілля компанії* (компанії з видобутку коксівного вугілля); *cell temperature* [242] – *клітинка температура* (температура акумуляторної батареї [222, с. 506]).

Неадекватний переклад словосполучень як семантичних єдностей – вживання словосполучень у не властивому для контексту та тематики значенні через неадекватний вибір еквівалента в українській мові – займає 13,8% від кількості лексичних помилок.

Словосполучення *primary responsibility* було перекладене як *переважна відповідальність* [265], проте правильним є варіант *головна відповідальність* [219, с. 163].

Термін *loss rate* [240] отримав переклад *втрата курсу*, хоча мало б бути *рівень втрат* [228, с. 270].

Помилки у перекладі фразеологічних сполучень становлять 1,6% від усієї кількості лексичних помилок.

*In any manner* [265] система переклала *або чином*, що є порушенням лексичних норм української мови. Правильний варіант – *у будь-який спосіб* [220, с. 521].

Машинний переклад звороту *all measures taken as a result thereof* [265] – *всі прийнятих в результаті нього заходи* є неправильним, оскільки *take measures* перекладається українською як *вжити заходів* [220, с. 918]. Тому правильний варіант перекладу звороту – *про всі заходи, вжиті у зв'язку з ним*.

Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті: *Corum Quality* [257] – *Corum якості*; *Corum Source* [257] – *Корум джерело*.

Лінгвістичний аналіз помилок у текстах машинного перекладу Yandex показав переважання граматичних помилок над лексичними – 66,9% та 33,1% відповідно. Причиною цьому є невикористання програмою граматичних правил у процесі роботи. Серед граматичних помилок домінують помилки у побудові словосполучень (29,9%). Випадки порушення морфологічних норм становлять 8,1% від загальної кількості помилок. Причиною виникнення лексичних помилок є недостатня повнота паралельних корпусів текстів, на основі яких працює програма.

Покращити якість роботи перекладача Yandex можна за допомогою вдосконалення алгоритмів функціонування системи та її подальшого навчання на основі паралельних корпусів текстів.

#### **4.7 Порівняння ефективності роботи систем Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation і Yandex в англо-українському напрямку перекладу**

Результати проведеної оцінки якості машинних перекладів систем Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation і Yandex

показали, що кожна із програм допускає чимало помилок на різних мовних рівнях. З метою визначення найефективнішої системи програмного забезпечення перекладу в англо-українському напрямку було виконано порівняння результатів оцінки якості зазначених систем. Результати проведеного аналізу наведені в Додатку И.

Найбільшу кількість помилок виявлено у текстах, перекладених програмою Babylon – 1343 помилки, серед яких переважають лексичні помилки – 762 одиниці (56,8% від загальної кількості помилок програми). Серед лексичних помилок домінують помилки у перекладі слів – 519 помилок.

Причиною виникнення лексичних помилок може бути відсутність відповідників слів і готових перекладів фраз у використовуваному програмою корпусі даних, а також неправильно вибраний варіант перекладу полісемантичних слів, оскільки комп'ютер не аналізує контекст і тематичне спрямування тексту в процесі роботи. Наявність у перекладі русизмів пояснюється використанням російської мови як проміжної ланки у процесі роботи системи.

Найбільшу кількість граматичних помилок знайдено у перекладах програми Free Translation – 587 одиниць, серед яких переважають синтаксичні помилки у побудові складних і простих речень, які становлять 28,0% від загальної кількості помилок програми. Наявність порушень граматичних норм у перекладених текстах пояснюється тим, що програма у процесі роботи не аналізує граматичні характеристики слів і не використовує граматичні правила.

Усунути виникнення лексичних та граматичних помилок у машинних перекладах можна за допомогою вдосконалення алгоритму роботи програми та подальшого навчання перекладача на більш повному двомовному корпусі текстів.

Лексико-граматичні та орфографічні помилки є найменш численними у машинних перекладах. Найбільшу кількість лексико-граматичних помилок – 42 помилку – виявлено у машинних перекладах програми Pragma 6.3 Business. Причиною їхнього виникнення є неправильна робота алгоритмів лексичного та синтаксичного аналізу, що призводить до вживання словосполучень у не властивому 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.

У машинних перекладах, виконаних програмами Babylon та Free Translation, виявлено по 2 орфографічні помилки. Причиною їхнього виникнення може бути присутність неправильних варіантів написання слів у корпусах текстів, які програми використовують у процесі роботи.

Оцінка результатів роботи систем програмного забезпечення перекладу Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex дозволяють зробити висновок про те, що типовими у всіх машинних перекладах є лексичні та граматичні помилки. Лексико-граматичні є менш численними, проте зустрічаються у машинних перекладах всіх програм. Найменш численними є орфографічні помилки, виявлені у текстах, перекладених програмами Babylon та Free Translation. Найбільшу кількість лексичних помилок знайдено в машинних перекладах текстів, виконаних програмою Babylon (1343). Можемо припустити, що це пов'язано з недостатнім обсягом двомовного корпусу текстів, закладених у програму. Однак у машинних перекладах програми Free Translation є найбільша кількість граматичних помилок, зокрема у побудові словосполучень і речень (477). Причина їхнього виникнення полягає у невикористанні програмою граматичних правил у процесі роботи. Найменше помилок виявлено у машинних перекладах програми Yandex, яка є найбільш ефективною для виконання перекладу з англійської мови українською.

#### **4.8 Порівняння ефективності роботи систем Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation і Yandex в англо-російському напрямку перекладу**

Результати проведеної оцінки якості машинних перекладів систем Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation і Yandex в англо-російському напрямку перекладу наведені в Додатку К.

Найбільшу кількість помилок виявлено у текстах, перекладених програмою Pragma 6.3 Business – 1119 помилок, серед яких переважають лексичні помилки – 601 помилка, що становить 53,7% від загальної кількості помилок програми. Лексичні помилки Pragma 6.3 Business на рівні слів і словосполучень є найбільш чисельними. Причинами виникнення цих помилок є неповнота автоматичного словника, його нездатність правильно вирішувати лексичну полісемію. Програма не має модуля транслітерації, тому вона залишила без перекладу власні назви. Незважаючи на тематичну розмітку словника, Pragma 6.3 Business неадекватно перекладає багатозначні слова. Недоліки роботи модуля морфологічного аналізу призвели до неправильного визначення частиномовної належності слів та їхнього неадекватного перекладу. Внаслідок відсутності у словнику готових відповідників словосполучень, система переклала їх дослівно. Вдосконалити роботу програми на лексичному рівні можна шляхом поповнення словника, що може зробити користувач, та шляхом внесення змін до алгоритмів роботи системи, що може здійснити лише розробник.

Найбільшу кількість граматичних помилок знайдено у перекладах програми Free Translation – 651 помилка, серед яких переважають синтаксичні помилки у побудові складних і простих речень, що становить 49,4% від загальної кількості помилок програми. Оскільки у роботі система не виконує синтаксичний аналіз речень та не використовує граматичні правила у процесі синтезу, попередити виникнення граматичних помилок можна лише за допомогою вдосконалення алгоритму роботи програми та подальшого навчання перекладача на нових двомовних корпусах текстів.

Лексико-граматичні та орфографічні помилки є найменш чисельними у машинних перекладах в англо-російському напрямку перекладу. Найбільшу кількість лексико-граматичних помилок – 42 помилки – виявлено у машинних перекладах програми Pragma 6.3 Business. Причиною їхнього виникнення є недоліки у роботі алгоритмів лексичного та синтаксичного аналізу, що призводять до неправильного вибору варіантів перекладу словосполучень та застосування необґрунтованих граматичних трансформацій.

У машинних перекладах, виконаних програмами Google Translate та Yandex, виявлено 2 та 1 орфографічну помилку відповідно. Причиною їхнього виникнення може бути наявність неправильних варіантів написання слів у корпусах текстів, які програми використовують у процесі роботи.

Результати роботи систем програмного забезпечення перекладу в англо-російському напрямку перекладу дозволяють зробити висновок про те, що типовими для усіх програм є лексичні та граматичні помилки. Проте їхня кількість варіюється від програми до програми. Лексико-граматичні є менш чисельними, проте трапляються у машинних перекладах всіх програм. Рідкісними є орфографічні помилки, виявлені у машинних перекладах лише двох програм із шести – Google та Yandex. Найбільшу кількість лексичних помилок знайдено в машинних перекладах текстів, виконаних програмою Pragma 6.3 Business (601 одиниця). Причиною їх виникнення є неповнота автоматичного англо-російського словника програми. У машинних перекладах системи Free Translation є найбільша кількість граматичних помилок, зокрема синтаксичних помилок, 651 одиниця. Причина їхнього виникнення полягає у невикористанні програмою граматичних правил у процесі роботи. Найменше помилок виявлено у машинних перекладах програми Yandex. Саме тому вона є найбільш ефективною для виконання перекладу з англійської мови російською.

#### **Висновки до розділу 4**

Результати роботи систем програмного забезпечення перекладу Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation та Yandex в англо-українському та англо-російському напрямках перекладу дозволяють зробити висновок про те, що типовими для всіх машинних перекладів є лексичні, граматичні та лексико-граматичні помилки. Орфографічні помилки трапляються 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%).

У машинних перекладах системи Pragma 6.3 Business переважають лексичні помилки (53,4% та 53,7% в англо-українському та англо-російському напрямках відповідно). Причиною їхнього виникнення є нездатність програми аналізувати контекст і тематичне спрямування тексту. Неповнота автоматичного словника призводить до появи неперекладених слів та неадекватно перекладених слів через невирішення полісемії, лексичної та лексико-граматичної омонімії, а також дослівного перекладу іншомовних висловів, які мають свої відповідники в українській та російській мовах. Незважаючи на тематичну розмітку словника, у машинних перекладах текстів різного змістового спрямування переважають помилки при вирішенні лексичної полісемії. Багатозначні слова отримують переклади, які не відповідають тематиці та контексту речень.

У машинних перекладах програми Babylon теж домінують лексичні помилки (56,8%). Причиною їхнього виникнення може бути недостатня повнота лінгвістичного забезпечення програми – паралельних корпусів текстів, на основі яких формується модель мови перекладу та обираються перекладні відповідники.

Серед помилок у перекладі слів трапляються синкретичні помилки. Так, неадекватне вирішення лексико-граматичної омонімії призводить до вибору нерелевантного перекладного відповідника.

Лексико-граматичні помилки теж мають синкретичний характер, адже порушують не лише лексичні, а й синтаксичні норми літературної мови. Вони полягають у вживанні словосполучень у не властивому для контексту та тематики значенні, а також відхиленні від нормативно-узусуального порядку слів, порушенні логіко-синтаксичного зв'язку між словами через неправильний вибір варіантів перекладу та граматичної трансформації. Проте вони не були виявлені у машинних перекладах програми Yandex.

У текстах, перекладених системами статистичного машинного перекладу (Google Translate, Free Translation, Yandex) та системою гібридного машинного перекладу (Bing Translator), переважають граматичні помилки, зокрема синтаксичні помилки (Google Translate – 50,0% та 49,2%, Bing Translator – 48,9% та 51,1%, Free



Translation – 52,0% та 49,0%, Yandex – 58,8% та 58,8% в англо-українському та англо-російському перекладах відповідно).

Причиною їхнього виникнення є невикористання програмами граматичних правил у процесі роботи. Помилки у побудові речень маю синкретичний характер. Неправильно визначена частиномовна належність слів (порушення морфологічних норм) призводить до порушення синтаксичних зв'язків між членами речення.

У машинних перекладах всіх програм в англо-українському напрямку перекладу серед помилок у побудові простих та складних речень переважають порушення синтаксичних зв'язків між членами речення та граматична неоднотипність однорідних членів речення.

У досліджуваних програмах спостерігаємо відмінності у рівні моделювання української та російської мов. Це проявляється у різній кількості помилок у машинних перекладах. Так, програми Pragma 6.3 Business, Google Translate, Free Translation та Yandex зробили меншу кількість помилок при перекладі з англійської мови українською. В англо-російських машинних перекладах програм Bing Translator та Babylon було знайдено меншу кількість помилок, ніж у відповідних англо-українських.

Найбільш ефективною системою для роботи з англо-українською та англо-російською мовними парами є програма Yandex, у машинних перекладах якої виявлено найменшу кількість помилок. Проте вона потребує подальшого вдосконалення. Адже у перекладі з англійської мови українською та з англійської мови російською переважають граматичні помилки (66,9% та 65,8 % відповідно). Оскільки програма Yandex не використовує у процесі роботи граматичні правила, можливий спосіб покращення якості її роботи вбачаємо у подальшому навчанні на двомовних корпусах текстів.

Основні положення розділу відображено у відповідних публікаціях автора [100; 101].

## 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. І. Бабіна, М. С. Блехман, В. П. Бурдаєв, Е. І. Корольов, О. Л. Шаміс, Ю. І. Шемакін);

2) системи програмного забезпечення перекладу охоплюють лінгвістичне забезпечення та алгоритми обробки мов (Ю. М. Марчук, Л. Л. Нелюбін, Р. Г. 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. В. Бірюков, М. С. Блехман, Ю. М. Марчук, П. М. Хроменков) та типології мовних помилок (Т. Г. Бондаренко, М. Г. 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.

Найбільшу кількість помилок було виявлено в текстах 10 тематик, до яких належать: будівництво (Civil and Erection Works), металургійна (Metallurgical Division), право (The North Atlantic Treaty), електроніка (BC-700 Microprocessor charger), машинобудування (Heavy Engineering), зв'язок (Deloitte analysis of top telecommunications trends for 2011), комерція (More than half of all computers aren't traditional computers anymore), текстиль (What Is the Difference between Tufted and Needlepunched Carpets?), гірничча (Mining Division), екологія (Shale gas extraction and climate change).

Орфографічні помилки становлять менше 1% і трапляються 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.

В англо-українському напрямку перекладу найбільшу кількість лексичних помилок знайдено в машинних перекладах текстів, виконаних програмою Babylon (1345).

Однак у машинних перекладах програми Free Translation простежується найбільша кількість граматичних помилок, зокрема у побудові словосполучень і речень (477). Причинами їх виникнення є недостатній обсяг двомовних корпусів текстів, закладених у програму, та невикористання програмами граматичних правил у процесі роботи.

Найменше помилок виявлено у машинних перекладах програми Yandex, яка є найбільш ефективною для виконання перекладу з англійської мови українською.

У текстах, перекладених системами статистичного машинного перекладу, переважають граматичні помилки: Google Translate – 59,3% та 57,5%, Free Translation – 64,0% та 61,8%, 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 трапляються синкретичні помилки, природа яких полягає в порушенні однією ненормативною одиницею різних мовних норм.

Неперекладені слова та дослівно перекладені словосполучення мають механічний характер і для їх усунення розробники не потребують ні творчих рішень, ні зміни алгоритмів роботи програмного забезпечення. Насамперед йдеться про якісне поповнення автоматичного словника, у результаті чого чого можна запобігти виникненню помилок на лексичному рівні.

Для оптимізації роботи систем програмного забезпечення перекладу слід збагачувати їхнє лінгвістичне забезпечення, удосконалювати алгоритми роботи і продовжувати їх навчання на основі багатомовних корпусів текстів.

Результати дослідження можуть бути використані розробниками систем машинного перекладу і користувачами для підвищення загальної якості машинного перекладу.

Перспективним напрямом подальших досліджень оцінки якості систем програмного забезпечення перекладу може бути, на наш погляд, зіставлення результатів роботи 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 ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие / [Большакова Е. И., Клышинский Э. С., Ландэ Д. В., Носков А. А. и др.]. — М. : МИЭМ, 2011. — 272 с.
2. Анастасова Е. А. Особенности герменевтики фразеологизмов современными электронными переводчиками / Е. А. Анастасова // Информационные управляющие системы и компьютерный мониторинг — 2011 : [сборник материалов I Всеукраинской научно-технической конференции студентов, аспирантов и молодых ученых “Информационные управляющие системы и компьютерный мониторинг” (ИУС и КМ — 2011)]. — Донецк : ДонНТУ, 2011. — Том II. — С. 185—188.
3. Антрушина Г. Б. Лексикология английского языка / Г. Б. Антрушина, О. В. Афанасьева, Н. Н. Морозова. — М. : Дрофа, 1999. — 288 с.
4. Бабина О. И. Роль личности переводчика в машинном переводе / Ольга Ивановна Бабина // Языковая личность переводчика: коллектив. моногр. / [отв. ред. Л. А. Нефедова]. — Челябинск : Изд-во Челяб. гос. ун-та, 2011. — 420 с. — С. 77—90.
5. Баранов А. Н. Введение в прикладную лингвистику : [учебное пособие] / Анатолий Николаевич Баранов. — М. : Едиториал УРСС, 2003. — 360 с.
6. Бархударов Л. С. Язык и перевод: вопросы общей и частной теории перевода / Леонид Степанович Бархударов. — М. : Междунар. отношения, 1975. — 240 с.
7. Беляева Л. Н. Информационное пространство филолога и принципы его организации [Электронный ресурс] / Лариса Николаевна Беляева // Известия РГПУ им. А. И. Герцена. — 2004. — № 9. — С. 17—31. — Режим доступа : <http://cyberleninka.ru/article/n/informatsionnoe-prostranstvo-filologa-i-printsipy-ego-organizatsii>.
8. Беляева Л. Н. Лингвистические автоматы в современных гуманитарных технологиях / Лариса Николаевна Беляева. — Спб. : ООО “Книжный Дом”, 2007. — 192 с.

9. Бірюков А. В. Оцінка якості машинного перекладу : автореф. дис. на здобуття наук. ступеня канд. філол. наук : 10.02.16 “Перекладознавство” / А. В. Бірюков. — К., 2008. — 20 с.
10. Бірюков А. В. Система оцінки якості автоматизованого перекладу [Електронний ресурс] / Артем Валентинович Бірюков // Вісник Сумського державного університету. — Серія Філологічні науки. — 2004. — №3 (62). — С. 42—48. — Режим доступу : <http://essuir.sumdu.edu.ua/handle/123456789/11023>.
11. Блехман М. С. Комп’ютерна лінгвістика / Михайло Самойлович Блехман. — Х. : ХГУ, 1997. — 153 с.
12. Бодик О. П. Сучасна українська літературна мова. Лексикологія. Фразеологія. Лексикографія : [навч. посіб] / О. П. Бодик, Т. М. Рудакова. — К. : Центр учбової літератури, 2011. — 416 с.
13. Бондаренко Т. Г. Типологія мовних помилок та їх усунення під час редагування журналістських матеріалів : автореф. дис. на здобуття наук. ступеня канд. філол. наук : 10.01.08 “Журналістика” / Т. Г. Бондаренко. — Київ : Б.в., 2003. — 20 с.
14. Бурдаєв В. П. Системи штучного інтелекту : конспект лекцій / Володимир Петрович Бурдаєв. — Х. : ХНЕУ, 2008. — 156 с.
15. Ванников Ю. В. Проблемы адекватности перевода. Типы адекватности, виды перевода и переводческая деятельность / Юрий Вениаминович Ванников // Текст и перевод. — М. : Наука, 1988. — С. 34—39.
16. Виноградов В. С. Введение в переводоведение (общие и лексические вопросы) / Венедикт Степанович Виноградов. — М. : Издательство института общего среднего образования РАО, 2001. — 224 с.
17. Вихованець І. Р. Теоретична морфологія української мови : Академічна граматики укр. мови / І. Р. Вихованець, К. Г. Городенська / [за ред. І. Вихованця]. — К. : Унів. вид-во “Пульсари”, 2004. — 400 с.
18. Вопросы моделирования языка и машинного перевода : [сборник научных трудов] / [под ред. Х. А. Радзикулова]. — Самарканд : Изд-во СамГУ, 1984. — 99 с.

19. Вусик Г. Л. Взаємозв'язок мови та мислення / Ганна Леонідівна Вусик // Філософська антропологія : культура, раціональність, комунікація : [матеріали Всеукр. наук.-практ. конф., (Донецьк, 17 квіт. 2009 р.)]. — Донецьк : ДОНІЗГ, 2009. — С. 89—95.
20. Вусик Г. Л. Основные жанры компьютерного общения / Ганна Леонидовна Вусик // Гуманитарий : Актуальные проблемы науки и образования. — Саранск : Мордовский государственный университет им. Н.П. Огарева, 2012. — №4. — С. 107—111.
21. Гайна Г. А. Системи штучного інтелекту / Георгій Анатолійович Гайна. — К. : КНУБА, 2007. — 207 с.
22. Гарбовский Н. К. Теория перевода : [учебник] / Николай Константинович Гарбовский. — М. : Изд-во Моск. ун-та, 2004. — 544 с.
23. Гладченко О. В. Штучний інтелект. Експертні та навчальні системи : [навч. посіб.] / Оксана Вікторівна Гладченко. — Ірпінь, 2008. — 84 с.
24. Глибовець М. М. Штучний інтелект : [підручник для студентів вищих навчальних закладів] / М. М. Глибовець, О. В. Олецкий. — К. : Видавничий дім “КМ Академія”, 2002. — 366 с.
25. Городецкий Б. Ю. Актуальные проблемы прикладной лингвистики / Борис Юрьевич Городецкий // Новое в зарубежной лингвистике. — М. : Радуга, 1983. — Вып. 12. — С. 5—22.
26. Горпинич В. О. Морфологія української мови / Володимир Олександрович Горпинич. — К. : ВЦ “Академія”, 2004. — 336 с.
27. Гуменюк О. О. Сучасний стан онлайн-перекладачів (німецька та українська мови) / Олександр Олександрович Гуменюк // Наукові записки [Ніжинського державного університету ім. Миколи Гоголя]. Сер.: Філологічні науки. — 2013. — Кн. 3. — С. 145—150.
28. Гусев В. В. Эмпатическая модель в формировании стратегии перевода / Владимир Викторович Гусев // Вестник МГЛУ. — 2002. — Вып. 480 : Перевод как когнитивная деятельность. — С. 26—41.



29. Дарчук Н. П. Комп'ютерна лінгвістика (автоматичне опрацювання тексту) : підручник / Наталія Петрівна Дарчук. — К. : Видавничо-поліграфічний центр “Київський університет”, 2008. — 351 с.
30. Демьянков В. З. Интерпретация, понимание и лингвистические аспекты их моделирования на ЭВМ / Валерий Закиевич Демьянков. — М. : Изд-во Моск. ун-та, 1989. — 172 с.
31. Доля В. Г. Комп'ютерні системи штучного інтелекту : навчальний посібник для студентів вищих навчальних закладів / Віктор Гнатович Доля. — К : Університет “Україна”, 2011. — 296с.
32. Дрейфус Х. Чего не могут вычислительные машины. Критика искусственного разума / Хьюберт Дрейфус. — М. : Прогресс, 1978. — 333 с.
33. Журавльова О. Переклад як складова процесу інтерпретації тексту / О. Журавльова // Науковий вісник Херсонського державного університету. Сер.: “Лінгвістика”. — Випуск X. — Херсон : Видавництво ХДУ, 2009. — 434 с. — С. 244—249.
34. Загнітко А. П. Теоретична граматики української мови : Синтаксис : [монографія] / Анатолій Панасович Загнітко. — Донецьк : ДонНУ, 2001. — 662 с.
35. Заде Л. А. Понятие лингвистической переменной и его применение к принятию приближенных решений / Лотфи Аскер Заде. — М. : Мир, 1976. — 165 с.
36. Зайцева Н. Ю. Ранние годы машинного перевода : воспоминания и биографии основоположников / Наталья Юрьевна Зайцева // Вестник Российской Академии Наук. — 2002. — Том 72, № 5. — С. 453—456.
37. Зайченко Ю. П. Основи проектування інтелектуальних систем / Юрій Петрович Зайченко. — К. : Слово, 2004. — 352 с.
38. Зубков М. Г. Українська мова / Микола Григорович Зубков. — Х. : ВД “ШКОЛА”, 2006. — 496 с.
39. Зубов А. В. Основы искусственного интеллекта для лингвистов : [учеб. пособие] / А. В. Зубов, И. И. Зубова. — М. : Университетская книга; Логос, 2007. — 320 с.

40. Иомдин Л. Л. Уроки машинного перевода для детей и взрослых / Леонид Лейбович Иомдин // Лингвистика для всех. Зимняя лингвистическая школа—2004. — Москва : НИИРО, 2004. — С. 56—68.
41. Исламов Р. С. Анализ современных систем машинного перевода типа SMT и RBMT / Р. С. Исламов, А. Г. Фомин // Филологические науки. Вопросы теории и практики. — Тамбов : Грамота, 2013. — № 3 (21) : в 2-х ч. Ч. I. — С. 69—73.
42. Інформаційні технології в перекладі : [навч. посібник] / [за заг. ред. Т. І. Коваль]. — К. : Видавничий центр КНЛУ, 2010. — 261 с.
43. Історія машинного перекладу: стислий огляд : [навч. посіб.] / [уклад. П. О. Бочан]; Чернів. нац. ун-т ім. Ю.Федьковича. — Чернівці : Рута, 2010. — 47 с.
44. Кавун С. В. Системи штучного інтелекту : [навч. посіб.] / С. В. Кавун, В. М. Коротченко. — Харків : Вид. ХНЕУ, 2007. — 320 с.
45. Карабан В. І. Переклад англійської наукової і технічної літератури. / В'ячеслав Іванович Карабан. — Вінниця : Нова Книга, 2001. — 271 с.
46. Карпіловська Є. А. Вступ до комп'ютерної лінгвістики / Євгенія Анатоліївна Карпіловська. — Донецьк : ТОВ "Юго-Восток, Лтд", 2003. — 184 с.
47. Карпіловська Є. А. Вступ до прикладної лінгвістики : комп'ютерна лінгвістика : [підручник] / Євгенія Анатоліївна Карпіловська. — Донецьк : ТОВ "Юго-Восток, Лтд", 2006. — 188 с.
48. Касевич В. Б. Элементы общей лингвистики / Вадим Борисович Касевич. — М. : Наука, 1977. — 183 с.
49. Кибрик А. Е. Очерки по общим и прикладным вопросам языкознания : Универсальное в языке / Александр Евгеньевич Кибрик. — М. : УРСС, 2002. — 336 с.
50. Ким Л. Г. Модель интерпретационного процесса и факторы, детерминирующие вариативность интерпретационного результата / Лидия Густовна Ким // Вестник Томского государственного университета. — 2009. — № 318. — С. 29—36.

51. Кісіль М. В. До проблеми машинного перекладу / Микола Васильович Кісіль // Актуальні проблеми соціальної комунікації : [збірник матеріалів конференції]. — Івано-Франківськ : В-во ІФНГУНГ, 2011. — С. 97—103.
52. Комиссаров В. Н. Лингвистика перевода / Вилен Наумович Комиссаров. — М. : Международные отношения, 1980. — 166 с.
53. Комиссаров В. Н. Общая теория перевода / Вилен Наумович Комиссаров. — М. : ЧеРо, 1999. — 136 с.
54. Комиссаров В. Н. Современное переводоведение : [учеб. пособие] / Вилен Наумович Комиссаров. — М. : ЭТС. — 2001. — 424 с.
55. Комиссаров В. Н. Теория перевода (лингвистические аспекты) : [учеб. для ин-тов и фак. иностр. яз.] / Вилен Наумович Комиссаров. — М. : Высш. шк., 1990. — 253 с.
56. Компанцева Л. Ф. Информационно-коммуникативные технологии : когнитивно-прагматический и социально-философский аспекты / Лариса Феликсовна Компанцева // Психолінгвістика : [зб. наук. пр.] / [гол. ред. О.О. Селіванова]. — Переяслав-Хмельницький : ПП “СКД”, 2009. — Вип. 3. — С. 125—130.
57. Компанцева Л. Ф. От классического языкознания — к Интернет-лингвистике / Лариса Феликсовна Компанцева // Studia Linguistica : [зб. наук. пр.] / [відп. ред. І.О. Голубовська]. — К. : Видавничо-поліграфічний центр “Київський університет”, 2010. — Вип. 4. — С. 24—30.
58. Комп’ютерний перекладач Pragma 6.x [Електронний ресурс]. — Режим доступу : <http://translate.ua/uk/pragma-6x?showall=1>.
59. Королев Э. И. Промышленные системы машинного перевода / Э. И. Королев, Л. Л. Нелюбин — М. : ВЦП, 1991. — 105 с.
60. Кочерган М. П. Загальне мовознавство / Михайло Петрович Кочерган. — К. : ВЦ “Академія”, 2008. — 464 с.
61. Куниловская М. А. Классификация переводческих ошибок и их электронная разметка в brat / Мария Анатольевна Куниловская // Проблемы теории, практики и методики перевода. — НГЛУ, 2013. — Выпуск 16, том 1. — С. 59—71.

62. Кузьменко Б. В. Системи штучного інтелекту : [навч. посіб.] / Б. В. Кузьменко, О. А. Чайковська. — К. : Альтерпрес, 2006. — 140 с.
63. Латышев Л. К. Курс перевода : Эквивалентность перевода и способы ее достижения / Лев Константинович Латышев. — М. : Международные отношения, 1981. — 248 с.
64. Левитан К. М. Юридический перевод : основы теории и практики / Константин Михайлович Левитан. — М. : Проспект, 2011. — 352 с.
65. Леонтьева Н. Н. Автоматическое понимание текстов: системы, модели, ресурсы : [учеб. пособие для студ. лингв. фак. вузов] / Нина Николаевна Леонтьева. — М. : Издательский центр “Академия”, 2006. — 304 с.
66. Люгер Дж.Ф. Искусственный интеллект : стратегии и методы решения сложных проблем / Джордж Ф. Люгер ; пер. с англ. — М. : Издательский дом “Вильямс”, 2003. — 864 с.
67. Мамедова М. Г. Машинный перевод : эволюция и основные аспекты моделирования / М. Г. Мамедова, З. Ю. Мамедова. — Баку : Издательство “INFORMASIYA TEXNOLOGIYALARI”, 2005. — 156 с.
68. Марчук Ю. Н. Компьютерная лингвистика / Юрий Николаевич Марчук. — М. : АСТ, Восток-Запад, 2007. — 320 с.
69. Марчук Ю.Н. Методы моделирования перевода / Юрий Николаевич Марчук. — М. : Наука, 1985. — 202 с.
70. Марчук Ю.Н. Модели перевода : [учеб. пособие для студ. учреждений высш. проф. образования] / Юрий Николаевич Марчук. — М. : Издательский центр “Академия”, 2010. — 176 с.
71. Марчук Ю.Н. Проблемы машинного перевода / Юрий Николаевич Марчук. — М. : Наука, 1983. — 232 с.
72. Масленникова О. Е. Основы искусственного интеллекта : [учеб. пособие] / О. Е. Масленникова, И. В. Гаврилова. — М. : ФЛИНТА, 2013. — 282 с.
73. Мельничайко В. Я. Українська мова : Письмовий екзамен. Як уникнути помилок / Володимир Ярославович Мельничайко. — Тернопіль : Навчальна книга — Богдан, 1998. — 64 с.

74. Міщенко А. Л. Лінгвістика фахових мов та сучасна модель науково-технічного перекладу : [монографія] / Алла Леонідівна Міщенко. — Вінниця : Нова Книга, 2013. — 448 с.
75. Міщенко А. Л. Машинний переклад у контексті сучасного науково-технічного перекладу / Алла Леонідівна Міщенко // Вісник ХНУ імені В. Н. Каразіна. Філологія. — 2013. — № 73. — С. 172—180.
76. Наер В. Л. Понимание и интерпретация (к основам интерпретации текста как аналитической деятельности) / Владимир Леонидович Наер // Сборник научных трудов МГЛУ. Проблемы современной стилистики. — М., 2001. — Вып. 459. — С. 3—13.
77. Научно-технический перевод / [Ванников Ю. В., Кудряшова Л. М., Марчук Ю. Н. и др.]. — М. : Наука, 1987. — 139 с.
78. Нелюбин Л. Л. Компьютерная лингвистика и машинный перевод : методическое пособие / Лев Львович Нелюбин. — М. : ВЦП, 1991. — 151 с.
79. Нелюбин Л. Л. Наука о переводе (история и теория с древнейших времен до наших дней) / Л. Л. Нелюбин, Г. Т. Хухуни. — М. : Флинта : МПСИ, 2006. — 416 с.
80. Нелюбин Л. Л. Перевод и прикладная лингвистика / Лев Львович Нелюбин. — М. : Высшая школа, 1983. — 207 с.
81. Новиков В. А. Трансфер в современных системах машинного перевода : дис. ... канд. филол. наук: 10.12.21 / Виктор Алексеевич Новиков. — М., 2000. — 194 с.
82. Новое в зарубежной лингвистике : Вып. XXIV. Компьютерная лингвистика : [Сост., ред. и вступ. ст. Б. Ю. Городецкого]. — М. : Прогресс, 1989. — 432 с.
83. Олійник Т. Прагматика перекладу : адекватність перекладу як критерій оцінки його якості / Т. Олійник // Сучасні проблеми лінгвістичних досліджень і методики викладання іноземних мов професійного спілкування у вищій школі : [зб. наук. праць] / [за ред. В. Т. Сулима, С. Н. Денисенко]. — Ч. 1. Лінгвостилістика. Лексична семантика. Фразеологія. — Львів : Видавничий центр ЛНУ імені Івана Франка, 2007. — С. 271—273.

84. О переводчике Bing [Електронний ресурс]. — Режим доступу : [www.bing.com/translator/help](http://www.bing.com/translator/help).
85. О сервисе Яндекс. Перевод [Електронний ресурс]. — Режим доступу : <https://translate.yandex.ua/help>.
86. Партико З. В. Загальне редагування / Зіновій Васильович Партико. — Львів : Афіша, 2006. — 416 с.
87. Партико З. В. Прикладна і комп'ютерна лінгвістика : Вступ до спеціальності : [навчальний посібник] / Зіновій Васильович Партико. — Львів : Афіша, 2008. — 224 с.
88. Пиотровский Р. Г. Инженерная лингвистика и проблемы искусственного интеллекта / Раймонд Генрихович Пиотровский // Лингвистические проблемы искусственного интеллекта. — М. : ИНИОН АН СССР, 1980. — 250 с. — С. 157—189.
89. Пиотровский Р. Г. Инженерная лингвистика и теория языка / Раймонд Генрихович Пиотровский. — Л. : Наука, Ленингр. отд-ние, 1979. — 112 с.
90. Пиотровский Р. Г. Лингвистический автомат (в исследовании и непрерывном обучении) / Раймонд Генрихович Пиотровский. — СПб : Изд-во РГПУ им. А. И. Герцена, 2008. — 256 с.
91. Пономарів О. Д. Стилїстика сучасної української мови : [підручник] / Олександр Данилович Пономарів. — К. : Либідь, 1993. — 248 с.
92. Поспелов Д. А. Из истории искусственного интеллекта : история искусственного интеллекта до середины 80-х годов / Дмитрий Александрович Поспелов // Новости искусственного интеллекта. — 1994. — №4. — С. 70—90.
93. Рождественский Ю. В. Введение в прикладную филологию. Языковая семиотика / Рождественский Ю. В., Волков А. А., Марчук Ю. Н. — М. : Изд-во Моск. гос. ун-та, 1987. — 116 с.
94. Саватєєва Л. В. Wordbuilding in Modern English / Саватєєва Л. В., Триш О. М., Кондратенко М. В. // Економічна стратегія і перспективи розвитку сфери торгівлі та послуг. — 2012. — Вип. 1(2). — С. 472—479.

95. Селіванова О. О. Сучасна лінгвістика : напрями та проблеми : [підручник] / Олена Олександрівна Селіванова. — Полтава : Довкілля-К, 2008. — 712 с.
96. Семенов А. Л. Современные информационные технологии и перевод / Аркадий Львович Семенов. — М. : Издательский центр “Академия”, 2008. — 224 с.
97. Соснина Е. П. О разработке и использовании российского учебного корпуса переводов / Екатерина Петровна Соснина // Труды международной конференции “Корпусная лингвистика — 2006”. — Спб : СпбГУ, 2006. — С. 365—373.
98. Спирін О. М. Початки штучного інтелекту / Олег Михайлович Спирін. — Житомир : ЖДПУ, 2001. — 91 с.
99. 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.
100. 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.
101. 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.
102. 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.
103. 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.

104. Стахмич Ю. С. Системы программного обеспечения перевода: когнитивно-интерпретационный аспект / Ю. С. Стахмич // Филология, искусствоведение и культурология в XXI веке : [материалы междунар. заочной науч.-практ. конф., (Новосибирск, 18 марта 2013 г.)]. — Новосибирск : Изд. “СибАК”, 2013. — С. 55—63.
105. Стахмич Ю. С. Системы 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.
106. Сухова А. В. Міжгалузева та внутрішньогалузева омонімія науково-технічних термінів та труднощі їх перекладу / А. В. Сухова, І. С. Степанішина // Проблеми та перспективи формування національної гуманітарно-технічної еліти : [зб. наук. праць] / [за ред. Л. Л. ТОВАЖНЯНСЬКОГО, О. Г. РОМАНОВСЬКОГО]. — Вип. 29 (33). — Харків : НТУ “ХП”, 2011. — С. 232—238.
107. Сучасна українська літературна мова : [підручник] / [за ред. А. П. Грищенка]. — К. : Вища школа, 2002. — 439 с.
108. Сучасна українська літературна мова : [підручник] / [за ред. М.Я. Плющ]. — К. : Вища школа, 2003. — 430 с.
109. Сучасна українська літературна мова : лексика і фразеологія / [за заг. ред. І. К. Білодіда]. — К. : Наук. думка, 1973. — 439 с.
110. Сучасна українська літературна мова : морфологія / [за заг. ред. І. К. Білодіда]. — К. : Наук. думка, 1969. — 583 с.
111. Сучасна українська літературна мова : синтаксис / [за заг. ред. І. К. Білодіда]. — К. : Наук. думка, 1972. — 516 с.
112. Сучасна українська мова : [підручник] / [Пономарів О. Д., Різун В. В., Шевченко Л. Ю. та ін.] / [за ред. О. Д. Пономарева]. — К. : Либідь, 2005. — 488 с.
113. Тетерлева Е. В. Типы переводческих ошибок / Е. В. Тетерлева, Ю. К. Попова // Проблемы романо-германской филологии, философии,



- педагогике и методике преподавания иностранных языков : [сб. науч. тр] / [науч. ред. Т. Н. Романова]. — Пермь : ПГПУ, 2009. — С. 64—71.
114. Тихомирова Е. На какой язык лучше всего переводит Google Translate? [Электронный ресурс] / Елена Тихомирова. — Режим доступа : <http://mozgorilla.com/analytics/na-kakoj-yazyk-luchshe-vsego-perevodit-google-translate/>.
115. Федяева Н. Д. Нормы в пространстве языка : [монография] / Наталья Дмитриевна Федяева. — М. : ФЛИНТА, 2011. — 172 с.
116. Финн В. К. Философские проблемы логики интеллектуальных систем [Электронный ресурс] / Виктор Константинович Финн // Новости искусственного интеллекта №1, 1999. — Режим доступа : <http://www.raai.org/library/library.shtml?ainews>.
117. Хроменков П. Н. Анализ и оценка эффективности современных систем машинного перевода: дис... канд. филол. наук : 10.12.21 / Павел Николаевич Хроменков. — М., 2000. — 265 с.
118. Цвиллинг М. Я. Когнитивные модели и перевод (к постановке проблемы) / Михаил Яковлевич Цвиллинг // Вестник МГЛУ. — 2002. — Вып. 480. — С. 21—26.
119. Шаляпина З. М. Автоматический перевод : эволюция и современные тенденции / Зоя Михайловна Шаляпина // Вопросы языкознания. — 1996. — № 2. — С. 105—117.
120. Шамис А. Л. Пути моделирования мышления : Активные синергетические нейронные сети, мышление и творчество, формальные модели поведения и “распознавания с пониманием” / Александр Львович Шамис . — М. : КомКнига, 2006. — 336 с.
121. Швейцер А. Д. Теория перевода : Статус, проблемы, аспекты / Александр Давидович Швейцер. — М. : Наука, 1988. — 216 с.
122. Шевченко А. И. Актуальные проблемы теории искусственного интеллекта / Анатолий Иванович Шевченко. — Київ : ІПШ “Наука і освіта”, 2003. — 228 с.

123. Шемакин Ю. И. Начала компьютерной лингвистики : [учеб. пособие] / Юрий Иванович Шемакин. — М. : Изд-во МГОУ, А/О “Росвузнаука”, 1992. — 113 с.
124. Шишков Д. П. Почему компьютеры в принципе не могут обрабатывать информацию, или может ли компьютер думать? / Д. П. Шишков // Искусственный интеллект : [международ. науч.-теорет. журн.] / [гол. ред. А. И. Шевченко]. — Донецьк. — 1996 — № 2. — С. 7—28.
125. Шкурятяна Н. Г. Сучасна українська мова : модульний курс / Н. Г. Шкурятяна, С. В. Шевчук. — К. : Вища школа, 2007. — 822 с.
126. Штерн І. Б. Вибрані топіки та лексикон сучасної лінгвістики / Ірен Борисівна Штерн. — К. : “АртЕк”, 1998. — 336 с.
127. Шульжук К. Ф. Синтаксис української мови / Каленик Федорович Шульжук. — К. : Видавничий центр “Академія”, 2004. — 408 с.
128. Ющук І. П. Українська мова : [підручник] / Іван Пилипович Ющук. — К. : Либідь, 2008. — 639 с.
129. Яковина В. Огляд та аналіз метрик оцінювання якості перекладу / В. Яковина, В. Масюкевич // Вісник Національного університету “Львівська політехніка”. Комп’ютерні науки та інформаційні технології. — 2013. — № 771. — С. 101—107.
130. Яндекс компания. Технологии / поисковые технологии. Машинный перевод [Электронный ресурс]. — Режим доступа : <http://company.yandex.ru/technologies/translation/>.
131. Aminzadeh A. Advocate : A Distributed Architecture for Speech-to-Speech Translation / A. Aminzadeh, W. Shen // Lincoln Laboratory Journal. — 2009. — Vol. 18. — Number 1. — P. 54—65.
132. Arnold D. Machine Translation: An Introductory Guide / [Arnold D., Balkan L., Meijer S. and others]. — London : NCC Blackwell, 1994. — 240 p.
133. Aziz W. PET : a Tool for Post-editing and Assessing Machine Translation [Электронный ресурс] / Aziz W., Sousa de S. C. M., Specia L. — Режим доступа : [http://www.lrec-conf.org/proceedings/lrec2012/pdf/985\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/985_Paper.pdf).

134. Babylon's Free Online Translation [Электронный ресурс]. — Режим доступа : <http://translation.babylon.com/>.
135. Banerjee S. METEOR : An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgements / S. Banerjee, A. Lavie // Proc. of the ACL 2005 Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization. — Michigan, 2005. — P. 65—72.
136. Benoit L. N. Quality Control in Human Translation : Use Cases and Specifications / [Электронный ресурс] / Le Ny Benoit. — Режим доступа : [http://transread.limsi.fr/Deliverable\\_3.1.pdf](http://transread.limsi.fr/Deliverable_3.1.pdf).
137. Bolshakov I. A. Computational Linguistics. Models, Resources, Applications // I. A. Bolshakov, A. Gelbukh. — Mexico : IPN, 2004. — 186 p.
138. Brown P. F. A statistical approach to machine translation / [Brown P. F., Cocke S. A., Della Pietra S. A. and others] // Computational Linguistics. 1990. — No. 16(2). — P. 79—85.
139. Brown P. The mathematics of statistical machine translation / [Brown P., Della Pietra S. A., Della Pietra V. J., Mercer R. L.] // Computational Linguistics. — 1993. — Vol. 19. — No. 2 — P. 263—311.
140. Capuat M. The trouble with SMT consistency / M. Capuat, M. Simard // Proceeding of the Seventh Workshop on Statistical Machine Translation. — Canada, Monreal. — P. 442—449.
141. Chiang D. A Hierarchical phrase-based translation / David Chiang // Computational Linguistics. — 2007. — Vol. 33. — No. 2. — P. 201—228.
142. Crego J. Ngram-based versus Phrase-based Statistical Machine Translation [Электронный ресурс] / [Crego J. M., Costa-jussa M. R., Marino J. B., Fonollosa J. A. R.]. — Режим доступа : <http://ssli.ee.washington.edu/people/katrin/mtrg/index.html/papers/iwslt05.pdf>.
143. Denkowski M. Meteor 1.3 : Automatic metric for reliable optimization and evaluation of machine translation systems / M. Denkowski, A. Lavie // Proceedings of the Sixth Workshop on Statistical Machine Translation. — 2011. — P. 85—91.

144. Deschamps A. Natural Language Processing : After a complicated history, is the maturity step finally reached? [Электронный ресурс] / Alix Deschamps. — Режим доступа : <http://mms.ecs.soton.ac.uk/2012/papers/26.pdf>.
145. Dhakar B. S. A Survey of Translation Quality of English to Hindi Online Translation Systems (Google and Bing) / [Электронный ресурс] / Dhakar B. S., Sanha S. K., Pandey K. K. — Режим доступа : <http://www.ijsrp.org/research-paper-1301/ijsrp-p1355.pdf>.
146. Doddington G. Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics / George Doddington // HLT '02 Proceedings of the second international conference on Human Language Technology Research. — San Francisco: Morgan Kaufmann Publishers Inc., 2002. — P. 138—145.
147. Early years in machine translation: memoirs and biographies of pioneers / [Ed. by W. J. Hutchins]. — Amsterdam—Philadelphia : John Benjamins Publishing Company, 2000. — 405 p.
148. Federico M. Assessing the Impact of Translation Errors on Machine Translation Quality with Mixed-effects Models / [Federico M., Negri M., Bentivogli L., Turchi M.] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), October 25—29, 2014, Doha, Qatar. — P. 1643—1653.
149. Frandsen M. W. IraqComm and FlexTrans : A Speech Translation System and Flexible Framework [Электронный ресурс] / Frandsen M. W., Riehemann S. Z., Precoda K. — Режим доступа : <http://doors.stanford.edu/~sr/frandsen-iraqcomm-flextrans-cisse.pdf>.
150. Gaspari F. Online and Free! Ten Years of Online Machine Translation : Origins, Developments, Current Use and Future Prospects / [Электронный ресурс] / F. Gaspari, J. Hutchins. — Режим доступа : <http://www.hutchinsweb.me.uk/MTS-2007.pdf>.
151. Gerasimov A. Review of Translation Quality Assurance Software [Электронный ресурс] / Andrei Gerasimov — Режим доступа : <http://www.translationdirectory.com/articles/article1253.htm>.

152. Girardi C. MT-EQuAl : a Toolkit for Human Assessment of Machine Translation Output / [Girardi C., Bentivogli L., Farajian M. A., Federico M.] // Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics : System Demonstrations, Dublin, Ireland, August 23—29 2014. — P. 120—123.
153. Glenn M. L. Management of Large Annotation Projects Involving Multiple Human Judges: a Case Study of GALE Machine Translation Post-editing / [Glenn M. L., Strassel S., Friedman L., Lee H.] // Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, 26 May—1 June 2008, Marrakech, Morocco. — P. 2957—2960.
154. Handbook of Natural Language Processing / [Ed. by R. Dale, H. Moise, H. Somers]. — New York : CRC Press, 2000. — 964 p.
155. Handbook of Natural Language Processing and Machine Translation : DARPA Global Autonomous Language Exploitation / [Ed. by J. Olive, C. Christianson, J. McCary]. — New York : Springer Science and Business Media, 2011. — 962 p.
156. Handbook of Natural Language Processing. Second Edition / [Ed. by N. Indurkha, F. Damerau]. — New York : CRC Press, 2010. — 704 p.
157. Hearne M. Statistical Machine Translation : A Guide for Linguists and Translators / M. Hearne, A. Way // Language and Linguistics Compass. — 2011. — P. 1—21.
158. How to use Google Translate [Электронный ресурс]. — Режим доступа : <https://support.google.com/translate/answer/2534527?hl=uk>.
159. Hutchins J. An Introduction to Machine Translation / J. Hutchins, H. L. Somers. — London : Academic Press, 1992. — 159 p.
160. Hutchins J. Machine Translation : General Overview / J. Hutchins // The Oxford Handbook of Computational Linguistics / [Ed. by Ruslan Mitkov]. — Oxford : University Press, 2009. — 786 p. — P. 501—511.
161. Hutchins J. The IAMT Certification initiative and defining translation system categories [Электронный ресурс] / John Hutchins. — Режим доступа : <http://hutchinsweb.me.uk/EAMT-2000.pdf>.
162. Hutchins J. Machine Translation : Past, Present, Future / John Hutchins. — N.Y. : Halsted Press, 1986. — 382 p.

163. Jones K. S. Computational Linguistics : Overview / Karen Sparck Jones // International Encyclopedia of Linguistics. — New York : Oxford University Press, 2003. — 525 p.
164. Kadhim A. K. An Evaluation of Online Machine Translation of Arabic into English News Headlines : Implications on Students' Learning Purposes [Электронный ресурс] / [Kadhim A. K., Habeeb S. L., Sapar A. A. and others]. — Режим доступа : <http://files.eric.ed.gov/fulltext/EJ1015412.pdf>.
165. Koponen M. Assessing Machine Translation Quality with Error Analysis [Электронный ресурс] / Maarit Koponen — Режим доступа : [http://www.skyl.fi/@Bin/40701/Koponen\\_MikaEL2010.pdf](http://www.skyl.fi/@Bin/40701/Koponen_MikaEL2010.pdf).
166. Keith A. The Oxford Handbook of the History of Linguistics / Allan Keith. — Oxford : Oxford University Press, 2013. — 924 p.
167. Knight K. Automating knowledge acquisition for machine translation / Kevin Knight // AI Magazine. — 1997. — 18(1) — P. 81—96.
168. Koehn P. Statistical Machine Translation / Philipp Koehn. — Cambridge : Cambridge University Press, 2010. — 433 p.
169. Koehn P. Statistical Phrase-Based Translation / Koehn P., Och F. J. , Marcu D. // Proceeding of HLT—NAACL 2003. Main Papers. — Edmonton, 2003. — P. 48—54.
170. Language Weaver Software Translation Solution Chosen by Babylon, Ltd. [Электронный ресурс]. — Режим доступа : <http://in.reuters.com/article/2008/09/15/idUS112710+15-Sep-2008+MW20080915>.
171. Languages and machines : computers in translation and linguistics. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council [Электронный ресурс]. — Режим доступа : [http://www.nap.edu/catalog.php?record\\_id=9547](http://www.nap.edu/catalog.php?record_id=9547).
172. Leusch G. N. CDER : Efficient MT Evaluation Using Block Movements / Leusch G. N., Ueffing N., Ney H. // Proc. of the 11th Conference of the European Chapter of the Association for Computational Linguistics. — 2006. — P. 241—248.
173. Lin C. Y. Automatic evaluation of summaries using n-gram co-occurrence statistics / C.Y. Lin, E. Hovy // Proceedings of the 2003 Conference of the North American

- Chapter of the Association for Computational Linguistics on Human Language Technology. — 2003. — Volume 1. — P. 71—78.
174. Liu C. Better Evaluation Metrics Lead to Better Machine Translation / Liu C., Dahlmeier D., Ng H.T. // Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. — Edinburgh, Scotland, UK, July 27—31, 2011. — P. 375—384.
175. Liu D. Syntactic Features for Evaluation Measures for Machine Translation and/or Summarization / D. Liu, D. Gildea // Proceedings of ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization. — 2005. — P. 25—32.
176. Lopez J. M. A Machine Translation Typology for MT Evaluations / J. M. Lopez, S. C. Roca [Электронный ресурс] — Режим доступа : <http://www.mt-archive.info/EAMT-2008-More-Lopez.pdf>.
177. Machine Translation group at Microsoft Research [Электронный ресурс]. — Режим доступа : <http://blogs.msdn.com/b/translation/archive/2007/09/22/machine-translation-group-at-microsoft-research.aspx>.
178. Madnani N. Re-examining Machine Translation Metrics for Paraphrase Identification / Madnani N., Tetreault J., Chodorow M. // Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies. — P. 182—190.
179. Makoushina J. Translation Quality Assurance Tools : Current State and Future Approaches [Электронный ресурс] / Julia Makoushina. — Режим доступа : <http://mt-archive.info/Aslib-2007-Makoushina.pdf>.
180. Marcu D. A phrase-based, joint probability model for statistical machine translation / D. Marcu, W. Wong // Proceedings of the Conference on Empirical Methods in Natural Language Processing, WMMNLP'02. — 2002. — P. 133—139.
181. Matsoukas S. Integrating Speech Recognition and Machine Translation [Электронный ресурс] / [Matsoukas S., Bulyko I., Xiang B. and others] — Режим доступа : [http://www.bbn.com/resources/pdf/icassp07\\_STT\\_MT\\_Integration.pdf](http://www.bbn.com/resources/pdf/icassp07_STT_MT_Integration.pdf).

182. Matusov V. On the Integration of Speech Recognition and Statistical Machine Translation [Электронный ресурс] / Matusov E., Santhak S., Ney H. — Режим доступа : [http://www.testar.org/publicazioni/scientific\\_publications/RWTH/RWTH\\_28\\_06/Matusov%20Kanthak%20Ney%20-%20Integration%20of%20ASR%20and%20SMT%20-%20Interspeech2005.pdf](http://www.testar.org/publicazioni/scientific_publications/RWTH/RWTH_28_06/Matusov%20Kanthak%20Ney%20-%20Integration%20of%20ASR%20and%20SMT%20-%20Interspeech2005.pdf).
183. Melamed I. Precision and Recall of Machine Translation / Melamed I., Green R., Turian J. // Proceedings of the HLT—NAACL 2003 : Short Papers. — Canada, Edmonton. — P. 61—63.
184. Milkowski M. Translation Quality Checking in Language Tool [Электронный ресурс] / Marcin Milkowski. — Режим доступа : [http://marcinmilkowski.pl/downloads/PALC\\_2011\\_LT.pdf](http://marcinmilkowski.pl/downloads/PALC_2011_LT.pdf).
185. Nagao M. A Framework of a Mechanical Translation between Japanese and English by Analogy Principle / Makoto Nagao // Artificial and Human Intelligence. — Amsterdam, 1984. — P. 135—139.
186. Nirenburg S. Progress in Machine Translation / Sergei Nirenburg. — Amsterdam : IOS Press, 1992. — 223 p.
187. Nomiya H. Machine Translation by Case Generalization / Hiroshi Nomiya // Proceedings of COLING 1992. — Nantes, 1992. — P. 714—720.
188. Olive J. Handbook of Natural Language Processing and Machine Translation : Processing and Machine Translation : DARPA Global Autonomous Language Exploitation / Joseph Olive. — Springer Science, 2011. — 936 p.
189. Papineni K. BLEU : a Method for Automatic Evaluation of Machine Translation / [Papineni K., Roukos S., Ward T., Zhu WJ.] // Proc. of the 40th Annual Meeting of the Association for Comp. Linguistics. — Philadelphia, PA., 2002. — P. 311—318.
190. Partridge D. The Foundations of Artificial Intelligence. A Sourcebook / D. Partridge, Y. Wilks. — New York : Cambridge University Press, 1990. — 498 p.
191. Pedersen T. A decision tree of bigrams is an accurate predictor of word senses // Ted Pedersen // Proceedings of Second Annual Meeting of NAC ACL. — Pittsburgh, PA, 2001. — P. 79—86.



192. Piotrovskij R. G. MT in the Former USSR and in the Newly Independent States (NIS). History, Romantic Era, Prosaic Time / Raimund G. piotrovskij // Early years in machine translation : memoirs and biographies of pioneers / [Ed. by W. J. Hutchins]. — Amsterdam — Philadelphia : John Benjamins Publishing Company, 2000. — 405 p. — P. 237—239.
193. Pragma Business [Электронный ресурс]. — Режим доступа : <http://www.translate.ua/uk/pragma-6x/pragma-business>.
194. Rajman M. Automatically predicting MT systems rankings compatible with Fluency, Adequacy or Informativeness scores / M. Rajman, A. Hartley // Proceedings of the Fourth Workshop on MT Evaluation, MT Summit VIII, Santiago de Compostela, September, 2001. — P. 29—34.
195. Rinsche A. Comparative MT Performance Evaluation : An Empirical Study / Adrian Rinsche // Proceedings of the Evaluators' Forum, April 21st—24th, 1991, Les Rasses, Vaud, Switzerland / [Ed. K. Falkedal]. — Geneva : ISSCO. — P. 169—180.
196. Saldanha G. Research Methodologies in Translation Studies / G. Saldanha, S. O'Brien. — N.Y. : Routledge, 2014. — 280 p.
197. Samuelsson C. Statistical Methods / Christen Samuelsson // Oxford Handbook of Computational Linguistics. — Oxford University Press, 2003. — P. 358—375.
198. Schmidt C. Using a Bilingual Context in Word-Based Statistical Machine Translation [Электронный ресурс] / Schmidt C., Vilar D., Ney H. — Режим доступа : <http://www.dfki.de/~davi01/papers/schmidt08:pris.pdf>.
199. Snover M. A Study of Translation Edit Rate with Targeted Human Annotation / [M. Snover, B. Dorr, R. Schwartz and others] // Proceedings of Association for Machine Translation in America (AMTA-2006). — Cambridge, Massachusetts, 2006. — P. 223—231.
200. Snover M. G. TER-Plus : Paraphrase, Semantic and Alignment Enhancements to Translation Edit Rate / [Snover M. G., Madnani N., Dorr B., Schwartz R.] // Machine Translation. — 2009. — No. 23(2). — P. 117—127.

201. Somers H. Machine Translation : Latest Developments / Harold Somers // The Oxford Handbook of Computational Linguistics / [Ed. by Ruslan Mitkov]. — Oxford : University Press, 2009. — 786 pp. — P. 512—528.
202. Somers H. Review article : example-based machine translation / Harold Somers // Machine Translation. — 1999. — 14(2). — P. 113—157.
203. Specia L. Fundamental and New Approaches to Statistical Machine Translation [Электронный ресурс] / Lucia Specia — Режим доступа : [http://pers-www.wlv.ac.uk/~in1316/resources/tutorial\\_smt.pdf](http://pers-www.wlv.ac.uk/~in1316/resources/tutorial_smt.pdf).
204. Tillmann C. Accelerated DP Based Search for Statistical Translation / [Tillmann C., Vogel S. Ney H. and others] // European Conference on Speech Communication and Technology. — Rhodes, Greece, 1997. — P. 2667—2670.
205. Tomita M. Another Stride Towards Knowledge-Based Machine Translation / M. Tomita, J. G. Carbonell // Proceedings of COLING 1986. — Bonn, 1986. — P. 633— 638.
206. Transparent Language launches FreeTranslation.com [Электронный ресурс]. — Режим доступа : <http://www.multilingual.com/newsDetail.php?id=886>.
207. Tripathi S. Approaches to machine translation / S. Tripathi, J. K. Sarkhel // Annals of Library and Information Studies. — 2010. — Vol. 57. — P. 388—393.
208. Trujillo A. Translation Engines : Techniques for Machine Translation / Arturo Trujillo. — London : Springer Science & Business Media, 1999. — 303 p.
209. Turian J. P. Evaluation of Machine Translation and Its Evaluation / Turian J. P., Shea L., Melamed I. D. // Proc. of MT Summit IX. — New Orleans, 2003. — P. 386—393.
210. Vilar D. Error Analysis of Statistical Machine Translation Output [Электронный ресурс] / [Vilar D., Xu J., D'Haro L. F., Ney H.]. — Режим доступа : [https://www-i6.informatik.rwth-aachen.de/publications/download/281/VilarDavidXuJiaDHaroLuisFernoNeyHermann-7BErrorAnalysisofMachineTranslationOutput %7D--2006.pdf](https://www-i6.informatik.rwth-aachen.de/publications/download/281/VilarDavidXuJiaDHaroLuisFernoNeyHermann-7BErrorAnalysisofMachineTranslationOutput%7D--2006.pdf).
211. Weaver W. Memorandum, July 1949 [Электронный ресурс] / Warren Weaver. — Режим доступа : <http://www.hutchinsweb.me.uk/MTNI-22-1999>.

212. Watanabe H. A Method for Distinguishing Exceptional and General Examples in Example-based Transfer Systems / Hideo Watanabe // Proceedings of COLING 94. — Kyoto, 1994. — P. 39—44.
213. Whitelock P. Shake-and-Bake Translation / P. Whitelock // Proceedings of COLING 92. — Nantes, 1992. — P. 784—785.
214. Wilks Y. Machine Translation : Its Scope and Limits / Yorick Wilks. — New York : Springer Science + Business Media LLC, 2009. — 264 p.
215. Yamada K. A syntax-based statistical translation model / K. Yamada, K. Knight // 39th Annual Meeting of the Association for Computational Linguistics. — 2001. — P. 523—530.
216. Zens R. Phrase-based statistical machine translation / Zens R., Och F., Hey H. // KI-2002 : Advances in artificial intelligence. — 2002. — Vol. LNAI 2479. — P. 18—32.
217. Zerfass A. Comparing Terminology Checkers. Examining acrocheck, Wordfast, Déjà Vu, ErrorSpy, Quintilian, TRADOS, Across, Transit and SDLX [Електронний ресурс] / Angelika Zerfass. — Режим доступу : <http://www.51lunwen.org/UploadFile/org201012271307368263/20101227130736890.pdf>.

### СПИСОК ЛЕКСИКОГРАФІЧНИХ ТА ДОВІДКОВИХ ДЖЕРЕЛ

218. Англо-український словник : У 2 т. — Близько 120 000 слів / [Склав М. І. Балла]. — К. : Освіта, 1996. — Т. 1. — 1996. — 752 с.
219. Англо-український словник : У 2 т. — Близько 120 000 слів / [Склав М. І. Балла]. — К. : Освіта, 1996. — Т. 2. — 1996. — 712 с.
220. Англо-український фразеологічний словник / [Уклад. К. Т. Баранцев]. — 2-ге вид., випр. — К. : Т-во “Знання”, КОО, 2005. — 1056 с.
221. Большой англо-русский политехнический словарь : В 2 т. Около 200 000 терминов / [С. М. Баринов, А. Б. Борковский, В. А. Владимиров и др.]. — М. : Русский язык, 1991. — Т. 1. — 1991. — 701 с.

222. Большой англо-русский политехнический словарь : В 2 т. Около 200 000 терминов / [С. М. Баринов, А. Б. Борковский, В. А. Владимиров и др.]. — М. : Русский язык, 1991. — Т. 1. — 1991. — 720 с.
223. Глосарій з навчальної дисципліни “Системи штучного інтелекту” для студентів напряму підготовки “Комп’ютерні науки” всіх форм навчання / [Укл. С. В. Кавун, В. С. Коротченко]. — Харків : Вид. ХНЕУ, 2008. — 136 с.
224. Селіванова О. О. Лінгвістична енциклопедія / Олена Олександрівна Селіванова. — Полтава : Довкілля-К, 2011. — 844 с.
225. Словник української мови: в 11 т. / [ред. колег. І. К. Білодід (голова) та ін.]. — К.: Наукова думка, 1970—1980. — Т. 4 : І—М / [ред. А. А. Бурячок, Г. М. Гнатюк, П. П. Доценко]. — К.: Наукова думка, 1973. — 840 с.
226. Українська мова : Енциклопедія / [редкол. : В. М. Русанівський, О. О. Тараненко, М. П. Зяблюк та ін.] — К. : Укр. енцикл. ім. М. П. Бажана, 2004. — 824 с.
227. Український правопис / НАН України, Ін-т мовознавства ім. О. О. Потебні; Ін-т української мови. — К. : Наук. думка, 2012. — 288 с
228. Яремійчук Р. Англо-український нафтогазовий словник / Роман Яремійчук. — К. : Українська книга, 1998. — 544 с.
229. CED — Collins English Dictionary [Електронний ресурс]. — Режим доступу : <http://www.collinsdictionary.com/>

### СПИСОК ДЖЕРЕЛ ІЛЮСТРАТИВНОГО МАТЕРІАЛУ

230. Авіація. In the future along with Boeing 737-800 NG [Електронний ресурс]. — Режим доступу : <http://www.utair.ua/en/news/4265783.html>.
231. Авто. New Ford Focus Delivers Industry-First Technology That Can Predict and Prevent Skids [Електронний ресурс]. — Режим доступу : <http://winner.ua/en/news/ournews/4/406/>.

232. Банківська справа. NEFCO and Ukrainian Bank Lviv sign new agreement [Електронний ресурс]. — Режим доступу : [http://www.nefco.org/news/nefco\\_and\\_ukrainian\\_bank\\_lviv\\_sign\\_new\\_agreement](http://www.nefco.org/news/nefco_and_ukrainian_bank_lviv_sign_new_agreement).
233. Біблія. Matthew 20 [Електронний ресурс]. — Режим доступу : [http://www.just1word.com/#!/bible/verse/matthew\\_20?version=nkjv](http://www.just1word.com/#!/bible/verse/matthew_20?version=nkjv).
234. Бізнес. Top 500 largest companies in Central and Eastern Europe – 2013 [Електронний ресурс]. — Режим доступу : [http://www.deloitte.com/view/en\\_UA/ua/press/ua-pressreleases-en/6392b457e2b41410VgnVCM1000003256f70aRCRD.htm](http://www.deloitte.com/view/en_UA/ua/press/ua-pressreleases-en/6392b457e2b41410VgnVCM1000003256f70aRCRD.htm).
235. Біологія. Embryonic Development: The First 4 Weeks. Chapter 3. Fertilization [Електронний ресурс]. — Режим доступу : [http://www.ehd.org/resources\\_bpd\\_illustrated.php?page=6](http://www.ehd.org/resources_bpd_illustrated.php?page=6).
236. Будівництво. Civil and Erection Works [Електронний ресурс]. — Режим доступу : <http://centrbud.ua/en/content/civil-and-erection-works>.
237. Військова. Tenth anniversary of NATO's Air Policing mission in the Baltic States [Електронний ресурс]. — Режим доступу : [http://www.nato.int/cps/en/natolive/news\\_106114.htm?selectedLocale=en](http://www.nato.int/cps/en/natolive/news_106114.htm?selectedLocale=en).
238. Гірнична. Mining Division [Електронний ресурс]. — Режим доступу : [http://www.metinvestholding.com/en/activity/raw\\_iron](http://www.metinvestholding.com/en/activity/raw_iron).
239. Деревообробна. Houses Made Of Logs [Електронний ресурс]. — Режим доступу : <http://aspekt-lviv.com/en/construction-of-wood/>.
240. Екологія. Shale gas extraction and climate change [Електронний ресурс]. — Режим доступу : <http://shalegas.in.ua/en/shale-gas-and-climate-change/>.
241. Економіка. Top 500 largest companies in Central and Eastern Europe — 2013 [Електронний ресурс]. — Режим доступу : [http://www.deloitte.com/view/en\\_UA/ua/press/ua-pressreleases-en/6392b457e2b41410VgnVCM1000003256f70aRCRD.htm](http://www.deloitte.com/view/en_UA/ua/press/ua-pressreleases-en/6392b457e2b41410VgnVCM1000003256f70aRCRD.htm).
242. Електроніка. BC-700 Microprocessor charger [Електронний ресурс]. — Режим доступу : [http://www.tme.eu/en/pages/Product\\_of\\_the\\_week:bc-700-microprocessor-charger.html](http://www.tme.eu/en/pages/Product_of_the_week:bc-700-microprocessor-charger.html).

243. Енергетика. Energy efficiency boosted in Ukrainian cities Vinnytsia and Zhytomyr [Електронний ресурс]. — Режим доступу : [http://www.nefco.org/news/energy\\_efficiency\\_boosted\\_in\\_ukrainian\\_cities\\_vinnytsia\\_and\\_zhytomyr](http://www.nefco.org/news/energy_efficiency_boosted_in_ukrainian_cities_vinnytsia_and_zhytomyr).
244. Загальна. Personnel management consulting services. From personnel to human capital [Електронний ресурс]. — Режим доступу : [http://www.deloitte.com/view/en\\_UA/ua/services/consulting/personnel-management/index.htm](http://www.deloitte.com/view/en_UA/ua/services/consulting/personnel-management/index.htm).
245. Залізниця. Train check-in [Електронний ресурс]. — Режим доступу : <http://gd.tickets.ua/en/pub/rail/rail-faq/chek-in/registration-rail.html>.
246. Зв'язок. Deloitte analysis of top telecommunications trends for 2011 [Електронний ресурс]. — Режим доступу : [http://www.deloitte.com/view/en\\_UA/ua/press/ua-pressreleases-en/cde4bb093042e210VgnVCM3000001c56f00aRCRD.htm](http://www.deloitte.com/view/en_UA/ua/press/ua-pressreleases-en/cde4bb093042e210VgnVCM3000001c56f00aRCRD.htm).
247. Інструкція. Panasonic Operating Instructions [Електронний ресурс]. — Режим доступу : <http://service.us.panasonic.com/OPERMANPDF/PVGS500-MULTI.PDF>.
248. Інтернет. Cybersecurity : Everybody's imperative [Електронний ресурс]. — Режим доступу : [http://www.deloitte.com/view/en\\_JO/jo/industries/public-sector/93f4d69972948310VgnVCM1000001956f00aRCRD.htm](http://www.deloitte.com/view/en_JO/jo/industries/public-sector/93f4d69972948310VgnVCM1000001956f00aRCRD.htm).
249. Інформатика. Seismic data processing and interpretation [Електронний ресурс]. — Режим доступу : <http://www.ukrgeofizika.kiev.ua/en/data-processing.html>.
250. Кіно. Festival Regulations 2014 [Електронний ресурс]. — Режим доступу : <http://www.molodist.com/en/regulations>.
251. Комерція. More than half of all computers aren't traditional computers anymore, according to Deloitte's Technology Predictions 2011 report [Електронний ресурс]. — Режим доступу : [http://www.deloitte.com/view/en\\_AU/au/news-research/media-releases/6f18e6af745cd210VgnVCM3000001c56f00aRCRD.htm](http://www.deloitte.com/view/en_AU/au/news-research/media-releases/6f18e6af745cd210VgnVCM3000001c56f00aRCRD.htm).
252. Косметика. Graftobian HD Make-up [Електронний ресурс]. — Режим доступу : <http://topcosmetics.ua/eng/partners/brands/tm-graftobian.html>.

253. Космонавтика. NASA Researches: DNA Building Blocks Can Be Made in Space [Електронний ресурс]. — Режим доступу : <http://www.nasa.gov/topics/solarsystem/features/dna-meteorites.html>.
254. Кулінарія. Exclusive Italian cuisine — always the best, always nearby! [Електронний ресурс]. — Режим доступу : <http://alaminute.ua/en/pastamia/about/>.
255. Логістика. Metinvest. Supply Chain Management. [Електронний ресурс]. — Режим доступу : [http://www.metinvestholding.com/en/activity/coal\\_coke](http://www.metinvestholding.com/en/activity/coal_coke).
256. Математика. Right-angled Triangle [Електронний ресурс]. — Режим доступу : <http://www.e-olimp.com/en/problems/1049>.
257. Машинобудування. Heavy Engineering [Електронний ресурс]. — Режим доступу : [http://www.scmholding.com/en/business/sectors/heavy\\_engineering/](http://www.scmholding.com/en/business/sectors/heavy_engineering/).
258. Медицина. Infectious diseases [Електронний ресурс]. — Режим доступу : [http://oxford-med.com.ua/en/services/gynecology/infectious\\_diseases/](http://oxford-med.com.ua/en/services/gynecology/infectious_diseases/).
259. Металургійна. Metallurgical Division [Електронний ресурс]. — Режим доступу : [http://www.metinvestholding.com/en/activity/steel\\_rolling](http://www.metinvestholding.com/en/activity/steel_rolling).
260. Морська. Counter-piracy operations [Електронний ресурс]. — Режим доступу : [http://www.nato.int/cps/en/natolive/topics\\_48815.htm?selectedLocale=en](http://www.nato.int/cps/en/natolive/topics_48815.htm?selectedLocale=en).
261. Музика. Freedom in music [Електронний ресурс]. — Режим доступу : <http://www.day.kiev.ua/en/article/culture/freedom-music>.
262. Нафтогаз. Shale revolution will lead to Gazprom's losses — expert [Електронний ресурс]. — Режим доступу : [http://www.ukrinform.ua/eng/news/shale\\_revolution\\_will\\_lead\\_to\\_gazproms\\_losses\\_expert\\_310221](http://www.ukrinform.ua/eng/news/shale_revolution_will_lead_to_gazproms_losses_expert_310221).
263. Подорож. Visa. Avoiding visa refusal? [Електронний ресурс]. — Режим доступу : <http://skytravel.kiev.ua/eng/skynews/news/923.html>.
264. Поліграфія. Production of photopolymer printing plates [Електронний ресурс]. — Режим доступу : <http://www.dpa.com.ua/en/fotopolimiernaia-piechatnaia-forma/>.
265. Право. The North Atlantic Treaty [Електронний ресурс]. — Режим доступу : [http://www.nato.int/cps/en/natolive/official\\_texts\\_17120.htm?selectedLocale=en](http://www.nato.int/cps/en/natolive/official_texts_17120.htm?selectedLocale=en).
266. Секс. 44% of Ukrainians consider sex to be an important part of their lives – social research by KIIS [Електронний ресурс]. — Режим доступу : <http://www.kiis.com.ua/?lang=eng&cat=reports&id=144&page=1>.

267. Сільське господарство. Pig Breeding [Електронний ресурс]. — Режим доступу : [http://www.tdp.od.ua/en/agriculture/pig\\_breeding/](http://www.tdp.od.ua/en/agriculture/pig_breeding/).
268. Спорт. The principal distinguishing features of playing footdoubleball in contrast to playing football [Електронний ресурс]. — Режим доступу : [http://www.footdoubleball.org.ua/.%D0%93%D0%BB%D0%B0%D0%B2%D0%BD%D0%B0%D1%8F/otlichija\\_pravil/tabid/69/language/en-US/Default.aspx](http://www.footdoubleball.org.ua/.%D0%93%D0%BB%D0%B0%D0%B2%D0%BD%D0%B0%D1%8F/otlichija_pravil/tabid/69/language/en-US/Default.aspx).
269. Текстиль. What Is the Difference between Tufted and Needlepunched Carpets? [Електронний ресурс]. — Режим доступу : [http://carpets.sintelon.rs/What\\_Is\\_the\\_Difference\\_between\\_Tufted\\_and\\_Needlepunched\\_Carpets-27-1-72](http://carpets.sintelon.rs/What_Is_the_Difference_between_Tufted_and_Needlepunched_Carpets-27-1-72).
270. Техніка. Ford Delivers Most Powerful Diesel Kuga Ever Alongside Lower CO<sub>2</sub>, New Colours and More Advanced Technology [Електронний ресурс]. — Режим доступу : <http://winner.ua/en/news/ournews/1/433/>.
271. Фізика. Faster-than-light neutrinos face time trial [Електронний ресурс]. — Режим доступу : [http://nauka.in.ua/en/news/articles/article\\_detail/6827](http://nauka.in.ua/en/news/articles/article_detail/6827).
272. Харчова. General milk characteristics [Електронний ресурс]. — Режим доступу : <http://lactalis.com.ua/en/healthy/milkFeatures>.
273. Хімія. Urea [Електронний ресурс]. — Режим доступу : <http://axiom-agro.com.ua/en/catalogue/urea-en>.



## ДОДАТКИ

## Додаток А

**Кількісне співвідношення помилок  
системи Pragma 6.3 Business**

<b>Типи помилок</b>	<b>Кількість помилок</b>	
<b>I. Лексичні помилки</b>	1759	53,5%
1. Помилки у перекладі слів	1218	37,0%
1.1 Вибір нерелевантних перекладних відповідників	924	28,1%
1.1.1 Неправильне вирішення лексичної полісемії	733	22,3%
1.1.2 Неправильне вирішення лексико-граматичної омонімії	173	5,3%
1.1.3 Неадекватний переклад власних назв та аббревіатур, які слід було залишити в оригінальному варіанті	8	0,2%
1.1.4 Неправильне вирішення лексичної омонімії	5	0,2%
1.1.5 Вибір нерелевантних перекладних відповідників через наявність помилки в оригіналі	3	0,1%
1.1.6 Вживання русизмів	2	0,1%
1.2 Неперекладені слова	294	8,9%
2. Помилки у перекладі словосполучень	541	16,5%
2.1 Дослівний переклад іншомовних висловів	247	7,5%
2.2 Неадекватний переклад словосполучень як семантичних єдностей	190	5,8%
2.3 Неадекватний переклад фразеологічних сполучень	95	2,9%
2.4 Неадекватний переклад словосполучень, які слід було залишити в оригінальному варіанті	9	0,3%
<b>II. Граматичні помилки</b>	1423	43,2%
1. Морфологічні помилки	465	14,1%
1.1 Помилки при знятті граматичної омонімії	392	11,9%
1.2 Ненормативне утворення пасивної конструкції дієслова	15	0,5%
1.3 Порушення категорії числа іменника	15	0,5%

1.4 Ненормативне утворення ступеня порівняння прикметника	11	0,3%
1.5 Порухення категорії часу дієслова	9	0,3%
1.6 Вживання ненормативної форми дієприкметника	8	0,2%
1.7 Ненормативне утворення відмінкової форми іменника	8	0,2%
1.8 Порухення категорії способу дієслова	4	0,1%
1.9 Вживання ненормативної форми дієприслівника	3	0,1%
2. Синтаксичні помилки	958	29,1%
2.1 Помилки у побудові словосполучень	358	10,9%
2.1.1 Порухення керування	119	3,6%
2.1.2 Відхилення від нормативно-узуального порядку слів, порухення логіко-синтаксичного зв'язку між словами через неправильний вибір граматичної трансформації	79	2,4%
2.1.3 Неправильний вибір прийменника	74	2,3%
2.1.4 Порухення узгодження між словами	47	1,4%
2.1.5 Відсутність необхідного прийменника	25	0,8%
2.1.6 Вживання зайвого прийменника	14	0,4%
2.2 Помилки у побудові речень	600	18,2%
2.2.1 Помилки у побудові простих речень	320	9,7%
2.2.1.1 Порухення синтаксичних зв'язків між членами речення	161	4,9%
2.2.1.2 Граматична неоднотипність однорідних членів речення	86	2,6%
2.2.1.3 Порухення координації між підметом і присудком	47	1,4%
2.2.1.4 Неправильний порядок слів	16	0,4%
2.2.1.5 Штучне ускладнення речення	4	0,1%
2.2.1.6 Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	2	0,1%
2.2.1.7 Вживання зайвих членів речення	2	0,1%
2.2.1.8 Пропущені члени речення	2	0,1%
2.2.2 Помилки у побудові складних речень	280	8,5%
2.2.2.1 Порухення синтаксичних зв'язків між членами речення	117	3,6%
2.2.2.2 Граматична неоднотипність однорідних членів речення	49	1,4%
2.2.2.3 Порухення координації між підметом і	47	1,4%

присудком		
2.2.2.4 Невідповідність анафоричного займенника співвідноному з ним іменникові в іншій частині речення	17	0,5%
2.2.2.5 Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення	16	0,5%
2.2.2.6 Штучне ускладнення речення	12	0,4%
2.2.2.7 Неправильний порядок слів	12	0,4%
2.2.2.8 Вживання зайвих членів речення	8	0,2%
2.2.2.9 Пропущені члени речення	2	0,1%
III. Лексико-граматичні помилки	100	3,1%
IV. Орфографічні помилки	5	0,2%
Усього	3287	100%

## Додаток Б

## Помилки системи Pragma 6.3 Business за тематиками

Назва тематики	Кількість помилок
1. Будівництво	147
2. Металургійна	122
3. Право	121
4. Електроніка	115
5. Машинобудування	112
6. Зв'язок	106
7. Комерція	105
8. Текстиль	104
9. Гірнична	96
10. Екологія	96
11. Морська	95
12. Авто	90
13. Деревообробна	88
14. Космонавтика	87
15. Кіно	82
16. Інтернет	78
17. Косметика	77
18. Бізнес	73
19. Економіка	72
20. Залізниця	70
21. Поліграфія	69
22. Енергетика	65
23. Інструкція	65
24. Авіація	64
25. Техніка	64
26. Нафтогаз	63
27. Банки	62
28. Логістика	61
29. Інформатика	60
30. Музика	57
31. Військова	56
32. Кулінарія	48
33. Сільське господарство	47
34. Математика	46
35. Секс	46
36. Біологія	43
37. Харчова	42
38. Фізика	41
39. Спорт	40
40. Медицина	38
41. Подорож	37
42. Біблія	33
43. Хімія	31
44. Загальна	28
Усього	3287

## Додаток В

**Кількісне співвідношення помилок  
системи Google Translate**

<b>Типи помилок</b>	<b>Кількість помилок</b>	
<b>I. Лексичні помилки</b>	266	39,5%
1. Помилки у перекладі слів	158	23,4%
1.1 Вибір нерелевантних перекладних відповідників	136	20,1%
1.1.1 Неправильне вирішення лексичної полісемії	100	14,8%
1.1.2 Неправильне вирішення лексико-граматичної омонімії	27	4,0%
1.1.3 Вживання русизмів	7	1,0%
1.1.4 Неправильне вирішення лексичної омонімії	2	0,3%
1.2 Неперекладені слова	22	3,3%
2. Помилки у перекладі словосполучень	108	16,1%
2.1 Дослівний переклад іншомовних висловів	59	8,8%
2.2 Неадекватний переклад словосполучень як семантичних єдностей	43	6,4%
2.3 Неадекватний переклад фразеологічних сполучень	6	0,9%
<b>II. Граматичні помилки</b>	400	59,3%
1. Морфологічні помилки	63	9,3%
1.1 Помилки при знятті граматичної омонімії	52	7,7%
1.2 Ненормативне утворення відмінкової форми іменника	5	0,8%
1.3 Порушення категорії числа іменника	3	0,4%
1.4 Порушення категорії часу дієслова	3	0,4%
2. Синтаксичні помилки	337	50,0%
2.1 Помилки у побудові словосполучень	171	25,4%
2.1.1 Порушення керування	102	15,1%
2.1.2 Порушення узгодження між словами	31	4,6%
2.1.3 Неправильний вибір прийменника	18	2,7%
2.1.4 Відхилення від нормативно-узуального порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір граматичної трансформації	12	1,8%

2.1.5 Вживання зайвого прийменника	5	0,7%
2.1.6 Відсутність необхідного прийменника	3	0,5%
2.2 Помилки у побудові речень	166	24,6%
2.2.1 Помилки у побудові простих речень	82	12,2%
2.2.1.1 Порушення синтаксичних зв'язків між членами речення	39	5,8%
2.2.1.2 Граматична неоднотипність однорідних членів речення	31	4,6%
2.2.1.3 Порушення координації між підметом і присудком	7	1,1%
2.2.1.4 Пропущені члени речення	5	0,7%
2.2.2 Помилки у побудові складних речень	84	12,4%
2.2.2.1 Порушення синтаксичних зв'язків між членами речення	30	4,5%
2.2.2.2 Порушення координації між підметом і присудком	24	3,6%
2.2.2.3 Граматична неоднотипність однорідних членів речення	18	2,7%
2.2.2.4 Невідповідність анафоричного займенника співвідносно з ним іменникові в іншій частині речення	5	0,7%
2.2.2.5 Пропущені члени речення	3	0,4%
2.2.2.6 Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення	2	0,3%
2.2.2.7 Неправильний порядок слів	1	0,1%
2.2.2.8 Штучне ускладнення речення	1	0,1%
III. Лексико-граматичні помилки	8	1,2%
IV. Орфографічні помилки	0	0,0%
Усього	674	100%

## Додаток Д

**Кількісне співвідношення помилок  
системи Bing Translator**

<b>Типи помилок</b>	<b>Кількість помилок</b>	
<b>I. Лексичні помилки</b>	371	39,8%
1. Помилки у перекладі слів	217	23,2%
1.1 Вибір нерелевантних перекладних відповідників	174	18,6%
1.1.1 Неправильне вирішення лексичної полісемії	124	13,3%
1.1.2 Неправильне вирішення лексико-граматичної омонімії	42	4,5%
1.1.3 Неправильне вирішення лексичної омонімії	5	0,5%
1.1.4 Неадекватний переклад власних назв та абревіатур, які слід було залишити в оригінальному варіанті	2	0,2%
1.1.5 Вживання русизмів	1	0,1%
1.2 Неперекладені слова	43	4,6%
2. Помилки у перекладі словосполучень	154	16,6%
2.1 Дослівний переклад іншомовних висловів	86	9,2%
2.2 Неадекватний переклад словосполучень як семантичних єдностей	50	5,4%
2.3 Неадекватний переклад фразеологічних сполучень	10	1,1%
2.4 Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті	8	0,9%
<b>II. Граматичні помилки</b>	548	58,7%
1. Морфологічні помилки	91	9,8%
1.1 Помилки при знятті граматичної омонімії	75	8,0%
1.2 Порушення категорії числа іменника	10	1,1%
1.3 Порушення категорії часу дієслова	3	0,3%
1.4 Ненормативне утворення відмінкової форми іменника	2	0,2%
1.5 Ненормативне утворення ступеня порівняння прикметника	1	0,1%
2. Синтаксичні помилки	457	48,9%
2.1 Помилки у побудові словосполучень	248	26,5%
2.1.1 Порушення керування	145	15,5%

2.1.2	Порушення узгодження між словами	63	6,8%
2.1.3	Неправильний вибір прийменника	18	1,9%
2.1.4	Відхилення від нормативно-узуального порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір граматичної трансформації	11	1,2%
2.1.5	Відсутність необхідного прийменника	8	0,8%
2.1.6	Вживання зайвого прийменника	3	0,3%
2.2	Помилки у побудові речень	209	22,4%
2.2.1	Помилки у побудові простих речень	104	11,1%
2.2.1.1	Порушення синтаксичних зв'язків між членами речення	52	5,6%
2.2.1.2	Граматична неоднотипність однорідних членів речення	32	3,4%
2.2.1.3	Порушення координації між підметом і присудком	14	1,5%
2.2.1.4	Пропущені члени речення	4	0,4%
2.2.1.5	Зайві члени речення	1	0,1%
2.2.1.6	Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	1	0,1%
2.2.2	Помилки у побудові складних речень	105	11,3%
2.2.2.1	Порушення синтаксичних зв'язків між членами речення	42	4,5%
2.2.2.2	Порушення координації між підметом і присудком	29	3,1%
2.2.2.3	Граматична неоднотипність однорідних членів речення	20	2,2%
2.2.2.4	Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	5	0,6%
2.2.2.5	Пропущені члени речення	4	0,4%
2.2.2.6	Неправильний порядок слів	3	0,3%
2.2.2.7	Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення	1	0,1%
2.2.2.8	Зайві члени речення	1	0,1%
	III. Лексико-граматичні помилки	14	1,5%
	IV. Орфографічні помилки	0	0,0%
	Усього	933	100%



## Додаток Е

**Кількісне співвідношення помилок  
системи Babylon**

<b>Типи помилок</b>	<b>Кількість помилок</b>	
<b>I. Лексичні помилки</b>	<b>762</b>	<b>56,7%</b>
1. Помилки у перекладі слів	517	38,5%
1.1 Вибір нерелевантних перекладних відповідників	286	21,3%
1.1.1 Неправильне вирішення лексичної полісемії	190	14,1%
1.1.2 Неправильне вирішення лексико-граматичної омонімії	89	6,6%
1.1.3 Неправильне вирішення лексичної омонімії	4	0,3%
1.1.4 Вживання русизмів	2	0,2%
1.1.5 Неадекватно перекладені власні назви та аббревіатури, які слід було залишити в оригінальному варіанті	1	0,1%
1.2 Неперекладені слова	231	17,2%
2. Помилки у перекладі словосполучень	245	18,2%
2.1 Дослівний переклад іншомовних висловів	108	8,0%
2.2 Неадекватний переклад словосполучень як семантичних єдностей	93	6,9%
2.3 Неправильний переклад фразеологічних сполучень	41	3,1%
2.4 Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті	3	0,2%
<b>II. Граматичні помилки</b>	<b>563</b>	<b>41,9%</b>
1. Морфологічні помилки	181	13,5%
1.1 Помилки при знятті граматичної омонімії	153	11,4%
1.2 Порушення категорії числа іменника	15	1,1%
1.3 Ненормативне утворення відмінкової форми іменника	5	0,4%
1.4 Ненормативне утворення ступеня порівняння прикметника	3	0,2%
1.5 Вживання ненормативної форми дієприслівника	3	0,2%
1.6 Порушення категорії часу дієслова	1	0,1%
1.7 Вживання ненормативної форми	1	0,1%

дієприкметника		
2. Синтаксичні помилки	382	28,4%
2.1 Помилки у побудові словосполучень	174	12,9%
2.1.1 Порушення керування	97	7,2%
2.1.2 Порушення узгодження між словами	32	2,4%
2.1.3 Неправильний вибір прийменника	27	2,0%
2.1.4 Відсутність необхідного прийменника	10	0,7%
2.1.5 Відхилення від нормативно-узуального порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір граматичної трансформації	6	0,4%
2.1.6 Вживання зайвого прийменника	2	0,2%
2.2 Помилки у побудові речень	208	15,5%
2.2.1 Помилки у побудові простих речень	98	7,3%
2.2.1.1 Порушення синтаксичних зв'язків між членами речення	55	4,1%
2.2.1.2 Граматична неоднотипність однорідних членів речення	31	2,3%
2.2.1.3 Порушення координації між підметом і присудком	11	0,8%
2.2.1.4 Неправильний порядок слів	1	0,1%
2.2.2 Помилки у побудові складних речень	110	8,2%
2.2.2.1 Порушення синтаксичних зв'язків між членами речення	45	3,4%
2.2.2.2 Порушення координації між підметом і присудком	27	2,0%
2.2.2.3 Граматична неоднотипність однорідних членів речення	20	1,4%
2.2.2.4 Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	9	0,6%
2.2.2.5 Пропущені члени речення	3	0,2%
2.2.2.6 Неправильний порядок слів	2	0,2%
2.2.2.7 Зайві члени речення	2	0,2%
2.2.2.8 Шгучне ускладнення речення	1	0,1%
2.2.2.9 Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення	1	0,1%
III. Лексико-граматичні помилки	16	1,2%
IV. Орфографічні помилки	2	0,2%
Усього	1343	100%

## Додаток Ж

**Кількісне співвідношення помилок  
системи Free Translation**

<b>Типи помилок</b>	<b>Кількість помилок</b>	
<b>I. Лексичні помилки</b>	327	35,6%
1. Помилки у перекладі слів	182	19,8%
1.1 Вибір нерелевантних перекладних відповідників	154	16,7%
1.1.1 Неправильне вирішення лексичної полісемії	117	12,7%
1.1.2 Неправильне вирішення лексико-граматичної омонімії	36	3,9%
1.1.3 Неадекватний переклад власних назв та аббревіатур, які слід було залишити в оригінальному варіанті	1	0,1%
1.2 Неперекладені слова	28	3,1%
2. Помилки у перекладі словосполучень	145	15,8%
2.1 Неадекватний переклад словосполучень як семантичних єдностей	78	8,5%
2.2 Дослівний переклад іншомовних висловів	44	4,8%
2.3 Неправильний переклад фразеологічних сполучень	15	1,6%
2.4 Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті	8	0,9%
<b>II. Граматичні помилки</b>	587	64,0%
1. Морфологічні помилки	110	12,0%
1.1 Помилки при знятті граматичної омонімії	93	10,1%
1.2 Порушення категорії числа іменника	6	0,7%
1.3 Порушення категорії часу дієслова	5	0,6%
1.4 Ненормативне утворення пасивної конструкції дієслова	4	0,4%
1.5 Ненормативне утворення відмінкової форми іменника	2	0,2%
2. Синтаксичні помилки	477	52,0%
2.1 Помилки у побудові словосполучень	220	24,0%
2.1.1 Порушення керування	109	11,9%
2.1.2 Порушення узгодження між словами	66	7,2%
2.1.3 Неправильний вибір прийменника	16	1,7%
2.1.4 Вживання зайвого прийменника	16	1,7%

2.1.5 Відсутність необхідного прийменника	7	0,8%
2.1.6 Відхилення від нормативно-узуального порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір граматичної трансформації	6	0,7%
2.2 Помилки у побудові речень	257	28,0%
2.2.1 Помилки у побудові простих речень	124	13,5%
2.2.1.1 Порушення синтаксичних зв'язків між членами речення	51	5,6%
2.2.1.2 Граматична неоднотипність однорідних членів речення	31	3,4%
2.2.1.3 Порушення координації між підметом і присудком	16	1,7%
2.2.1.4 Пропущені члени речення	11	1,2%
2.2.1.5 Неправильний порядок слів	9	1,0%
2.2.1.6 Зайві члени речення	5	0,5%
2.2.1.7 Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	1	0,1%
2.2.2 Помилки у побудові складних речень	133	14,5%
2.2.2.1 Порушення координації між підметом і присудком	44	4,8%
2.2.2.2 Порушення синтаксичних зв'язків між членами речення	43	4,7%
2.2.2.3 Граматична неоднотипність однорідних членів речення	20	2,2%
2.2.2.4 Пропущені члени речення	8	0,9%
2.2.2.5 Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	5	0,6%
2.2.2.6 Вживання сполучника, що суперечить смисловим зв'язкам між частинами речення	4	0,4%
2.2.2.7 Неправильний порядок слів	4	0,4%
2.2.2.8 Зайві члени речення	4	0,4%
2.2.2.9 Штучне ускладнення речення	1	0,1%
III. Лексико-граматичні помилки	2	0,2%
IV. Орфографічні помилки	2	0,2%
Усього	918	100%

## Додаток 3

**Кількісне співвідношення помилок  
системи Yandex**

<b>Типи помилок</b>	<b>Кількість помилок</b>	
<b>I. Лексичні помилки</b>	189	33,1%
1. Помилки у перекладі слів	107	18,7%
1.1 Вибір нерелевантних перекладних відповідників	87	15,2%
1.1.1 Неправильне вирішення лексичної полісемії	70	12,3%
1.1.2 Неправильне вирішення лексико-граматичної омонімії	15	2,6%
1.1.3 Вживання русизмів	2	0,3%
1.2 Неперекладені слова	20	3,5%
2. Помилки у перекладі словосполучень	82	14,4%
2.1 Неадекватний переклад словосполучень як семантичних єдностей	26	4,6%
2.2 Дослівний переклад іншомовних висловів	51	8,9%
2.3 Неадекватний переклад фразеологічних сполучень	3	0,5%
2.4 Неадекватний переклад власних назв, які слід було залишити в оригінальному варіанті	2	0,4%
<b>II. Граматичні помилки</b>	382	66,9%
1. Морфологічні помилки	46	8,1%
1.1 Помилки при знятті граматичної омонімії	32	5,6%
1.2 Порушення категорії числа іменника	7	1,2%
1.3 Ненормативне утворення відмінкової форми іменника	3	0,5%
1.4 Порушення категорії часу дієслова	3	0,5%
1.5 Ненормативне утворення ступеня порівняння прикметника	1	0,2%
2. Синтаксичні помилки	336	58,8%
2.1 Помилки у побудові словосполучень	171	29,9%
2.1.1 Порушення керування	102	17,9%
2.1.2 Порушення узгодження між словами	33	5,8%
2.1.3 Неправильний вибір прийменника	14	2,5%
2.1.4 Вживання зайвого прийменника	9	1,6%
2.1.5 Відсутність необхідного прийменника	7	1,2%
2.1.6 Відхилення від нормативно-uzuального	6	0,9%

порядку слів, порушення логіко-синтаксичного зв'язку між словами через неправильний вибір граматичної трансформації		
2.2 Помилки у побудові речень	165	28,9%
2.2.1 Помилки у побудові простих речень	82	14,4%
2.2.1.1 Порушення синтаксичних зв'язків між членами речення	38	6,7%
2.2.1.2 Граматична неоднотипність однорідних членів речення	20	3,5%
2.2.1.3 Порушення координації між підметом і присудком	11	1,9%
2.2.1.4 Зайві члени речення	6	1,1%
2.2.1.5 Пропущені члени речення	4	0,7%
2.2.1.6 Неправильний порядок слів	2	0,3%
2.2.1.7 Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	1	0,2%
2.2.2 Помилки у побудові складних речень	83	14,5%
2.2.2.1 Порушення синтаксичних зв'язків між членами речення	36	6,3%
2.2.2.2 Порушення координації між підметом і присудком	19	3,3%
2.2.2.3 Граматична неоднотипність однорідних членів речення	16	2,8%
2.2.2.4 Пропущені члени речення	5	0,9%
2.2.2.5 Невідповідність анафоричного займенника співвідносному з ним іменникові в іншій частині речення	4	0,7%
2.2.2.6 Зайві члени речення	2	0,3%
2.2.2.7 Шгучне ускладнення речення	1	0,2%
III. Лексико-граматичні помилки	0	0,0%
IV. Орфографічні помилки	0	0,0%
Усього	571	100%

## Додаток И

### Помилки систем

#### Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation i Yandex (англо-український напрям перекладу)

Типи помилок	Pragma 6.3 Business		Google Translate		Bing Translator		Babylon		Free Translation		Yandex	
	Кількість	Відсоток	Кількість	Відсоток	Кількість	Відсоток	Кількість	Відсоток	Кількість	Відсоток	Кількість	Відсоток
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:												
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%

## Додаток К

## Помилки систем

**Pragma 6.3 Business, Google Translate, Bing Translator, Babylon, Free Translation і Yandex**  
(англо-російський напрям перекладу)

Типи помилок	Pragma 6.3 Business		Google Translate		Bing Translator		Babylon		Free Translation		Yandex	
	Кількість	Відсоток	Кількість	Відсоток	Кількість	Відсоток	Кількість	Відсоток	Кількість	Відсоток	Кількість	Відсоток
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:	189	16,8%	158	23,2%	208	26,1%	257	24,5%	277	26,3%	167	28,5%
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%